with the biotinylated anti-Tac, thus decreasing fluorescence more.

Biological properties of the humanized antibody

For optimal use in treatment of human disease, the humanized antibody should be able to destroy T-cells in the body that express the IL-2 receptor. One mechanism by which antibodies may destroy target cells is antibody-dependent cell-mediated cytotoxicity, abbreviated ADCC (Fundamental Immunology, Paul, W., Ed., Raven Press, New York (1984), at pg. 681), in which the antibody forms a bridge between the target cell and an effector cell such as a macrophage that can lyse the target. To determine whether the humanized antibody and the original mouse anti-Tac antibody can mediate ADCC, a chromium release assay was performed by standard methods. Specifically, human leukemia HUT-102 cells, which express the IL-2 receptor, were incubated with ^{51}Cr to allow them to absorb this radionuclide. The HUT-102 cells were then incubated with an excess of either anti-Tac or humanized anti-Tac antibody. The HUT-102 cells were next incubated for 4 hrs with either a 30:1 or 100:1 ratio of effector cells, which were normal purified human peripheral blood mononuclear cells that had been activated by incubation for about 20 hrs with human recombinant IL-2. Release of ^{51}Cr, which indicated lysis of the target HUT-102 cells, was measured and the background subtracted (Table 1). The results show that at either ratio of effector cells, anti-Tac did not lyse a significant number of the target cells (less than 5%), while the humanized antibody did (more than 20%). Hence, the humanized antibody is likely to be more efficacious than the original mouse antibody in treating T-cell leukemia or other T-cell mediated diseases.

TABLE 1

Percent ^{51}Cr release after ADCC

	Effector: Target ratio	
	30:1	100:1
Antibody		
Anti-Tac	4%	< 1%
Humanized anti-Tac	24%	23%

From the foregoing, it will be appreciated that the human-like immunoglobulins of the present invention offer numerous advantages of other human IL-2 receptor-specific antibodies. In comparison to anti-Tac mouse monoclonal antibodies, the present human-like immunoglobulin can be more economically produced and contain substantially less foreign amino acid sequences. This reduced likelihood of antigenicity after injection into a human patient represents a significant therapeutic improvement.

Although the present invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

WE CLAIM:

1. A composition comprising a substantially pure human-like immunoglobulin specifically reactive with p55 Tac protein.

2. A composition according to Claim 1, wherein the immunoglobulin comprises two pairs of light/heavy chain dimers, wherein each chain comprises a variable region and a constant region.

3. A composition according to Claim 2, wherein a variable region of at least one chain comprises at least about 75 amino acids from a human immunoglobulin variable region framework.

4. A composition comprising a substantially pure human-like immunoglobulin capable of inhibiting binding of human interleukin-2 (IL-2) to a human IL-2 receptor.

5. A composition according to Claims 1 or 4, wherein the immunoglobulin exhibits a binding affinity to a human IL-2 receptor of about 10^8 M^{-1} or stronger.

6. A composition according to Claims 1 or 4, wherein the immunoglobulin comprises complementarity determining regions from one immunoglobulin and framework regions from at least one different immunoglobulin.

7. A recombinant immunoglobulin composition comprising a human-like framework and one or more foreign complementarity determining regions not naturally associated with the framework, wherein said immunoglobulin is capable of binding to a human interleukin-2 receptor.

8. A composition according to Claim 7, wherein one or more of the foreign CDR's are substantially homologous to a CDR from an immunoglobulin reactive with human p55 Tac protein.

9. A composition according to Claim 7, wherein all of the foreign CDR's are located on heavy chains of the immunoglobulin.

10. A composition according to Claim 7, wherein the immunoglobulin is an IgG_1 immunoglobulin isotype.

11. A composition according to Claim 7, wherein the mature light and heavy variable region protein sequences are substantially homologous to the sequences in Figures 3 and 4.

12. A human-like immunoglobulin having two pairs of light chain/heavy chain dimers and capable of specifically reacting with an epitope on a human interleukin-2 receptor with an affinity of at least about 10^8 M^{-1}, said light and heavy chains comprising complementarity determining regions (CDR's) and human-like framework regions, wherein the CDR's are from different immunoglobulin molecules than the framework regions.

13. An immunoglobulin according to Claim 12, which binds to an epitope located on a p55 Tac protein.

14. An immunoglobulin according to Claim 12, which is capable of blocking the binding of interleukin-2 (IL-2) to human IL-2 receptors.

15. An immunoglobulin according to Claim 12, wherein the human-like framework regions comprise amino acids sequences from at least two human immunoglobulins.

16. An immunoglobulin according to Claim 12, wherein the CDR's are from a mouse immunoglobulin.

17. A humanized immunoglobulin capable of binding to human interleukin-2 receptors, said immunoglobulin comprising one or more complementarity determining regions (CDR's) from anti-Tac antibody in a human-like framework.

18. A humanized immunoglobulin according to Claim 17, wherein the human framework is substantially homologous to an Eu immunoglobulin framework.

19. A humanized immunoglobulin according to Claim 17, having a mature heavy chain variable sequence as shown in Figure 3, and a mature light chain sequence as shown in Figure 4.

20. A humanized immunoglobulin according to Claim 17 which is capable of blocking the binding of IL-2 to interleukin-2 receptors on human T-cells.

21. A method of treating T-cell mediated disorders in a human patient, said method comprising administering to said patient a therapeutically effective dose of an immunoglobulin according to Claims 1, 5, 12, or 17.

22. An immunoglobulin according to Claims 1, 5, 12, or 17 which was produced in a myeloma or hybridoma cell.

23. A human-like immunoglobulin heavy chain comprising a human-like heavy chain framework region and a hypervariable region which is substantially identical to a monoclonal antibody heavy chain hypervariable region secreted by the cell line designated A.T.C.C. Accession No. CRL 9688.

24. A human-like immunoglobulin light chain comprising a human light chain framework region and a hypervariable region which is substantially identical to a monoclonal antibody light chain hypervariable region secreted by the cell line designated A.T.C.C. Accession No. CRL 9688.

25. A polynucleotide molecule comprising a first sequence coding for human-like immunoglobulin framework regions and a second sequence coding for one or more mouse immunoglobulin complementarity determining regions, wherein upon expression said polynucleotide encodes an immunoglobulin specifically reactive with p55 Tac protein and capable of blocking the binding of interleukin-2 (IL-2) to the IL-2 receptor on human T-cells.

26. A cell line transfected with a polynucleotide of Claim 25.

NOVEL IL-2 RECEPTOR-SPECIFIC HUMAN IMMUNOGLOBULINS

ABSTRACT OF THE DISCLOSURE

Human-like immunoglobulins specifically reactive with human IL-2 receptors are prepared employing recombinant DNA technology for use in, e.g., treatment of T-cell mediated disorders.

ATTORNEY DOCKET NO.

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:
My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

NOVEL IL-2 RECEPTOR-SPECIFIC HUMAN IMMUNOGLOBULINS

the specification of which ☒ is attached hereto or ☐ was filed on _____ as Application Serial No. _____ and was amended on _____ (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a). I claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)

COUNTRY	APPLICATION NUMBER	DATE OF FILING	PRIORITY CLAIMED UNDER 35 U.S.C. 119
			Yes____ No____
			Yes____ No____

I claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION SERIAL NO.	DATE OF FILING	STATUS
		☐ Patented ☐ Pending ☐ Abandoned
		☐ Patented ☐ Pending ☐ Abandoned

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) who are partners and associates in the firm of Townsend and Townsend to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. William M. Smith, Reg. No. 30,223
Steven W. Parmelee, Reg. No. 31,990
James M. Heslin, Reg. No. 29,541

SEND CORRESPONDENCE TO:	DIRECT TELEPHONE CALLS TO (name, registration number, and telephone number)
William M. Smith TOWNSEND and TOWNSEND Steuart Street Tower, One Market Plaza San Francisco, CA 94105	William M. Smith, Reg. 30,223 ☐ (415) 543-9600 or ☒ (415) 326-2400

201	FULL NAME OF INVENTOR	Last Name: Queen	First Name: Cary		Middle Name or Initial: L.
	RESIDENCE & CITIZENSHIP	City: Palo Alto	State or Foreign Country: California		Country of Citizenship: USA
	POST OFFICE ADDRESS	Post Office Address: 1300 Oak Creek Dr.	City: Palo Alto	State or Country: California	Zip Code: 94304
202	FULL NAME OF INVENTOR	Last Name: Selick	First Name: Harold		Middle Name or Initial: Edwin
	RESIDENCE & CITIZENSHIP	City: Belmont	State or Foreign Country: California		Country of Citizenship: USA
	POST OFFICE ADDRESS	Post Office Address: 1673 Sunnyslope Ave.	City: Belmont	State or Country: California	Zip Code: 94002
203	FULL NAME OF INVENTOR	Last Name:	First Name:		Middle Name or Initial:
	RESIDENCE & CITIZENSHIP	City:	State or Foreign Country:		Country of Citizenship:
	POST OFFICE ADDRESS	Post Office Address:	City:	State or Country:	Zip Code:

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 201	Signature of Inventor 202	Signature of Inventor 203
	Harold E. Selick	
Date: 12/28/88	Date: 12/28/88	Date:

TST 144 (12 87)

Atty. Docket [illegible]

VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS
(37 CFR 1.9(f) and 1.27(c)) — SMALL BUSINESS CONCERN

Applicant or Patentee: Cary L. Queen and Harold Edwin Selick
Serial No.: _____ Filing Date: December 28, 1988
Patent No.: _____ Issued: _____
For: NOVEL IL-2 RECEPTOR-SPECIFIC HUMAN IMMUNOGLOBULINS

I hereby declare that I am

[ ] the owner of the small business concern identified below:
[X] an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF CONCERN Protein Design Labs, Inc., a Delaware Corporation
ADDRESS OF CONCERN 3181 Porter Drive
Palo Alto, California 94304

I hereby declare that the above identified small business concern qualifies as a small business concern as defined in 13 CFR 121.3-18, and reproduced in 37 CFR 1.9(d), for purposes of paying reduced fees under section 41(a) and (b) of Title 35, United States Code, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby declare that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention, entitled NOVEL IL-2 RECEPTOR-SPECIFIC HUMAN IMMUNOGLOBULINS by inventor(s) Cary L. Queen and Harold Edwin Selick described in

[X] the application filed herewith
[ ] application serial no. _____, filed _____.
[ ] patent no. _____, issued _____.

If the rights held by the above identified small business concern are not exclusive, each individual, concern or organization having rights to the invention is listed below* and no rights to the invention are held by any person, other than the inventor, who could not qualify as a small business concern under 37 CFR 1.9(d) or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

*NOTE: Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR 1.27)

NAME _____
ADDRESS _____
[ ] INDIVIDUAL [ ] SMALL BUSINESS CONCERN [ ] NONPROFIT ORGANIZATION

NAME _____
ADDRESS _____
[ ] INDIVIDUAL [ ] SMALL BUSINESS CONCERN [ ] NONPROFIT ORGANIZATION

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING Laurence Jay Korn
TITLE OF PERSON OTHER THAN OWNER President
ADDRESS OF PERSON SIGNING 3181 Porter Drive, Palo Alto, California 94304

SIGNATURE Laurence Jay Korn DATE 28/12/88

ver 10

FIGURE 1

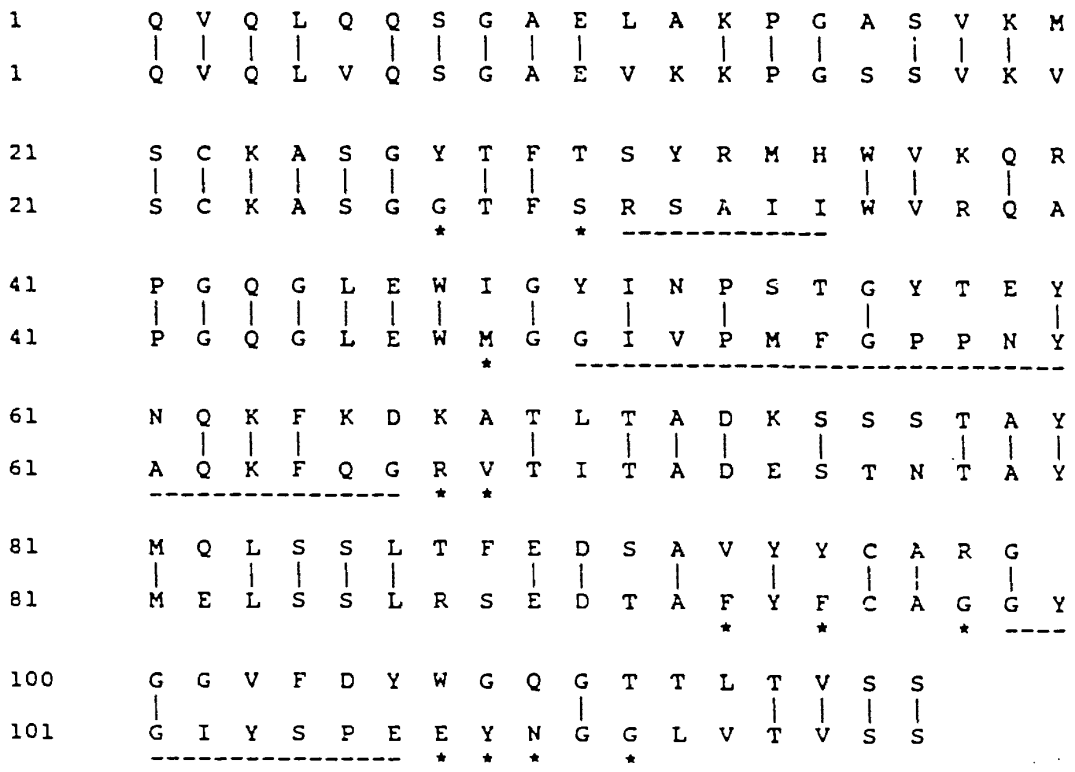

1 Q V Q L Q Q S G A E L A K P G A S V K M
1 Q V Q L V Q S G A E V K K P G S S V K V
21 S C K A S G Y T F T S Y R M H W V K Q R
21 S C K A S G G T F S R S A I I W V R Q A
41 P G Q G L E W I G Y I N P S T G Y T E Y
41 P G Q G L E W M G G I V P M F G P P N Y
61 N Q K F K D K A T L T A D K S S S T A Y
61 A Q K F Q G R V T I T A D E S T N T A Y
81 M Q L S S L T F E D S A V Y Y C A R G
81 M E L S S L R S E D T A F Y F C A G G Y
100 G G V F D Y W G Q G T T L T V S S
101 G I Y S P E E Y N G G L V T V S S

FIGURE 2

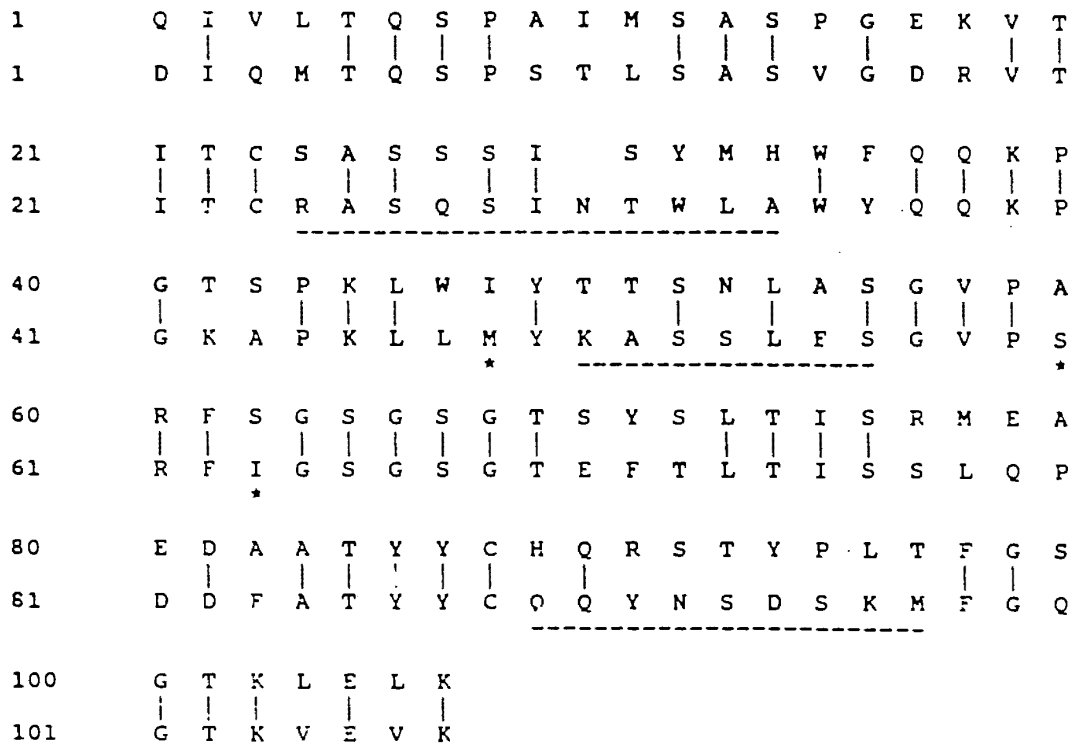

1 Q I V L T Q S P A I M S A S P G E K V T
1 D I Q M T Q S P S T L S A S V G D R V T
21 I T C S A S S S I S Y M H W F Q Q K P
21 I T C R A S Q S I N T W L A W Y Q Q K P
40 G T S P K L W I Y T T S N L A S G V P A
41 G K A P K L L M Y K A S S L F S G V P S
60 R F S G S G S G T S Y S L T I S R M E A
61 R F I G S G S G T E F T L T I S S L Q P
80 E D A A T Y Y C H Q R S T Y P L T F G S
81 D D F A T Y Y C Q Q Y N S D S K M F G Q
100 G T K L E L K
101 G T K V E V K

FIGURE 3

```
        10         20         30         40         50         60
TCTAGATGGGATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGTACCGCGGGCGTGCACT
      M  G  W  S  W  I  F  L  F  L  L  S  G  T  A  G  V  H

        70         80         90        100        110        120
CTCAGGTCCAGCTTGTCCAGTCTGGGGCTGAAGTCAAGAAACCTGGCTCGAGCGTGAAGG
S  Q  V  Q  L  V  Q  S  G  A  E  V  K  K  P  G  S  S  V  K

       130        140        150        160        170        180
TCTCCTGCAAGGCTTCTGGCTACACCTTTACTAGCTACAGGATGCACTGGGTAAGGCAGG
V  S  C  K  A  S  G  Y  T  F  T  S  Y  R  M  H  W  V  R  Q

       190        200        210        220        230        240
CCCCTGGACAGGGTCTGGAATGGATTGGATATATTAATCCGTCGACTGGGTATACTGAAT
A  P  G  Q  G  L  E  W  I  G  Y  I  N  P  S  T  G  Y  T  E

       250        260        270        280        290        300
ACAATCAGAAGTTCAAGGACAAGGCAACAATTACTGCAGACGAATCCACCAATACAGCCT
Y  N  Q  K  F  K  D  K  A  T  I  T  A  D  E  S  T  N  T  A

       310        320        330        340        350        360
ACATGGAACTGAGCAGCCTGAGATCTGAGGACACCGCAGTCTATTACTGTGCAAGAGGGG
Y  M  E  L  S  S  L  R  S  E  D  T  A  V  Y  Y  C  A  R  G

       370        380        390        400        410        420
GGGGGGTCTTTGACTACTGGGGCCAAGGAACCCTGGTCACAGTCTCCTCAGGTGAGTCCT
G  G  V  F  D  Y  W  G  Q  G  T  L  V  T  V  S  S

       430
TAAAACCTCTAGA
```

FIGURE 4

```
         10         20         30         40         50         60
TCTAGATGGAGACCGATACCCTCCTGCTATGGGTCCTCCTGCTATGGGTCCCAGGATCAA
      M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S

         70         80         90        100        110        120
CCGGAGATATTCAGATGACCCAGTCTCCATCTACCCTCTCTGCTAGCGTCGGGGATAGGG
T  G  D  I  Q  M  T  Q  S  P  S  T  L  S  A  S  V  G  D  R

        130        140        150        160        170        180
TCACCATAACCTGCTCTGCCAGCTCAAGTATAAGTTACATGCACTGGTACCAGCAGAAGC
V  T  I  T  C  S  A  S  S  S  I  S  Y  M  H  W  Y  Q  Q  K

        190        200        210        220        230        240
CAGGCAAAGCTCCCAAGCTTCTAATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTG
P  G  K  A  P  K  L  L  I  Y  T  T  S  N  L  A  S  G  V  P

        250        260        270        280        290        300
CTCGCTTCAGTGGCAGTGGATCTGGGACCGAGTTCACCCTCACAATCAGCTCTCTGCAGC
A  R  F  S  G  S  G  S  G  T  E  F  T  L  T  I  S  S  L  Q

        310        320        330        340        350        360
CAGATGATTTCGCCACTTATTACTGCCATCAAAGGAGTACTTACCCACTCACGTTCGGTC
P  D  D  F  A  T  Y  Y  C  H  Q  R  S  T  Y  P  L  T  F  G

        370        380        390        400
AGGGGACCAAGGTGGAGGTCAAACGTAAGTACACTTTTCTAGA
Q  G  T  K  V  E  V  K
```

FIGURE 5

A

HES12 AGCTTCTAGATGGGATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGTACCGCGGGCGTG
CACTCTCAGGTCCAGCTTGTCCAGTCTGGGGCTGAAGTCAAGAAACCTGGCTCGAGCGTG
AAGGTC

HES13 CCCAGTCGACGGATTAATATATCCAATCCATTCCAGACCCTGTCCAGGGGCCTGCCTTAC
CCAGTGCATCCTGTAGCTAGTAAAGGTGTAGCCAGAAGCCTTGCAGGAGACCTTCACGCT
CGAGCCAGG

HES14 TATATTAATCCGTCGACTGGGTATACTGAATACAATCAGAAGTTCAAGGACAAGGCAACA
ATTACTGCAGACGAATCCACCAATACAGCCTACATGGAACTGAGCAGCCTGAGATCTGAG
GACA

HES15 ATATCGTCTAGAGGTTTTAAGGACTCACCTGAGGAGACTGTGACCAGGGTTCCTTGGCCC
CAGTAGTCAAAGACCCCCCCCCCCTCTTGCACAGTAATAGACTGCGGTGTCCTCAGATCTC
AGGCTGCT

B

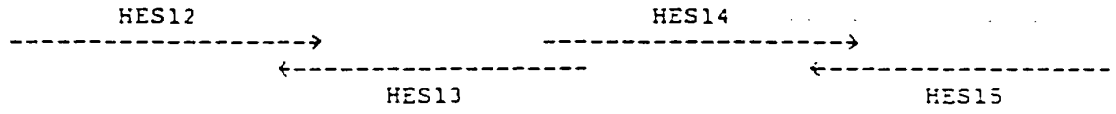


FIGURE 6

A

JFD1 CAAATCTAGATGGAGACCGATACCCTCCTGCTATGGGTCCTCCTGCTATGGGTCCCAGGA
TCAACCGGAGATATTCAGATGACCCAGTCTCCATCTACCCTCTCTGCTAGCGTCGGGGAT

JFD2 ATAAATTAGAAGCTTGGGAGCTTTGCCTGGCTTCTGCTGGTACCAGTGCATGTAACTTAT
ACTTGAGCTGGCAGAGCAGGTTATGGTGACCCTATCCCCGACGCTAGCAGAGAG

JFD3 GCTCCCAAGCTTCTAATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTGCTCGCTTC
AGTGGCAGTGGATCTGGGACCGAGTTCACCCTCACAATCAGCTCTCTGCAGCCAGATGAT
TTC

JFD4 TATATCTAGAAAAGTGTACTTACGTTTGACCTCCACCTTGGTCCCCTGACCGAACGTGAG
TGGGTAAGTACTCCTTTGATGGCAGTAATAAGTGGCGAAATCATCTGGCTGCAGAGAGCT
GA

B

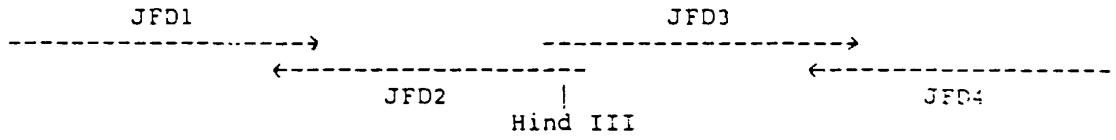


FIGURE 7

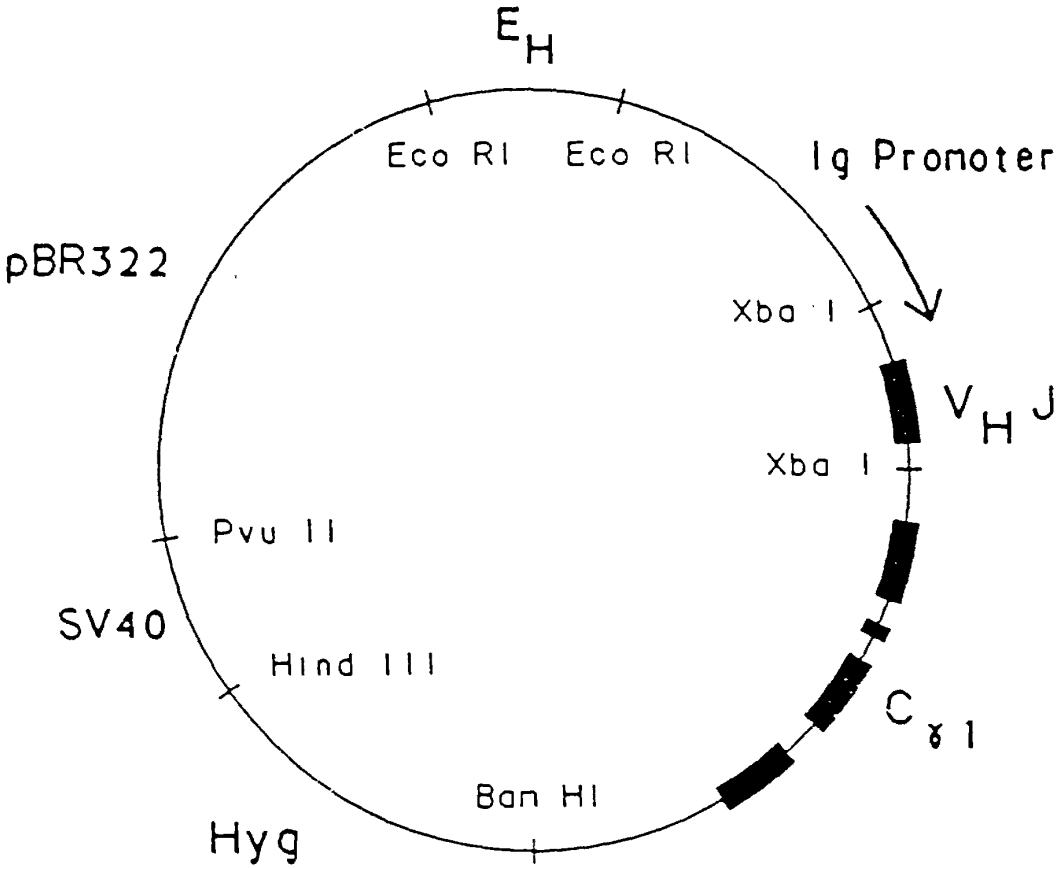

E_H
Eco RI
Eco RI
Ig Promoter
pBR322
Xba I
$V_H J$
Xba I
Pvu II
SV40
Hind III
$C_{\gamma 1}$
Ban HI
Hyg

FIGURE 8

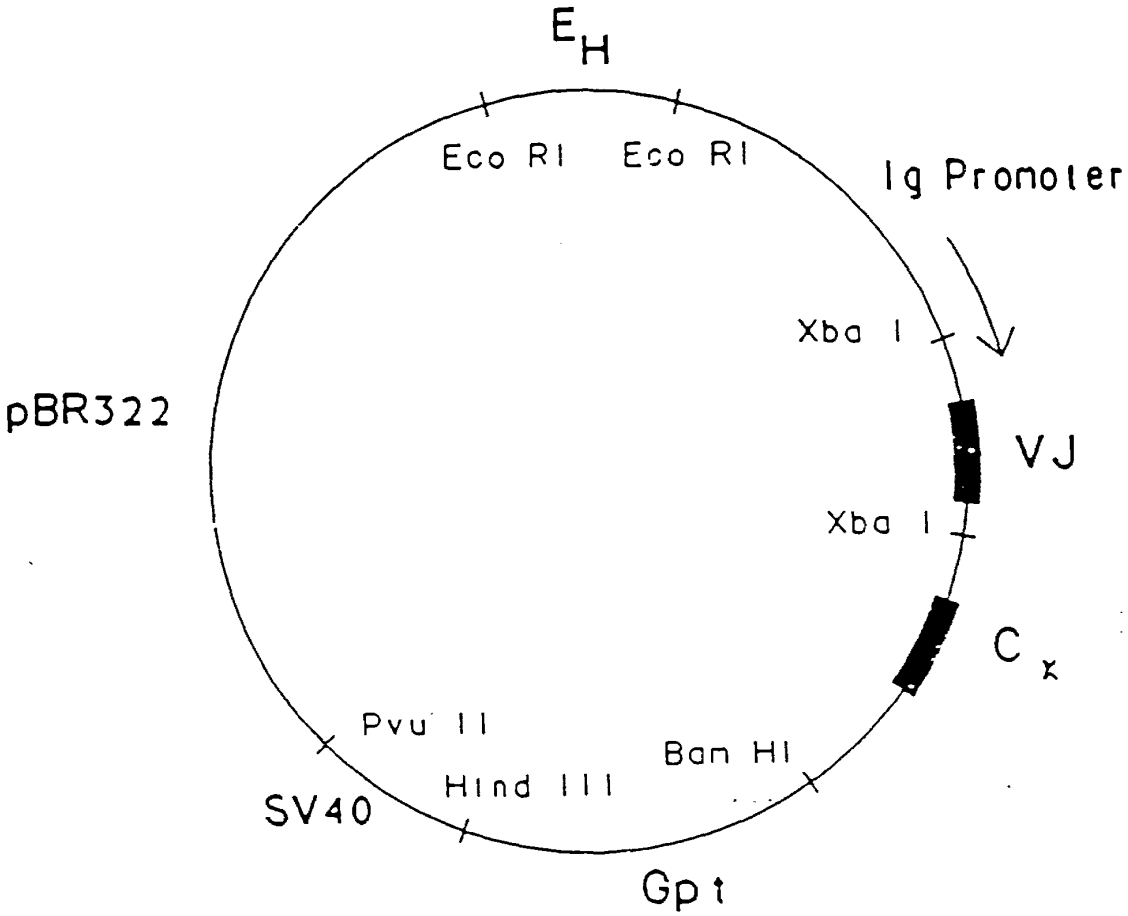

E H
Eco RI
Eco RI
Ig Promoter
Xba I
pBR322
VJ
Xba I
C x
Pvu II
Ban HI
Hind III
SV40
Gpt

FIGURE 9

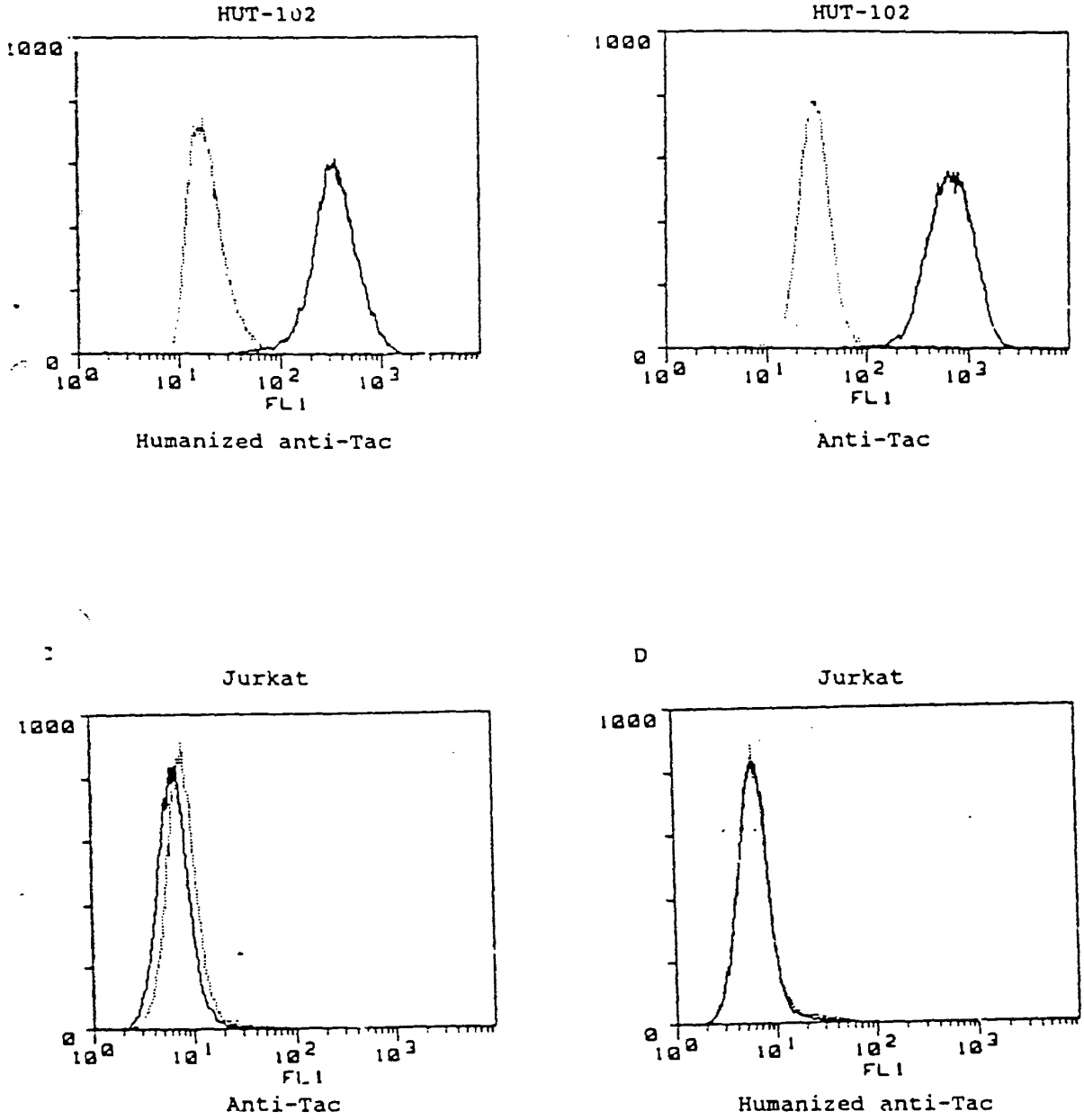
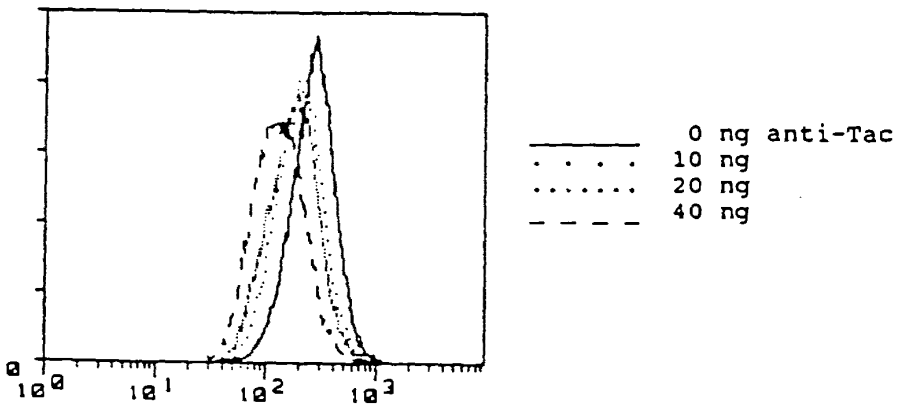
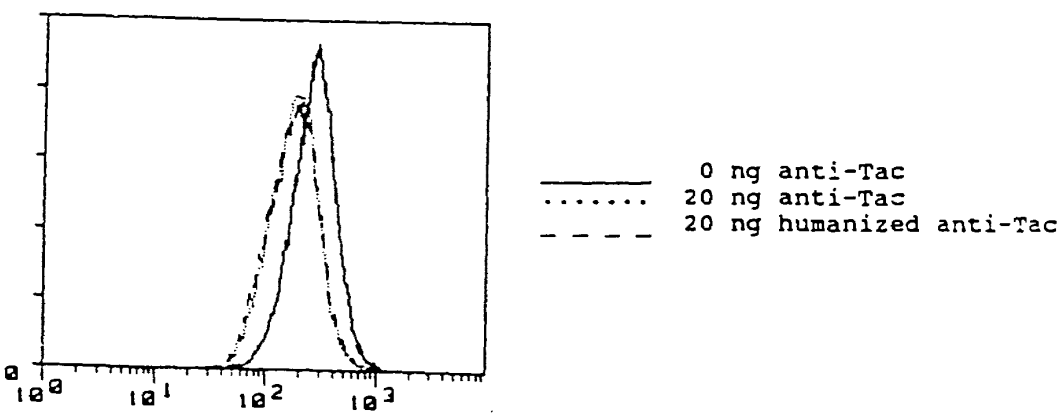

HUT-102
1000
0
10^0
10^1
10^2
10^3
FL1
Humanized anti-Tac
B
HUT-102
1000
0
10^0
10^1
10^2
10^3
FL1
Anti-Tac
Jurkat
1000
0
10^0
10^1
10^2
10^3
FL1
Anti-Tac
D
Jurkat
1000
0
10^0
10^1
10^2
10^3
FL1
Humanized anti-Tac

FIGURE 10

A


0
10^0
10^1
10^2
10^3
0 ng anti-Tac
10 ng
20 ng
40 ng

B


0
10^0
10^1
10^2
10^3
0 ng anti-Tac
20 ng anti-Tac
20 ng humanized anti-Tac

PCT/US 89/05857

SERIAL NUMBER (Series of 1987)	PATENT DATE	PATENT NUMBER
07/310252		

SERIAL NUMBER	FILING DATE	CLASS	SUBCLASS	GROUP ART UNIT	EXAMINER
07/310,252	02/13/89	435		135	

APPLICANTS

CARY L. QUEEN, PALO ALTO, CA; HAROLD E. SELICK, BELMONT, CA.

CONTINUING DATA********
VERIFIED THIS APPLN IS A CIP OF 07/290,975 12/28/88

REC'D 28 DEC 1989
WIPO PCT

FOREIGN/PCT APPLICATIONS********
VERIFIED

PRIORITY DOCUMENT

FOREIGN FILING LICENSE GRANTED 03/03/89 ***** SMALL ENTITY *****

Foreign priority claimed / 35 USC 119 conditions met / Verified and Acknowledged	AS FILED	STATE OR COUNTRY	SHEETS DRWGS.	TOTAL CLAIMS	INDEP. CLAIMS	FILING FEE RECEIVED	ATTORNEY'S DOCKET NO.
☐yes ☐no / ☐yes ☐no / Examiner's Initials	→	CA	10	28	3	$ 279.00	118239

ADDRESS
WILLIAM M. SMITH
TOWNSEND AND TOWNSEND
STEUART STREET TOWER
ONE MARKET PLAZA
SAN FRANCISCO, CA 94105

TITLE
DESIGNING IMPROVED HUMANIZED IMMUNOGLOBULINS

...COMM. Pat. & TM Office — PTO-436A (rev. 10-78)

This is to certify that annexed hereto is a true copy from the records of the United States Patent and Trademark Office of the application as originally filed which is identified above.

By authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

H. L. Jackson

Certifying Officer

Date ... EB 1990

PATENT APPLICATION SERIAL NO. 07/310252

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

S 70016 02/17 09 310252 20 1430 035 201 [illegible]

PTO-1556
(5/87)

DESIGNING IMPROVED HUMANIZED IMMUNOGLOBULINS

CROSS-REFERENCE TO RELATED APPLICATION

This is a continuation-in-part application of commonly assigned patent application U.S.S.N. 290,975, filed December 28, 1988, which is incorporated herein by reference.

Field of the Invention

The present invention relates generally to the combination of recombinant DNA and monoclonal antibody technologies for developing novel therapeutic agents and, more particularly, to the production of non-immunogenic antibodies having strong affinity for a predetermined antigen.

Background of the Invention

The advent of monoclonal antibody technology in the mid 1970's heralded a new age of medicine. For the first time, researchers and clinicians had access to essentially unlimited quantities of uniform antibodies capable of binding to a predetermined antigenic site and having various immunological effector functions. These proteins, known as "monoclonal antibodies" were thought to hold great promise in, e.g., the removal of harmful cells in vivo. Indeed, the clinical value of monoclonal antibodies seemed limitless for this use alone.

Unfortunately, the development of appropriate therapeutic products based on these proteins has been severely hampered by a number of drawbacks inherent in monoclonal antibody production. For example, most monoclonal antibodies are mouse derived, and thus do not fix human complement well. They also lack other important immunoglobulin functional characteristics when used in humans.

Perhaps most importantly, non-human monoclonal antibodies contain substantial stretches of amino acid sequences that will be immunogenic when injected into a human

patient. Numerous studies have shown that after injection of a foreign antibody, the immune response mounted by a patient can be quite strong, essentially eliminating the antibody's therapeutic utility after an initial treatment. Moreover, as increasing numbers of different mouse or other antigenic (to humans) monoclonal antibodies can be expected to developed to treat various diseases, after the first or second treatments with any non-human antibodies, subsequent treatments, even for unrelated therapies, can be ineffective or even dangerous in themselves.

While the production of so called "chimeric antibodies" (e.g., mouse variable regions joined to human constant regions) has proven somewhat successful, a significant immunogenicity problem remains. Moreover, efforts to immortalize human B-cells or generate human hybridomas capable of producing human immunoglobulins against a desired antigen have been generally unsuccessful, particularly with many important human antigens. Most recently, recombinant DNA technology has been utilized to produce immunoglobulins which have human framework regions combined with complementarity determining regions (CDR's) from a donor mouse or rat immunoglobulin (see, e.g., EPO Publication No. 0239400, which is incorporated herein by reference). These new proteins are called "humanized immunoglobulins" and the process by which the donor immunoglobulin is converted into a human-like immunoglobulin by combining its CDR's with a human framework is called "humanization". Humanized antibodies are important because they bind to the same antigen as the original antibodies, but are less immunogenic when injected into humans.

However, a major problem with present humanization procedures has been a loss of affinity for the antigen, usually by at least 2 to 3-fold (Jones et al., Nature, 321:522-525 (1986)) and in some instances as much as 10-fold or more, especially when the antigen is a protein (Verhoeyen et al., Science, 239:1534-1536 (1988)). Loss of any affinity is, of course, highly undesirable. At the least, it means that more of the humanized antibody will have to be injected

into the patient, at higher cost and greater risk of adverse effects. Even more critically, an antibody with reduced affinity may have poorer biological functions, such as complement lysis, antibody-dependent cellular cytotoxicity, or virus neutralization. For example, the loss of affinity in the partially humanized antibody HuVHCAMP may have caused it to lose all ability to mediate complement lysis (see, Riechmann et al., Nature, 332:323-327 (1988); Table 1).

Thus, there is a need for improved means for producing humanized antibodies specifically reactive with strong affinity to a predetermined antigen. These humanized immunoglobulins should remain substantially non-immunogenic in humans, yet be easily and economically produced in a manner suitable for therapeutic formulation and other uses. The present invention fulfills these and other needs.

Summary of the Invention

The present invention provides novel methods for designing humanized immunoglobulin chains having one or more complementarity determining regions (CDR's) from a donor immunoglobulin and a framework region from a human immunoglobulin, the preferred methods comprising first comparing the framework or variable region amino acid sequence of the donor immunoglobulin to corresponding sequences in a collection of human immunoglobulin chains, and selecting as the human immunoglobulin one of the more homologous sequences from the collection. The human immunoglobulin, or acceptor immunoglobulin, sequence is typically selected from a collection of at least 10 to 20 immunoglobulin chain sequences, and usually will have the highest homology to the donor immunoglobulin sequence of any sequence in the collection. The human immunoglobulin framework sequence will typically have about 65 to 70% homology or more to the donor immunoglobulin framework sequences. The donor immunoglobulin may be either a heavy chain or light chain (or both), and the human collection will contain the same kind of chain. A humanized light and heavy chain can be used to form a complete humanized immunoglobulin

or antibody, having two light/heavy chain pairs, with or without partial or full-length human constant regions and other proteins.

In another embodiment of the present invention, either in conjunction with the above comparison step or separately, additional amino acids in an acceptor immunoglobulin chain may be replaced with amino acids form the CDR-donor immunoglobulin chain. More specifically, further optional substitutions of a human framework amino acid of the acceptor immunoglobulin with a corresponding amino acid from a donor immunoglobulin will be made at positions in the immunoglobulins where:

(a) the amino acid in the human framework region of an acceptor immunoglobulin is rare for that position and the corresponding amino acid in the donor immunoglobulin is common for that position in human immunoglobulin sequences; or

(b) the amino acid is immediately adjacent to one of the CDR's; or

(c) the amino acid is predicted to be within about 3Å of the CDR's in a three-dimensional immunoglobulin model and capable of interacting with the antigen or with the CDR's of the humanized immunoglobulin.

The humanized immunoglobulin chain will typically comprise at least about 3 amino acids from the donor immunoglobulin in addition to the CDR's, usually at least one of which is immediately adjacent to a CDR in the donor immunoglobulin. The heavy and light chains may each be designed by using any one or all three of the position criteria.

When combined into an intact antibody, the humanized light and heavy chains of the present invention will be substantially non-immunogenic in humans and retain substantially the same affinity as the donor immunoglobulin to the antigen (such as a protein or other compound containing an epitope). These affinity levels can vary from about 10^8 M^{-1} or higher, and may be within about 4 fold of the donor immunoglobulin's original affinity to the antigen.

Once designed, the immunoglobulins, including binding fragments and other immunoglobulin forms, of the present invention may be produced readily by a variety of recombinant DNA or other techniques. Preferably, polynucleotides encoding the desired amino acid sequences are produced synthetically or by joining appropriate nucleic acid sequences for expression in a suitable host (e.g., cell culture). The humanized immunoglobulins will be particularly useful in treating human disorders susceptible to monoclonal antibody therapy, but find a variety of other uses as well.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1. Comparison of sequences of anti-Tac heavy chain (upper lines) and Eu heavy chain (lower lines). The 1-letter code for amino acids is used. The first amino acid on each line is numbered at the left. Identical amino acids in the two sequences are connected by lines. The 3 CDR's are underlined. Other amino acid positions for which the anti-Tac amino acid rather than the Eu amino acid was used in the humanized anti-Tac heavy chain are denoted by an *.

Figure 2. Comparison of sequences of anti-Tac light chain (upper lines) and Eu light chain (lower lines). The single-letter code for amino acids is used. The first amino acid on each line is numbered at the left. Identical amino acids in the two sequences are connected by lines. The 3 CDR's are underlined. Other amino acid positions for which the anti-Tac amino acid rather than the Eu amino acid was used in the humanized anti-Tac heavy chain are denoted by an *.

Figure 3. Nucleotide sequence of the gene for the humanized anti-Tac heavy chain variable region gene. The translated amino acid sequence for the part of the gene encoding protein is shown underneath the nucleotide sequence. The nucleotides TCTAGA at the beginning and end of the gene are Xba I sites. The mature heavy chain sequence begins with amino acid #20 Q.

Figure 4. Nucleotide sequence of the gene for the humanized anti-Tac light chain variable region gene. The translated amino acid sequence for the part of the gene encoding protein is shown underneath the nucleotide sequence. The nucleotides TCTAGA at the beginning and end of the gene are Xba I sites. The mature light chain sequence begins with amino acid #21 D.

Figure 5. A. Sequences of the four oligonucleotides used to synthesize the humanized anti-Tac heavy chain gene, printed 5' to 3'. B. Relative positions of the oligonucleotides. The arrows point in the 3' direction for each oligonucleotide.

Figure 6. (A) Sequences of the four oligonucleotides used to synthesize the humanized anti-Tac light chain gene, printed 5' to 3'. (B) Relative positions of the oligonucleotides. The arrows point in the 3' direction for each oligonucleotide. The position of a Hind III site in the overlap of JFD2 and JFD3 is shown.

Figure 7. Schematic diagram of the plasmid pHuGTAC1 used to express the humanized anti-Tac heavy chain. Relevant restriction sites are shown, and coding regions of the heavy chain are displayed as boxes. The direction of transcription from the immunoglobulin (Ig) promoter is shown by an arrow. E_H = heavy chain enhancer, Hyg = hygromycin resistance gene.

Figure 8. Schematic diagram of the plasmid pHuLTAC used to express the humanized anti-Tac light chain. Relevant restriction sites are shown, and coding regions of the light chain are displayed as boxes. The direction of transcription from the Ig promoter is shown by an arrow.

Figure 9. Fluorocytometry of HUT-102 and Jurkat cells stained with anti-Tac antibody or humanized anti-Tac antibody followed respectively by fluorescein-conjugated goat anti-mouse Ig antibody or goat anti-human Ig antibody, as labeled. In each panel, the dotted curve shows the results when the first antibody was omitted, and the solid curve the results when first and second (conjugated) antibodies were included as described.

Figure 10. (A) Fluorocytometry of HUT-102 cells stained with 0-40 ng of anti-Tac as indicated, then with biotinylated anti-Tac, and then with phycoerythrin-conjugated avidin. (B) Fluorocytometry of HUT-102 cells stained with the indicated antibody, then with biotinylated anti-Tac, and then with phycoerythrin-conjugated avidin.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, novel means of designing humanized immunoglobulins capable of specifically binding to a predetermined antigen with strong affinity are provided. These improved methods produce immunoglobulins that are substantially non-immunogenic in humans but have binding affinities of at least about 10^8 M^{-1}, preferably 10^9 M^{-1} to 10^{10} M^{-1}, or stronger. The humanized immunoglobulins will have a human framework and have one or more complementary determining regions (CDR's), plus a limited number of other amino acids, from a donor immunoglobulin specifically reactive with an antigen. The immunoglobulins can be produced economically in large quantities and find use, for example, in the treatment of various human disorders by a variety of techniques.

In order that the invention may be more completely understood, several definitions are set forth. As used herein, the term "immunoglobulin" refers to a protein having one or more polypeptides substantially encoded by immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Full-length immunoglobulin "light chains" (about 25 Kd, about 214 amino acids) are encoded by a variable region gene at the NH2-terminus (about 110 amino acids) and a kappa or lambda constant region gene at the COOH - terminus. Full-length immunoglobulin "heavy chains" (about 50 Kd, about 446 amino acids), are similarly encoded by a variable region gene (encoding about 116 amino acids) and one of the other

aforementioned constant region genes, e.g., gamma (encoding about 330 amino acids).

One form of immunoglobulin constitutes the basic structural unit of an antibody. This form is a tetramer and consists of two identical pairs of immunoglobulin chains, each pair having one light and one heavy chain. In each pair, the light and heavy chain variable regions are together responsible for binding to an antigen, and the constant regions are responsible for the antibody effector functions. In addition to antibodies, immunoglobulins may exist in a variety of other forms (including less than full-length that retain the desired activities), including, for example, Fv, Fab, and F(ab')2, as well as single chain antibodies (e.g., Huston et al., Proc. Nat. Acad. Sci. U.S.A., 85:5879-5883 (1988) and Bird et al., Science, 242:423-426 (1988), which are incorporated herein by reference). (See, generally, Hood et al., "Immunology", Benjamin, N.Y., 2nd ed. (1984), and Hunkapiller and Hood, Nature, 323:15-16 (1986), which are incorporated herein by reference).

An immunoglobulin light or heavy chain variable region consists of a "framework" region interrupted by three hypervariable regions, also called CDR's. The extent of the framework region and CDR's have been precisely defined (see, "Sequences of Proteins of Immunological Interest," E. Kabat et al., U.S. Department of Health and Human Services, (1983); which is incorporated herein by reference). The sequences of the framework regions of different light or heavy chains are relatively conserved within a species. The framework region of an antibody, that is the combined framework regions of the constituent light and heavy chains, serves to position and align the CDR's. The CDR's are primarily responsible for binding to an epitope of an antigen.

Chimeric antibodies are antibodies whose light and heavy chain genes have been constructed, typically by genetic engineering, from immunoglobulin variable and constant region genes belonging to different species. For example, the variable segments of the genes from a mouse monoclonal antibody may be joined to human constant segments, such as

gamma 1 and gamma 3. A typical therapeutic chimeric antibody is thus a hybrid protein composed of the variable or antigen-binding domain from a mouse antibody and the constant or effector domain from a human antibody (e.g., A.T.C.C. Accession No. CRL 9688 secretes an anti-Tac chimeric antibody), although other mammalian species may be used.

As used herein, the term "humanized" immunoglobulin refers to an immunoglobulin comprising a substantially human framework region and one or more CDR's from a non-human (usually a mouse or rat) immunoglobulin. The non-human immunoglobulin providing the CDR's is called the "donor" and a human immunoglobulin providing the framework is called the "acceptor". Constant regions need not be present, but if they are, they must be substantially homologous to human immunoglobulin constant regions, i.e., at least about 85-90%, preferably about 95% or more identical. Hence, all parts of a humanized immunoglobulin, except possibly the CDR's, are substantially homologous to corresponding parts of natural human immunoglobulin sequences. A "humanized antibody" is an antibody comprising a humanized light chain and a humanized heavy chain immunoglobulin. For example, a humanized antibody would not encompass a typical chimeric antibody as defined above, e.g., because the entire variable region of a chimeric antibody is non-human. One says that the donor antibody has been "humanized", by the process of "humanization", because the resultant humanized antibody is expected to bind to the same antigen as the donor antibody that provides the CDR's.

Humanized immunoglobulins, including humanized antibodies, have been constructed by means of genetic engineering. Most humanized immunoglobulins that have been previously described (Jones et al., op. cit.; Verhoeyen et al., op. cit.; Riechmann et al., op. cit.) have comprised a framework that is identical to the framework of a particular human immunoglobulin chain, the acceptor, and three CDR's from a non-human donor immunoglobulin chain. In one case (Riechmann et al., op. cit.), two additional amino acids in the framework were changed to be the same as amino acids in

other human framework regions. The present invention includes criteria by which a limited number of amino acids in the framework of a humanized immunoglobulin chain are chosen to be the same as the amino acids at those positions in the donor rather than in the acceptor, in order to increase the affinity of an antibody comprising the humanized immunoglobulin chain.

The present invention is based in part on the model that two contributing causes of the loss of affinity in prior means of producing humanized antibodies (using as examples mouse antibodies as the source of CDR's) are:

(1) When the mouse CDR's are combined with the human framework, the amino acids in the framework close to the CDR's become human instead of mouse. Without intending to be bound by theory, we believe that these changed amino acids may slightly distort the CDR's, because they create different electrostatic or hydrophobic forces than in the donor mouse antibody, and the distorted CDR's may not make as effective contacts with the antigen as the CDR's did in the donor antibody;

(2) Also, amino acids in the original mouse antibody that are close to, but not part of, the CDR's (i.e., still part of the framework), may make contacts with the antigen that contribute to affinity. These amino acids are lost when the antibody is humanized, because all framework amino acids are made human.

To avoid these problems, and to produce humanized antibodies that have a very strong affinity for a desired antigen, the present invention uses the following four criteria for designing humanized immunoglobulins. These criteria may be used singly, or when necessary in combination, to achieve the desired affinity or other characteristics.

Criterion I: As acceptor, use a framework from a particular human immunoglobulin that is unusually homologous to the donor immunoglobulin to be humanized, or use a consensus framework from many human antibodies. For example,

comparison of the sequence of a mouse heavy (or light) chain variable region against human heavy (or light) variable regions in a data bank (for example, the National Biomedical Research Foundation Protein Identification Resource) shows that the extent of homology to different human regions varies greatly, typically from about 40% to about 60-70%. By choosing as the acceptor immunoglobulin one of the human heavy (respectively light) chain variable regions that is most homologous to the heavy (respectively light) chain variable region of the donor immunoglobulin, fewer amino acids will be changed in going from the donor immunoglobulin to the humanized immunoglobulin. Hence, and again without intending to be bound by theory, it is believed that there is a smaller chance of changing an amino acid near the CDR's that distorts their conformation. Moreover, the precise overall shape of a humanized antibody comprising the humanized immunoglobulin chain may more closely resemble the shape of the donor antibody, also reducing the chance of distorting the CDR's.

Typically, one of the 3-5 most homologous heavy chain variable region sequences in a representative collection of at least about 10 to 20 distinct human heavy chains will be chosen as acceptor to provide the heavy chain framework, and similarly for the light chain. Preferably, one of the 1-3 most homologous variable regions will be used. The selected acceptor immunoglobulin chain will most preferably have at least about 65% homology in the framework region to the donor immunoglobulin.

Regardless of how the acceptor immunoglobulin is chosen, higher affinity may be achieved by selecting a small number of amino acids in the framework of the humanized immunoglobulin chain to be the same as the amino acids at those positions in the donor rather than in the acceptor. The following criteria define what amino acids may be so selected. Preferably, at most or all amino acid positions satisfying one of these criteria, the donor amino acid will in fact be selected.

Criterion II: If an amino acid in the framework of the human acceptor immunoglobulin is unusual (i.e., "rare", which as used herein indicates an amino acid occurring at that position in no more than about 10% of human heavy (respectively light) chain V region sequences in a representative data bank), and if the donor amino acid at that position is typical for human sequences (i.e., "common", which as used herein indicates an amino acid occurring in at least about 25% of sequences in a representative data bank), then the donor amino acid rather than the acceptor may be selected. This criterion helps ensure that an atypical amino acid in the human framework does not disrupt the antibody structure. Moreover, by replacing an unusual amino acid with an amino acid from the donor antibody that happens to be typical for human antibodies, the humanized antibody may be made less immunogenic.

Criterion III: In the positions immediately adjacent to the 3 CDR's in the humanized immunoglobulin chain, the donor amino acid rather than acceptor amino acid may be selected. These amino acids are particularly likely to interact with the amino acids in the CDR's and, if chosen from the acceptor, distort the donor CDR's and reduce affinity. Moreover, the adjacent amino acids may interact directly with the antigen (Amit et al., Science, 233, 747-753 (1986), which is incorporated herein by reference) and selecting these amino acids from the donor may be desirable to keep all the antigen contacts that provide affinity in the original antibody.

Criterion IV: A 3-dimensional model, typically of the original donor antibody, shows that certain amino acids outside of the CDR's are close to the CDR's and have a good probability of interacting with amino acids in the CDR's by hydrogen bonding, Van der Waals forces, hydrophobic interactions, etc. At those amino acid positions, the donor amino acid rather than the acceptor immunoglobulin amino acid may be selected. Amino acids according to this criterion will

generally have a side chain atom within about 3 angstrom units of some site in the CDR's and must contain atoms that could interact with the CDR atoms according to established chemical forces, such as those listed above. Computer programs to create models of proteins such as antibodies are generally available and well known to those skilled in the art (see, Loew et al., Int. J. Quant. Chem., Quant. Biol. Symp., 15:55-66 (1988); Bruccoleri et al., Nature, 335, 564-568 (1988); Chothia et al., Science, 233:755-758 (1986), all of which are incorporated herein by reference). These do not form part of the invention. Indeed, because all antibodies have similar structures, the known antibody structures, which are available from the Brookhaven Protein Data Bank, can be used if necessary as rough models of other antibodies. Commercially available computer programs can be used to display these models on a computer monitor, to calculate the distance between atoms, and to estimate the likelihood of different amino acids interacting (see, Ferrin et al., J. Mol. Graphics, 6:13-27 (1988)).

Humanized antibodies generally have at least three potential advantages over mouse or in some cases chimeric antibodies for use in human therapy:

1) Because the effector portion is human, it may interact better with the other parts of the human immune system (e.g., destroy the target cells more efficiently by complement- dependent cytotoxicity (CDC) or antibody-dependent cellular cytotoxicity (ADCC)).

2) The human immune system should not recognize the framework or constant region of the humanized antibody as foreign, and therefore the antibody response against such an injected antibody should be less than against a totally foreign mouse antibody or a partially foreign chimeric antibody.

3) Injected mouse antibodies have been reported to have a half-life in the human circulation much shorter than the half-life of normal antibodies (D. Shaw et al., J. Immunol., 138:4534-4538 (1987)). Injected humanized antibodies will presumably have a half-life more similar to

naturally occurring human antibodies, allowing smaller and less frequent doses to be given.

In one aspect, the present invention is directed to designing humanized immunoglobulins that are produced by expressing recombinant DNA segments encoding the heavy and light chain CDR's from a donor immunoglobulin capable of binding to a desired antigen, such as the human IL-2 receptor, to DNA segments encoding acceptor human framework regions. Exemplary DNA sequences designed in accordance with the present invention and, which on expression code for the polypeptide chains comprising heavy and light chain CDR's with substantially human framework regions, are shown in Figures 3 and 4, respectively. Due to codon degeneracy and non-critical amino acid substitutions, other DNA sequences can be readily substituted for those sequences, as detailed below. In general, the criteria of the present invention find applicability to designing substantially any humanized immunoglobulin.

The DNA segments will typically further include an expression control DNA sequence operably linked to the humanized immunoglobulin coding sequences, including naturally-associated or heterologous promoter regions. Preferably, the expression control sequences will be eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells, but control sequences for prokaryotic hosts may also be used. Once the vector has been incorporated into the appropriate host, the host is maintained under conditions suitable for high level expression of the nucleotide sequences, and, as desired, the collection and purification of the humanized light chains, heavy chains, light/heavy chain dimers or intact antibodies, binding fragments or other immunoglobulin forms may follow (see, S. Beychok, Cells of Immunoglobulin Synthesis, Academic Press, N.Y., (1979), which is incorporated herein by reference.

Human constant region DNA sequences can be isolated in accordance with well known procedures from a variety of human cells, but preferably immortalized B-cells (see, Kabat

op. cit. and WP87/02671). The CDR's for producing the immunoglobulins of the present invention will be similarly derived from monoclonal antibodies capable of binding to the predetermined antigen, such as the human IL-2 receptor, and produced by well known methods in any convenient mammalian source including, mice, rats, rabbits, or other vertebrate capable of producing antibodies. Suitable source cells for the constant region and framework DNA sequences, and host cells for immunoglobulin expression and secretion, can be obtained from a number of sources, such as the American Type Culture Collection ("Catalogue of Cell Lines and Hybridomas," Fifth edition (1985) Rockville, Maryland, U.S.A., which is incorporated herein by reference).

In addition to the humanized immunoglobulins specifically described herein, other "substantially homologous" modified immunoglobulins to the native sequences can be readily designed and manufactured utilizing various recombinant DNA techniques well known to those skilled in the art. For example, the framework regions can vary specifically from the sequences in Figures 3 and 4 at the primary structure level by several amino acid substitutions, terminal and intermediate additions and deletions, and the like. Moreover, a variety of different human framework regions may be used singly or in combination as a basis for the humanized immunoglobulins of the present invention. In general, modifications of the genes may be readily accomplished by a variety of well-known techniques, such as site-directed mutagenesis (see, Gillman and Smith, Gene, 8:81-97 (1979) and S. Roberts et al., Nature, 328:731-734 (1987), both of which are incorporated herein by reference).

Alternatively, polypeptide fragments comprising only a portion of the primary antibody structure may be produced, which fragments possess one or more immunoglobulin activities (e.g., complement fixation activity). Also because like many genes, the immunoglobulin-related genes contain separate functional regions, each having one or more distinct biological activities, the genes may be fused to functional regions from other genes (e.g., enzymes, see,

commonly assigned U.S.S.N. 132,387, filed Dec. 15, 1987, which is incorporated herein by reference) to produce fusion proteins (e.g., immunotoxins) having novel properties. The nucleic acid sequences of the present invention capable of ultimately expressing the desired humanized antibodies can be formed from a variety of different polynucleotides (genomic or cDNA, RNA, synthetic oligonucleotides, etc.) and components (e.g., V, J, D, and C regions), as well as by a variety of different techniques. Joining appropriate genomic sequences is presently the most common method of production, but cDNA sequences may also be utilized (see, European Patent Publication No. 0239400 and L. Reichmann et al., Nature, 332:323-327 (1988), both of which are incorporated herein by reference).

As stated previously, the DNA sequences will be expressed in hosts after the sequences have been operably linked to (i.e., positioned to ensure the functioning of) an expression control sequence. These expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosomal DNA. Commonly, expression vectors will contain selection markers, e.g., tetracycline or neomycin, to permit detection of those cells transformed with the desired DNA sequences (see, e.g., U.S. Patent 4,704,362, which is incorporated herein by reference).

E. coli is one prokaryotic host useful particularly for cloning the DNA sequences of the present invention. Other microbial hosts suitable for use include bacilli, such as Bacillus subtilus, and other enterobacteriaceae, such as Salmonella, Serratia, and various Pseudomonas species. In these prokaryotic hosts, one can also make expression vectors, which will typically contain expression control sequences compatible with the host cell (e.g., an origin of replication). In addition, any number of a variety of well-known promoters will be present, such as the lactose promoter system, a tryptophan (trp) promoter system, a beta-lactamase promoter system, or a promoter system from phage lambda. The promoters will typically control expression, optionally with an operator sequence, and have ribosome binding site

sequences and the like, for initiating and completing transcription and translation.

Other microbes, such as yeast, may also be used for expression. Saccharomyces is a preferred host, with suitable vectors having expression control sequences, such as promoters, including 3-phosphoglycerate kinase or other glycolytic enzymes, and an origin of replication, termination sequences and the like as desired.

In addition to microorganisms, mammalian tissue cell culture may also be used to express and produce the polypeptides of the present invention (see, Winnacker, "From Genes to Clones," VCH Publishers, N.Y., N.Y. (1987), which is incorporated herein by reference). Eukaryotic cells are actually preferred, because a number of suitable host cell lines capable of secreting intact immunoglobulins have been developed in the art, and include the CHO cell lines, various COS cell lines, HeLa cells, myeloma cell lines, etc, but preferably transformed B-cells or hybridomas. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter, an enhancer (Queen et al., Immunol. Rev., 89:49-68 (1986), which is incorporated herein by reference), and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites, and transcriptional terminator sequences. Preferred expression control sequences are promoters derived from immunoglobulin genes, SV40, Adenovirus, Bovine Papilloma Virus, and the like.

The vectors containing the DNA segments of interest (e.g., the heavy and light chain encoding sequences and expression control sequences) can be transferred into the host cell by well-known methods, which vary depending on the type of cellular host. For example, calcium chloride transfection is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment or electroporation may be used for other cellular hosts. (See, generally, Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, (1982), which is incorporated herein by reference.)

Once expressed, the whole antibodies, their dimers, individual light and heavy chains, or other immunoglobulin forms of the present invention, can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis and the like (see, generally, R. Scopes, "Protein Purification", Springer-Verlag, N.Y. (1982)). Substantially pure immunoglobulins of at least about 90 to 95% homogeneity are preferred, and 98 to 99% or more homogeneity most preferred, for pharmaceutical uses. Once purified, partially or to homogeneity as desired, the polypeptides may then be used therapeutically (including extracorporeally) or in developing and performing assay procedures, immunofluorescent stainings, and the like. (See, generally, Immunological Methods, Vols. I and II, Lefkovits and Pernis, eds., Academic Press, New York, N.Y. (1979 and 1981)).

The antibodies of the present invention will typically find use individually in treating substantially any disease susceptible to monoclonal antibody-based therapy. In particular, the immunoglobulins can be used for passive immunization or the removal of unwanted cells or antigens, such as by complement mediated lysis, all without substantial immune reactions (e.g., anaphylactic shock) associated with many prior antibodies. For example, where the cell linked to a disease has been identified as IL-2 receptor bearing, then humanized antibodies that bind to the human IL-2 receptor are suitable (see, U.S.S.N. 085,707, entitled "Treating Human Malignancies and Disorders," which is incorporated herein by reference). For such a humanized immunoglobulin, typical disease states suitable for treatment include graft versus host disease and transplant rejection in patients undergoing an organ transplant, such as heart, lungs, kidneys, liver, etc. Other diseases include autoimmune diseases, such as Type I diabetes, multiple sclerosis, rheumatoid arthritis, systemic lupus erythematosus, and myasthenia gravis.

The method of producing humanized antibodies of the present invention can be used to humanize a variety of donor

antibodies, especially monoclonal antibodies reactive with markers on cells responsible for a disease. For example, suitable antibodies bind to antigens on T-cells, such as those grouped into the so-called "Clusters of Differentiation," as named by the First International Leukocyte Differentiation Workshop, Leukocyte Typing, Bernard et al., Eds., Springer- Verlag, N.Y. (1984), which is incorporated herein by reference.

The antibodies of the present invention can also be used as separately administered compositions given in conjunction with chemotherapeutic or immunosuppressive agents. Typically, the agents will include cyclosporin A or a purine analog (e.g., methotrexate, 6-mercaptopurine, or the like), but numerous additional agents (e.g., cyclophosphamide, prednisone, etc.) well-known to those skilled in the art may also be utilized.

A preferred pharmaceutical composition of the present invention comprises the use of the subject antibodies in immunotoxins. Immunotoxins are characterized by two components and are particularly useful for killing selected cells in vitro or in vivo. One component is a cytotoxic agent which is usually fatal to a cell when attached or absorbed. The second component, known as the "delivery vehicle," provides a means for delivering the toxic agent to a particular cell type, such as cells comprising a carcinoma. The two components are commonly chemically bonded together by any of a variety of well-known chemical procedures. For example, when the cytotoxic agent is a protein and the second component is an intact immunoglobulin, the linkage may be by way of heterobifunctional cross-linkers, e.g., SPDP, carbodiimide, glutaraldehyde, or the like. Production of various immunotoxins is well-known with the art, and can be found, for example in "Monoclonal Antibody-Toxin Conjugates: Aiming the Magic Bullet," Thorpe et al., Monoclonal Antibodies in Clinical Medicine, Academic Press, pp. 168-190 (1982), which is incorporated herein by reference.

A variety of cytotoxic agents are suitable for use in immunotoxins. Cytotoxic agents can include radionuclides,

such as Iodine-131, Yttrium-90, Rhenium-188, and Bismuth-212; a number of chemotherapeutic drugs, such as vindesine, methotrexate, adriamycin, and cisplatinm; and cytotoxic proteins such as ribosomal inhibiting proteins like pokeweed antiviral protein, Pseudomonas exotoxin A, ricin, diphtheria toxin, ricin A chain, etc., or an agent active at the cell surface, such as the phospholipase enzymes (e.g., phospholipase C). (See, generally, commonly assigned U.S.S.N. 290,968 (Townsend and Townsend Docket No. 11823-7-2) filed in U.S.P.T.O. on December 28, 1988, "Chimeric Toxins," Olsnes and Phil, Pharmac. Ther., 25:355-381 (1982), and "Monoclonal Antibodies for Cancer Detection and Therapy," eds. Baldwin and Byers, pp. 159-179, 224-266, Academic Press (1985), all of which are incorporated herein by reference.)

The delivery component of the immunotoxin will include the humanized immunoglobulins of the present invention. Intact immunoglobulins or their binding fragments, such as Fab, are preferably used. Typically, the antibodies in the immunotoxins will be of the human IgM or IgG isotype, but other mammalian constant regions may be utilized as desired.

For diagnostic purposes, the antibodies may either be labeled or unlabeled. Unlabeled antibodies can be used in combination with other labeled antibodies (second antibodies) that are reactive with the humanized antibody, such as antibodies specific for human immunoglobulin constant regions. Alternatively, the antibodies can be directly labeled. A wide variety of labels may be employed, such as radionuclides, fluors, enzymes, enzyme substrates, enzyme co-factors, enzyme inhibitors, ligands (particularly haptens), etc.. Numerous types of immunoassays are available and are well known to those skilled in the art.

The following examples are offered by way of illustration, not by limitation.

EXPERIMENTAL

Design of genes for humanized light and heavy chains

The sequence of the human antibody Eu (Sequences of Proteins of Immunological Interest, E. Kabat et al., U.S. Dept. of Health and Human Services, 1983) was used to provide the framework of the humanized antibody, because the amino acid sequence of the heavy chain variable region of anti-Tac is more homologous to the heavy chain of this antibody than to any other complete heavy chain variable region sequence in the National Biomedical Foundation Protein Identification Resource.

To select the sequence of the humanized heavy chain, the anti-Tac heavy chain sequence (see, commonly assigned U.S.S.N.'s 186,862 and 223,037, which are incorporated herein by reference) was aligned with the sequence of the Eu heavy chain (Figure 1). At each position, the Eu amino acid was selected for the humanized sequence, unless that position fell in any one of the following categories, in which case the anti-Tac amino acid was selected:

(1) The position fell within a complementarity determining region (CDR), as defined by Kabat, et al., op. cit. (amino acids 31-35, 50-66, 99-106);

(2) The Eu amino acid was rare for human heavy chains at that position, whereas the anti-Tac amino acid was common for human heavy chains at that position (amino acids 27, 93, 95, 98, 107-109, 111);

(3) The position was immediately adjacent to a CDR in the amino acid sequence of the anti-Tac heavy chain (amino acids 30 and 67); or

(4) 3-dimensional modeling of the anti-Tac antibody suggested that the amino acid was physically close to the antigen binding region (amino acids 48 and 68).

Amino acid #27 is listed in category (4) because the acceptor Eu amino acid Gly is rare, and the donor anti-Tac amino acid Tyr is chemically similar to the amino acid Phe, which is common, but the substitution was actually made because #27

also fell in category (4). Although some amino acids fell in more than one of these categories, they are only listed in one. Categories (2) - (4) correspond to criteria (2) - (4) described above.

To select the sequence of the humanized light chain, the anti-Tac light chain sequence was aligned with the sequence of the Eu light chain (Figure 2). The Eu amino acid was selected at each position, unless the position again fell into one of the categories (1) - (4) (with light chain replacing heavy chain in the category definitions):

(1) CDR's (amino acids 24-34, 50-56, 89-97);

(2) Anti-Tac amino acid more typical than Eu (amino acids 48 and 63);

(3) Adjacent to CDR's (no amino acids; Eu and anti-Tac were already the same at all these positions); or

(4) Possible 3-dimensional proximity to binding region (amino acid 60).

The actual nucleotide sequence of the heavy (Figure 3) and light chain (Figure 4) genes were selected as follows:

(1) The nucleotide sequences code for the amino acid sequences chosen as described above;

(2) 5' of these coding sequences, the nucleotide sequences code for a leader (signal) sequence, namely the leader of the light chain of the antibody MOPC 63 and the leader of the heavy chain of the antibody PCH 108A (Kabat et al., op. cit.). These leader sequences were chosen as typical of antibodies;

(3) 3' of the coding sequences, the nucleotide sequences are the sequences that follow the mouse light chain J5 segment and the mouse heavy chain J2 segment, which are part of the anti-Tac sequences. These sequences are included because they contain splice donor signals; and

(4) At each end of the sequence is an Xba I site to allow cutting at the Xba I sites and cloning into the Xba I site of a vector.

Construction of humanized light and heavy chain genes

To synthesize the heavy chain, four oligonucleotides HES12, HES13, HES14, HES15 (Figure 5A) were synthesized using an Applied Biosystems 380B DNA synthesizer. Two of the oligonucleotides are part of each strand of the heavy chain, and each oligonucleotide overlaps the next one by about 20 nucleotides to allow annealing (Figure 5B). Together, the oligonucleotides cover the entire humanized heavy chain variable region (Figure 3) with a few extra nucleotides at each end to allow cutting at the Xba I sites. The oligonucleotides were purified from polyacrylamide gels.

Each oligonucleotide was phosphorylated using ATP and T4 polynucleotide kinase by standard procedures (see, Maniatis, op. cit.). To anneal the phosphorylated oligonucleotides, they were suspended together in 40 ul of TA (33 mM Tris acetate, pH 7.9, 66 mM potassium acetate, 10 mM magnesium acetate) at a concentration of about 3.75 uM each, heated to 95 deg for 4 min. and cooled slowly to 4 deg. To synthesize the complete gene from the oligonucleotides by synthesizing the opposite strand of each oligonucleotide (Figure 5B), the following components were added in a final volume of 100ul:

10 ul	annealed oligonucleotides
0.16 mM each	deoxyribonucleotide
0.5 mM	ATP
0.5 mM	DTT
100 ug/ml	BSA
3.5 ug/ml	T4 g43 protein (DNA polymerase)
25 ug/ml	T4 g44/62 protein (polymerase accessory protein)
25 ug/ml	45 protein (polymerase accessory protein)

The mixture was incubated at 37 deg for 30 min. Then 10 u of T4 DNA ligase was added and incubation at 37 deg resumed for 30 min. The polymerase and ligase were inactivated by incubation of the reaction at 70 deg for

15 min. To digest the gene with Xba I, to the reaction was added 50 ul of 2x TA containing BSA at 200 ug/ml and DTT at 1 mM, 43 ul of water, and 50 u of Xba I in 5 ul. The reaction was incubated for 3 hr at 37 deg, and run on a gel. The 431 bp Xba I fragment was purified from a gel and cloned into the Xba I site of the plasmid pUC19 by standard methods. Four plasmid isolates were purified and sequenced using the dideoxy method. One of these had the correct sequence (Figure 3).

To synthesize the light chain, four oligonucleotides JFD1, JFD2, JFD3, JFD4 (Figure 6A) were synthesized. Two of the oligonucleotides are part of each strand of the light chain, and each oligonucleotide overlaps the next one by about 20 nucleotides to allow annealing (Figure 6B). Together, the oligonucleotides cover the entire humanized light chain variable region (Figure 4) with a few extra nuclectides at each end to allow cutting at the Xba I sites. The oligonucleotides were purified from polyacrylamide gels.

The light chain gene was synthesized from these oligonuclectides in two parts. 0.5 ug each of JFD1 and JFD2 were combined in 20 ul sequenase buffer (40 mM Tris-HCl, pH 7.5, 20 mM magnesium chloride, 50 mM sodium chloride), heated at 70 deg for 3 min and allowed to cool slowly tc 23 deg in order for the oligonucleotides to anneal. JFD3 and JFD4 were treated in the same way. Each reaction was made 10 mM in DTT and 0.5 mM in each deoxyribonucleotide and 6.5 u of sequenase (US Biochemicals) was added, in a final volume of 24 ul, and incubated for 1 hr at 37 deg to synthesize the opposite strands of the oligonucleotides. Xba I and Hind III were added to each reaction to digest the DNA (there is a Hind III site in the region where JFD2 and JFD3 overlap and therefore in each of the synthesized DNAs; Figure 6B). The reactions were run on polyacrylamide gels, and the Xba I - Hind III fragments were purified and cloned into pUC18 by standard methods. Several plasmid isolates for each fragment were sequenced by the dideoxy method, and correct ones chosen.

Construction of plasmids to express humanized light and heavy chains

The heavy chain Xba I fragment was isolated from the pUC19 plasmid in which it had been inserted and then inserted into the Xba I site of the vector pVγ1 (see, commonly assigned U.S.S.N. 223,037) in the correct orientation by standard methods, to produce the plasmid pHuGTAC1 (Figure 7). This plasmid will express high levels of a complete heavy chain when transfected into an appropriate host cell.

The two light chain Xba I - Hind III fragments were isolated from the pUC18 plasmids in which they had been inserted. The vector plasmid pVκ1 (see, commonly assigned U.S.S.N. 223,037) was cut with Xba I, dephosphorylated and ligated with the two fragments by standard methods. The desired reaction product has the circular form: vector - Xba I - fragment 1 - Hind III - fragment 2 - Xba I - vector. Several plasmid isolates were analyzed by restriction mapping and sequencing, and one with this form chosen. This plasmid, pHuLTAC (Figure 8), therefore contains the complete humanized light chain (Figure 4) and will express high levels of the light chain when transfected into an appropriate host cell.

Synthesis and affinity of humanized antibody

The plasmids pHuGTAC1 and pHuLTAC were transfected into mouse Sp2/0 cells, and cells that integrated the plasmids were selected on the basis of resistance to mycophenolic acid and/or hygromycin B conferred by the gpt and hyg genes on the plasmids (Figures 7,8) by standard methods. To verify that these cells secreted antibody that binds to the IL-2 receptor, supernatant from the cells was incubated with HUT-102 cells that are known to express the IL-2 receptor. After washing, the cells were incubated with fluorescein-conjugated goat anti-human antibody, washed, and analyzed for fluorescence on a FACSCAN cytofluorometer. The results (Figure 9A), clearly show that the humanized antibody binds to these cells, but not to Jurkat T-cells that do not express the IL-2 receptor (Figure 9D). As controls, the

original mouse anti-Tac antibody was also used to stain these cells (Figure 9B,C), giving similar results.

For further experiments, cells producing the humanized antibody were injected into mice, and the resultant ascites collected. Humanized antibody was purified to substantial homogeneity from the ascites by passage through an affinity column of goat anti-human immunoglobulin antibody, prepared on an Affigel-10 support (Bio-Rad Laboratories, Inc., Richmond, CA) according to standard techniques. To determine the affinity of the humanized antibody relative to the original anti-Tac antibody, a competitive binding experiment was performed. About 5×10^5 HUT-102 cells were incubated with known quantities (10 - 40 ng) of the anti-Tac antibody and the humanized anti-Tac antibody for 10 min at 4 deg. Then 100 ng of biotinylated anti-Tac was added to the cells and incubated for 30 min at 4 deg. This quantity of anti-Tac had previously been determined to be sufficient to saturate the binding sites on the cells, but not to be in large excess. Then the cells were washed twice with 2 ml of phosphate buffered saline (PBS) containing 0.1% sodium azide. The cells were then incubated for 30 min at 4 deg with 250 ng of phycoerythrin-conjugated avidin, which bound to the biotinylated anti-Tac already bound to the cells. The cells were washed again as above, fixed in PBS containing 1% paraformaldehyde, and analyzed for fluorescence on a FACSCAN cytofluorometer.

Use of increasing amounts (10 - 40 ng) of the anti-Tac antibody as competitor in the first step decreased the amount of biotinylated anti-Tac that could bind to the cells in the second step, and therefore the amount of phycoerythrin-conjugated avidin that bound in the last step, thus decreasing fluorescence (Figure 10A). Equivalent amounts (20 ng) of anti-Tac, and humanized anti-Tac used as competitor decreased the fluorescence to approximately the same degree (Figure 10B). This shows that these antibodies have approximately the same affinity, because if one had greater affinity, it would have more effectively competed

with the bictinylated anti-Tac, thus decreasing fluorescence more.

Biological properties of the humanized antibody

For optimal use in treatment of human disease, the humanized antibody should be able to destroy T-cells in the body that express the IL-2 receptor. One mechanism by which antibodies may destroy target cells is antibody-dependent cell-mediated cytotoxicity, abbreviated ADCC (Fundamental Immunology, Paul, W., Ed., Raven Press, New York (1984), at pg. 681), in which the antibody forms a bridge between the target cell and an effector cell such as a macrophage that can lyse the target. To determine whether the humanized antibody and the original mouse anti-Tac antibody can mediate ADCC, a chromium release assay was performed by standard methods. Specifically, human leukemia HUT-102 cells, which express the IL-2 receptor, were incubated with ^{51}Cr to allow them to absorb this radionuclide. The HUT-102 cells were then incubated with an excess of either anti-Tac or humanized anti-Tac antibody. The HUT-102 cells were next incubated for 4 hrs with either a 30:1 or 100:1 ratio of effector cells, which were normal purified human peripheral blood mononuclear cells that had been activated by incubation for about 20 hrs with human recombinant IL-2. Release of ^{51}Cr, which indicated lysis of the target HUT-102 cells, was measured and the background subtracted (Table 1). The results show that at either ratio of effector cells, anti-Tac did not lyse a significant number of the target cells (less than 5%), while the humanized antibody did (more than 20%). Hence, the humanized antibody is likely to be more efficacious than the original mouse antibody in treating T-cell leukemia or other T-cell mediated diseases.

TABLE 1

Percent ^{51}Cr release after ADCC

	Effector: Target ratio	
	30:1	100:1
Antibody		
Anti-Tac	4%	< 1%
Humanized anti-Tac	24%	23%

Although the present invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

WE CLAIM:

1. A method of designing a humanized immunoglobulin (Ig) chain having one or more complementarity determining regions (CDR's) from a donor Ig and a framework region from a human Ig, said method comprising: comparing the framework or variable region amino acid sequence of the donor Ig with corresponding sequences in a collection of human Ig chains; and selecting to provide the human Ig framework one of the about three most homologous sequences from the collection.

2. A method according to Claim 1, wherein the human Ig sequence is selected from a collection of at least about ten to twenty Ig chain sequences.

3. A method according to Claim 1, wherein the human Ig chain sequence selected has the highest homology in the collection to the donor Ig sequence.

4. A method according to Claim 1, wherein the human Ig framework sequence selected is at least about 65% homologous to the donor Ig framework sequence.

5. A method according to Claim 1, wherein the immunoglobulin chain is a heavy chain.

6. A method according to Claim 1, wherein the humanized Ig chain comprises a human constant region.

7. An immunoglobulin comprising two light/heavy chain pairs, wherein at least one chain is designed in accordance with Claim 1.

8. A method of designing a humanized immunoglobulin chain having a framework region from a human acceptor immunoglobulin and complementarity determining regions (CDR's) from a donor immunoglobulin capable of binding to an antigen, said method comprising the steps of substituting at least one human framework amino acid of the acceptor immunoglobulin with a corresponding amino acid from the donor immunoglobulin at a position in the immunoglobulins where:

(a) the amino acid in the human framework region of the acceptor immunoglobulin is rare for said position and the corresponding amino acid in the donor immunoglobulin is common for said position in human immunoglobulin sequences; or

(b) the amino acid is immediately adjacent to one of the CDR's; or

(c) the amino acid is predicted to have a side chain atom within about 3Å of the CDR's in a three-dimensional immunoglobulin model and to be capable of interacting with the antigen or with the CDR's of the humanized immunoglobulin.

9. A method according to Claim 8, wherein the humanized immunoglobulin chain comprises in addition to the CDR's at least three amino acids from the donor immunoglobulin chosen by criteria (a), (b) or (c).

10. A method according to Claim 9, wherein at least one of the amino acids substituted from the donor is immediately adjacent a CDR.

11. A method according to Claim 9, wherein said humanized immunoglobulin chain is a heavy chain.

12. An immunoglobulin comprising two light/heavy chain pairs, wherein at least one chain is designed in accordance with Claim 8.

13. An immunoglobulin according to Claim 12, which is specifically reactive with an antigen at an affinity of at least about 10^8 M^{-1} or stronger.

14. An immunoglobulin according to Claim 12, wherein the designed chain is a light chain comprising about 214 amino acids.

15. An immunoglobulin according to Claim 12, wherein the designed chain is a heavy chain comprising about 446 amino acids.

16. A DNA sequence which upon expression encodes a humanized immunoglobulin chain according to Claim 1 or Claim 8.

17. A method for improving the affinity of a humanized immunoglobulin (Ig) to an antigen, by replacing amino acids of the human Ig framework with amino acids from the donor Ig framework at positions where:

(a) the amino acid in the human framework region of the first immunoglobulin is rare for said position and the corresponding amino acid in the donor immunoglobulin is common for said position in human immunoglobulin sequences; or

(b) the amino acid is immediately adjacent to one of the CDR's; or

(c) the amino acid is predicted to have a side chain atom within about 3Å of the CDR's in a three-dimensional immunoglobulin model and to be capable of interacting with the antigen or the CDR's of the humanized immunoglobulin.

18. A method according to Claim 17, wherein the additional amino acids comprise up to three amino acids, each of which is immediately adjacent to one of the CDR's in the second Ig.

19. A method according to Claim 17, wherein the additional amino acids comprise one amino acid immediately adjacent to a CDR.

20. A method according to Claim 17, wherein the additional amino acids comprise at least two amino acids from the donor Ig which are predicted by modelling to be capable of interacting with the antigen or the CDR's.

21. A method according to Claim 20, wherein said two or more amino acids are predicted to be within about 3Å of the donor Ig CDR's.

22. A method according to Claim 17, wherein the humanized Ig has an affinity to the antigen within about 2 to 3 fold of the donor Ig.

23. A method according to Claim 17, wherein the antigen is a protein.

24. A method of producing a humanized immunoglobulin containing a heavy chain and a light chain designed in accordance with Claim 17, said method comprising:

culturing a host capable of expressing said heavy chain, said light chain, or both, under conditions suitable for production of said chains; and

recovering from the culture said humanized immunoglobulin.

25. A polynucleotide composition comprising a DNA sequence coding for a humanized immunoglobulin designed in accordance with Claim 17.

26. A method of producing an improved humanized immunoglobulin comprising expressing the polynucleotide composition of Claim 25.

27. A cell transformed with a polynucleotide composition according to Claim 25.

28. A composition comprising a humanized immunoglobulin secreted by a cell line according to Claim 24.

DESIGNING IMPROVED HUMANIZED IMMUNOGLOBULINS

ABSTRACT OF THE DISCLOSURE

Novel methods for designing humanized immunoglobulins having one or more complementarity determining regions (CDR's) from a donor immunoglobulin and a framework region from a human immunoglobulin comprising first comparing the framework or variable region amino acid sequence of the donor immunoglobulin to corresponding sequences in a collection of human immunoglobulin chains, and selecting as the human immunoglobulin one of the more homologous sequences from the collection. Each humanized immunoglobulin chain may comprise about 3 or more amino acids from the donor immunoglobulin in addition to the CDR's, usually at least one of which is immediately adjacent to a CDR in the donor immunoglobulin. The heavy and light chains may each be designed by using any one or all three additional position criteria. When combined into an intact antibody, the humanized immunoglobulins of the present invention will be substantially non-immunogenic in humans and retain substantially the same affinity as the donor immunoglobulin to the antigen, such as a protein or other compound containing an epitope.

11923-9
ATTORNEY DOCKET NO

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:
My residence, post office address and citizenship are as stated below next to my name. I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

DESIGNING IMPROVED HUMANIZED IMMUNOGLOBULINS

the specification of which ☒ is attached hereto or ☐ was filed on _____ as Application Serial No. _____ and was amended on _____ (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a). I claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

COUNTRY	APPLICATION NUMBER	DATE OF FILING	PRIORITY CLAIMED UNDER 35 U.S.C. 119
			Yes____ No____
			Yes____ No____

I claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION SERIAL NO.	DATE OF FILING	STATUS
290,975	December 28, 1988	☐ Patented ☒ Pending ☐ Abandoned
		☐ Patented ☐ Pending ☐ Abandoned

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) who are partners and associates in the firm of Townsend and Townsend to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

William M. Smith, Reg. No. 30,223
James M. Heslin, Reg. No. 29,541
Steve W. Parmelee, Reg. No. 31,990

SEND CORRESPONDENCE TO:	DIRECT TELEPHONE CALLS TO (name, registration number, and telephone number)
William M. Smith TOWNSEND and TOWNSEND Steuart Street Tower, One Market Plaza San Francisco, CA 94105	William M. Smith, Reg. 30,223 ☐ (415) 543-9600 or ☒ (415) 326-2400

201 FULL NAME OF INVENTOR	Last Name: Queen	First Name: Cary		Middle Name or Initial: L.
201 RESIDENCE & CITIZENSHIP	City: Palo Alto	State or Foreign Country: California		Country of Citizenship: USA
201 POST OFFICE ADDRESS	Post Office Address: 1300 Oak Creek Dr.	City: Palo Alto	State or Country: California	Zip Code: 94304
202 FULL NAME OF INVENTOR	Last Name: Selick	First Name: Harold		Middle Name or Initial: Edwin
202 RESIDENCE & CITIZENSHIP	City: Belmont	State or Foreign Country: California		Country of Citizenship: USA
202 POST OFFICE ADDRESS	Post Office Address: 1673 Sunnyslope Ave.	City: Belmont	State or Country: California	Zip Code: 94002
203 FULL NAME OF INVENTOR	Last Name:	First Name:		Middle Name or Initial:
203 RESIDENCE & CITIZENSHIP	City:	State or Foreign Country:		Country of Citizenship:
203 POST OFFICE ADDRESS	Post Office Address:	City:	State or Country:	Zip Code:

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 201	Signature of Inventor 202	Signature of Inventor 203
	Harold E. Selick	
Date: 2/1/89	Date: 2/1c/89	Date:

GT 14A 12 87

TOWNSEND AND TOWNSEND

Atty. Docket No. [illegible]

VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS
(37 CFR 1.9(f) and 1.27(c)) — SMALL BUSINESS CONCERN

Applicant or Patentee: Cary L. Queen and Harold Edwin Selick
Serial No.: Not yet assigned Filing Date: February 13, 1989
Patent No.: Issued:
For: DESIGNING IMPROVED HUMANIZED IMMUNOGLOBULINS

I hereby declare that I am

[ ] the owner of the small business concern identified below:
[X] an official of the small business concern empowered to act on behalf of the concern identified below

NAME OF CONCERN Protein Design Labs, Inc.
ADDRESS OF CONCERN 3181 Porter Drive
Palo Alto, California 94304

I hereby declare that the above identified small business concern qualifies as a small business concern as defined in 13 CFR 121.3-18, and reproduced in 37 CFR 1.9(d), for purposes of paying reduced fees under section 41(a) and (b) of Title 35, United States Code, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby declare that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention, entitled DESIGNING IMPROVED HUMANIZED IMMUNOGLOBULINS by (inventor(s)) Cary L. Queen and Harold Edwin Selick described in

[X] the application filed herewith
[ ] application serial no. , filed
[ ] patent no. , issued

If the rights held by the above identified small business concern are not exclusive, each individual, concern or organization having rights to the invention is listed below* and no rights to the invention are held by any person, other than the inventor, who could not qualify as a small business concern under 37 CFR 1.9(d) or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

*NOTE: Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR 1.27)

NAME
ADDRESS
[ ] INDIVIDUAL [ ] SMALL BUSINESS CONCERN [ ] NONPROFIT ORGANIZATION

NAME
ADDRESS
[ ] INDIVIDUAL [ ] SMALL BUSINESS CONCERN [ ] NONPROFIT ORGANIZATION

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING Shirley L. Clayton
TITLE OF PERSON OTHER THAN OWNER Chief Financial Officer
ADDRESS OF PERSON SIGNING Protein Design Labs, Inc., 3181 Porter Drive, Palo Alto, CA 94304

SIGNATURE Shirley L. Clayton DATE 2/13/89

PAT 19

FIGURE 1

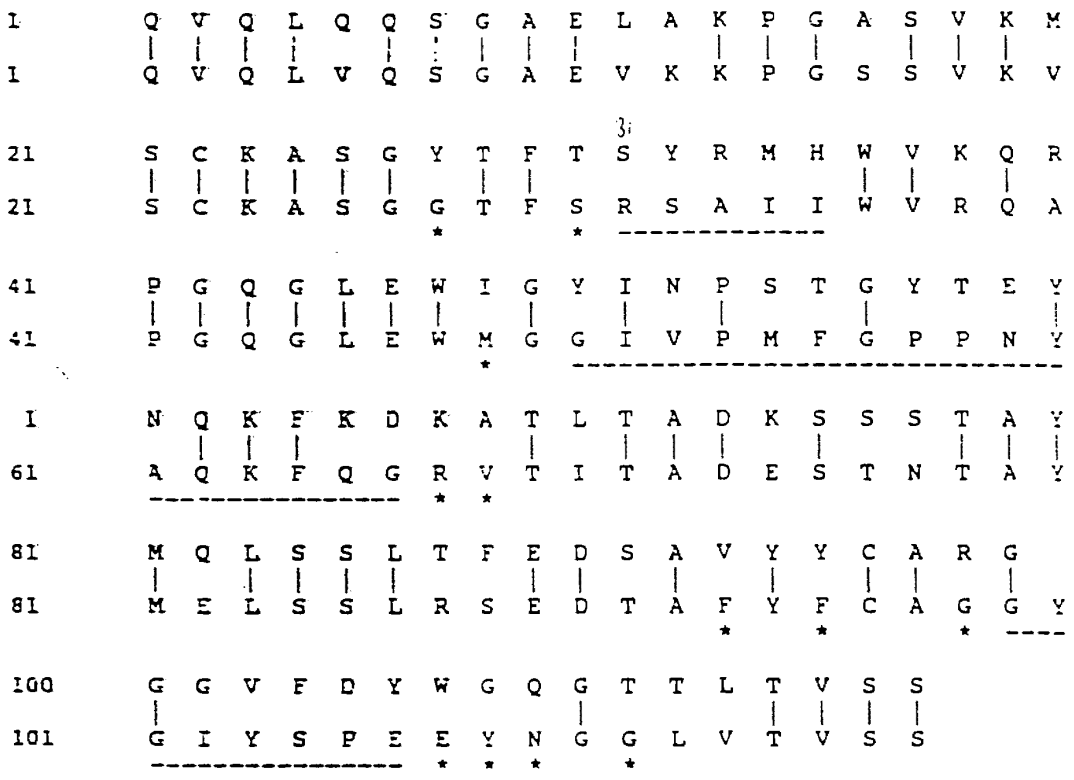

1 Q V Q L Q Q S G A E L A K P G A S V K M
1 Q V Q L V Q S G A E V K K P G S S V K V
21 S C K A S G Y T F T S Y R M H W V K Q R
21 S C K A S G G T F S R S A I I W V R Q A
41 P G Q G L E W I G Y I N P S T G Y T E Y
41 P G Q G L E W M G G I V P M F G P P N Y
I N Q K F K D K A T L T A D K S S S T A Y
61 A Q K F Q G R V T I T A D E S T N T A Y
81 M Q L S S L T F E D S A V Y Y C A R G
81 M E L S S L R S E D T A F Y F C A G G Y
100 G G V F D Y W G Q G T T L T V S S
101 G I Y S P E E Y N G G L V T V S S

FIGURE 2

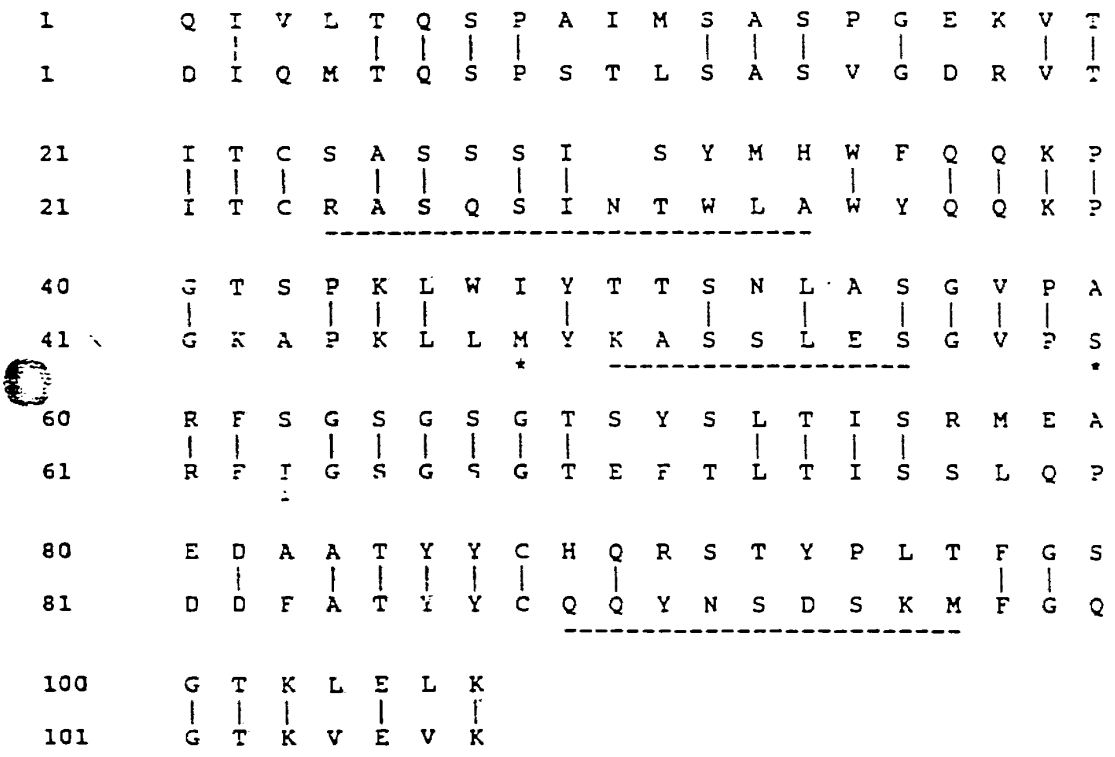

1 Q I V L T Q S P A I M S A S P G E K V T
1 D I Q M T Q S P S T L S A S V G D R V T
21 I T C S A S S S I S Y M H W F Q Q K P
21 I T C R A S Q S I N T W L A W Y Q Q K P
40 G T S P K L W I Y T T S N L A S G V P A
41 G K A P K L L M Y K A S S L E S G V P S
60 R F S G S G S G T S Y S L T I S R M E A
61 R F I G S G S G T E F T L T I S S L Q P
80 E D A A T Y Y C H Q R S T Y P L T F G S
81 D D F A T Y Y C Q Q Y N S D S K M F G Q
100 G T K L E L K
101 G T K V E V K

FIGURE 3

```
        10         20         30         40         50         60
TCTAGATGGGATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGTACCGCGGGCGTGCACT
      M  G  W  S  W  I  F  L  F  L  L  S  G  T  A  G  V  H

        70         80         90        100        110        120
CTCAGGTCCAGCTTGTCCAGTCTGGGGCTGAAGTCAAGAAACCTGGCTCGAGCGTGAAGG
S  Q  V  Q  L  V  Q  S  G  A  E  V  K  K  P  G  S  S  V  K

       130        140        150        160        170        180
TCTCCTGCAAGGCTTCTGGCTACACCTTTACTAGCTACAGGATGCACTGGGTAAGGCAGG
V  S  C  K  A  S  G  Y  T  F  T  S  Y  R  M  H  W  V  R  Q

       190        200        210        220        230        240
CCCCTGGACAGGGTCTGGAATGGATTGGATATATT. ATCCGTCGACTGGGTATACTGAAT
A  P  G  Q  G  L  E  W  I  G  Y  I  N  P  S  T  G  Y  T  E

       250        260        270        280        290        300
ACAATCACAAGTTCAAGGACAAGGCAACAATTACTGCAGACGAATCCACCAATACAGCCT
Y  N  Q  K  F  K  D  K  A  T  I  T  A  D  E  S  T  N  T  A

       310        320        330        340        350        360
ACATGGAACTGAGCAGCCTGAGATCTGAGGACACCGCAGTCTATTACTGTGCAAGAGGGG
Y  M  E  L  S  S  L  R  S  E  D  T  A  V  Y  Y  C  A  R  G

       370        380        390        400        410        420
GGGGGGTCTTTGACTACTGGGGCCAAGGAACCCTGGTCACAGTCTCCTCAGGTGAGTCCT
G  G  V  F  D  Y  W  G  Q  G  T  L  V  T  V  S  S

       430
TAAAACCTCTAGA
```

FIGURE 4

```
        10         20         30         40         50         60
TCTAGATGGAGACCGATACCCTCCTGCTATGGGTCCTCCTGCTATGGGTCCCAGGATCAA
      M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S

        70         80         90        100        110        120
CCGGACATATTCAGATGACCCAGTCTCCATCTACCCTCTCTGCTAGCGTCGGGGATAGGG
T  G  D  I  Q  M  T  Q  S  P  S  T  L  S  A  S  V  G  D  R

       130        140        150        160        170        180
TCACCATAACCTGCTCTGCCAGCTCAAGTATAAGTTACATGCACTGGTACCAGCAGAAGC
V  T  I  T  C  S  A  S  S  S  I  S  Y  M  H  W  Y  Q  Q  K

       190        200        210        220        230        240
CAGGCAAAGCTCCCAAGCTTCTAATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTG
P  G  K  A  P  K  L  L  I  Y  T  T  S  N  L  A  S  G  V  P

       250        260        270        280        290        300
CTCGCTTCAGTGGCAGTGGATCTGGGACCGAGTTCACCCTCACAATCAGCTCTCTGCAGC
A  R  F  S  G  S  G  S  G  T  E  F  T  L  T  I  S  S  L  Q

       310        320        330        340        350        360
CAGATGATTTCGCCACTTATTACTGCCATCAAAGGAGTACTTACCCACTCACGTTCGGTC
P  D  D  F  A  T  Y  Y  C  H  Q  R  S  T  Y  P  L  T  F  G

       370        380        390        400
AGGGGACCAAGGTGGAGGTCAAACGTAAGTACACTTTTCTAGA
Q  G  T  K  V  E  V  K
```

FIGURE 5

A

HES12 AGCTTCTAGATGGGATGGAGCTGGATCTTTCTCTTCCTCCTGTCAGGTACCGCGGGCGTG
CACTCTCAGGTCCAGCTTGTCCAGTCTGGGGCTGAAGTCAAGAAACCTGGCTCGAGCGTG
AAGGTC

HES13 CCCAGTCGACGGATTAATATATCCAATCCATTCCAGACCCTGTCCAGGGGCCTGCCTTAC
CCAGTGCATCCTGTAGCTAGTAAAGGTGTAGCCAGAAGCCTTGCAGGAGACCTTCACGCT
CGAGCCAGG

HES14 TATATTAATCCGTCGACTGGGTATACTGAATACAATCAGAAGTTCAAGGACAAGGCAACA
ATTACTGCAGACGAATCCACCAATACAGCCTACATGGAACTGAGCAGCCTGAGATCTGAG
GACA

HES15 ATATCGTCTAGAGGTTTTAAGGACTCACCTGAGGAGACTGTGACCAGGGTTCCTTGGCCC
CAGTAGTCAAAGACCCCCCCCCCTCTTGCACAGTAATAGACTGCGGTGTCCTCAGATCTC
AGGCTGCT

B

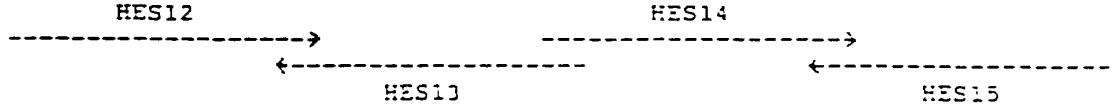


FIGURE 6

A

JFD1 CAAATCTAGATGGAGACCGATACCCTCCTGCTATGGGTCCTCCTGCTATGGGTCCCAGGA
TCAACCGGAGATATTCAGATGACCCAGTCTCCATCTACCCTCTCTGCTAGCGTCGGGGAT

JFD2 ATAAATTAGAAGCTTGGGAGCTTTGCCTGGCTTCTGCTGGTACCAGTGCATGTAACTTAT
ACTTGAGCTGGCAGAGCAGGTTATGGTGACCCTATCCCCGACGCTAGCAGAGAG

JFD3 GCTCCCAAGCTTCTAATTTATACCACATCCAACCTGGCTTCTGGAGTCCCTGCTCGCTTC
AGTGGCAGTGGATCTGGGACCGAGTTCACCCTCACAATCAGCTCTCTGCAGCCAGATGAT
TTC

JFD4 TATATCTAGAAAAGTGTACTTACGTTTGACCTCCACCTTGGTCCCCTGACCGAACGTGAG
TGGGTAAGTACTCCTTTGATGGCAGTAATAAGTGGCGAAATCATCTGGCTGCAGAGAGCT
GA

B

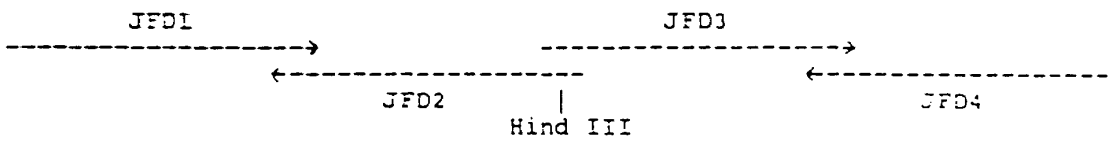


FIGURE 7

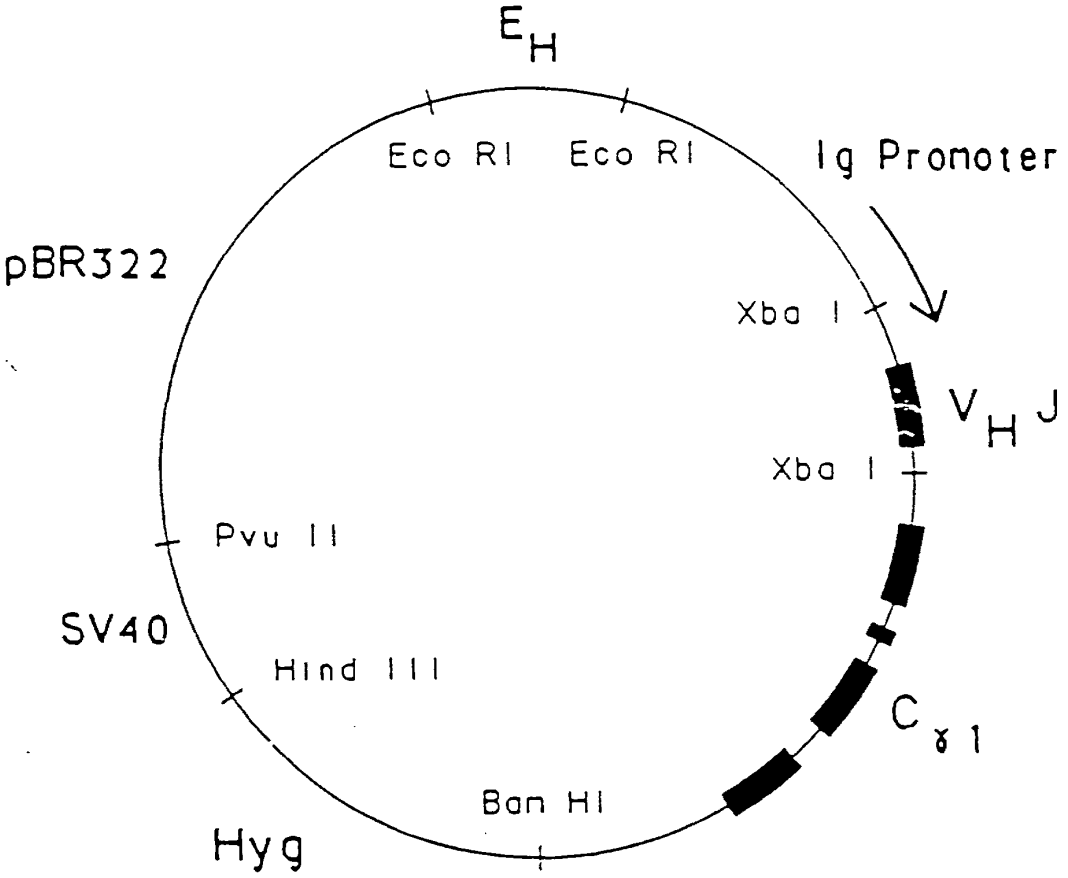

E H
Eco RI
Eco RI
Ig Promoter
pBR322
Xba I
V H J
Xba I
Pvu II
SV40
Hind III
C γ 1
Ban HI
Hyg

FIGURE 8

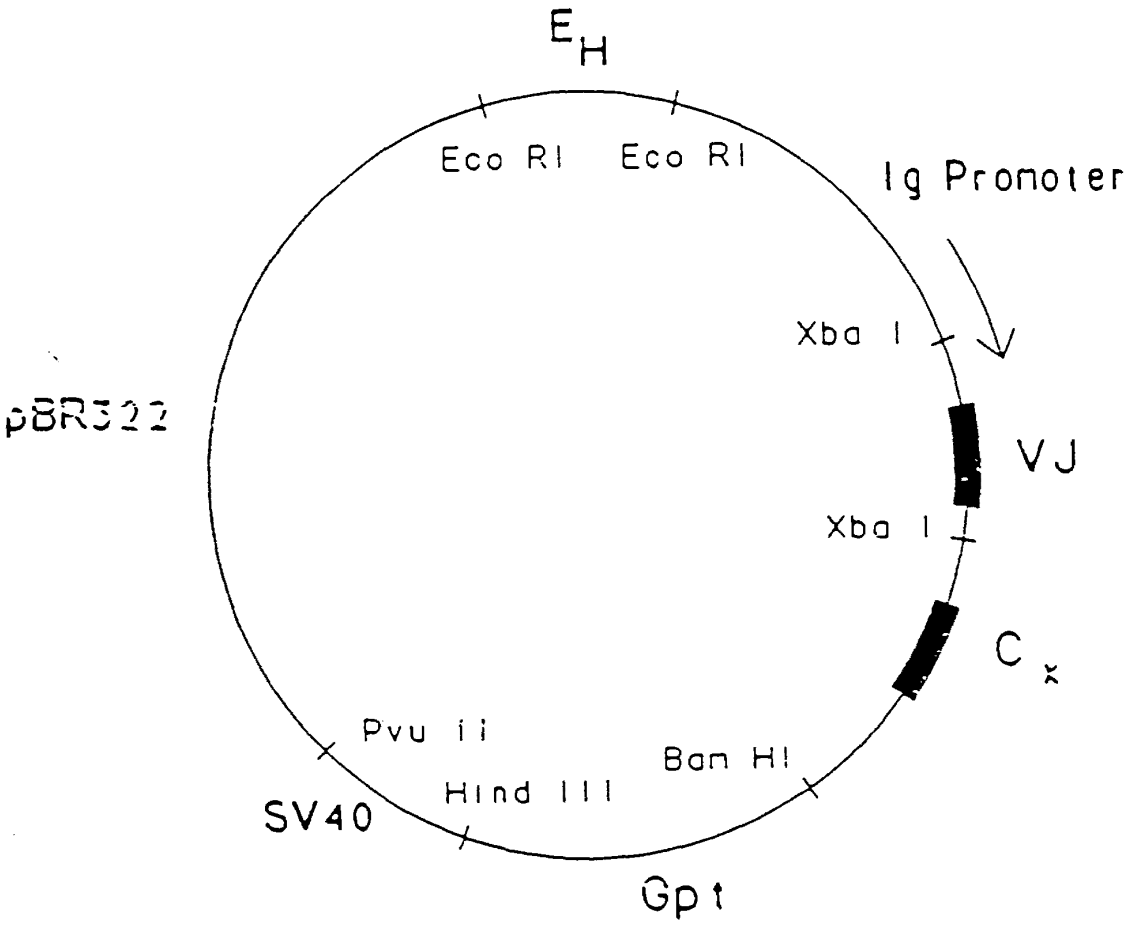

E H
Eco RI
Eco RI
Ig Promoter
Xba I
VJ
Xba I
C κ
pBR322
Pvu II
Ban HI
Hind III
SV40
Gpt

FIGURE 9

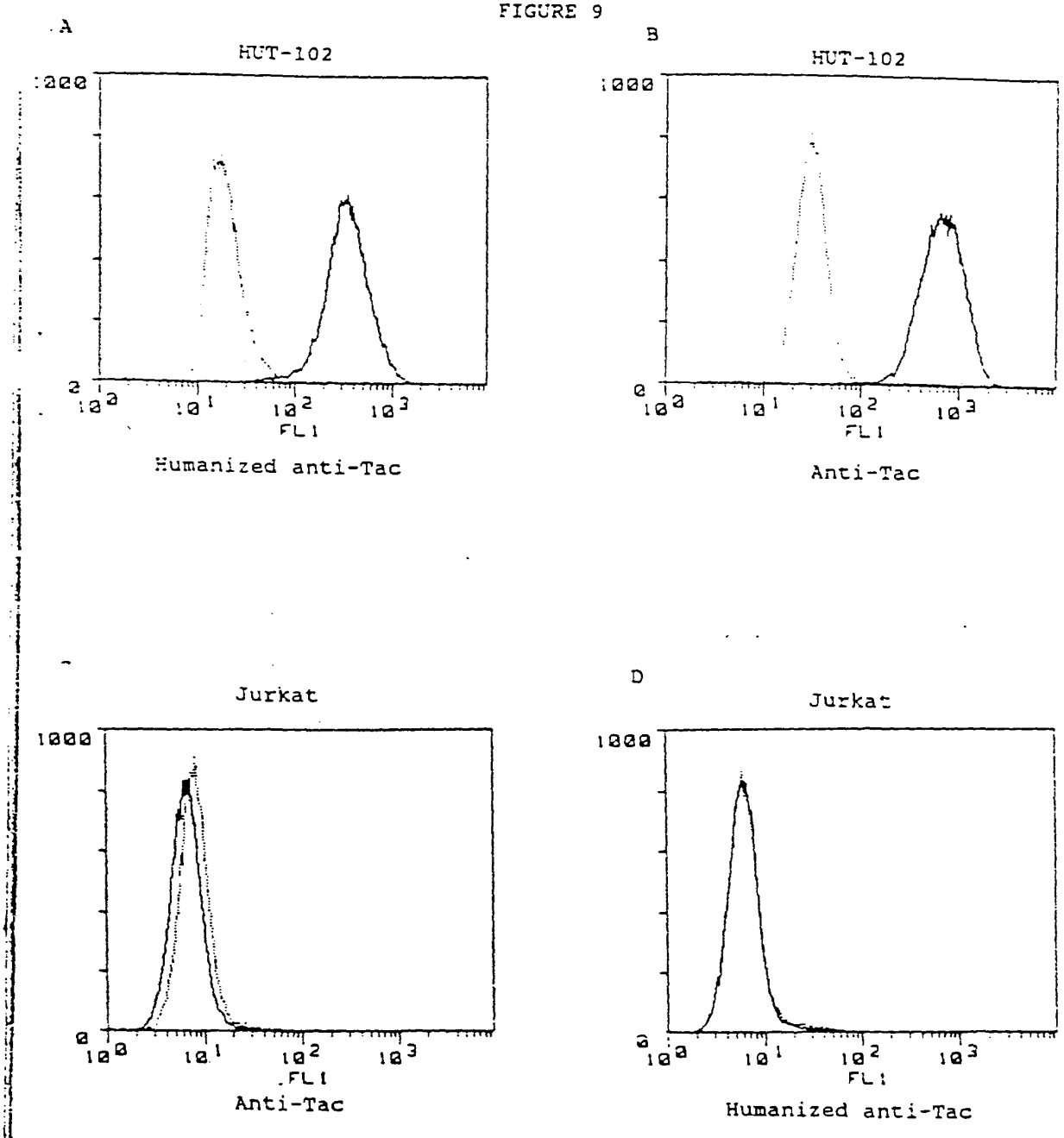
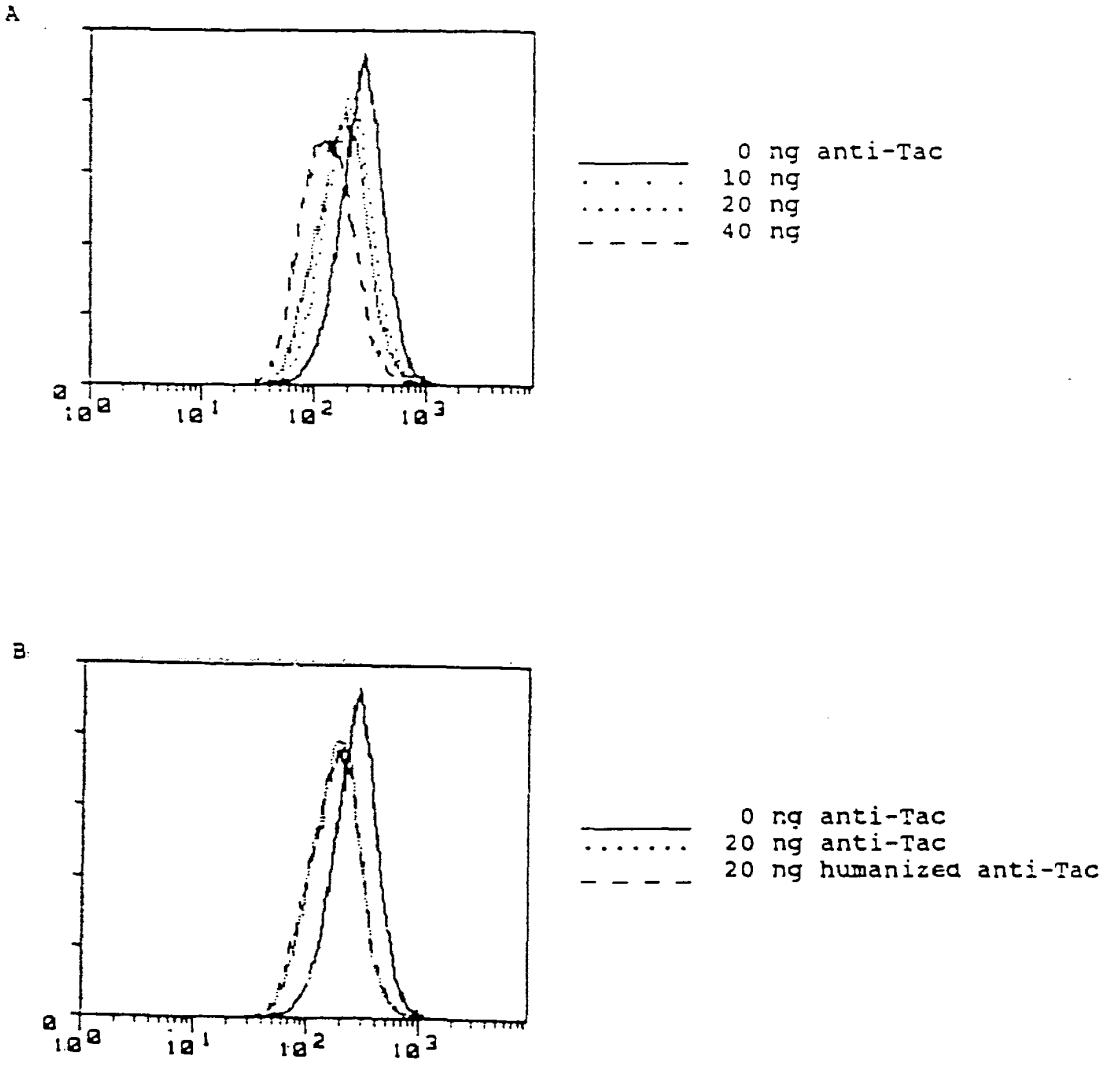

A
HUT-102
1000
0
10^0 10^1 10^2 10^3
FL1
Humanized anti-Tac
B
HUT-102
1000
0
10^0 10^1 10^2 10^3
FL1
Anti-Tac
C
Jurkat
1000
0
10^0 10^1 10^2 10^3
FL1
Anti-Tac
D
Jurkat
1000
0
10^0 10^1 10^2 10^3
FL1
Humanized anti-Tac

FIGURE 10


A
0
10^0
10^1
10^2
10^3
0 ng anti-Tac
10 ng
20 ng
40 ng
B
0
10^0
10^1
10^2
10^3
0 ng anti-Tac
20 ng anti-Tac
20 ng humanized anti-Tac

#46/ZK
02/11/99

Patent Docket P0709P1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of Paul J. Carter et al. Serial No.: 08/146,206 Filed: November 17, 1993 For: METHOD FOR MAKING HUMANIZED ANTIBODIES	Group Art Unit: 1642 Examiner: J. Reeves CERTIFICATE OF HAND DELIVERY I hereby certify that this correspondence is beinghand delivered in an envelope addressed to: Assistant Commissioner of Patents, Washington, D.C. 20231 on February 1, 1999 R. H. Mitchell

SUPPLEMENTAL INFORMATION DISCLOSURE STATEMENT

03/26/1999 TGRAY1 00000002 070630 08146206
01 FC:126 240.00 CH

Assistant Commissioner of Patents
Washington, D.C. 20231

Sir:

Applicants submit herewith patents, publications or other information (attached hereto and listed on the attached Form PTO-1449) of which they are aware, which they believe may be material to the examination of this application and in respect of which there may be a duty to disclose in accordance with 37 CFR §1.56.

This Information Disclosure Statement:

(a) ☐ accompanies the new patent application submitted herewith. 37 CFR §1.97(a).

(b) ☐ is filed within three months after the filing date of the application or within three months after the date of entry of the national stage of a PCT application as set forth in 37 CFR§1.491.

(c) ☐ as far as is known to the undersigned, is filed before the mailing date of a first Office action on the merits.

(d) ☐ is filed after the first Office Action and more than three months after the application's filing date or PCT national stage date of entry filing but, as far as is known to the undersigned, prior to the mailing date of either a final rejection or a notice of allowance, whichever occurs first, and is accompanied by either the fee ($240) set forth in 37 CFR §1.17(p) or a statement as specified in 37 CFR §1.97(e), as checked below. Should any fee be due, the U.S. Patent and Trademark Office is hereby authorized to charge Deposit Account No. 07-0630 in the amount of $240.00 to cover the cost of this Information Disclosure Statement. Any deficiency or overpayment

should be charged or credited to this deposit account. A duplicate of this sheet is enclosed.

(e) () is filed after the mailing date of either a final rejection or a notice of allowance, whichever occurred first, and is accompanied by the fee ($130) set forth in 37 CFR §1.17(i) and a statement as specified in 37 CFR §1.97(e), as checked below. This document is to be considered as a petition requesting consideration of the information disclosure statement. The U.S. Patent and Trademark Office is hereby authorized to charge Deposit Account No. 07-0630 in the amount of $130.00 to cover the cost of this Information Disclosure Statement. Any deficiency or overpayment should be charged or credited to this deposit account. A duplicate of this sheet is enclosed.

(f) (x) is filed after the mailing date of a final rejection, but a request to withdraw the finality thereof under 37 CFR § 1.129(a) was submitted on August 24, 1998. The U.S. Patent and Trademark Office is hereby authorized to charge Deposit Account No. 07-0630 to cover the cost of this Information Disclosure Statement in the event that any fees are due. A duplicate of this sheet is enclosed.

(If either of boxes (d) or (e) is checked above, the following statement under 37 CFR §1.97(e) may need to be completed.) The undersigned states that:

() Each item of information contained in the information disclosure statement was cited in a communication mailed from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of this information disclosure statement.

() No item of information contained in this information disclosure statement was cited in a communication mailed from a foreign patent office in a counterpart foreign application and, to the knowledge of the undersigned after making reasonable inquiry, was known to any individual designated in 37 CFR §1.56(c) more than three months prior to the filing of this information disclosure statement.

A list of the patent(s) or publication(s) is set forth on the attached Form PTO-1449 (Modified).

A copy of the items on PTO-1449 is supplied herewith:

(x) each () none () only those listed below:

A concise explanation of relevance of the items listed on PTO-1449 is:

(x) not given

() given for each listed item

() given for only non-English language listed item(s) (Required)

() in the form of an English language copy of a Search Report from a foreign patent

office, issued in a counterpart application, which refers to the relevant portions of the references.

The Examiner is reminded that a "concise explanation of the relevance" of the submitted prior art "may be nothing more than identification of the particular figure or paragraph of the patent or publication which has some relation to the claimed invention," MPEP §609.

While the information and references disclosed in this Information Disclosure Statement may be "material" pursuant to 37 CFR §1.56, it is not intended to constitute an admission that any patent, publication or other information referred to therein is "prior art" for this invention unless specifically designated as such.

In accordance with 37 CFR §1.97(g), the filing of this Information Disclosure Statement shall not be construed to mean that a search has been made or that no other material information as defined in 37 CFR §1.56(a) exists. It is submitted that the Information Disclosure Statement is in compliance with 37 CFR §1.98 and MPEP §609 and the Examiner is respectfully requested to consider the listed references.

Respectfully submitted,
GENENTECH, INC.

Date: January 29, 1999 By: _____
Wendy M. Lee
Reg. No. 40,378

1 DNA Way
So. San Francisco, CA 94080-4990
Phone: (650) 225-1994
Fax: (650) 952-9881

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 460 167 B1**

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification: **11.10.95**

(51) Int. Cl.[6]: **C12P 21/08**, C12N 15/13, A61K 39/395, C07K 16/18, C12N 5/10, C12N 15/62

(21) Application number: **91901433.2**

(22) Date of filing: **21.12.90**

(86) International application number:
PCT/GB90/02017

(87) International publication number:
WO 91/09967 (11.07.91 91/15)

Divisional applications 94104042.0, 94202090.0.

(54) HUMANISED ANTIBODIES.

(30) Priority: **21.12.89 GB 8928874**

(43) Date of publication of application:
11.12.91 Bulletin 91/50

(45) Publication of the grant of the patent:
11.10.95 Bulletin 95/41

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IT LI LU NL SE

(56) References cited:
EP-A- 0 239 400
EP-A- 0 323 806
EP-A- 0 328 404
EP-A- 0 403 156

PROCEEDINGS OF THE NATL. ACADEMY OF SCIENCES USA, vol. 86, December 1989; C. QUEEN et al., pp. 10029-10033&NUM;

NATURE, vol. 332, March 1988; L. RIECHMANN et al., pp. 323-327&NUM;

(73) Proprietor: **CELLTECH THERAPEUTICS LIMITED**
216 Bath Road
Slough,
Berkshire SL1 4EN (GB)

(72) Inventor: **ADAIR, John, Robert**
23 George Road
Stokenchurch
High Wycombe
Buckinghamshire HP14 3RN (GB)
Inventor: **ATHWAL, Diljeet, Singh**
Flat 35
Knollys House
Tavistock Square
London WC1 (GB)
Inventor: **EMTAGE, John, Spencer**
49 Temple Mill Island
Temple
Marlow
Buckinghamshire SL7 1SQ (GB)

EP 0 460 167 B1

NATURE, vol. 321, May 1986; P.T. JONES et al., pp. 522-525&NUM;

NATURE, vol. 328, August 1987; S. ROBERTS et al., pp. 731-734&NUM;

SCIENCE, vol. 239, 1988; M. VERHOEYEN et al., pp. 1534-1536&NUM;

NATURE, vol. 341, October 1989; E.S. WARD et al., pp. 544-546&NUM;

(74) Representative: Mercer, Christopher Paul et al
Carpmaels & Ransford
43, Bloomsbury Square
London WC1A 2RA (GB)

Description

Field of the Invention

The present invention relates to humanised antibody molecules, to processes for their production using recombinant DNA technology, and to their therapeutic uses.

The term "humanised antibody molecule" is used to describe a molecule having an antigen binding site derived from an immunoglobulin from a non-human species, and remaining immunoglobulin-derived parts of the molecule being derived from a human immunoglobulin. The antigen binding site typically comprises complementarity determining regions (CDRs) which determine the binding specificity of the antibody molecule and which are carried on appropriate framework regions in the variable domains. There are 3 CDRs (CDR1, CDR2 and CDR3) in each of the heavy and light chain variable domains.

In the description, reference is made to a number of publications by number. The publications are listed in numerical order at the end of the description.

Background of the Invention

Natural immunoglobulins have been known for many years, as have the various fragments thereof, such as the Fab, $(Fab')_2$ and Fc fragments, which can be derived by enzymatic cleavage. Natural immunoglobulins comprise a generally Y-shaped molecule having an antigen-binding site towards the end of each upper arm. The remainder of the structure, and particularly the stem of the Y, mediates the effector functions associated with immunoglobulins.

Natural immunoglobulins have been used in assay, diagnosis and, to a more limited extent, therapy. However, such uses, especially in therapy, were hindered until recently by the polyclonal nature of natural immunoglobulins. A significant step towards the realisation of the potential of immunoglobulins as therapeutic agents was the discovery of procedures for the production of monoclonal antibodies (MAbs) of defined specificity (1).

However, most MAbs are produced by hybridomas which are fusions of rodent spleen cells with rodent myeloma cells. They are therefore essentially rodent proteins. There are very few reports of the production of human MAbs.

Since most available MAbs are of rodent origin, they are naturally antigenic in humans and thus can give rise to an undesirable immune response termed the HAMA (Human Anti-Mouse Antibody) response. Therefore, the use of rodent MAbs as therapeutic agents in humans is inherently limited by the fact that the human subject will mount an immunological response to the MAb and will either remove it entirely or at least reduce its effectiveness. In practice, MAbs of rodent origin may not be used in patients for more than one or a few treatments as a HAMA response soon develops rendering the MAb ineffective as well as giving rise to undesirable reactions. For instance, OKT3 a mouse IgG2a/k MAb which recognises an antigen in the T-cell receptor-CD3 complex has been approved for use in many countries throughout the world as an immunosuppressant in the treatment of acute allograft rejection [Chatenoud et al (2) and Jeffers et al (3)]. However, in view of the rodent nature of this and other such MAbs, a significant HAMA response which may include a major anti-idiotype component, may build up on use. Clearly, it would be highly desirable to diminish or abolish this undesirable HAMA response and thus enlarge the areas of use of these very useful antibodies.

Proposals have therefore been made to render non-human MAbs less antigenic in humans. Such techniques can be generically termed "humanisation" techniques. These techniques typically involve the use of recombinant DNA technology to manipulate DNA sequences encoding the polypeptide chains of the antibody molecule.

Early methods for humanising MAbs involved production of chimeric antibodies in which an antigen binding site comprising the complete variable domains of one antibody is linked to constant domains derived from another antibody. Methods for carrying out such chimerisation procedures are described in EP0120694 (Celltech Limited), EP0125023 (Genentech Inc. and City of Hope), EP-A-0 171496 (Res. Dev. Corp. Japan), EP-A-0 173 494 (Stanford University), and WO 86/01533 (Celltech Limited). This latter Celltech application (WO 86/01533) discloses a process for preparing an antibody molecule having the variable domains from a mouse MAb and the constant domains from a human immunoglobulin. Such humanised chimeric antibodies, however, still contain a significant proportion of non-human amino acid sequence, i.e. the complete non-human variable domains, and thus may still elicit some HAMA response, particularly if administered over a prolonged period [Begent et al (ref. 4)].

In an alternative approach, described in EP-A-0239400 (Winter), the complementarity determining regions (CDRs) of a mouse MAb have been grafted onto the framework regions of the variable domains of a human immunoglobulin by site directed mutagenesis using long oligonucleotides. The present invention relates to humanised antibody molecules prepared according to this alternative approach, i.e. CDR-grafted humanised antibody molecules. Such CDR-grafted humanised antibodies are much less likely to give rise to a HAMA response than humanised chimeric antibodies in view of the much lower proportion of non-human amino acid sequence which they contain.

The earliest work on humanising MAbs by CDR-grafting was carried out on MAbs recognising synthetic antigens, such as the NP or NIP antigens. However, examples in which a mouse MAb recognising lysozyme and a rat MAb recognising an antigen on human T-cells were humanised by CDR-grafting have been described by Verhoeyen et al (5) and Riechmann et al (6) respectively. The preparation of CDR-grafted antibody to the antigen on human T cells is also described in WO 89/07452 (Medical Research Council).

In Riechmann et al/Medical Research Council it was found that transfer of the CDR regions alone [as defined by Kabat refs. (7) and (8)] was not sufficient to provide satisfactory antigen binding activity in the CDR-grafted product. Riechmann et al found that it was necessary to convert a serine residue at position 27 of the human sequence to the corresponding rat phenylalanine residue to obtain a CDR-grafted product having improved antigen binding activity. This residue at position 27 of the heavy chain is within the structural loop adjacent to CDR1. A further construct which additionally contained a human serine to rat tyrosine change at position 30 of the heavy chain did not have a significantly altered binding activity over the humanised antibody with the serine to phenylalanine change at position 27 alone. These results indicate that changes to residues of the human sequence outside the CDR regions, in particular in the structural loop adjacent to CDR1, may be necessary to obtain effective antigen binding activity for CDR-grafted antibodies which recognise more complex antigens. Even so the binding affinity of the best CDR-grafted antibodies obtained was still significantly less than the original MAb.

Very recently Queen et al (9) have described the preparation of a humanised antibody that binds to the interleukin 2 receptor, by combining the CDRs of a murine MAb (anti-Tac) with human immunoglobulin framework and constant regions. The human framework regions were chosen to maximise homology with the anti-Tac MAb sequence. In addition computer modelling was used to identify framework amino acid residues which were likely to interact with the CDRs or antigen, and mouse amino acids were used at these positions in the humanised antibody.

In WO 90/07861 Queen et al propose four criteria for designing humanised immunoglobulins. The first criterion is to use as the human acceptor the framework from a particular human immunoglobulin that is unusually homologous to the non-human donor immunoglobulin to be humanised, or to use a consensus framework from many human antibodies. The second criterion is to use the donor amino acid rather than the acceptor if the human acceptor residue is unusual and the donor residue is typical for human sequences at a specific residue of the framework. The third criterion is to use the donor framework amino acid residue rather than the acceptor at positions immediately adjacent to the CDRs. The fourth criterion is to use the donor amino acid residue at framework positions at which the amino acid is predicted to have a side chain atom within about 3 Å of the CDRs in a three-dimensional immunoglobulin model and to be capable of interacting with the antigen or with the CDRs of the humanised immunoglobulin. It is proposed that criteria two, three or four may be applied in addition or alternatively to criterion one, and may be applied singly or in any combination.

WO 90/07861 describes in detail the preparation of a single CDR-grafted humanised antibody, a humanised antibody having specificity for the p55 Tac protein of the IL-2 receptor. The combination of all four criteria, as above, were employed in designing this humanised antibody, the variable region frameworks of the human antibody Eu (7) being used as acceptor. In the resultant humanised antibody the donor CDRs were as defined by Kabat et al (7 and 8) and in addition the mouse donor residues were used in place of the human acceptor residues, at positions 27, 30, 48, 66, 67, 89, 91, 94, 103, 104, 105 and 107 in the heavy chain and at positions 48, 60 and 63 in the light chain, of the variable region frameworks. The humanised anti-Tac antibody obtained is reported to have an affinity for p55 of 3×10^9 M^{-1}, about one-third of that of the murine MAb.

We have further investigated the preparation of CDR-grafted humanised antibody molecules and have identified a hierarchy of positions within the framework of the variable regions (i.e. outside both the Kabat CDRs and structural loops of the variable regions) at which the amino acid identities of the residues are important for obtaining CDR-grafted products with satisfactory binding affinity. This has enabled us to establish a protocol for obtaining satisfactory CDR-grafted products which may be applied very widely irrespective of the level of homology between the donor immunoglobulin and acceptor framework. The set of residues which we have identified as being of critical importance does not coincide with the residues

identified by Queen et al (9).

Summary of the Invention

Accordingly, in a first aspect, the invention provides an antibody molecule having affinity for a predetermined antigen and comprising a composite heavy chain and a complementary light chain, said composite heavy chain having a variable domain comprising acceptor antibody heavy chain framework residues and donor antibody heavy chain antigen-binding residues, said donor antibody having affinity for said predetermined antigen, wherein, according to the Kabat numbering system, in said composite heavy chain, amino acid residues 5, 8, 10, 12 to 17, 19, 21, 22, 40, 42 to 44, 66, 68, 70, 74, 77, 79, 81, 83 to 85, 90, 92, 105, 109, 111 and 113 at least are acceptor residues and amino acid residues 23, 24, 31 to 35, 49 to 58, 71, 73, 78 and 95 to 102 at least are donor residues.

Preferably amino acid residues 26 to 30 and 59 to 65 in said composite heavy chain are additionally donor residues.

If desired, at least one of amino acid residues 1, 3, and 76 in said composite heavy chain are additionally donor residues.

If further desired at least one of amino acid residues 36, 94, 104, 106 and 107 in said composite heavy chain are additionally donor residues and preferably at least one of amino acid residues 2, 4, 6, 38, 48, 67 and 69 in said composite heavy chain are additionally donor residues.

Preferably, amino acid residues 7, 9, 11, 18, 20, 25, 37, 39, 41, 45, 47, 48, 72, 75, 80, 82, 86 to 89, 91, 93, 103, 108, 110 and 112 in said composite heavy chain are additionally acceptor residues.

Advantageously, said complementary light chain is a composite light chain having a variable domain comprising acceptor antibody light chain framework residues and donor antibody light chain antigen-binding residues, said donor antibody having affinity for said predetermined antigen, wherein, according to the Kabat numbering system, in said composite light chain, amino acid residues 5, 7 to 9, 11, 13 to 18, 20, 22, 23, 39, 41 to 43, 57, 59, 61, 72, 74 to 79, 81, 82, 84, 86, 88, 100, 104, 106 and 107 at least are acceptor residues and amino acid residues 24 to 34, 46, 48, 50 to 56, 58, 71 and 89 to 97 at least are donor residues.

Preferably, amino acid residues 1, 3 and 47 in said composite light chain are additionally donor residues.

If desired, amino acid residues 36, 44, 47, 85 and 87 in said composite light chain are additionally donor residues.

If further desired, at least one of amino acid residues 2, 4, 6, 49, 62, 64 to 69, 98, 99, 101 and 102 in said composite light chain are additionally donor residues.

Preferably, at least one of amino acid residues 1, 3, 10, 12, 21, 40, 60, 63, 70, 73, 80, 103 and 105 in said composite light chain are additionally donor residues.

In a second aspect, the invention provides a therapeutic or diagnostic composition comprising the antibody molecule as defined above in combination with a pharmaceutically acceptable carrier, diluent or excipient.

In a third aspect, the invention provides a method for producing a recombinant antigen binding molecule having affinity for a predetermined antigen comprising the steps of:

[1] determining the amino acid sequence of the variable domain of the heavy chain of a donor antibody which has affinity for said predetermined antigen;

[2] determining the amino acid sequence of the variable domain of the heavy chain of a non-specific acceptor antibody;

[3] providing a composite heavy chain for an antibody molecule, said composite heavy chain having acceptor framework residues and donor antigen binding residues wherein, according to the Kabat numbering system, amino acid residues 5, 8, 10, 12 to 17, 19, 21, 22, 40, 42 to 44, 66, 68, 70, 74, 77, 79, 81, 83 to 85, 90, 92, 105, 109, 111 and 113 at least are acceptor residues and amino acid residues 23, 24, 31 to 35, 49 to 58, 71, 73 and 78 and 95 to 102 at least are donor residues;

[4] associating the heavy chain produced in step [3] with a complementary light chain to form an antibody molecule;

[5] determining the affinity of the antibody molecule formed in step [4] for said predetermined antigen;

[6] if the affinity determined in step [5] is not equivalent to that of the donor antibody, providing a heavy chain as described in [3] above but in which amino acid residues 71, 73 and 78 are additionally donor residues;

[7] associating the heavy chain produced in step [6] with a complementary light chain to form an antibody molecule;

[8] determining the affinity of the antibody molecule formed in step [7] for said predetermined antigen;
[9] if the affinity determined in step [8] is not equivalent to that of the donor antibody, providing a heavy chain as described in [6] above but in which amino acid residues 26 to 30 are additionally donor residues;
[10] associating the heavy chain produced in step [9] with a complementary light chain to form an antibody molecule;
[11] determining the affinity of the antibody molecule formed in step [10] for said predetermined antigen;
[12] if the affinity determined in step [11] is not equivalent to that of the donor antibody, providing a heavy chain as described in [9] above but in which at least one of amino acid residues 1, 3, and 76 are additionally donor residues;
[13] associating the heavy chain produced in step [12] with a complementary light chain to form an antibody molecule;
[14] determining the affinity of the antibody molecule formed in step [13] for said predetermined antigen;
[15] if the affinity determined in step [14] is not equivalent to that of the donor antibody, providing a heavy chain as described in [12] above but in which at least one of amino acid residues 36, 94, 104, 106, 107 are additionally donor residues;
[16] associating the heavy chain produced in step [15] with a complementary light chain to form an antibody molecule.
[17] determining the affinity of the antibody molecule formed in step [16] for said predetermined antigen;
[18] if the affinity determined in step [17] is not equivalent to that of the donor antibody, providing a heavy chain as described in [15] above but in which at least one of amino acid residues 2, 4, 6, 38, 48, 67 and 69 are additionally donor residues; and
[19] associating the heavy chain produced in step [18] with a complementary light chain to form an antibody molecule.

Preferably, the method further comprises the steps of:
[1] determining the amino acid sequence of the variable domain of the light chain of said donor antibody which has affinity for said predetermined antigen;
[2] determining the amino acid sequence of the variable domain of the light chain of a non-specific acceptor antibody;
[3] providing a composite light chain for an antibody molecule, said composite light chain having acceptor framework residues and donor antigen binding residues wherein, according to the Kabat numbering system, amino acid residues 5, 7 to 9, 11, 13 to 18, 20, 22, 23, 39, 41 to 43, 57, 59, 61, 72, 74 to 79 to 81, 82, 84, 86, 88, 100, 104 and 106 to 109 at least are acceptor residues and amino acid residues 24 to 34, 46, 48, 50 to 56, 58, 71 and 89 to 97 at least are donor residues;
[4] associating the light chain produced in step [3] with a complementary heavy chain to form an antibody molecule;
[5] determining the affinity of the antibody molecule formed in step [4] for said predetermined antigen;
[6] if the affinity determined in step [5] is not equivalent to that of the donor antibody, providing a light chain as described in [3] above but in which amino acid residues 1, 2, 3 and 47 are additionally donor residues;
[7] associating the light chain produced in step [6] with a complementary heavy chain to form an antigen-binding molecule;
[8] determining the affinity of the antigen-binding molecule formed in step [7] for said predetermined antigen;
[9] if the affinity determined in step [8] is not equivalent to that of the donor antibody, providing a light chain as described in [6] above but in which amino acid residues 36, 44, 47, 85 and 87 are additionally donor residues;
[10] associating the light chain produced in step [9] with a complementary heavy chain to form an antibody molecule;
[11] determining the affinity of the antibody molecule formed in step [10] for said predetermined antigen;
[12] if the affinity determined in step [11] is not equivalent to that of the donor antibody, providing a light chain as described in [9] above but in which at least one of amino acid residues 2, 4, 6, 49, 62, 64 to 69, 98, 99, 101 are additionally donor residues; and
[13] associating the light chain produced in step [9] with a complementary heavy chain to form an antibody molecule.

In the present application, reference is made to antibody molecules comprising acceptor framework and donor antigen binding regions. It will be appreciated that the invention is widely applicable to the CDR-grafting of antibodies in general. Thus, the donor and acceptor antibodies may be derived from animals of

the same species and even the same antibody class or sub-class. More usually, however, the donor and acceptor antibodies are derived from animals of different species. Typically the donor antibody is a non-human antibody, such as a rodent MAb, and the acceptor antibody is a human antibody.

The residue designations given above and elsewhere in the present application are numbered according to the Kabat numbering [refs. (7) and (8)]. Thus the residue designations do not always correspond directly with the linear numbering of the amino acid residues. The actual linear amino acid sequence may contain fewer or additional amino acids than in the strict Kabat numbering corresponding to a shortening of, or insertion into, a structural component, whether framework or CDR, of the basic variable domain structure. For example, the heavy chain variable region of the anti-Tac antibody described by Queen et al. (9) contains a single amino acid insert (residue 52a) after residue 52 of CDR2 and a three amino acid insert (residues 82a, 82b and 82c) after framework residue 82, in the Kabat numbering. The correct Kabat numbering of residues may be determined for a given antibody by alignment at regions of homology of the sequence of the antibody with a "standard" Kabat numbered sequence.

The humanised antibody molecules of the present invention may comprise: a complete antibody molecule, having full length heavy and light chains; a fragment thereof, such as a Fab, $(Fab')_2$ or FV fragment; or a single chain antibody, e.g. a single chain FV in which heavy and light chain variable regions are joined by a peptide linker; or any other CDR-grafted molecule with the same specificity as the original donor antibody. Similarly the CDR-grafted heavy and light chain variable region may be combined with other antibody domains as appropriate.

Also the heavy or light chains or humanised antibody molecules of the present invention may have attached to them an effector or reporter molecule. For instance, it may have a macrocycle, for chelating a heavy metal atom, or a toxin, such as ricin, attached to it by a covalent bridging structure. Alternatively, the procedures of recombinant DNA technology may be used to produce an immunoglobulin molecule in which the Fc fragment or CH3 domain of a complete immunoglobulin molecule has been replaced by, or has attached thereto by peptide linkage, a functional non-immunoglobulin protein, such as an enzyme or toxin molecule.

Any appropriate acceptor variable region framework sequences may be used having regard to class/type of the donor antibody from which the antigen binding regions are derived. Preferably, the type of acceptor framework used is of the same/similar class/type as the donor antibody. Conveniently, the framework may be chosen to maximise/optimise homology with the donor antibody sequence particularly at positions close or adjacent to the CDRs. However, a high level of homology between donor and acceptor sequences is not important for application of the present invention. The present invention identifies a hierarchy of framework residue positions at which donor residues may be important or desirable for obtaining a CDR-grafted antibody product having satisfactory binding properties. The CDR-grafted products usually having binding affinities of at least 10^5 M^{-1}, or especially in the range 10^8-10^{12} M^{-1}. In principle, the present invention is applicable to any combination of donor and acceptor antibodies irrespective of the level of homology between their sequences. A protocol for applying the invention to any particular donor-acceptor antibody pair is given hereinafter. Examples of human frameworks which may be used are KOL, NEWM, REI, EU, LAY and POM (refs. 4 and 5) and the like; for instance KOL and NEWM for the heavy chain and REI for the light chain and EU, LAY and POM for both the heavy chain and the light chain.

Also the constant region domains of the products of the invention may be selected having regard to the proposed function of the antibody in particular the effector functions which may be required. For example, the constant region domains may be human IgA, IgE, IgG or IgM domains. In particular, IgG human constant region domains may be used, especially of the IgG1 and IgG3 isotypes, when the humanised antibody molecule is intended for therapeutic uses, and antibody effector functions are required. Alternatively, IgG2 and IgG4 isotypes may be used when the humanised antibody molecule is intended for therapeutic purposes and antibody effector functions are not required, e.g. for simple blocking of lymphokine activity.

However, the remainder of the antibody molecules need not comprise only protein sequences from immunoglobulins. For instance, a gene may be constructed in which a DNA sequence encoding part of a human immunoglobulin chain is fused to a DNA sequence encoding the amino acid sequence of a functional polypeptide such as an effector or reporter molecule.

Preferably the CDR-grafted antibody heavy and light chain and antibody molecule products are produced by recombinant DNA technology.

The general methods by which vectors may be constructed, transfection methods and culture methods are well known per se and form no part of the invention. Such methods are shown, for instance, in references 10 and 11.

The DNA sequences which encode the donor amino acid sequence may be obtained by methods well known in the art. For example the donor coding sequences may be obtained by genomic cloning, or cDNA cloning from suitable hybridoma cell lines. Positive clones may be screened using appropriate probes for the heavy and light chain genes in question. Also PCR cloning may be used.

DNA coding for acceptor, e.g. human acceptor, sequences may be obtained in any appropriate way. For example DNA sequences coding for preferred human acceptor frameworks such as KOL, REI, EU and NEWM, are widely available to workers in the art.

The standard techniques of molecular biology may be used to prepare DNA sequences coding for the CDR-grafted products. Desired DNA sequences may be synthesised completely or in part using oligonucleotide synthesis techniques. Site-directed mutagenesis and polymerase chain reaction (PCR) techniques may be used as appropriate. For example oligonucleotide directed synthesis as described by Jones et al (ref. 20) may be used. Also oligonucleotide directed mutagenesis of a pre-exising variable region as, for example, described by Verhoeyen et al (ref. 5) or Riechmann et al (ref. 6) may be used. Also enzymatic filling in of gapped oligonucleotides using T_4 DNA polymerase as, for example, described by Queen et al (ref. 9) may be used.

Any suitable host cell/vector system may be used for expression of the DNA sequences coding for the CDR-grafted heavy and light chains. Bacterial e.g. E. coli, and other microbial systems may be used, in particular for expression of antibody fragments such as FAb and $(Fab')_2$ fragments, and especially FV fragments and single chain antibody fragments e.g. single chain FVs. Eucaryotic e.g. mammalian host cell expression systems may be used for production of larger CDR-grafted antibody products, including complete antibody molecules. Suitable mammalian host cells include CHO cells and myeloma or hybridoma cell lines.

For production of products comprising both heavy and light chains, the cell line may be transfected with two vectors, the first vector may contain an operon encoding a light chain-derived polypeptide and the second vector containing an operon encoding a heavy chain-derived polypeptide. Preferably, the vectors are identical, except in so far as the coding sequences and selectable markers are concerned, so as to ensure as far as possible that each polypeptide chain is equally expressed. Alternatively, a single vector may be used, the vector including the sequences encoding both light chain- and heavy chain-derived polypeptides.

The DNA in the coding sequences for the light and heavy chains may comprise cDNA or genomic DNA or both. However, it is preferred that the DNA sequence encoding the heavy or light chain comprises at least partially, genomic DNA, preferably a fusion of cDNA and genomic DNA.

The present invention is applicable to antibodies of any appropriate specificity. Advantageously, however, the invention may be applied to the humanisation of non-human antibodies which are used for in vivo therapy or diagnosis. Thus the antibodies may be site-specific antibodies such as tumour-specific or cell surface-specific antibodies, suitable for use in in vivo therapy or diagnosis, e.g. tumour imaging. Examples of cell surface-specific antibodies are anti-T cell antibodies, such as anti-CD3, and CD4 and adhesion molecules, such as CR3, ICAM and ELAM. The antibodies may have specificity for interleukins (including lymphokines, growth factors and stimulating factors), hormones and other biologically active compounds, and receptors for any of these. For example, the antibodies may have specificity for any of the following: Interferons α, β, γ or δ, IL1, IL2, IL3, or IL4, etc., TNF, GCSF, GMCSF, EPO, hGH, or insulin, etc.

A preferred protocol for obtaining CDR-grafted antibody molecules in accordance with the present invention is set out below together with the rationale by which we have derived this protocol. This protocol and rationale are given without prejudice to the generality of the invention as hereinbefore described and defined.

Protocol

It is first of all necessary to sequence the DNA coding for the heavy and light chain variable regions of the donor antibody, to determine their amino acid sequences. It is also necessary to choose appropriate acceptor heavy and light chain variable regions, of known amino acid sequence. The CDR-grafted chain is then designed starting from the basis of the acceptor sequence. It will be appreciated that in some cases the donor and acceptor amino acid residues may be identical at a particular position and thus no change of acceptor framework residue is required.

1. As a first step donor residues are substituted for acceptor residues in the CDRs. For this purpose the CDRs are preferably defined as follows:

Heavy chain - CDR1: residues 26-35
- CDR2: residues 50-65

- CDR3: residues 95-102

Light chain
- CDR1: residues 24-34
- CDR2: residues 50-56
- CDR3: residues 89-97

The positions at which donor residues are to be substituted for acceptor in the framework are then chosen as follows, first of all with respect to the heavy chain and subsequently with respect to the light chain.

2. Heavy Chain

2.1 Choose donor residues at all of positions 23, 24, 49, 71, 73 and 78 of the heavy chain or all of positions 23, 24 and 49 (71, 73 and 78 are always either all donor or all acceptor).

2.2 Check that the following have the same amino acid in donor and acceptor sequences, and if not preferably choose the donor: 2, 4, 6, 25, 36, 37, 39, 47, 48, 93, 94, 103, 104, 106 and 107.

2.3 To further optimise affinity consider choosing donor residues at one, some or any of:

i. 1, 3
ii. 72, 76
iii. If 48 is different between donor and acceptor sequences, consider 69
iv. If at 48 the donor residue is chosen, consider 38 and 46
v. If at 69 the donor residue is chosen, consider 80 and then 20
vi. 67
vii. If at 67 the donor residue is chosen, consider 82 and then 18
viii. 91
ix. 88
x. 9, 11, 41, 87, 108, 110, 112

3. Light Chain

3.1 Choose donor at 46, 48, 58 and 71

3.2 Check that the following have the same amino acid in donor and acceptor sequences, if not preferably choose donor: 2, 4, 6, 35, 38, 44, 47, 49, 62, 64-69 inclusive, 85, 87, 98, 99, 101 and 102

3.3 To further optimise affinity consider choosing donor residues at one, some or any of:

i. 1, 3
ii. 63
iii. 60, if 60 and 54 are able to form potential saltbridge
iv. 70, if 70 and 24 are able to form potential saltbridge
v. 73, and 21 if 47 is different between donor and acceptor
vi. 37, and 45 if 47 is different between donor and acceptor
vii. 10, 12, 40, 80, 103, 105

Rationale

In order to transfer the binding site of an antibody into a different acceptor framework, a number of factors need to be considered.

1. The extent of the CDRs

The CDRs (Complementary Determining Regions) were defined by Wu and Kabat (refs. 4 and 5) on the basis of an analysis of the variability of different regions of antibody variable regions. Three regions per domain were recognised. In the light chain the sequences are 24-34, 50-56, 89-97 (numbering according to Kabat (ref. 4), Eu Index) inclusive and in the heavy chain the sequences are 31-35, 50-65 and 95-102 inclusive.

When antibody structures became available it became apparent that these CDR regions corresponded in the main to loop regions which extended from the β barrel framework of the light and heavy variable domains. For H1 there was a discrepancy in that the loop was from 26 to 32 inclusive and for H2 the loop was 52 to 56 and for L2 from 50 to 53. However, with the exception of H1 the CDR regions encompassed the loop regions and extended into the β strand frameworks. In H1 residue 26 tends to be a serine and 27 a phenylalanine or tyrosine, residue 29 is a phenylalanine in most cases. Residues 28 and 30 which are surface residues exposed to solvent might be involved in antigen-binding. A prudent definition of the H1 CDR therefore would include residues 26-35 to include both the loop region and the hypervariable residues 33-35.

It is of interest to note the example of Riechmann et al (ref. 3), who used the residue 31-35 choice for CDR-H1. In order to produce efficient antigen binding, residue 27 also needed to be recruited from the donor (rat) antibody.

2. Non-CDR residues which contribute to antigen binding

By examination of available X-ray structures we have identified a number of residues which may have an effect on net antigen binding and which can be demonstrated by experiment. These residues can be sub-divided into a number of groups.

2.1 Surface residues near CDR [all numbering as in Kabat et al (ref. 7)].

2.1.1. Heavy Chain - Key residues are 23, 71 and 73. Other residues which may contribute to a lesser extent are 1, 3 and 76. Finally 25 is usually conserved but the murine residue should be used if there is a difference.

2.1.2 Light Chain - Many residues close to the CDRs, e.g. 63, 65, 67 and 69 are conserved. If conserved none of the surface residues in the light chain are likely to have a major effect. However, if the murine residue at these positions is unusual, then it would be of benefit to analyse the likely contribution more closely. Other residues which may also contribute to binding are 1 and 3, and also 60 and 70 if the residues at these positions and at 54 and 24 respectively are potentially able to form a salt bridge i.e. 60 + 54; 70 + 24.

2.2 Packing residues near the CDRs.

2.2.1. Heavy Chain - Key residues are 24, 49 and 78. Other key residues would be 36 if not a tryptophan, 94 if not an arginine, 104 and 106 if not glycines and 107 if not a threonine. Residues which may make a further contribution to stable packing of the heavy chain and hence improved affinity are 2, 4, 6, 38, 46, 67 and 69. 67 packs against the CDR residue 63 and this pair could be either both mouse or both human. Finally, residues which contribute to packing in this region but from a longer range are 18, 20, 80, 82 and 86. 82 packs against 67 and in turn 18 packs against 82. 80 packs against 69 and in turn 20 packs against 80. 86 forms an H bond network with 38 and 46. Many of the mouse-human differences appear minor e.g. Leu-Ile, but could have an minor impact on correct packing which could translate into altered positioning of the CDRs.

2.2.2. Light Chain - Key residues are 48, 58 and 71. Other key residues would be 6 if not glutamine, 35 if not tryptophan, 62 if not phenylalanine or tryosine, 64, 66, 68, 99 and 101 if not glycines and 102 if not a threonine. Residues which make a further contribution are 2, 4, 37, 45 and 47. Finally residues 73 and 21 and 19 may make long distance packing contributions of a minor nature.

2.3. Residues at the variable domain interface between heavy and light chains - In both the light and heavy chains most of the non-CDR interface residues are conserved. If a conserved residue is replaced by a residue of different character, e.g. size or charge, it should be considered for retention as the murine residue.

2.3.1. Heavy Chain - Residues which need to be considered are 37 if the residue is not a valine but is of larger side chain volume or has a charge or polarity. Other residues are 39 if not a glutamine, 45 if not a leucine, 47 if not a tryptophan, 91 if not a phenylalanine or tyrosine, 93 if not an alanine and 103 if not a tryptophan. Residue 89 is also at the interface but is not in a position where the side chain could be of great impact.

2.3.2. Light Chain - Residues which need to be considered are 36, if not a tyrosine, 38 if not a glutamine, 44 if not a proline, 46, 49 if not a tyrosine, residue 85, residue 87 if not a tyrosine and 98 if not a phenylalanine.

2.4. Variable-Constant region interface - The elbow angle between variable and constant regions may be affected by alterations in packing of key residues in the variable region against the constant region which may affect the position of V_L and V_H with respect to one another. Therefore it is worth noting the residues likely to be in contact with the constant region. In the heavy chain the surface residues potentially in contact with the variable region are conserved between mouse and human antibodies therefore the variable region contact residues may influence the V-C interaction. In the light chain the amino acids found at a number of the constant region contact points vary, and the V & C regions are not in such close proximity as the heavy chain. Therefore the influences of the light chain V-C interface may be minor.

2.4.1. Heavy Chain - Contact residues are 7, 11, 41, 87, 108, 110, 112.

2.4.2. Light Chain - In the light chain potentially contacting residues are 10, 12, 40, 80, 83, 103 and 105.

The above analysis coupled with our considerable practical experimental experience in the CDR-grafting of a number of different antibodies have lead us to the protocol given above.

The present invention is now described, by way of example only, with reference to the accompanying Figures 1 - 13.

Brief Description of the Figures

Figure 1 shows DNA and amino acid sequences of the OKT3 light chain;
Figure 2 shows DNA and amino acid sequences of the OKT3 heavy chain;
Figure 3 shows the alignment of the OKT3 light variable region amino acid sequence with that of the light variable region of the human antibody REI;
Figure 4 shows the alignment of the OKT3 heavy variable region amino acid sequence with that of the heavy variable region of the human antibody KOL;
Figure 5 shows the heavy variable region amino acid sequences of OKT3, KOL and various corresponding CDR grafts;
Figure 6 shows the light variable region amino acid sequences of OKT3, REI and various corresponding CDR grafts;
Figure 7 shows a graph of binding assay results for various grafted OKT3 antibodies'
Figure 8 shows a graph of blocking assay results for various grafted OKT3 antibodies;
Figure 9 shows a similar graph of blocking assay results;
Figure 10 shows similar graphs for both binding assay and blocking assay results;
Figure 11 shows further similar graphs for both binding assay and blocking assay results;
Figure 12 shows a graph of competition assay results for a minimally grafted OKT3 antibody compared with the OKT3 murine reference standard, and
Figure 13 shows a similar graph of competition assay results comparing a fully grafted OKT3 antibody with the murine reference standard.

DETAILED DESCRIPTION OF EMBODIMENTS OF THE INVENTION

EXAMPLE 1

CDR-GRAFTING OF OKT3

MATERIAL AND METHODS

1. INCOMING CELLS

Hybridoma cells producing antibody OKT3 were provided by Ortho (seedlot 4882.1) and were grown up in antibiotic free Dulbecco's Modified Eagles Medium (DMEM) supplemented with glutamine and 5% foetal calf serum, and divided to provide both an overgrown supernatant for evaluation and cells for extraction of RNA. The overgrown supernatant was shown to contain 250 ug/mL murine IgG2a/kappa antibody. The supernatant was negative for murine lambda light chain and IgG1, IgG2b, IgG3, IgA and IgM heavy chain. 20mL of supernatant was assayed to confirm that the antibody present was OKT3.

2. MOLECULAR BIOLOGY PROCEDURES

Basic molecular biology procedures were as described in Maniatis et al (ref. 9) with, in some cases, minor modifications. DNA sequencing was performed as described in Sanger et al (ref. 11) and the Amersham International Plc sequencing handbook. Site directed mutagenesis was as described in Kramer et al (ref. 12) and the Anglian Biotechnology Ltd. handbook. COS cell expression and metabolic labelling studies were as described in Whittle et al (ref. 13)

3. RESEARCH ASSAYS

3.1. ASSEMBLY ASSAYS

Assembly assays were performed on supernatants from transfected COS cells to determine the amount of intact IgG present.

3.1.1. COS CELLS TRANSFECTED WITH MOUSE OKT3 GENES

The assembly assay for intact mouse IgG in COS cell supernatants was an ELISA with the following format:

96 well microtitre plates were coated with $F(ab')_2$ goat anti-mouse IgG Fc. The plates were washed in water and samples added for 1 hour at room temperature. The plates were washed and $F(ab')_2$ goat anti-mouse IgG $F(ab')_2$ (HRPO conjugated) was then added. Substrate was added to reveal the reaction. UPC10, a mouse IgG2a myeloma, was used as a standard.

3.1.2. COS AND CHO CELLS TRANSFECTED WITH CHIMERIC OR CDR-GRAFTED OKT3 GENES

The assembly assay for chimeric or CDR-grafted antibody in COS cell supernatants was an ELISA with the following format:

96 well microtitre plates were coated with F(ab')$_2$ goat anti-human IgG Fc. The plates were washed and samples added and incubated for 1 hour at room temperature. The plates were washed and monoclonal mouse anti-human kappa chain was added for 1 hour at room temperature.

The plates were washed and F(ab')$_2$ goat anti-mouse IgG Fc (HRPO conjugated) was added. Enzyme substrate was added to reveal the reaction.

Chimeric B72.3 (IgG4) (ref. 13) was used as a standard. The use of a monoclonal anti-kappa chain in this assay allows grafted antibodies to be read from the chimeric standard.

3.2. ASSAY FOR ANTIGEN BINDING ACTIVITY

Material from COS cell supernatants was assayed for OKT3 antigen binding activity onto CD3 positive cells in a direct assay. The procedure was as follows:

HUT 78 cells (human T cell line, CD3 positive) were maintained in culture. Monolayers of HUT 78 cells were prepared onto 96 well ELISA plates using poly-L-lysine and glutaraldehyde. Samples were added to the monolayers for 1 hour at room temperature.

The plates were washed gently using PBS. F(ab')$_2$ goat anti-human IgG Fc (HRPO conjugated) or F(ab')$_2$ goat anti-mouse IgG Fc (HRPO conjugated) was added as appropriate for humanised or mouse samples. Substrate was added to reveal the reaction.

The negative control for the cell-based assay was chimeric B72.3. The positive control was mouse Orthomune OKT3 or chimeric OKT3, when available. This cell-based assay was difficult to perform, and an alternative assay was developed for CDR-grafted OKT3 which was more sensitive and easier to carry out.

In this system CDR-grafted OKT3 produced by COS cells was tested for its ability to bind to the CD3-positive HPB-ALL (human peripheral blood acute lymphocytic leukemia) cell line. It was also tested for its ability to block the binding of murine OKT3 to these cells. Binding was measured by the following procedure: HPB-ALL cells were harvested from tissue culture. Cells were incubated at 4°C for 1 hour with various dilutions of test antibody, positive control antibody, or negative control antibody. The cells were washed once and incubated at 4°C for 1 hour with an FITC-labelled goat anti-human IgG (Fc-specific, mouse absorbed). The cells were washed twice and analysed by cytofluorography. Chimeric OKT3 was used as a positive control for direct binding. Cells incubated with mock- transfected COS cell supernatant, followed by the FITC-labelled goat anti-human IgG, provided the negative control. To test the ability of CDR-grafted OKT3 to block murine OKT3 binding, the HPB-ALL cells were incubated at 4°C for 1 hour with various dilutions of test antibody or control antibody. A fixed saturating amount of FITC OKT3 was added. The samples were incubated for 1 hour at 4°C, washed twice and analysed by cytofluorography.

FITC-labelled OKT3 was used as a positive control to determine maximum binding. Unlabelled murine OKT3 served as a reference standard for blocking. Negative controls were unstained cells with or without mock-transfected cell supernatant. The ability of the CDR-grafted OKT3 light chain to bind CD3-positive cells and block the binding of murine OKT3 was initially tested in combination with the chimeric OKT3 heavy chain. The chimeric OKT3 heavy chain is composed of the murine OKT3 variable region and the human IgG4 constant region. The chimeric heavy chain gene is expressed in the same expression vector used for the CDR-grafted genes. The CDR-grafted light chain expression vector and the chimeric heavy chain expression vector were co-transfected into COS cells. The fully chimeric OKT3 antibody (chimeric light chain and chimeric heavy chain) was found to be fully capable of binding to CD3 positive cells and blocking the binding of murine OKT3 to these cells.

3.3 DETERMINATION OF RELATIVE BINDING AFFINITY

The relative binding affinities of CDR-grafted anti-CD3 monoclonal antibodies were determined by competition binding (ref. 6) using the HPB-ALL human T cell line as a source of CD3 antigen, and fluorescein-conjugated murine OKT3 (Fl-OKT3) of known binding affinity as a tracer antibody. The binding affinity of Fl-OKT3 tracer antibody was determined by a direct binding assay in which increasing amounts of Fl-OKT3 were incubated with HPB-ALL ($5x10^5$) in PBS with 5% foetal calf serum for 60 min. at 4°C. Cells were washed, and the fluorescence intensity was determined on a FACScan flow cytometer calibrated with quantitative microbead standards (Flow Cytometry Standards, Research Triangle Park, NC). Fluorescence intensity per antibody molecule (F/P ratio) was determined by using microbeads which have a predetermined number of mouse IgG antibody binding sites (Simply Cellular beads, Flow Cytometry Standards). F/P equals the fluorescence intensity of beads saturated with Fl-OKT3 divided by the number of binding sites per bead. The amount of bound and free Fl-OKT3 was calculated from the mean fluorescence intensity per cell, and the ratio of bound/free was plotted against the number of moles of antibody bound. A linear fit was used to determine the affinity of binding (absolute value of the slope).

For competitive binding, increasing amounts of competitor antibody were added to a sub-saturating dose of FI-OKT3 and incubated with $5x10^5$ HPB-ALL in 200 ml of PBS with 5% foetal calf serum, for 60 min at 4°C. The fluorescence intensities of the cells were measured on a FACScan flow cytometer calibrated with quantitative microbead standards. The concentrations of bound and free FI-OKT3 were calculated. The affinities of competing antibodies were calculated from the equation [X]-[OKT3] = (1/Kx) - (1/Ka), where Ka is the affinity of murine OKT3, Kx is the affinity of competitor X, [ ] is the concentration of competitor antibody at which bound/free binding is R/2, and R is the maximal bound/free binding.

4. cDNA LIBRARY CONSTRUCTION

4.1. mRNA PREPARATION AND cDNA SYNTHESIS

OKT3 producing cells were grown as described above and 1.2×10^9 cells harvested and mRNA extracted using the guanidinium/LiCl extraction procedure. cDNA was prepared by priming from Oligo-dT to generate full length cDNA. The cDNA was methylated and EcoR1 linkers added for cloning.

4.2. LIBRARY CONSTRUCTION

The cDNA library was ligated to pSP65 vector DNA which had been EcoR1 cut and the 5' phosphate groups removed by calf intestinal phosphatase (EcoR1/CIP). The ligation was used to transform high transformation efficiency Escherichia coli (E.coli) HB101. A cDNA library was prepared. 3600 colonies were screened for the light chain and 10000 colonies were screened for the heavy chain.

5. SCREENING

E.coli colonies positive for either heavy or light chain probes were identified by oligonucleotide screening using the oligonucleotides:

5' TCCAGATGTTAACTGCTCAC for the light chain, which is complementary to a sequence in the mouse kappa constant region, and 5' CAGGGGCCAGTGGATGGATAGAC for the heavy chain which is complementary to a sequence in the mouse IgG2a constant CH1 domain region. 12 light chain and 9 heavy chain clones were identified and taken for second round screening. Positive clones from the second round of screening were grown up and DNA prepared. The sizes of the gene inserts were estimated by gel electrophoresis and inserts of a size capable of containing a full length cDNA were subcloned into M13 for DNA sequencing.

6. DNA SEQUENCING

Clones representing four size classes for both heavy and light chains were obtained in M13. DNA sequence for the 5' untranslated regions, signal sequences, variable regions and 3' untranslated regions of full length cDNAs [Figures 1(a) and 2(a)] were obtained and the corresponding amino acid sequences predicted [(Figures 1(b) and 2(b)]. In Figure 1(a) the untranslated DNA regions are shown in uppercase, and in both Figures 1 and 2 the signal sequences are underlined.

7. CONSTRUCTION OF cDNA EXPRESSION VECTORS

Celltech expression vectors are based on the plasmid pEE6hCMV (ref. 14). A polylinker for the insertion of genes to be expressed has been introduced after the major immediate early promoter/enhancer of the human Cytomegalovirus (hCMV). Marker genes for selection of the plasmid in transfected eukaryotic cells can be inserted as BamH1 cassettes in the unique BamH1 site of pEE6 hCMV; for instance, the neo marker to provide pEE6 hCMV neo. It is usual practice to insert the neo and gpt markers prior to insertion of the gene of interest, whereas the GS marker is inserted last because of the presence of internal EcoR1 sites in the cassette.

The selectable markers are expressed from the SV40 late promoter which also provides an origin of replication so that the vectors can be used for expression in the COS cell transient expression system.

The mouse sequences were excised from the M13 based vectors described above as EcoR1 fragments and cloned into either pEE6-hCMV-neo for the heavy chain and into EE6-hCMV-gpt for the light chain to yield vectors pJA136 and pJA135 respectively.

8. EXPRESSION OF cDNAS IN COS CELLS

Plasmids pJA135 and pJA136 were co-transfected into COS cells and supernatant from the transient expression experiment was shown to contain assembled antibody which bound to T-cell enriched lymphocytes. Metabolic labelling experiments using ^{35}S methionine showed expression and assembly of heavy and light chains.

9. CONSTRUCTION OF CHIMERIC GENES

Construction of chimeric genes followed a previously described strategy [Whittle et al (ref. 13)]. A restriction site near the 3' end of the variable domain sequence is identified and used to attach an oligonucleotide adapter coding for the remainder of the mouse variable region and a suitable restriction site for attachment to the constant region of choice.

9.1. LIGHT CHAIN GENE CONSTRUCTION

The mouse light chain cDNA sequence contains an AvaI site near the 3' end of the variable region [Fig. 1(a)]. The majority of the sequence of the variable region was isolated as a 396 bp. EcoR1-AvaI fragment. An oligonucleotide adapter was designed to replace the remainder of the 3' region of the variable region from the AvaI site and to include the 5' residues of the human constant region up to and including a unique Nar1 site which had been previously engineered into the constant region.

A Hind111 site was introduced to act as a marker for insertion of the linker.

The linker was ligated to the V_L fragment and the 413 bp EcoR1-Nar1 adapted fragment was purified from the ligation mixture.

The constant region was isolated as an Nar1-BamH1 fragment from an M13 clone NW361 and was ligated with the variable region DNA into an EcoR1/BamH1/C1P pSP65 treated vector in a three way reaction to yield plasmid JA143. Clones were isolated after transformation into E.coli and the linker and junction sequences were confirmed by the presence of the Hind111 site and by DNA sequencing.

9.2 LIGHT CHAIN GENE CONSTRUCTION - VERSION 2

The construction of the first chimeric light chain gene produces a fusion of mouse and human amino acid sequences at the variable-constant region junction. In the case of the OKT3 light chain the amino acids at the chimera junction are:

```
.....Leu-Glu-Ile-Asn-Arg/   -/Thr-Val-Ala   -Ala
                  VARIABLE         CONSTANT
```

This arrangement of sequence introduces a potential site for Asparagine (Asn) linked (N-linked) glycosylation at the V-C junction. Therefore, a second version of the chimeric light chain oligonucleotide adapter was designed in which the threonine (Thr), the first amino acid of the human constant region, was replaced with the equivalent amino acid from the mouse constant region, Alanine (Ala).

An internal Hind111 site was not included in this adapter, to differentiate the two chimeric light chain genes.

The variable region fragment was isolated as a 376 bp EcoR1-AvaI fragment. The oligonucleotide linker was ligated to Nar1 cut pNW361 and then the adapted 396bp constant region was isolated after recutting the modified pNW361 with EcoR1. The variable region fragment and the modified constant region fragment were ligated directly into EcoR1/C1P treated pEE6hCMVneo to yield pJA137. Initially all clones examined had the insert in the incorrect orientation. Therefore, the insert was re-isolated and recloned to turn the insert round and yield plasmid pJA141. Several clones with the insert in the correct orientation were obtained and the adapter sequence of one was confirmed by DNA sequencing

9.3. HEAVY CHAIN GENE CONSTRUCTION

9.3.1. CHOICE OF HEAVY CHAIN GENE ISOTYPE

The constant region isotype chosen for the heavy chain was human IgG4.

9.3.2. GENE CONSTRUCTION

The heavy chain cDNA sequence showed a Ban1 site near the 3' end of the variable region [Fig. 2-(a)]. The majority of the sequence of the variable region was isolated as a 426bp. EcoR1/C1P/Ban1 fragment. An oligonucleotide adapter was designated to replace the remainder of the 3' region of the variable region from the Ban1 site up to and including a unique HindIII site which had been previously engineered into the first two amino acids of the constant region.

The linker was ligated to the V_H fragment and the EcoR1-Hind111 adapted fragment was purified from the ligation mixture.

The variable region was ligated to the constant region by cutting pJA91 with EcoR1 and Hind111 removing the intron fragment and replacing it with the V_H to yield pJA142. Clones were isolated after transformation into E.coli JM101 and the linker and junction sequences were confirmed by DNA sequencing. (N.B. The Hind111 site is lost on cloning).

10. CONSTRUCTION OF CHIMERIC EXPRESSION VECTORS

10.1. neo AND gpt VECTORS

The chimeric light chain (version 1) was removed from pJA143 as an EcoR1 fragment and cloned into EcoR1/C1P treated pEE6hCMVneo expression vector to yield pJA145. Clones with the insert in the

correct orientation were identified by restriction mapping.

The chimeric light chain (version 2) was constructed as described above.

The chimeric heavy chain gene was isolated from pJA142 as a 2.5Kbp EcoR1/BamH1 fragment and cloned into the EcoR1/Bcl1/C1P treated vector fragment of a derivative of pEE6hCMVgpt to yield plasmid pJA144.

10.2. GS SEPARATE VECTORS

GS versions of pJA141 and pJA144 were constructed by replacing the neo and gpt cassettes by a BamH1/Sa11/C1P treatment of the plasmids, isolation of the vector fragment and ligation to a GS-containing fragment from the plasmid pRO49 to yield the light chain vector pJA179 and the heavy chain vector pJA180.

10.3. GS SINGLE VECTOR CONSTRUCTION

Single vector constructions containing the cL (chimeric light), cH (chimeric heavy) and GS genes on one plasmid in the order cL-cH-GS, or cH-cL-GS and with transcription of the genes being head to tail e.g. cL>cH>GS were constructed. These plasmids were made by treating pJA179 or pJA180 with BamH1/C1P and ligating in a Bgl11/Hind111 hCMV promoter cassette along with either the Hind111/BamH1 fragment from pJA141 into pJA180 to give the cH-cL-GS plasmid pJA182 or the Hind111/BamH1 fragment from pJA144 into pJA179 to give the cL-cH-GS plasmid pJA181.

11. EXPRESSION OF CHIMERIC GENES

11.1. EXPRESSION IN COS CELLS

The chimeric antibody plasmid pJA145 (cL) and pJA144 (cH) were co-transfected into COS cells and supernatant from the transient expression experiment was shown to contain assembled antibody which bound to the HUT 78 human T-cell line. Metabolic labelling experiments using ^{35}S methionine showed expression and assembly of heavy and light chains. However the light chain mobility seen on reduced gels suggested that the potential glycosylation site was being glycosylated. Expression in COS cells in the presence of tunicamycin showed a reduction in size of the light chain to that shown for control chimeric antibodies and the OKT3 mouse light chain. Therefore JA141 was constructed and expressed. In this case the light chain did not show an aberrant mobility or a size shift in the presence or absence of tunicamycin. This second version of the chimeric light chain, when expressed in association with chimeric heavy (cH) chain, produced antibody which showed good binding to HUT 78 cells. In both cases antigen binding was equivalent to that of the mouse antibody.

11.2 EXPRESSION IN CHINESE HAMSTER OVARY (CHO) CELLS

Stable cell lines have been prepared from plasmids PJA141/pJA144 and from pJA179/pJA180, pJA181 and pJA182 by transfection into CHO cells.

12. CDR-GRAFTING

The approach taken was to try to introduce sufficient mouse residues into a human variable region framework to generate antigen binding activity comparable to the mouse and chimeric antibodies.

12.1. VARIABLE REGION ANALYSIS

From an examination of a small database of structures of antibodies and antigen-antibody complexes it is clear that only a small number of antibody residues make direct contact with antigen. Other residues may contribute to antigen binding by positioning the contact residues in favourable configurations and also by inducing a stable packing of the individual variable domains and stable interaction of the light and heavy chain variable domains. The residues chosen for transfer can be identified in a number of ways:

(a) By examination of antibody X-ray crystal structures the antigen binding surface can be predominantly located on a series of loops, three per domain, which extend from the B-barrel framework.

(b) By analysis of antibody variable domain sequences regions of hypervariability [termed the Complementarity Determining Regions (CDRs) by Wu and Kabat (ref: 5)] can be identified: In the most but not all cases these CDRs correspond to, but extend a short way beyond, the loop regions noted above.

(c) Residues not identified by (a) and (b) may contribute to antigen binding directly or indirectly by affecting antigen binding site topology, or by inducing a stable packing of the individual variable domains and stabilising the inter-variable domain interaction. These residues may be identified either by superimposing the sequences for a given antibody on a known structure and looking at key residues for their contribution, or by sequence alignment analysis and noting "idiosyncratic" residues followed by examination of their structural location and likely effects.

12.1.1. LIGHT CHAIN

Figure 3 shows an alignment of sequences for the human framework region RE1 and the OKT3 light

variable region. The structural loops (LOOP) and CDRs (KABAT) believed to correspond to the antigen binding region are marked. Also marked are a number of other residues which may also contribute to antigen binding as described in 13.1(c). Above the sequence in Figure 3 the residue type indicates the spatial location of each residue side chain, derived by examination of resolved structures from X-ray crystallography analysis. The key to this residue type designation is as follows:

N - near to CDR (From X-ray Structures)
P - Packing B - Buried Non-Packing
S - Surface E - Exposed
I - Interface * - Interface - Packing/Part Exposed
? - Non-CDR Residues which may require to be left as Mouse sequence.

Residues underlined in Figure 3 are amino acids. RE1 was chosen as the human framework because the light chain is a kappa chain and the kappa variable regions show higher homology with the mouse sequences than a lambda light variable region, e.g. KOL (see below). RE1 was chosen in preference to another kappa light chain because the X-ray structure of the light chain has been determined so that a structural examination of individual residues could be made.

12.1.2. HEAVY CHAIN

Similarly Figure 4 shows an alignment of sequences for the human framework region KOL and the OKT3 heavy variable region. The structural loops and CDRs believed to correspond to the antigen binding region are marked. Also marked are a number of other residues which may also contribute to antigen binding as described in 12.1(c). The residue type key and other indicators used in Figure 4 are the same as those used in Figure 3. KOL was chosen as the heavy chain framework because the X-ray structure has been determined to a better resolution than, for example, NEWM and also the sequence alignment of OKT3 heavy variable region showed a slightly better homology to KOL than to NEWM.

12.2. DESIGN OF VARIABLE GENES

The variable region domains were designed with mouse variable region optimal codon usage [Grantham and Perrin (ref. 15)] and used the B72.3 signal sequences [Whittle et al (ref. 13)]. The sequences were designed to be attached to the constant region in the same way as for the chimeric genes described above. Some constructs contained the "Kozak consensus sequence" [Kozak (ref. 16)] directly linked to the 5' of the signal sequence in the gene. This sequence motif is believed to have a beneficial role in translation initiation in eukaryotes.

12.3. GENE CONSTRUCTION

To build the variable regions, various strategies are available. The sequence may be assembled by using oligonucleotides in a manner similar to Jones et al (ref. 17) or by simultaneously replacing all of the CDRs or loop regions by oligonucleotide directed site specific mutagenesis in a manner similar to Verhoeyen et al (ref. 2). Both strategies were used and a list of constructions is set out in Tables 1 and 2 and Figures 4 and 5. It was noted in several cases that the mutagenesis approach led to deletions and rearrangements in the gene being remodelled, while the success of the assembly approach was very sensitive to the quality of the oligonucleotides.

13. CONSTRUCTION OF EXPRESSION VECTORS

Genes were isolated from M13 or SP65 based intermediate vectors and cloned into pEE6hCMVneo for the light chains and pEE6hCMVgpt for the heavy chains in a manner similar to that for the chimeric genes as described above.

TABLE 1 CDR-GRAFTED GENE CONSTRUCTS

CODE	MOUSE SEQUENCE CONTENT	METHOD OF CONSTRUCTION	KOZAK SEQUENCE -	KOZAK SEQUENCE +
LIGHT CHAIN	ALL HUMAN FRAMEWORK REI			
121	26-32, 50-56, 91-96 inclusive	SDM and gene assembly	+	n.d.
121A	26-32, 50-56, 91-96 inclusive +1, 3, 46, 47	Partial gene assembly	n.d.	+
121B	26-32, 50-56, 91-96 inclusive + 46, 47	Partial gene assembly	n.d.	+
221	24-24, 50-56, 91-96 inclusive	Partial gene assembly	+	+
221A	24-34, 50-56, 91-96 inclusive +1, 3, 46, 47	Partial gene assembly	+	+
221B	24-34, 50-56, 91-96 inclusive +1, 3	Partial gene assembly	+	+
221C	24-34, 50-56, 91-96 inclusive	Partial gene assembly	+	+
HEAVY CHAIN	ALL HUMAN FRAMEWORK KOL			
121	26-32, 50-56, 95-100B inclusive	Gene assembly	n.d.	+
131	26-32, 50-58, 95-100B inclusive	Gene assembly	n.d.	+
141	26-32, 50-65, 95-100B inclusive	Partial gene assembly	+	n.d.
321	26-35, 50-56, 95-100B inclusive	Partial gene assembly	+	n.d.
331	26-35, 50-58, 95-100B inclusive	Partial gene assembly	+	
		Gene assembly		+
341	26-35, 50-65, 95-100B inclusive	SDM	+	
		Partial gene assembly		+
341A	26-35, 50-65, 95-100B inclusive +6, 23, 24, 48, 49, 71, 73, 76, 78, 88, 91 (+63 - human)	Gene assembly	n.d.	+
341B	26-35, 50-65, 95-100B inclusive + 48, 49, 71, 73, 76, 78, 88, 91 (+63 + human)	Gene assembly	n.d.	+

KEY

n.d.	not done
SDM	Site directed mutagenesis
Gene assembly	Variable region assembled entirely from oligonucleotides
Partial gene assembly	Variable region assembled by combination of restriction fragments either from other genes originally created by SDM and gene assembly or by oligonucleotide assembly of part of the variable region and reconstruction with restriction fragments from other genes originally created by SDM and gene assembly

14. EXPRESSION OF CDR-GRAFTED GENES

14.1. PRODUCTION OF ANTIBODY CONSISTING OF GRAFTED LIGHT (gL) CHAINS WITH MOUSE HEAVY (mH) OR CHIMERIC HEAVY (cH) CHAINS

All gL chains, in association with mH or cH produced reasonable amounts of antibody. Insertion of the Kozak consensus sequence at a position 5' to the ATG (kgL constructs) however, led to a 2-5 fold

improvement in net expression. Over an extended series of experiments expression levels were raised from approximately 200ng/ml to approximately 500 ng/ml for kgL/cH or kgL/mH combinations.

When direct binding to antigen on HUT 78 cells was measured, a construct designed to include mouse sequence based on loop length (gL121) did not lead to active antibody in association with mH or cH. A construct designed to include mouse sequence based on Kabat CDRs (gL221) demonstrated some weak binding in association with mH or cH. However, when framework residues 1, 3, 46, 47 were changed from the human to the murine OKT3 equivalents based on the arguments outlined in Section 12.1 antigen binding was demonstrated when both of the new constructs, which were termed 121A and 221A were co-expressed with cH. When the effects of these residues were examined in more detail, it appears that residues 1 and 3 are not major contributing residues as the product of the gL221B gene shows little detectable binding activity in association with cH. The light chain product of gL221C, in which mouse sequences are present at 46 and 47, shows good binding activity in association with cH.

14.2 PRODUCTION OF ANTIBODY CONSISTING OF GRAFTED HEAVY (gH) CHAINS WITH MOUSE LIGHT (mL) OR CHIMERIC LIGHT (cL) CHAINS

Expression of the gH genes proved to be more difficult to achieve than for gL. First, inclusion of the Kozak sequence appeared to have no marked effect on expression of gH genes. Expression appears to be slightly improved but not to the same degree as seen for the grafted light chain.

Also, it proved difficult to demonstrate production of expected quantities of material when the loop choice (amino acid 26-32) for CDR1 is used, e.g. gH121, 131, 141 and no conclusions can be drawn about these constructs.

Moreover, co-expression of the gH341 gene with cL or mL has been variable and has tended to produce lower amounts of antibody than the cH/cL or mH/mL combinations. The alterations to gH341 to produce gH341A and gH341B lead to improved levels of expression.

This may be due either to a general increase in the fraction of mouse sequence in the variable region, or to the alteration at position 63 where the residue is returned to the human amino acid Valine (Val) from Phenylalanine (Phe) to avoid possible internal packing problems with the rest of the human framework. This arrangement also occurs in gH331 and gH321.

When gH321 or gH331 were expressed in association with cL, antibody was produced but antibody binding activity was not detected.

When the more conservative gH341 gene was used antigen binding could be detected in association with cL or mL, but the activity was only marginally above the background level.

When further mouse residues were substituted based on the arguments in 12.1, antigen binding could be clearly demonstrated for the antibody produced when kgH341A and kgH341B were expressed in association with cL.

14.3 PRODUCTION OF FULLY CDR-GRAFTED ANTIBODY

The kgL221A gene was co-expressed with kgH341, kgH341A or kgH341B. For the combination kgH221A/kgH341 very little material was produced in a normal COS cell expression.

For the combinations kgL221A/kgH341A or kgH221A/kgH341B amounts of antibody similar to gL/cH was produced.

In several experiments no antigen binding activity could be detected with kgH221A/gH341 or kgH221A/kgH341 combinations, although expression levels were very low.

Antigen binding was detected when kgL221A/kgH341A or kgH221A/kgH341B combinations were expressed. In the case of the antibody produced from the kgL221A/kgH341A combination the antigen binding was very similar to that of the chimeric antibody.

An analysis of the above results is given below.

15. DISCUSSION OF CDR-GRAFTING RESULTS

In the design of the fully humanised antibody the aim was to transfer the minimum number of mouse amino acids that would confer antigen binding onto a human antibody framework.

15.1. LIGHT CHAIN

15.1.1. EXTENT OF THE CDRs

For the light chain the regions defining the loops known from structural studies of other antibodies to contain the antigen contacting residues, and those hypervariable sequences defined by Kabat et al (refs. 4 and 5) as Complementarity Determining Regions (CDRs) are equivalent for CDR2. For CDR1 the hypervariable region extends from residues 24-34 inclusive while the structural loop extends from 26-32 inclusive. In the case of OKT3 there is only one amino acid difference between the two options, at amino acid 24, where the mouse sequence is a serine and the human framework RE1 has glutamine. For CDR3 the loop extends from residues 91-96 inclusive while the

Kabat hypervariability extends from residues 89-97 inclusive. For OKT3 amino acids 89, 90 and 97 are the same between OKT3 and RE1 (Fig. 3). When constructs based on the loop choice for CDR1 (gL121) and the Kabat choice (gL221) were made and co-expressed with mH or cH no evidence for antigen binding activity could be found for gL121, but trace activity could be detected for the gL221, suggesting that a single extra mouse residue in the grafted variable region could have some detectable effect. Both gene constructs were reasonably well expressed in the transient expression system.

15.1.2. FRAMEWORK RESIDUES

The remaining framework residues were then further examined, in particular amino acids known from X-ray analysis of other antibodies to be close to the CDRs and also those amino acids which in OKT3 showed differences from the consensus framework for the mouse subgroup (subgroup VI) to which OKT3 shows most homology. Four positions 1, 3, 46 and 47 were identified and their possible contribution was examined by substituting the mouse amino acid for the human amino acid at each position.

Therefore gL221A (gL221 + D1Q, Q3V, L46R, L47W, see Figure 3 and Table 1) was made, cloned in EE6hCMVneo and co-expressed with cH (pJA144). The resultant antibody was well expressed and showed good binding activity. When the related genes gL221B (gL221 + D1Q, Q3V) and gL221C (gL221 + L46R, L47W) were made and similarly tested, while both genes produced antibody when co-expressed with cH, only the gL221C/cH combination showed good antigen binding. When the gL121A (gL121 + D1Q, Q3V, L46R, L47W) gene was made and co-expressed with cH, antibody was produced which also bound to antigen.

15.2. HEAVY CHAIN

15.2.1. EXTENT OF THE CDRs

For the heavy chain the loop and hypervariability analyses agree only in CDR3. For CDR1 the loop region extends from residues 26-32 inclusive whereas the Kabat CDR extends from residues 31-35 inclusive. For CDR2 the loop region is from 50-58 inclusive while the hypervariable region covers amino acids 50-65 inclusive. Therefore humanised heavy chains were constructed using the framework from antibody KOL and with various combinations of these CDR choices, including a shorter choice for CDR2 of 50-56 inclusive as there was some uncertainty as to the definition of the end point for the CDR2 loop around residues 56 to 58. The genes were co-expressed with mL or cL initially. In the case of the gH genes with loop choices for CDR1 e.g. gH121, gH131, gH141 very little antibody was produced in the culture supernatants. As no free light chain was detected it was presumed that the antibody was being made and assembled inside the cell but that the heavy chain was aberrant in some way, possibly incorrectly folded, and therefore the antibody was being degraded internally. In some experiments trace amounts of antibody could be detected in ^{35}S labelling studies.

As no net antibody was produced, analysis of these constructs was not pursued further.

When, however, a combination of the loop choice and the Kabat choice for CDR1 was tested (mouse amino acids 26-35 inclusive) and in which residues 31 (Ser to Arg), 33 (Ala to Thr), and 35 (Tyr to His) were changed from the human residues to the mouse residue and compared to the first series, antibody was produced for gH321, kgH331 and kgH341 when co-expressed with cL. Expression was generally low and could not be markedly improved by the insertion of the Kozak consensus sequence 5' to the ATG of the signal sequence of the gene, as distinct from the case of the gL genes where such insertion led to a 2-5 fold increase in net antibody production. However, only in the case of gH341/mL or kgH341/cL could marginal antigen binding activity be demonstrated. When the kgH341 gene was co-expressed with kgL221A, the net yield of antibody was too low to give a signal above the background level in the antigen binding assay.

15.2.2. FRAMEWORK RESIDUES

As in the case of the light chain the heavy chain frameworks were re-examined. Possibly because of the lower initial homology between the mouse and human heavy variable domains compared to the light chains, more amino acid positions proved to be of interest. Two genes kgH341A and kgH341B were constructed, with 11 or 8 human residues respectively substituted by mouse residues compared to gH341, and with the CDR2 residue 63 returned to the human amino acid potentially to improve domain packing. Both showed antigen binding when combined with cL or kgL221A, the kgH341A gene with all 11 changes appearing to be the superior choice.

15.3 INTERIM CONCLUSIONS

It has been demonstrated, therefore, for OKT3 that to transfer antigen binding ability to the humanised antibody, mouse residues outside the CDR regions defined by the Kabat hypervariability or structural

loop choices are required for both the light and heavy chains. Fewer extra residues are needed for the light chain, possibly due to the higher initial homology between the mouse and human kappa variable regions.

Of the changes seven (1 and 3 from the light chain and 6, 23, 71, 73 and 76 from the heavy chain) are predicted from a knowledge of other antibody structures to be either partly exposed or on the antibody surface. It has been shown here that residues 1 and 3 in the light chain are not absolutely required to be the mouse sequence; and for the heavy chain the gH341B heavy chain in combination with the 221A light chain generated only weak binding activity. Therefore the presence of the 6, 23 and 24 changes are important to maintain a binding affinity similar to that of the murine antibody. It was important, therefore, to further study the individual contribution of othe other 8 mouse residues of the kgH341A gene compared to kgH341.

16. FURTHER CDR-GRAFTING EXPERIMENTS

Additional CDR-grafted heavy chain genes were prepared substantially as described above. With reference to Table 2 the further heavy chain genes were based upon the gh341 (plasmid pJA178) and gH341A (plasmid pJA185) with either mouse OKT3 or human KOL residues at 6, 23, 24, 48, 49, 63, 71, 73, 76, 78, 88 and 91, as indicated. The CDR-grafted light chain genes used in these further experiments were gL221, gL221A, gL221B and gL221C as described above.

TABLE 2

OKT3 HEAVY CHAIN CDR GRAFTS

1. gH341 and derivatives

RES NUM	6	23	24	48	49	63	71	73	76	78	88	91	
OKT3vh	Q	K	A	I	G	F	T	K	S	A	A	Y	
gH341	E	S	S	V	A	F	R	N	N	L	G	F	JA178
gH341A	Q	K	A	I	G	V	T	K	S	A	A	Y	JA185
gH341E	Q	K	A	I	G	V	T	K	S	A	G	G	JA198
gH341*	Q	K	A	I	G	V	T	K	N	A	G	F	JA207
gH341*	Q	K	A	I	G	V	R	N	N	A	G	F	JA209
gH341D	Q	K	A	I	G	V	T	K	N	L	G	F	JA197
gH341*	Q	K	A	I	G	V	R	N	N	L	G	F	JA199
gH341C	Q	K	A	V	A	F	R	N	N	L	G	F	JA184
gH341*	Q	S	A	I	G	V	T	K	S	A	A	Y	JA203
gH341*	E	S	A	I	G	V	T	K	S	A	A	Y	JA205
gH341B	E	S	S	I	G	V	T	K	S	A	A	Y	JA183
gH341*	Q	S	A	I	G	V	T	K	S	A	G	F	JA204
gH341*	E	S	A	I	G	V	T	K	S	A	G	F	JA206
gH341*	Q	S	A	I	G	V	T	K	N	A	G	F	JA208
KOL	E	S	S	V	A		R	N	N	L	G	F	

OKT3 LIGHT CHAIN CDR GRAFTS

2. gL221 and derivatives

RES NUM	1	3	46	47	
OKT3vl	Q	V	R	W	
GL221	D	Q	L	L	DA221
gL221A	Q	V	R	W	DA221A
gL221B	Q	V	L	L	DA221B
GL221C	D	Q	R	W	DA221C
RE1	D	Q	L	L	

MURINE RESIDUES ARE UNDERLINED

The CDR-grafted heavy and light chain genes were co-expressed in COS cells either with one another in various combinations but also with the corresponding murine and chimeric heavy and light chain genes substantially as described above. The resultant antibody products were then assayed in binding and blocking assays with HPB-ALL cells as described above.

The results of the assays for various grafted heavy chains co-expressed with the gL221C light chain are given in Figures 7 and 8 (for the JA184, JA185, JA197 and JA198 constructs - see Table 2), in Figure 9 (for the JA183, JA184, JA185 and JA197 constructs) in Figure 10 (for the chimeric, JA185, JA199, JA204, JA205, JA207, JA208 and JA209 constructs) and in Figure 11 (for the JA183, JA184, JA185, JA198, JA203, JA205 and JA206 constructs).

The basic grafted product without any human to murine changes in the variable frameworks, i.e. gL221 co-expressed with gh341 (JA178), and also the "fully grafted" product, having most human to murine changes in the grafted heavy chain framework, i.e. gL221C co-expressed with gh341A (JA185), were assayed for relative binding affinity in a competition assay against murine OKT3 reference standard, using HPB-ALL cells. The assay used was as described above in section 3.3. The results obtained are given in Figure 12 for the basic grafted product and in Figure 13 for the fully grafted product. These results indicate that the basic grafted product has neglibible binding ability as compared with the OKT3 murine reference standard; whereas the "fully grafted" product has a binding ability very similar to that of the OKT3 murine reference standard.

The binding and blocking assay results indicate the following:
The JA198 and JA207 constructs appear to have the best binding characteristics and similar binding abilities, both substantially the same as the chimeric and fully grafted gH341A products. This indicates that positions 88 and 91 and position 76 are not highly critical for maintaining the OKT3 binding ability; whereas at least some of positions 6, 23, 24, 48, 49, 71, 73 and 78 are more important.

This is borne out by the finding that the JA209 and JA199, although of similar binding ability to one another, are of lower binding ability than the JA198 and JA207 constructs. This indicates the importance of having mouse residues at positions 71, 73 and 78, which are either completely or partially human in the JA199 and JA209 constructs respectively.

Moreover, on comparing the results obtained for the JA205 and JA183 constructs it is seen that there is a decrease in binding going from the JA205 to the JA183 constructs. This indicates the importance of retaining a mouse residue at position 23, the only position changed between JA205 and JA183.

These and other results lead us to the conclusion that of the 11 mouse framework residues used in the gH341A (JA185) construct, it is important to retain mouse residues at all of positions 6, 23, 24, 48 and 49, and possibly for maximum binding affinity at 71, 73 and 78.

Similar Experiments were carried out to CDR-graft a number of the rodent antibodies including antibodies having specificity for CD4 (OKT4), ICAM-1 (R6-5), TAG72 (B72.3), and TNFα(61E71, 101.4, hTNF1, hTNF2 and hTNF3).

EXAMPLE 2

CDR-GRAFTING OF A MURINE ANTI-CD4 T CELL RECEPTOR ANTIBODY, OKT4A

Anti OKT4A CDR-grafted heavy and light chain genes were prepared, expressed and tested substantially as described above in Example 1 for CDR-grafted OKT3. The CDR grafting of OKT4A is described in detail in Ortho patent application PCT/GB 90 of even date herewith entitled "Humanised Antibodies". The disclosure of this Ortho patent application PCT/GB 90 is incorporated herein by reference. A number of CDR-grafted OKT4 antibodies have been prepared. Presently the CDR-grafted OKT4A of choice is the combination of the grafted light chain LCDR2 and the grafted heavy chain HCDR10.

THE LIGHT CHAIN

The human acceptor framework used for the grafted light chains was RE1. The preferred LCDR2 light chain has human to mouse changes at positions 33, 34, 38, 49 and 89 in addition to the structural loop CDRs. Of these changed positions, positions 33, 34 and 89 fall within the preferred extended CDRs of the present invention (positions 33 and 34 in CDR1 and position 89 in CDR3). The human to murine changes at positions 38 and 49 corresponds to positions at which the amino acid residues are preferably donor murine amino acid residues in accordance with the present invention.
A comparison of the amino acid sequences of the donor murine light chain variable domain and the RE1 human acceptor light chain variable further reveals that the murine and human residues are identical at all of positions 46, 48 and 71 and at all of positions 2, 4, 6, 35, 36, 44, 47, 62, 64-69, 85, 87, 98, 99 and 101 and 102. However the amino acid residue at position 58 in LCDR2 is the human RE1 framework residue not the mouse OKT4 residue as would be preferred in accordance with the present invention.

THE HEAVY CHAIN

The human acceptor framework used for the grafted heavy chains was KOL.
The preferred CDR graft HCDR10 heavy chain has human to mouse changes at positions 24, 35, 57, 58, 60, 88 and 91 in addition to the structural loop CDRs.
Of these positions, positions 35 (CDR1) and positions 57, 58 and 60 (CDR2) fall within the preferred extended CDRs of the present invention. Also the human to mouse change at position 24 corresponds to a position at which the amino acid residue is a donor murine residue in accordance with the present invention. Moreover, the human to mouse changes at positions 88 and 91 correspond to positions at which the amino acid residues are optionally donor murine residues.
Moreover, a comparison of the murine OKT4A and human KOL heavy chain variable amino acid sequences reveals that the murine and human residues are identical at all of positions 23, 49, 71, 73 and 78 and at all of positions 2, 4, 6, 25, 36, 37, 39, 47, 48, 93, 94, 103, 104, 106 and 107.
Thus the OKT4A CDR-grafted heavy chain HCDR10 corresponds to a particularly preferred embodiment according to the present invention.

EXAMPLE 3

CDR-GRAFTING OF AN ANTI-MUCIN SPECIFIC MURINE ANTIBODY, B72.3

The cloning of the genes coding for the anti-mucin specific murine monoclonal antibody B72.3 and the preparation of B72.3 mouse-human chimeric antibodies has been described previously (ref. 13 and WO 89/01783). CDR-grafted versions of B72.3 were prepared as follows.

(a) B72.3 Light Chain

CDR-grafting of this light chain was accomplished by direct transfer of the murine CDRs into the framework of the human light chain RE1. The regions transferred were:

CDR Number	Residues
1	24-34
2	50-56
3	90-96

The activity of the resulting grafted light chain was assessed by co-expression in COS cells, of genes for the combinations:

B72.3 cH/B72.3 cL

and

B72.3 cH/B72.3 gL

Supernatants were assayed for antibody concentration and for the ability to bind to microtitre plates coated with mucin. The results obtained indicated that, in combination with the B72.3 cH chain, B72.3 cL and B72.3 gL had similar binding properties.

Comparison of the murine B72.3 and REI light chain amino acid sequences reveals that the residues are identical at positions 46, 58 and 71 but are different at position 48. Thus changing the human residue to the donor mouse residue at position 48 may further improve the binding characteristics of the CDR-grafted light chain, (B72.3 gL) in accordance with the present invention.

(b) B72.3 heavy chain

i. Choice of framework

At the outset it was necessary to make a choice of human framework. Simply put, the question was as follows: Was it necessary to use the framework regions from an antibody whose crystal structure was known or could the choice be made on some other criteria?
For B72.3 heavy chain, it was reasoned that, while knowledge of structure was important, transfer of the CDRs from mouse to human frameworks might be facilitated if the overall homology between the donor and receptor frameworks was maximised. Comparison of the B72.3 heavy chain sequence with those in Kabat (ref. 4) for human heavy chains showed clearly that B72.3 had poor homology for KOL and NEWM (for which crystal structures are available) but was very homologous to the heavy chain for EU.
On this basis, EU was chosen for the CDR-grafting and the following residues transferred as CDRs.

CDR Number	Residues
1	27-36
2	50-63
3	93-102

Also it was noticed that the FR4 region of EU was unlike that of any other human (or mouse) antibody. Consequently, in the grafted heavy chain genes this was also changed to produce a "consensus" human sequence. (Preliminary experiments showed that grafted heavy chain genes containing the EU FR4 sequence expressed very poorly in transient expression systems.)

ii. Results with grafted heavy chain genes

Expression of grafted heavy chain genes containing all human framework regions with either gL or cL genes produced a grafted antibody with little ability to bind to mucin. The grafted antibody had about 1% the activity of the chimeric antibody. In these experiments, however, it was noted that the activity of the grafted antibody could be increased to ~ 10% of B72.3 by exposure to pHs of 2-3.5.

This observation provided a clue as to how the activity of the grafted antibody could be improved without acid treatment. It was postulated that acid exposure brought about the protonation of an acidic residue (pKa of aspartic acid = 3.86 and of glutamine acid = 4.25) which in turn caused a change in structure of the CDR loops, or allowed better access of antigen.

From comparison of the sequences of B72.3 (ref. 13) and EU (refs. 4 and 5), it was clear that, in going from the mouse to human frameworks, only two positions had been changed in such a way that acidic residues had been introduced. These positions are at residues 73 and 81, where K to E and Q to E changes had been made, respectively.

Which of these positions might be important was determined by examining the crystal structure of the KOL antibody. In KOL heavy chain, position 81 is far removed from either of the CDR loops.

Position 73, however, is close to both CDRs 1 and 3 of the heavy chain and, in this position it was possible to envisage that a K to E change in this region could have a detrimental effect on antigen binding.

iii. Framework changes in B72.3 gH gene

On the basis of the above analysis, E73 was mutated to a lysine (K). It was found that this change had a dramatic effect on the ability of the grafted Ab to bind to mucin. Further the ability of the grafted B72.3 produced by the mutated gH/gL combination to bind to mucin was similar to that of the B72.3 chimeric antibody.

iv. Other framework changes

In the course of the above experiments, other changes were made in the heavy chain framework regions. Within the accuracy of the assays used, none of the changes, either alone or together, appeared beneficial.

v. Other

All assays used measured the ability of the grafted Ab to bind to mucin and, as a whole, indicated that the single framework change at position 73 is sufficient to generate an antibody with similar binding properties to B72.3. Comparison of the B72.3 murine and EU heavy chain sequences reveals that the mouse and human residues are identical at positions 23, 24, 71 and 78.

Thus the mutated CDR-grafted B72.3 heavy chain corresponds to a preferred embodiment of the present invention.

EXAMPLE 4

CDR-GRAFTING OF A MURINE ANTI-ICAM-1 MONOCLONAL ANTIBODY

A murine antibody, R6-5-D6 (EP 0314863) having specificity for Intercellular Adhesion Molecule 1 (ICAM-1) was CDR-grafted substantially as described above in previous examples. This work is described in greater detail in co-pending application, British Patent Application No. 9009549.8, the disclosure of which is incorporated herein by reference.

The human EU framework was used as the acceptor framework for both heavy and light chains. The CDR-grafted antibody currently of choice is provided by co-expression of grafted light chain gL221A and grafted heavy chain gH341D which has a binding affinity for ICAM 1 of about 75% of that of the corresponding mouse-human chimeric antibody.

LIGHT CHAIN

gL221A has murine CDRs at positions 24-34 (CDR1), 50-56 (CDR2) and 89-97 (CDR3). In addition several framework residues are also the murine amino acid. These residues were chosen after consideration of the possible contribution of these residues to domain packing and stability of the conformation of the antigen binding region. The residues which have been retained as mouse are at positions 2, 3, 48 (?), 60, 84, 85 and 87. Comparison of the murine anti-ICAM 1 and human EU light chain amino acid sequences reveals that the murine and human residues are identical at positions 46, 58 and 71.

HEAVY CHAIN

gH341D has murine CDRs at positions 26-35 (CDR1), 50-56 (CDR2) and94-100B (CDR3). In addition murine residues were used in gH341D at positions 24, 48, 69, 71, 73, 80, 88 and 91. Comparison of the murine anti-ICAM 1 and human EU heavy chain amino acid sequences are identical at positions 23, 49 and 78.

EXAMPLE 5

CDR-Grafting of murine anti-TNFa antibodies

A number of murine anti-TNFa monoclonal antibodies were CDR-grafted substantially as described above in previous examples. These antibodies include the murine monoclonal antibodies designated 61 E71, hTNF1, hTNF3 and 101.4 A brief summary of the CDR-grafting of each of these antibodies is given below.

61E71

A similar analysis as described above (Example 1, Section 12.1.) was done for 61E71 and for the heavy chain 10 residues were identified at 23, 24, 48, 49, 68, 69, 71, 73, 75 and 88 as residues to potentially retain as murine. The human frameworks chosen for CDR-grafting of this antibody, and the hTNF3 and 101.4 antibodies were RE1 for the light chain and KOL for the heavy chain. Three genes were built, the first of which contained 23, 24, 48, 49, 71 and 73 [gH341(6)] as murine residues. The second gene also had 75 and 88 as murine residues [gH341(8)] while the third gene additionally had 68, 69, 75 and 88 as murine residues [gH341(10)]. Each was co-expressed with gL221, the minimum grafted light chain (CDRs only). The gL221/gH341(6) and gL221/gH341(8) antibodies both bound as well to TNF as murine 61E71. The gL221/gH341(10) antibody did not express and this combination was not taken further.
Subsequently the gL221/gH341(6) antibody was assessed in an L929 cell competition assay in which the antibody competes against the TNF receptor on L929 cells for binding to TNF in solution. In this assay the gL221/gH341(6) antibody was approximately 10% as active as murine 61E71.

hTNF1

hTNF1 is a monoclonal antibody which recognises an epitope on human TNF- . The EU human framework was used for CDR-grafting of both the heavy and light variable domains.

Heavy Chain

In the CDR-grafted heavy chain (ghTNF1) mouse CDRs were used at positions 26-35 (CDR1), 50-65 (CDR2) and 95-102 (CDR3). Mouse residues were also used in the frameworks at positions 48, 67, 69, 71, 73, 76, 89, 91, 94 and 108. Comparison of the TNF1 mouse and EU human heavy chain residues reveals that these are identical at positions 23, 24, 29 and 78.

Light Chain

In the CDR-grafted light chain (gLhTNF1) mouse CDRs wre used at positions 24-34 (CDR1), 50-56 (CDR2) and 89-97 (CDR3). In addition mouse residues were used in the frameworks at positions 3, 42, 48, 49, 83, 106 and 108. Comparison of the hTNF1 mouse and EU human light chain residues reveals that these are identical at positions 46, 58 and 71.

The grafted hTNF1 heavy chain was co-expressed with the chimeric light chain and the binding ability of the product compared with that of the chimeric light chain/chimeric heavy chain product in a TNF binding assay. The grafted heavy chain product appeared to have binding ability for TNF slightly better than the fully chimeric product.

Similarly, a grafted heavy chain/grafted light chain product was co-expressed and compared with the fully chimeric product and found to have closely similar binding properties to the latter product.

hTNF3

hTNF3 recognises an epitope on human TNF-α. The sequence of hTNF3 shows only 21 differences compared to 61E71 in the light and heavy chain variable regions, 10 in the light chain (2 in the CDRs at positions 50, 96 and 8 in the framework at 1, 19, 40, 45, 46, 76, 103 and 106) and 11 in the heavy chain (3 in the CDR regions at positions 52, 60 and 95 and 8 in the framework at 1, 10, 38, 40, 67, 73, 87 and 105). The light and heavy chains of the 61E71 and hTNF3 chimeric antibodies can be exchanged without loss of activity in the direct binding assay. However 61E71 is an order of magnitude less able to compete with the TNF receptor on L929 cells for TNF-a compared to hTNF3. Based on the 61E71 CDR grafting data gL221 and gH341(+23, 24, 48, 49 71 and 73 as mouse) genes have been built for hTNF3 and tested and the resultant grafted antibody binds well to TNF-a, but competes very poorly in the L929 assay. It is possible that in this case also the framework residues identified for OKT3 programme may improve the competitive binding ability of this antibody.

101.4

101.4 is a further murine monoclonal antibody able to recognise human TNF-a. The heavy chain of this antibody shows good homology to KOL and so the CDR-grafting has been based on RE1 for the light chain and KOL for the heavy chain. Several grafted heavy chain genes have been constructed with conservative choices for the CDR's (gH341) and which have one or a small number of non-CDR residues at positions 73, 78 or 77-79 inclusive, as the mouse amino acids. These have been co-expressed with cL or gL221. In all cases binding to TNF equivalent to the chimeric antibody is seen and when co-expressed with cL the resultant antibodies are able to compete well in the L929 assay. However, with gL221 the resultant antibodies are at least an order of magnitude less able to compete for TNF against the TNF receptor on L929 cells.

Mouse residues at other positions in the heavy chain, for example, at 23 and 24 together or at 76 have been demonstrated to provide no improvement to the competitive ability of the grafted antibody in the L929 assay.

A number of other antibodies including antibodies having specificity for interleukins e.g. IL1 and cancer markers such as carcinoembryonic antigen (CEA) e.g. the monoclonal antibody A5B7 (ref. 21), have been successfully CDR-grafted according to the present invention.

References

1. Kohler & Milstein, Nature, 265, 295-497, 1975.
2. Chatenoud et al, (1986), J. Immunol. 137, 830-838.
3. Jeffers et al, (1986), Transplantation, 41, 572-578.
4. Begent et al, Br. J. Cancer 62: 487 (1990).
5. Verhoeyen et al, Science, 239, 1534-1536, 1988.
6. Riechmann et al, Nature, 332, 323-324, 1988.
7. Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M., Gottesman, K.S., 1987, in Sequences of Proteins of Immunological Interest, US Department of Health and Human Services, NIH, USA.
8. Wu, T.T., and Kabat, E.A., 1970, J. Exp. Med. 132 211-250.
9. Queen et al, (1989), Proc. Natl. Acad. Sci. USA, 86, 10029-10033 and WO 90/07861
10. Maniatis et al, Molecular Cloning, Cold Spring Harbor, New York, 1989.
11. Primrose and Old, Principles of Gene Manipulation, Blackwell, Oxford, 1980.
12. Sanger, F., Nicklen, S., Coulson, A.R., 1977, Proc. Natl. Acad. Sci. USA, 74 5463
13. Kramer, W., Drutsa, V., Jansen, H.-W., Kramer, B., Plugfelder, M., Fritz, H.-J., 1984, Nucl. Acids Res. 12, 9441
14. Whittle, N., Adair, J., Lloyd, J.C., Jenkins, E., Devine, J., Schlom, J., Raubitshek, A., Colcher, D., Bodmer, M., 1987, Protein Engineering 1, 499.

15. Sikder, S.S., Akolkar, P.N., Kaledas, P.M., Morrison, S.L., Kabat, E.A., 1985, J. Immunol. 135, 4215.
16. Wallick, S.C., Kabat, E.A., Morrison, S.L., 1988, J. Exp. Med. 168, 1099
17. Bebbington, C.R., Published International Patent Application WO 89/01036.
18. Granthan and Perrin 1986, Immunology Today 7, 160.
19. Kozak, M., 1987, J. Mol. Biol. 196, 947.
20. Jones, T.P., Dear, P.H., Foote, J., Neuberger, M.S., Winter, G., 1986, Nature, 321, 522
21. Harwood et al, Br. J. Cancer, 54, 75-82 (1986).

Claims

1. An antibody molecule having affinity for a predetermined antigen and comprising a composite heavy chain and a complementary light chain, said composite heavy chain having a variable domain comprising acceptor antibody heavy chain framework residues and donor antibody heavy chain antigen-binding residues, said donor antibody having affinity for said predetermined antigen, wherein, according to the Kabat numbering system, in said composite heavy chain, amino acid residues 5, 8, 10, 12 to 17, 19, 21, 22, 40, 42 to 44, 66, 68, 70, 74, 77, 79, 81, 83 to 85, 90, 92, 105, 109, 111 and 113 at least are acceptor residues and amino acid residues 23, 24, 31 to 35, 49 to 58, 71, 73, 78 and 95 to 102 at least are donor residues.

2. The antibody molecule of claim 1, wherein amino acid residues 26 to 30 and 59 to 65 in said composite heavy chain are additionally donor residues.

3. The antibody molecule of claim 1 or claim 2, wherein at least one of amino acid residues 1, 3, and 76 in said composite heavy chain are additionally donor residues.

4. The antibody molecule of any one of claims 1 to 3, wherein at least one of amino acid residues 36, 94, 104, 106 and 107 in said composite heavy chain are additionally donor residues.

5. The antibody molecule of claim 4, wherein at least one of amino acid residues 2, 4, 6, 38, 48, 67 and 69 in said composite heavy chain are additionally donor residues.

6. The antibody molecule of any one of claims 1 to 5, wherein amino acid residues 7, 9, 11, 18, 20, 25, 37, 39, 41, 45, 47, 48, 72, 75, 80, 82, 86 to 89, 91, 93, 103, 108, 110 and 112 in said composite heavy chain are additionally acceptor residues.

7. The antibody molecule of any one of claims 1 to 6, wherein said complementary light chain is a composite light chain having a variable domain comprising acceptor antibody light chain framework residues and donor antibody light chain antigen-binding residues, said donor antibody having affinity for said predetermined antigen, wherein, according to the Kabat numbering system, in said composite light chain, amino acid residues 5, 7 to 9, 11, 13 to 18, 20, 22, 23, 39, 41 to 43, 57, 59, 61, 72, 74 to 79, 81, 82, 84, 86, 88, 100, 104, 106 and 107 at least are acceptor residues and amino acid residues 24 to 34, 46, 48, 50 to 56, 58, 71 and 89 to 97 at least are donor residues.

8. The antibody molecule of claim 7, wherein amino acid residues 1, 3 and 47 in said composite light chain are additionally donor residues.

9. The antibody molecule of claim 7 or claim 8, wherein amino acid residues 36, 44, 47, 85 and 87 in said composite light chain are additionally donor residues.

10. The antibody molecule of any one of claims 7 to 9, wherein at least one of amino acid residues 2, 4, 6, 49, 62, 64 to 69, 98, 99, 101 and 102 in said composite light chain are additionally donor residues.

11. The antibody molecule of any one of claims 7 to 11, wherein at least one of amino acid residues 1, 3, 10, 12, 21, 40, 60, 63, 70, 73, 80, 103 and 105 in said composite light chain are additionally donor residues.

12. A therapeutic or diagnostic composition comprising the antibody molecule of any one of claims 1 to 11 in combination with a pharmaceutically acceptable carrier, diluent or excipient.

13. A method for producing a recombinant antigen binding molecule having affinity for a predetermined antigen comprising the steps of:

[1] determining the amino acid sequence of the variable domain of the heavy chain of a donor antibody which has affinity for said predetermined antigen;

[2] determining the amino acid sequence of the variable domain of the heavy chain of a non-specific acceptor antibody;

[3] providing a composite heavy chain for an antibody molecule, said composite heavy chain having acceptor framework residues and donor antigen binding residues wherein, according to the Kabat numbering system, amino acid residues 5, 8, 10, 12 to 17, 19, 21, 22, 40, 42 to 44, 66, 68, 70, 74, 77, 79, 81, 83 to 85, 90, 92, 105, 109, 111 and 113 at least are acceptor residues and amino acid residues 23, 24, 31 to 35, 49 to 58, 71, 73, 78 and 95 to 102 at least are donor residues;

[4] associating the heavy chain produced in step [3] with a complementary light chain to form an antibody molecule;

[5] determining the affinity of the antibody molecule formed in step [4] for said predetermined antigen;

[6] if the affinity determined in step [5] is not equivalent to that of the donor antibody, providing a heavy chain as described in [3] above but in which amino acid residues 71, 73 and 78 are additionally donor residues;

[7] associating the heavy chain produced in step [6] with a complementary light chain to form an antibody molecule;

[8] determining the affinity of the antibody molecule formed in step [7] for said predetermined antigen;

[9] if the affinity determined in step [8] is not equivalent to that of the donor antibody, providing a heavy chain as described in [6] above but in which amino acid residues 26 to 30 are additionally donor residues;

[10] associating the heavy chain produced in step [9] with a complementary light chain to form an antibody molecule;

[11] determining the affinity of the antibody molecule formed in step [10] for said predetermined antigen;

[12] if the affinity determined in step [11] is not equivalent to that of the donor antibody, providing a heavy chain as described in [9] above but in which at least one of amino acid residues 1, 3, and 76 are additionally donor residues;

[13] associating the heavy chain produced in step [12] with a complementary light chain to form an antibody molecule;

[14] determining the affinity of the antibody molecule formed in step [13] for said predetermined antigen;

[15] if the affinity determined in step [14] is not equivalent to that of the donor antibody, providing a heavy chain as described in [12] above but in which at least one of amino acid residues 36, 94, 104, 106, 107 are additionally donor residues;

[16] associating the heavy chain produced in step [15] with a complementary light chain to form an antibody molecule.

[17] determining the affinity of the antibody molecule formed in step [16] for said predetermined antigen;

[18] if the affinity determined in step [17] is not equivalent to that of the donor antibody, providing a heavy chain as described in [15] above but in which at least one of amino acid residues 2, 4, 6, 38, 48, 67 and 69 are additionally donor residues; and

[19] associating the heavy chain produced in step [18] with a complementary light chain to form an antibody molecule.

14. The method of claim 13, further comprising the steps of:

[1] determining the amino acid sequence of the variable domain of the light chain of said donor antibody which has affinity for said predetermined antigen;

[2] determining the amino acid sequence of the variable domain of the light chain of a non-specific acceptor antibody;

[3] providing a composite light chain for an antibody molecule, said composite light chain having acceptor framework residues and donor antigen binding residues wherein, according to the Kabat numbering system, amino acid residues 5, 7 to 9, 11, 13 to 18, 20, 22, 23, 39, 41 to 43, 57, 59, 61, 72, 74 to 79 to 81, 82, 84, 86, 88, 100, 104 and 106 to 109 at least are acceptor residues and amino

acid residues 24 to 34, 46, 48, 50 to 56, 58, 71 and 89 to 97 at least are donor residues;
[4] associating the light chain produced in step [3] with a complementary heavy chain to form an antibody molecule;
[5] determining the affinity of the antibody molecule formed in step [4] for said predetermined antigen;
[6] if the affinity determined in step [5] is not equivalent to that of the donor antibody, providing a light chain as described in [3] above but in which amino acid residues 1, 2, 3 and 47 are additionally donor residues;
[7] associating the light chain produced in step [6] with a complementary heavy chain to form an antigen-binding molecule;
[8] determining the affinity of the antigen-binding molecule formed in step [7] for said predetermined antigen;
[9] if the affinity determined in step [8] is not equivalent to that of the donor antibody, providing a light chain as described in [6] above but in which amino acid residues 36, 44, 47, 85 and 87 are additionally donor residues;
[10] associating the light chain produced in step [9] with a complementary heavy chain to form an antibody molecule;
[11] determining the affinity of the antibody molecule formed in step [10] for said predetermined antigen;
[12] if the affinity determined in step [11] is not equivalent to that of the donor antibody, providing a light chain as described in [9] above but in which at least one of amino acid residues 2, 4, 6, 49, 62, 64 to 69, 98, 99, 101 are additionally donor residues; and
[13] associating the light chain produced in step [9] with a complementary heavy chain to form an antibody molecule.

Patentansprüche

1. Ein Antikörpermolekül, das Affinität für ein vorherbestimmtes Antigen besitzt und eine zusammengesetzte schwere Kette und eine komplementäre leichte Kette umfaßt, wobei diese zusammengesetzte schwere Kette einen variablen Bereich aufweist, der Rahmenreste von Akzeptor-Antikörper-schwerer-Kette und Antigen-Bindungsreste von Donor-Antikörper-schwerer-Kette enthält, wobei dieser Donor-Antikörper Affinität für das genannte vorherbestimmte Antigen besitzt, in welchem Antikörpermolekül, gemäß dem Kabat-Numerierungssystem, in der genannten zusammengesetzten schweren Kette mindestens die Aminosäurereste 5, 8, 10, 12 bis 17, 19, 21, 22, 40, 42 bis 44, 66, 68, 70, 74, 77, 79, 81, 83 bis 85, 90, 92, 105, 109, 111 und 113 Akzeptor-Reste und mindestens die Aminosäurereste 23, 24, 31 bis 35, 49 bis 58, 71, 73, 78 und 95 bis 102 Donor-Reste sind.

2. Das Antikörpermolekül nach Anspruch 1, in welchem die Aminosäurereste 26 bis 30 und 59 bis 65 in der genannten zusammengesetzten schweren Kette zusätzlich Donor-Reste sind.

3. Das Antikörpermolekül nach Anspruch 1 oder Anspruch 2, in welchem mindestens einer der Aminosäurereste 1, 3 und 76 in der genannten zusammengesetzten schweren Kette zusätzlich ein Donor-Rest ist.

4. Das Antikörpermolekül nach einem der Ansprüche 1 bis 3, in welchem mindestens einer der Aminosäurereste 36, 94, 104, 106 und 107 in der genannten zusammengesetzten schweren Kette zusätzlich ein Donor-Rest ist.

5. Das Antikörpermolekül nach Anspruch 4, in welchem mindestens einer der Aminosäurereste 2, 4, 6, 38, 48, 67 und 69 in der genannten zusammengesetzten schweren Kette zusätzlich ein Donor-Rest ist.

6. Das Antikörpermolekül nach einem der Ansprüche 1 bis 5, in welchem die Aminosäurereste 7, 9, 11, 18, 20, 25, 37, 39, 41, 45, 47, 48, 72, 75, 80, 82, 86 bis 89, 91, 93, 103, 108, 110 und 112 in der genannten zusammengesetzten schweren Kette zusätzlich Akzeptor-Reste sind.

7. Das Antikörpermolekül nach einem der Ansprüche 1 bis 6, in welchem die komplementäre leichte Kette eine zusammengesetzte leichte Kette mit einem variablen Bereich ist, der Rahmenreste von Akzeptor-Antikörper-leichter-Kette und Antigen-Bindungsreste von Donor-Antikörpers-leichter-Kette enthält, wobei

dieser Donor-Antikörper Affinität für das genannte vorherbestimmte Antigen besitzt, in welchem Antikörpermolekül, gemäß dem Kabat-Numerierungssystem, in der genannten zusammengesetzten leichten Kette mindestens die Aminosäurereste 5, 7 bis 9, 11, 13 bis 18, 20, 22, 23, 39, 41 bis 43, 57, 59, 61, 72, 74 bis 79, 81, 82, 84, 86, 88, 100, 104, 106 und 107 Akzeptor-Reste und mindestens die Aminosäurereste 24 bis 34, 46, 48, 50 bis 56, 58, 71 und 89 bis 97 Donor-Reste sind.

8. Das Antikörpermolekül nach Anspruch 7, in welchem die Aminosäurereste 1, 3 und 47 in der genannten zusammengesetzten leichten Kette zusätzlich Donor-Reste sind.

9. Das Antikörpermolekül nach Anspruch 7 oder Anspruch 8, in welchem die Aminosäurereste 36, 44, 47, 85 und 87 in der genannten zusammengesetzten leichten Kette zusätzlich Donor-Reste sind.

10. Das Antikörpermolekül nach einem der Ansprüche 7 bis 9, in welchem mindestens einer der Aminosäurereste 2, 4, 6, 49, 62, 64 bis 69, 98, 99, 101 und 102 in der genannten zusammengesetzten leichten Kette zusätzlich ein Donor-Rest ist.

11. Das Antikörpermolekül nach einem der Ansprüche 7 bis 11, in welchem mindestens einer der Aminosäurereste 1, 3, 10, 12, 21, 40, 60, 63, 70, 73, 80, 103 und 105 in der genannten zusammengesetzten leichten Kette zusätzlich ein Donor-Rest ist.

12. Eine therapeutische oder diagnostische Zusammensetzung, die das Antikörpermolekül nach einem der Ansprüche 1 bis 11 in Kombination mit einem pharmazeutisch verwendbaren Träger, Verdünnungsmittel oder Vehikel enthält.

13. Ein Verfahren zur Herstellung eines rekombinanten Antigen-Bindungsmoleküls mit Affinität für ein vorherbestimmtes Antigen, welches Verfahren folgende Schritte umfaßt:

(1) Bestimmung der Aminosäuresequenz des variablen Bereichs der schweren Kette eines Donor-Antikörpers mit Affinität für das genannte vorherbestimmte Antigen;

(2) Bestimmung der Aminosäuresequenz des variablen Bereichs der schweren Kette eines nichtspezifischen Akzeptor-Antikörpers;

(3) Bereitstellung einer zusammengesetzten schweren Kette für ein Antikörpermolekül, wobei diese zusammengesetzte schwere Kette Akzeptor-Rahmenreste und Donor-Antigen-Bindungsreste aufweist, worin, nach dem Kabat-Numerierungssystem mindestens die Aminosäurereste 5, 8, 10, 12 bis 17, 19, 21, 22, 40, 42 bis 44, 66, 68, 70, 74, 77, 79, 81, 83 bis 85, 90, 92, 105, 109, 111 und 113 Akzeptor-Reste und mindestens die Aminosäurereste 23, 24, 31 bis 35, 49 bis 58, 71, 73, 78 und 85 bis 102 Donor-Reste sind;

(4) Assoziierung der in Stufe (3) hergestellten schweren Kette mit einer komplementären leichten Kette zur Bildung eines Antikörpermoleküls;

(5) Bestimmung der Affinität des in Stufe (4) gebildeten Antikörpermoleküls für das genannte vorherbestimmte Antigen;

(6) wenn die in Stufe (5) bestimmte Affinität nicht der des Donor-Antikörpers äquivalent ist, Bereitstellung einer schweren Kette nach der Beschreibung des obigen Punktes (3), in welcher die Aminosäurereste 71, 73 und 78 zusätzlich Donor-Reste sind;

(7) Assoziierung der in Stufe (6) hergestellten schweren Kette mit einer komplementären leichten Kette zur Bildung eines Antikörpermoleküls;

(8) Bestimmung der Affinität des in Stufe (7) gebildeten Antikörpermoleküls für das genannte Antigen;

(9) wenn die in Stufe (8) bestimmte Affinität der des Donor-Antikörpers nicht äquivalent ist, Bereitstellung einer schweren Kette nach der Beschreibung des obigen Punktes (6), in welcher jedoch die Aminosäurereste 26 bis 30 zusätzlich Donor-Reste sind;

(10) Assoziierung der in Stufe (9) hergestellten schweren Kette mit einer komplementären leichten Kette zur Bildung eines Antikörpermoleküls

(11) Bestimmung der Affinität des in Stufe (10) gebildeten Antikörpermoleküls für das genannte vorherbestimmte Antigen;

(12) wenn die in Stufe (11) bestimmte Affinität der des Donor-Antikörpers nicht äquivalent ist, Bereitstellung einer schweren Kette nach der Beschreibung des obigen Punktes (9), in welcher jedoch mindestens einer der Aminosäurereste 1, 3 und 76 zusätzlich ein Donor-Rest ist;

(13) Assoziierung der in Stufe (12) hergestellten schweren Kette mit einer komplementären leichten Kette zur Bildung eines Antikörpermoleküls;
(14) Bestimmung der Affinität des in Stufe (13) gebildeten Antikörpermoleküls für das genannte vorherbestimmte Antigen;
(15) wenn die in Stufe (14) bestimmte Affinität der des Donor-Antikörpers nicht äquivalent ist, Bereitstellung einer schweren Kette nach der Beschreibung des obigen Punktes (12), in welcher jedoch mindestens einer der Aminosäurereste 36, 94, 104, 106, 107 zusätzlich ein Donor-Rest ist;
(16) Assoziierung der in Stufe (15) hergestellten schweren Kette mit einer komplementären leichten Kette zur Bildung eines Antikörpermoleküls;
(17) Bestimmung der Affinität des in Stufe (16) gebildeten Antikörpermoleküls für das genannte vorherbestimmte Antigen;
(18) wenn die in Stufe (17) bestimmte Affinität der des Donor-Antikörpers nicht äquivalent ist, Bereitstellung einer schweren Kette nach der Beschreibung des obigen Punktes (15), in welcher jedoch mindestens einer der Aminosäurereste 2, 4, 6, 38, 48, 67 und 69 zusätzlich ein Donor-Rest ist;
(19) Assoziierung der in Stufe (18) hergestellten schweren Kette mit einer komplementären leichten Kette zur Bildung eines Antikörpermoleküls.

14. Das verfahren nach Anspruch 13, das außerdem folgende Schritte umfaßt:
(1) Bestimmung der Aminosäuresequenz des variablen Bereichs der leichten Kette des genannten Donor-Antikörpers, der Affinität für das genannte vorherbestimmte Antigen hat;
(2) Bestimmung der Aminosäuresequenz des variablen Bereichs der leichten Kette eines nichtspezifischen Akzeptor-Antikörpers;
(3) Bereitstellung einer zusammengesetzten leichten Kette für ein Antikörpermolekül, wobei diese zusammengesetzte leichte Kette Akzeptor-Rahmenreste und Donor-Antigen-Bindungsreste aufweist, worin, nach dem Kabat-Numerierungssystem, mindestens die Aminosäurereste 5, 7 bis 9, 11, 13 bis 18, 20, 22, 23, 39, 41 bis 43, 57, 59, 61, 72, 74 bis 79 bis 81, 82, 84, 86, 88, 100, 104 und 106 bis 109 Akzeptor-Reste und mindestens die Aminosäurereste 24 bis 34, 46, 48, 50 bis 56, 58, 71 und 89 bis 97 Donor-Reste sind;
(4) Assoziierung der in Stufe (3) hergestellten leichten Kette mit einer komplementären schweren Kette zur Bildung eines Antikörpermoleküls;
(5) Bestimmung der Affinität des in Stufe (4) gebildeten Antikörpermoleküls für das genannte vorherbestimmte Antigen;
(6) wenn die in Stufe (5) bestimmte Affinität der des Donor-Antikörpers nicht äquivalent ist, Bereitstellung einer leichten Kette nach der Beschreibung des obigen Punktes (3), in welcher jedoch die Aminosäurereste 1, 2, 3 und 47 zusätzlich Donor-Reste sind;
(7) Assoziierung der in Stufe (6) hergestellten leichten Kette mit einer komplementären schweren Kette zur Bildung eines Antigen-Bindungsmoleküls;
(8) Bestimmung der Affinität des in Stufe (7) gebildeten Antigen-Bindungsmoleküls für das genannte vorherbestimmte Antigen;
(9) wenn die in Stufe (8) bestimmte Affinität der des Donor-Antikörpers nicht äquivalent ist, Bereitstellung einer leichten Kette nach der Beschreibung des obigen Punktes (6), in welcher jedoch die Aminosäurereste 36, 44, 47, 85 und 87 zusätzlich Donor-Reste sind;
(10) Assoziierung der in Stufe (9) hergestellten leichten Kette mit einer komplementären schweren Kette zur Bildung eines Antikörpermoleküls;
(11) Bestimmung der Affinität des in Stufe (10) gebildeten Antikörpermoleküls für das genannte vorherbestimmte Antigen;
(12) wenn die in Stufe (11) bestimmte Affinität der des Donor-Antikörpers nicht äquivalent ist, Bereitstellung einer leichten Kette nach der Beschreibung des obigen Punktes (9), in welcher jedoch mindestens einer der Aminosäurereste 2, 4, 6, 49, 62, 64 bis 69, 98, 99, 101 zusätzlich ein Donor-Rest ist;
(13) Assoziierung der in Stufe (9) hergestellten leichten Kette mit einer komplementären schweren Kette zur Bildung eines Antikörpermoleküls.

Revendications

1. Molécule d'anticorps ayant une affinité pour un antigène prédéterminé et comprenant une chaîne lourde composite et une chaîne légère complémentaire, ladite chaîne lourde composite possédant un

domaine variable comprenant des résidus d'armature de chaîne lourde d'anticorps accepteur et des résidus de liaison à l'antigène de chaîne lourde d'anticorps donneur, ledit anticorps donneur ayant une affinité pour ledit antigène prédéterminé, dans laquelle, conformément au système de numérotation de Kabat, dans ladite chaîne lourde composite, les résidus d'acides aminés 5, 8, 10, 12 à 17, 19, 21, 22, 40, 42 à 44, 66, 68, 70, 74, 77, 79, 81, 83 à 85, 90, 92, 105, 109, 111 et 113 au moins sont des résidus accepteurs et les résidus d'acides aminés 23, 24, 31 à 35, 49 à 58, 71, 73, 78 et 95 à 102 au moins sont des résidus donneurs.

2. Molécule d'anticorps selon la revendication 1, dans laquelle les résidus d'acides aminés 26 à 30 et 59 à 65 dans ladite chaîne lourde composite sont en outre des résidus donneurs.

3. Molécule d'anticorps selon la revendication 1 ou la revendication 2, dans laquelle au moins un des résidus d'acides aminés 1, 3 et 76 dans ladite chaîne lourde composite est en outre un résidu donneur.

4. Molécule d'anticorps selon l'une quelconque des revendications 1 à 3, dans laquelle au moins un des résidus d'acides aminés 36, 94, 104, 106 et 107 dans ladite chaîne lourde composite est en outre un résidu donneur.

5. Molécule d'anticorps selon la revendication 4, dans laquelle au moins un des résidus d'acides aminés 2, 4, 6, 38, 48, 67 et 69 dans ladite chaîne lourde composite est en outre un résidu donneur.

6. Molécule d'anticorps selon l'une quelconque des revendications 1 à 5, dans laquelle les résidus d'acides aminés 7, 9, 11, 18, 20, 25, 37, 39, 41, 45, 47, 48, 72, 75, 80, 82, 86 à 89, 91, 93, 103, 108, 110 et 112 dans ladite chaîne lourde composite sont en outre des résidus accepteurs.

7. Molécule d'anticorps selon l'une quelconque des revendications 1 à 6, dans laquelle ladite chaîne légère complémentaire est une chaîne légère composite possédant un domaine variable comprenant des résidus d'armature de chaîne légère d'anticorps accepteur et des résidus de liaison à l'antigène de chaîne légère d'anticorps donneur, ledit anticorps donneur ayant une affinité pour ledit antigène prédéterminé, dans laquelle, conformément au système de numérotation de Kabat, dans ladite chaîne légère composite, les résidus d'acides aminés 5, 7 à 9, 11, 13 à 18, 20, 22, 23, 39, 41 à 43, 57, 59, 61, 72, 74 à 79, 81, 82, 84, 86, 88, 100, 104, 106 et 107 au moins sont des résidus accepteurs et les résidus d'acides aminés 24 à 34, 46, 48, 50 à 56, 58, 71 et 89 à 97 au moins sont des résidus donneurs.

8. Molécule d'anticorps selon la revendication 7, dans laquelle les résidus d'acides aminés 1, 3 et 47 dans ladite chaîne légère composite sont en outre des résidus donneurs.

9. Molécule d'anticorps selon la revendication 7 ou la revendication 8, dans laquelle les résidus d'acides aminés 36, 44, 47, 85 et 87 dans ladite chaîne légère composite sont en outre des résidus donneurs.

10. Molécule d'anticorps selon l'une quelconque des revendications 7 à 9, dans laquelle au moins un des résidus d'acides aminés 2, 4, 6, 49, 62, 64 à 69, 98, 99, 101 et 102 dans ladite chaîne légère composite est en outre un résidu donneur.

11. Molécule d'anticorps selon l'une quelconque des revendications 7 à 11, dans laquelle au moins un des résidus d'acides aminés 1, 3, 10, 12, 21, 40, 60, 63, 70, 73, 80, 103 et 105 dans ladite chaîne légère composite est en outre un résidu donneur.

12. Composition thérapeutique ou de diagnostic comprenant la molécule d'anticorps selon l'une quelconque des revendications 1 à 11 en combinaison avec un véhicule, diluant ou excipient pharmaceutiquement acceptable.

13. Procédé de production d'une molécule recombinante de liaison à l'antigène ayant une affinité pour un antigène prédéterminé comprenant les étapes qui consistent à :

[1] déterminer la séquence d'acides aminés du domaine variable de la chaîne lourde d'un anticorps donneur qui a une affinité pour ledit antigène prédéterminé;

[2] déterminer la séquence d'acides aminés du domaine variable de la chaîne lourde d'un anticorps accepteur non spécifique;
[3] produire une chaîne lourde composite pour une molécule d'anticorps, ladite chaîne lourde composite possédant des résidus d'armature accepteurs et des résidus de liaison à l'antigène donneurs, dans laquelle, conformément au système de numérotation de Kabat, les résidus d'acides aminés 5, 8, 10, 12 à 17, 19, 21, 22, 40, 42 à 44, 66, 68, 70, 74, 77, 79, 81, 83 à 85, 90, 92, 105, 109, 111 et 113 au moins sont des résidus accepteurs et les résidus d'acides aminés 23, 24, 31 à 35, 49 à 58, 71, 73, 78 et 95 à 102 au moins sont des résidus donneurs;
[4] associer la chaîne lourde produite dans l'étape [3] avec une chaîne légère complémentaire pour former une molécule d'anticorps;
[5] déterminer l'affinité de la molécule d'anticorps formée dans l'étape [4] pour ledit antigène prédéterminé;
[6] si l'affinité déterminée dans l'étape [5] n'est pas équivalente à celle de l'anticorps donneur, produire une chaîne lourde telle que décrite dans l'étape [3] ci-dessus mais dans laquelle les résidus d'acides aminés 71, 73 et 78 sont en outre des résidus donneurs;
[7] associer la chaîne lourde produite dans l'étape [6] avec une chaîne légère complémentaire pour former une molécule d'anticorps;
[8] déterminer l'affinité de la molécule d'anticorps formée dans l'étape [7] pour ledit antigène prédéterminé;
[9] si l'affinité déterminée dans l'étape [8] n'est pas équivalente à celle de l'anticorps donneur, produire une chaîne lourde telle que décrite dans l'étape [6] ci-dessus mais dans laquelle les résidus d'acides aminés 26 à 30 sont en outre des résidus donneurs;
[10] associer la chaîne lourde produite dans l'étape [9] avec une chaîne légère complémentaire pour former une molécule d'anticorps;
[11] déterminer l'affinité de la molécule d'anticorps formée dans l'étape [10] pour ledit antigène prédéterminé;
[12] si l'affinité déterminée dans l'étape [11] n'est pas équivalente à celle de l'anticorps donneur, produire une chaîne lourde telle que décrite dans l'étape [9] ci-dessus mais dans laquelle au moins un des résidus d'acides aminés 1, 3 et 76 est en outre un résidu donneur;
[13] associer la chaîne lourde produite dans l'étape [12] avec une chaîne légère complémentaire pour former une molécule d'anticorps;
[14] déterminer l'affinité de la molécule d'anticorps formée dans l'étape [13] pour ledit antigène prédéterminé;
[15] si l'affinité déterminée dans l'étape [14] n'est pas équivalente à celle de l'anticorps donneur, produire une chaîne lourde telle que décrite dans l'étape [12] ci-dessus mais dans laquelle au moins un des résidus d'acides aminés 36, 94, 104, 106, 107 est en outre un résidu donneur;
[16] associer la chaîne lourde produite dans l'étape [15] avec une chaîne légère complémentaire pour former une molécule d'anticorps;
[17] déterminer l'affinité de la molécule d'anticorps formée dans l'étape [16] pour ledit antigène prédéterminé;
[18] si l'affinité déterminée dans l'étape [17] n'est pas équivalente à celle de l'anticorps donneur, produire une chaîne lourde telle que décrite dans l'étape [15] ci-dessus mais dans laquelle au moins un des résidus d'acides aminés 2, 4, 6, 38, 48, 67 et 69 est en outre un résidu donneur; et
[19] associer la chaîne lourde produite dans l'étape [18] avec une chaîne légère complémentaire pour former une molécule d'anticorps.

14. Procédé selon la revendication 13, comprenant aussi les étapes qui consistent à:
[1] déterminer la séquence d'acides aminés du domaine variable de la chaîne légère dudit anticorps donneur qui a une affinité pour ledit antigène prédéterminé;
[2] déterminer la séquence d'acides aminés du domaine variable de la chaîne légère d'un anticorps accepteur non spécifique;
[3] fournir une chaîne légère composite pour une molécule d'anticorps, ladite chaîne légère composite possédant des résidus d'armature accepteurs et des résidus de liaison à l'antigène donneurs, dans laquelle, conformément au système de numérotation de Kabat, les résidus d'acides aminés 5, 7 à 9, 11, 13 à 18, 20, 22, 23, 39, 41 à 43, 57, 59, 61, 72, 74 à 79 à 81, 82, 84, 86, 88, 100, 104 et 106 à 109 au moins sont des résidus accepteurs et les résidus d'acides aminés 24 à 34, 46, 48, 50 à 56, 58, 71 et 89 à 97 au moins sont des résidus donneurs;

[4] associer la chaîne légère produite dans l'étape [3] avec une chaîne lourde complémentaire pour former une molécule d'anticorps;
[5] déterminer l'affinité de la molécule d'anticorps formée dans l'étape [4] pour ledit antigène prédéterminé;
[6] si l'affinité déterminée dans l'étape [5] n'est pas équivalente à celle de l'anticorps donneur, produire une chaîne légère telle que décrite dans l'étape [3] ci-dessus mais dans laquelle les résidus d'acides aminés 1, 2, 3 et 47 sont en outre des résidus donneurs;
[7] associer la chaîne légère produite dans l'étape [6] avec une chaîne lourde complémentaire pour former une molécule de liaison à l'antigène;
[8] déterminer l'affinité de la molécule de liaison à l'antigène formée dans l'étape [7] pour ledit antigène prédéterminé;
[9] si l'affinité déterminée dans l'étape [8] n'est pas équivalente à celle de l'anticorps donneur, produire une chaîne légère telle que décrite dans l'étape [6] ci-dessus mais dans laquelle les résidus d'acides aminés 36, 44, 47, 85 et 87 sont en outre des résidus donneurs;
[10] associer la chaîne légère produite dans l'étape [9] avec une chaîne lourde complémentaire pour former une molécule d'anticorps;
[11] déterminer l'affinité de la molécule d'anticorps formée dans l'étape [10] pour ledit antigène prédéterminé;
[12] si l'affinité déterminée dans l'étape [11] n'est pas équivalente à celle de l'anticorps donneur, produire une chaîne légère telle que décrite dans l'étape [9] ci-dessus mais dans laquelle au moins un des résidus d'acides aminés 2, 4, 6, 49, 62, 64 à 69, 98, 99, 101 est en outre un résidu donneur; et
[13] associer la chaîne légère produite dans l'étape [9] avec une chaîne lourde complémentaire pour former une molécule d'anticorps.

```
  1  GAATTCCCAA AGACAAAatg gattttcaag tgcagatttt cagcttcctg
 51  ctaatcagtg cctcagtcat aatatccaga ggacaaattg ttctcaccca
101  gtctccagca atcatgtctg catctccagg ggagaaggtc accatgacct
151  gcagtgccag ctcaagtgta agttacatga actggtacca gcagaagtca
201  ggcacctccc ccaaaagatg gatttatgac acatccaaac tggcttctgg
251  agtccctgct cacttcaggg gcagtgggtc tgggacctct tactctctca
301  caatcagcgg catggaggct gaagatgctg ccacttatta ctgccagcag
351  tggagtagta acccattcac gttcggctcg gggacaaagt tggaaataaa
401  ccgggctgat actgcaccaa ctgtatccat cttcccacca tccagtgagc
451  agttaacatc tggaggtgcc tcagtcgtgt gcttcttgaa caacttctac
501  cccaaagaca tcaatgtcaa gtggaagatt gatggcagtg aacgacaaaa
551  tggcgtcctg aacagttgga ctgatcagga cagcaaagac agcacctaca
601  gcatgagcag caccctcacg ttgaccaagg acgagtatga acgacataac
651  agctatacct gtgaggccac tcacaagaca tcaacttcac ccattgtcaa
701  gagcttcaac aggaatgagt gttAGAGACA AAGGTCCTGA GACGCCACCA
751  CCAGCTCCCA GCTCCATCCT ATCTTCCCTT CTAAGGTCTT GGAGGCTTCC
801  CCACAAGCGC tTACCACTGT TGCGGTGCTC tAAACCTCCT CCCACCTCCT
851  TCTCCTCCTC CTCCCTTTCC TTGGCTTTTA TCATGCTAAT ATTTGCAGAA
901  AATATTCAAT AAAGTGAGTC TTTGCCTTGA AAAAAAAAAA AAA
```

Fig. 1(a)

```
  1  MDFOVOIFSF LLISASVIIS RGQIVLTQSP AIMSASPGEK VTMTCSASSS
 51  VSYMNWYQQK SGTSPKRWIY DTSKLASGVP AHFRGSGSGT SYSLTISGME
101  AEDAATYYCQ QWSSNPFTFG SGTKLEINRA DTAPTVSIFP PSSEQLTSGG
151  ASVVCFLNNF YPKDINVKWK IDGSERQNGV LNSWTDQDSK DSTYSMSSTL
201  TLTKDEYERH NSYTCEATHK TSTSPIVKSF NRNEC*
```

Fig. 1(b)

```
   1  GAATTCCCCT CTCCACAGAC ACTGAAAACT CTGACTCAAC ATGGAAAGGC
  51  ACTGGATCTT TCTACTCCTG TTGTCAGTAA CTGCAGGTGT CCACTCCCAG
 101  GTCCAGCTGC AGCAGTCTGG GGCTGAACTG GCAAGACCTG GGGCCTCAGT
 151  GAAGATGTCC TGCAAGGCTT CTGGCTACAC CTTTACTAGG TACACGATGC
 201  ACTGGGTAAA ACAGAGGCCT GGACAGGGTC TGGAATGGAT TGGATACATT
 251  AATCCTAGCC GTGGTTATAC TAATTACAAT CAGAAGTTCA AGGACAAGGC
 301  CACATTGACT ACAGACAAAT CCTCCAGCAC AGCCTACATG CAACTGAGCA
 351  GCCTGACATC TGAGGACTCT GCAGTCTATT ACTGTGCAAG ATATTATGAT
 401  GATCATTACT GCCTTGACTA CTGGGGCCAA GGCACCACTC TCACAGTCTC
 451  CTCAGCCAAA ACAACAGCCC CATCGGTCTA TCCACTGGCC CCTGTGTGTG
 501  GAGATACAAC TGGCTCCTCG GTGACTCTAG GATGCCTGGT CAAGGGTTAT
 551  TTCCCTGAGC CAGTGACCTT GACCTGGAAC TCTGGATCCC TGTCCAGTGG
 601  TGTGCACACC TTCCCAGCTG TCCTGCAGTC TGACCTCTAC ACCCTCAGCA
 651  GCTCAGTGAC TGTAACCTCG AGCACCTGGC CCAGCCAGTC CATCACCTGC
 701  AATGTGGCCC ACCCGGCAAG CAGCACCAAG GTGGACAAGA AAATTGAGCC
 751  CAGAGGGCCC ACAATCAAGC CCTGTCCTCC ATGCAAATGC CCAGCACCTA
 801  ACCTCTTGGG TGGACCATCC GTCTTCATCT TCCCTCCAAA GATCAAGGAT
 851  GTACTCATGA TCTCCCTGAG CCCCATAGTC ACATGTGTGG TGGTGGATGT
 901  GAGCGAGGAT GACCCAGATG TCCAGATCAG CTGGTTTGTG AACAACGTGG
 951  AAGTACACAC AGCTCAGACA CAAACCCATA GAGAGGATTA CAACAGTACT
1001  CTCCGGGTGG TCAGTGCCCT CCCCATCCAG CACCAGGACT GGATGAGTGG
1051  CAAGGAGTTC AAATGCAAGG TCAACAACAA AGACCTCCCA GCGCCCATCG
1101  AGAGAACCAT CTCAAAACCC AAAGGGTCAG TAAGAGCTCC ACAGGTATAT
1151  GTCTTGCCTC CACCAGAAGA AGAGATGACT AAGAAACAGG TCACTCTGAC
1201  CTGCATGGTC ACAGACTTCA TGCCTGAAGA CATTTACGTG GAGTGGACCA
1251  ACAACGGGAA AACAGAGCTA AACTACAAGA ACACTGAACC AGTCCTGGAC
1301  TCTGATGGTT CTTACTTCAT GTACAGCAAG CTGAGAGTGG AAAAGAAGAA
1351  CTGGGTGGAA AGAAATAGCT ACTCCTGTTC AGTGGTCCAC GAGGGTCTGC
1401  ACAATCACCA CACGACTAAG AGCTTCTCCC GGACTCCGGG TAAATGAGCT
1451  CAGCACCCAC AAAACTCTCA GGTCCAAAGA GACACCCACA CTCATCTCCA
1501  TGCTTCCCTT GTATAAATAA AGCACCCAGC AATGCCTGGG ACCATGTAAA
1551   AAAAAAAAAA AAAGGAATTC
```

Fig. 2(a)

OKT 3 HEAVY CHAIN PROTEIN SEQUENCE DEDUCED FROM DNA SEQUENCE

```
  1  MERHWIFLLL LSVTAGVHSQ VQLQQSGAEL ARPGASVKMS CKASGYTFTR
 51  YTMHWVKQRP GQGLEWIGYI NPSRGYTNYN QKFKDKATLT TDKSSSTAYM
101  QLSSLTSEDS AVYYCARYYD DHYCLDYWGQ GTTLTVSSAK TTAPSVYPLA
151  PVCGDTTGSS VTLGCLVKGY FPEPVTLTWN SGSLSSGVHT FPAVLQSDLY
201  TLSSSVTVTS STWPSQSITC NVAHPASSTK VDKKIEPRGP TIKPCPPCKC
251  PAPNLLGGPS VFIFPPKIKD VLMISLSPIV TCVVVDVSED DPDVQISWFV
301  NNVEVHTAQT QTHREDYNST LRVVSALPIQ HQDWMSGKEF KCKVNNKDLP
351  APIERTISKP KGSVRAPQVY VLPPPEEEMT KKQVTLTCMV TDFMPEDIYV
401  EWTNNGKTEL NYKNTEPVLD SDGSYFMYSK LRVEKKNWVE RNSYSCSVVH
451  EGLHNHHTTK SFSRTPGK*
```

Fig. 2(b)

```
          1                    23                  42
          NN      N              N        N    N
RES TYPE  SBspSPESssBSbSsSssPSPSPsPSsse*s*p*Pi^ISsSe
Okt3vl    QIVLTQSPAIMSASPGEKVTMTCSASS.SVSYMNWYQQKSGT
REI       DIQMTQSPSSLSASVGDRVTITCQASQDIIKYLNWYQQTPGK
          ? ?
            CDR1   (LOOP)          *******
            CDR1   (KABAT)        ***********
```

```
                    56                          85
          N  NN
RES TYPE  *IsiPpIeesesssSBEsePsPSBSSEsPspsPsseesSPePb
Okt3vl    SPKRWIYDTSKLASGVPAHFRGSGSGTSYSLTISGMEAEDAAT
REI       APKLLIYEASNLQAGVPSRFSGSGSGTDYTFTISSLQPEDIAT
          ?  ??                        ?  ?
               *******  CDR2 (LOOP/KABAT)
```

```
                     102   108
RES TYPE  PiPIPies**iPIIsPPSPSPSS
Okt3vl    YYCQQWSSNPFTFGSGTKLEINR
REIvl     YYCQQYQSLPYTFGQGTKLQITR
                     ?     ?
             ******        CDR3 (LOOP)
           *********       CRD3 (KABAT)
```

Fig. 3

```
          NN N                  23 26    32 35 N39  43
RES TYPE  SESPs^SBssS^sSSsSpSpSPsPSEbSBssBePiPIpiesss
Okt3h     QVQLQQSGAELARPGASVKMSCKASGYTFTRYTMHWVKQRPGQ
KOL       QVQLVESGGGVVQPGRSLRLSCSSSGFIFSSYAMYWVRQAPGK
               ?                ??
                                    ******    CDR1 (LOOP)
                                        ***** CDR1 (KABAT)

                  52a      60  65     N N  N    82abc     89
RES TYPE IIeIppp^sssssssss^ps^pSSsbSpseSsSseSp^pSpsSBssS^ePb
Okt3vh   GLEWIGYINPSRGYTNTNQKFKDKATLTTDKSSSTAYMQLSSLTSEDSAV
KOL      GLEWVAIIWDDGSDQHYADSVKGRFTISRDNSKNTLFLQMDSLRPEDTGV
            ??          .          ? ?  ?  ?            ?
               ************        CDR2  (LOOP)
               ******************** CDR2  (KABAT)

           92 N                  107    113
RES TYPE  PiPIEissssiiisssbibi*EIPIP*spSBSS
Okt3vh    YYCARYYDDHY.......CLDYWGQGTTLTVSS
KOL       YFCARDGGHGFCSSASCFGPDYWGQGTPVTVSS
               ****************** CRD3 (KABAT/LOOP)
```

Fig. 4

OKT 3 HEAVY CHAIN CDR GRAFTS

1. gh341 and derivatives

```
          1                          26        35  39  43
Okt3vh    QVQLQQSGAELARPGASVKMSCKASGYTFTRYTMHWVKQRPGQ
gH341     QVQLVESGGGVVQPGRSLRLSCSSSGYTFTRYTMHWVRQAPGK   JA178
gH341A    QVQLVQSGGGVVQPGRSLRLSCKASGYTFTRYTMHWVRQAPGK   JA185

gH341E    QVQLVQSGGGVVQPGRSLRLSCKASGYTFTRYTMHWVRQAPGK   JA198
gH341*    QVQLVQSGGGVVQPGRSLRLSCKASGYTFTRYTMHWVRQAPGK   JA207
gH341*    QVQLVQSGGGVVQPGRSLRLSCKASGYTFTRYTMHWVRQAPGK   JA209
gH341D    QVQLVQSGGGVVQPGRSLRLSCKASGYTFTRYTMHWVRQAPGK   JA197
gH341*    QVQLVQSGGGVVQPGRSLRLSCKASGYTFTRYTMHWVRQAPGK   JA199
gH341C    QVQLVQSGGGVVQPGRSLRLSCKASGYTFTRYTMHWVRQAPGK   JA184

gH341*    QVQLVQSGGGVVQPGRSLRLSCSASGYTFTRYTMHWVRQAPGK   JA203
gH341*    QVQLVESGGGVVQPGRSLRLSCSASGYTFTRYTMHWVRQAPGK   JA205
gH341B    QVQLVESGGGVVQPGRSLRLSCSSSGYTFTRYTMHWVRQAPGK   JA183
gH341*    QVQLVQSGGGVVQPGRSLRLSCSASGYTFTRYTMHWVRQAPGK   JA204
gH341*    QVQLVESGGGVVQPGRSLRLSCSASGYTFTRYTMHWVRQAPGK   JA206
gH341*    QVQLVQSGGGVVQPGRSLRLSCSASGYTFTRYTMHWVRQAPGK   JA208
KOL       QVQLVESGGGVVQPGRSLRLSCSSSGFIFSSYAMYWVRQAPGK
```

Fig. 5(i)

```
        44    50              65                  83
Okt3vh  GLEWIGYINPSRGYTNYNQKFKDKATLTTDKSSSTAYMQLSSLT
gH341   GLEWVAYINPSRGYTNYNQKFKDRFTISRDNSKNTLFLQMDSLR JA178
gH341A  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKSTAFLQMDSLR JA185

gH341E  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKSTAFLQMDSLR JA198
gH341*  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKNTAFLQMDSLR JA207
gH341*  GLEWIGYINPSRGYTNYNQKVKDRFTISRDNSKNTAFLQMDSLR JA209
gH341D  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKNTLFLQMDSLR JA197
gH341*  GLEWIGYINPSRGYTNYNQKVKDRFTISRDNSKNTLFLQMDSLR JA199
gH341C  GLEWVAYINPSRGYTNYNQKFKDRFTISRDNSKNTLFLQMDSLR JA184

gH341*  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKSTAFLQMDSLR JA207
gH341*  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKSTAFLQMDSLR JA205
gH341B  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKSTAFLQMDSLR JA183
gH341*  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKSTAFLQMDSLR JA204
gH341*  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKSTAFLQMDSLR JA206
gH341*  GLEWIGYINPSRGYTNYNQKVKDRFTISTDKSKNTAFLQMDSLR JA208
KOL     GLEWVAIIWDDGSDQHYADSVKGRFTISRDNSKNTLFLQMDSLR
```

Fig. 5(ii)

	84 95 102 113	
Okt3vh	SEDSAVYYCARYYDDHY.....CLDYWGQGTTLTVSS	
gH341	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA178
gH341A	PEDTAVYYCARYYDDHY.....CLDYWGQGTTLTVSS	JA185
gH341E	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA198
gH341*	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA207
gH341D	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA197
gH341*	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA209
gH341*	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA199
gH341C	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA184
gH341*	PEDTAVYYCARYYDDHY.....CLDYWGQGTTLTVSS	JA203
gH341*	PEDTAVYYCARYYDDHY.....CLDYWGQGTTLTVSS	JA205
gH341B	PEDTAVYYCARYYDDHY.....CLDYWGQGTTLTVSS	JA183
gH341*	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA204
gH341*	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA206
gH341*	PEDTGVYFCARYYDDHY.....CLDYWGQGTTLTVSS	JA208
KOL	PEDTGVYFCARDGGHGFCSSASCFGPDYWGQGTPVTVSS	

Fig. 5(iii)

OKT3 LIGHT CHAIN CDR GRAFTING

1. gL221 and derivatives

```
          1                        24        34      42
Okt3v1    QIVLTQSPAIMSASPGEKVTMTCSASS.SVSYMNWYQQKSGT
gL221     DIQMTQSPSSLSASVGDRVTITCSASS.SVSYMNWYQQTPGK
gL221A    QIVMTQSPSSLSASVGDRVTITCSASS.SVSYMNWYQQTPGK
gL221B    QIVMTQSPSSLSASVGDRVTITCSASS.SVSYMNWYQQTPGK
gL221C    DIQMTQSPSSLSASVGDRVTITCSASS.SVSYMNWYQQTPGK
REI       DIQMTQSPSSLSASVGDRVTITCQASQDIIKYLNWYQQTPGK

          43     50     56                          85
Okt3v1    SPKRWIYDTSKLASGVPAHFRGSGSGTSYSLTISGMEAEDAAT
gL221     APKLLIYDTSKLASGVPSRFSGSGSGTDYTFTISSLQPEDIAT
gL221A    APKRWIYDTSKLASGVPSRFSGSGSGTDYTFTISSLQPEDIAT
gL221B    APKRWIYDTSKLASGVPSRFSGSGSGTDYTFTISSLQPEDIAT
gL221C    APKRWIYDTSKLASGVPSRFSGSGSGTDYTFTISSLQPEDIAT
REI       APKLLIYEASNLQAGVPSRFSGSGSGTDYTFTISSLQPEDIAT

         86   91   96          108
Okt3v1   YYCQQWSSNPFTFGSGTKLEINR
gL221    YYCQQWSSNPFTFGQGTKLQITR
gL221A   YYCQQWSSNPFTFGQGTKLQITR
gL221B   YYCQQWSSNPFTFGQGTKLQITR
gL221C   YYCQQWSSNPFTFGQGTKLQITR
REI      YYCQQYQSLPYTFGQGTKLQITR
```

CDR'S ARE UNDERLINED

FRAMEWORK RESIDUES INCLUDED IN THE GENE ARE DOUBLE UNDERLINED

Fig. 6

Fig. 7

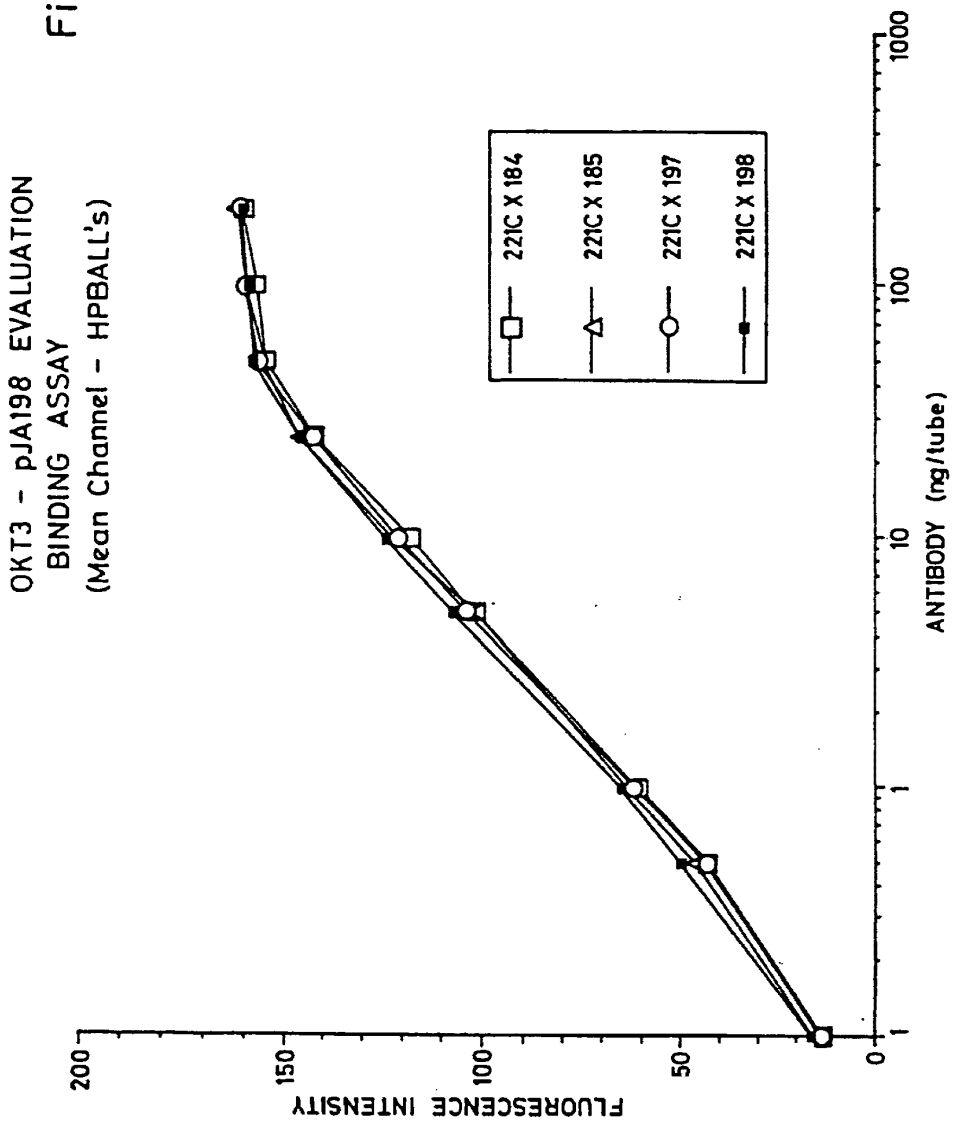
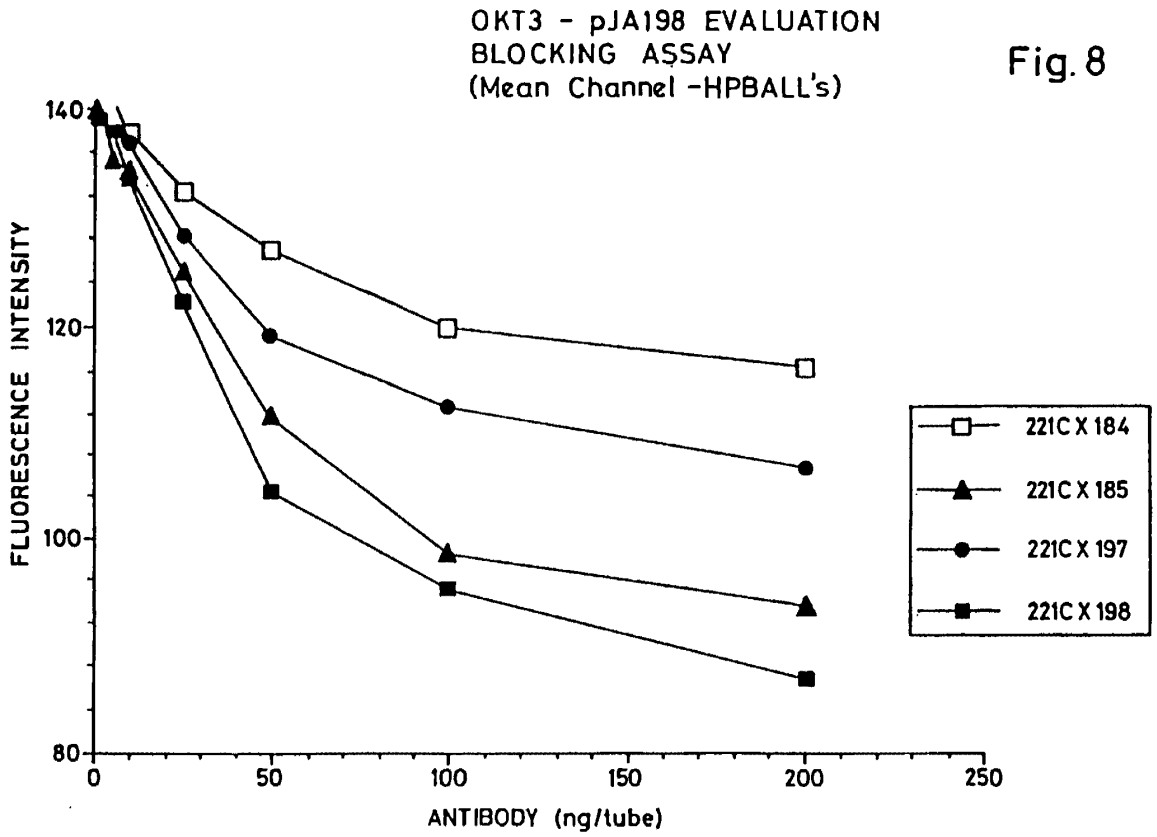


Fig. 8


OKT3 - pJA198 EVALUATION
BLOCKING ASSAY
(Mean Channel -HPBALL's)
FLUORESCENCE INTENSITY
140
120
100
80
0
50
100
150
200
250
ANTIBODY (ng/tube)
221C X 184
221C X 185
221C X 197
221C X 198

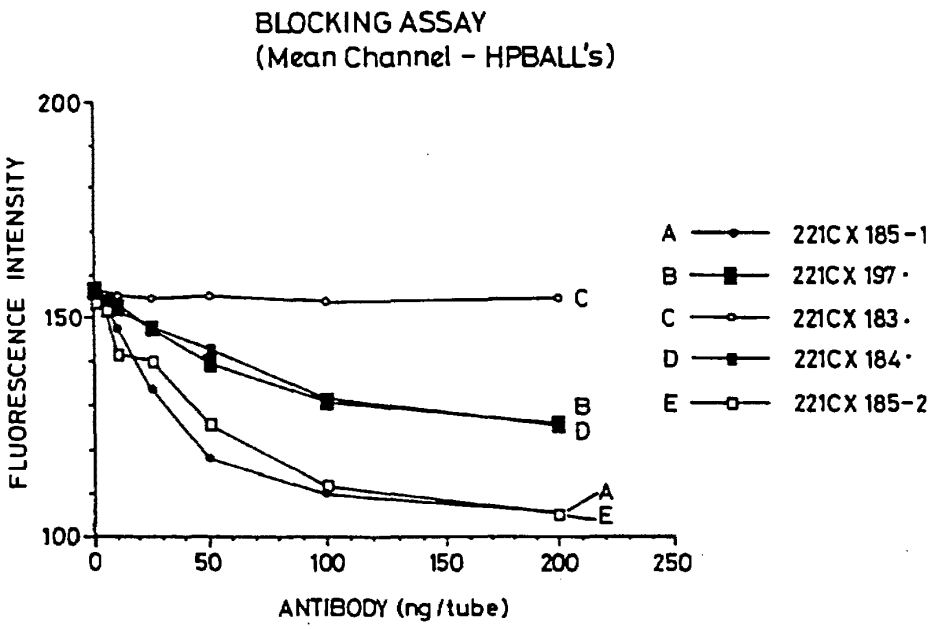
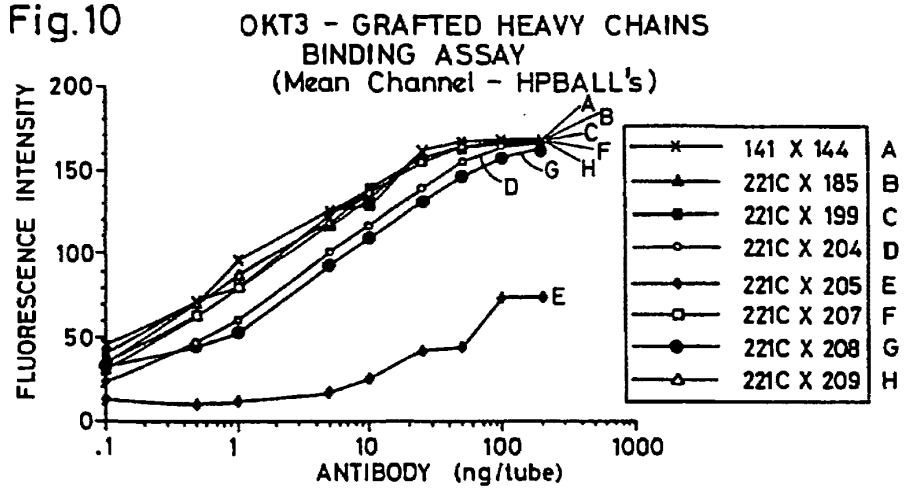
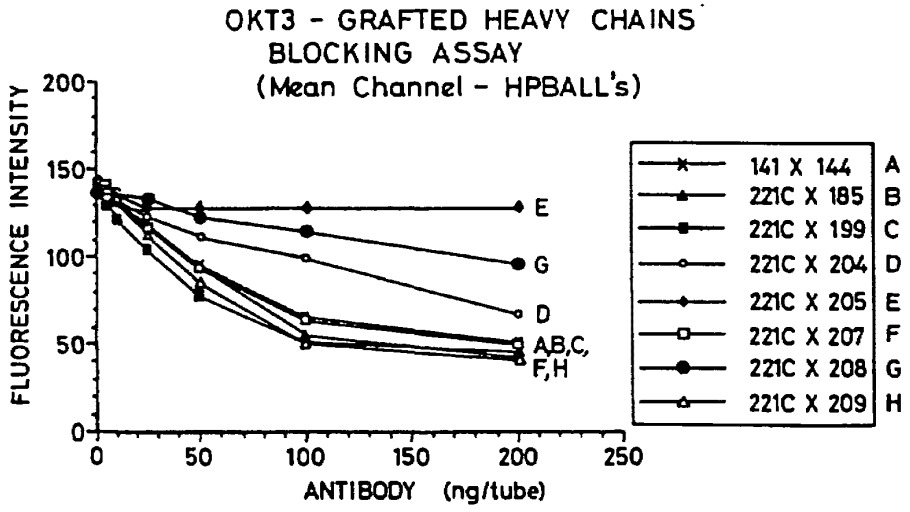


Fig. 9

Fig.10





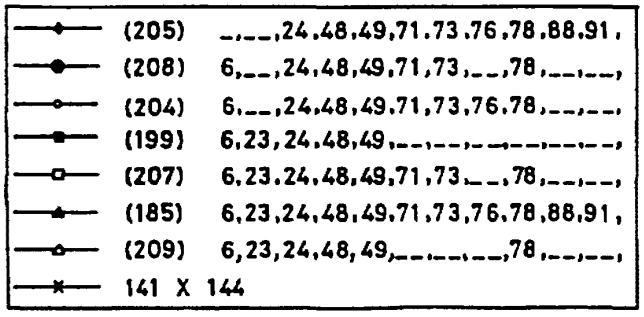
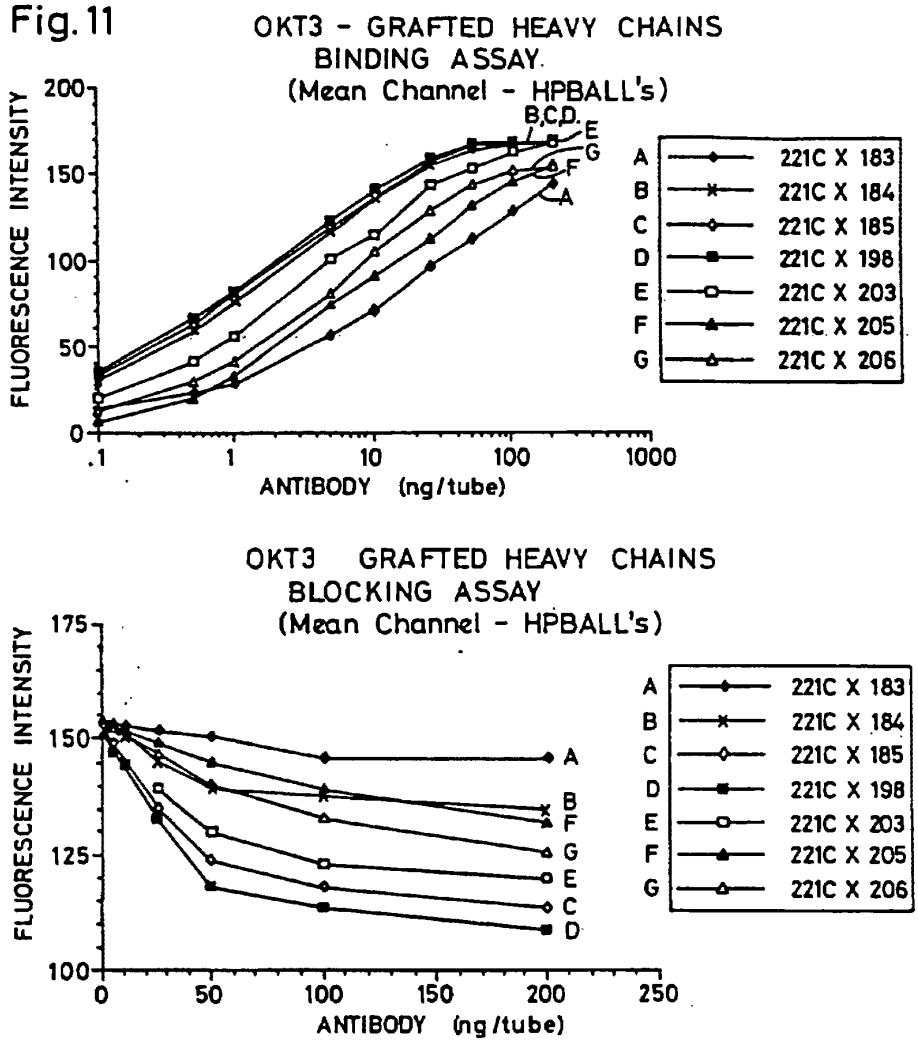


Fig. 11



(183)	_._._._,48,49,71,73,76,78,88,91,
(205)	_._._,24,48,49,71,73,76,78,88,91,
(184)	6,23,24,_._._._._._._._._._._._,
(206)	_._._,24,48,49,71,73,76,78,_._._,
(203)	6,_._,24,48,49,71,73,76,78,88,91,
(185)	6,23,24,48,49,71,73,76,78,88,91,
(198)	6,23,24,48,49,71,73,76,78,_._._,

Fig. 12

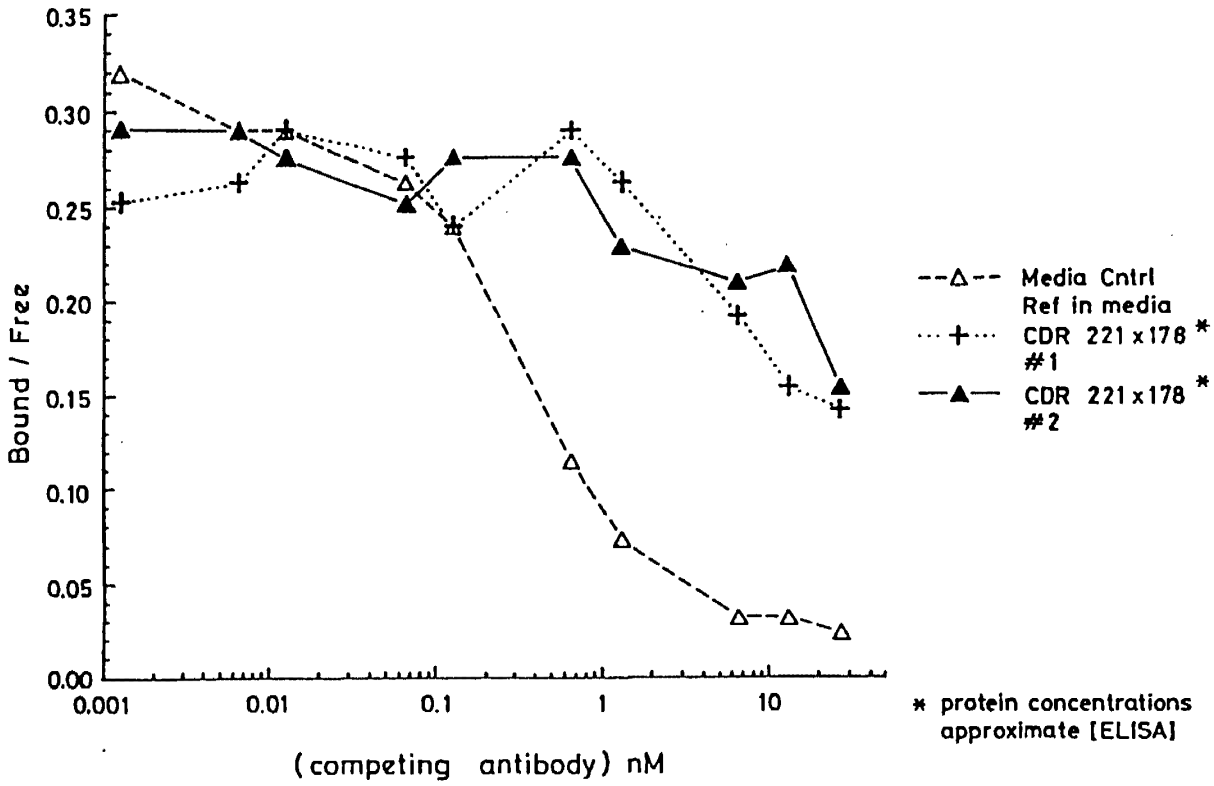
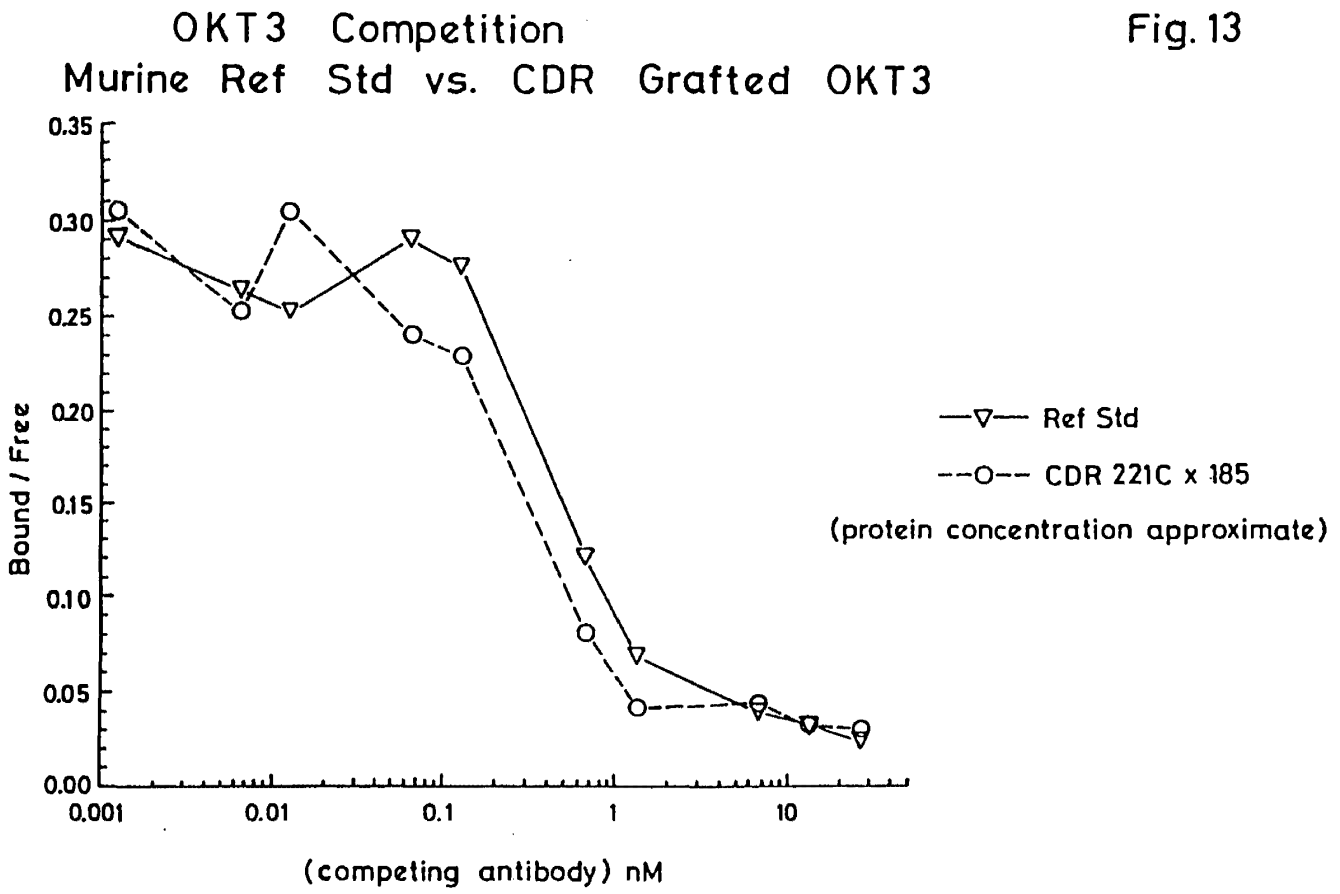

OKT3 Competition
Murine Ref Std vs. CDR Grafted OKT3
Bound / Free
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.001
0.01
0.1
1
10
(competing antibody) nM
Media Cntrl
Ref in media
CDR 221 x 178 * #1
CDR 221 x 178 * #2
* protein concentrations approximate [ELISA]

Fig. 13


OKT3 Competition
Murine Ref Std vs. CDR Grafted OKT3
Bound / Free
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.001
0.01
0.1
1
10
(competing antibody) nM
—▽— Ref Std
--O-- CDR 221C x 185
(protein concentration approximate)

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 519 596 A1**

(12) **EUROPEAN PATENT APPLICATION**

(21) Application number: 92304225.3

(51) Int. Cl.5: **C12N 15/13**, C12P 21/08

(22) Date of filing: 12.05.92

(30) Priority: 17.05.91 US 702217

(43) Date of publication of application:
23.12.92 Bulletin 92/52

(84) Designated Contracting States:
CH DE FR GB IT LI NL

(71) Applicant: **MERCK & CO. INC.**
126, East Lincoln Avenue P.O. Box 2000
Rahway New Jersey 07065-0900(US)
Applicant: **NATIONAL INSTITUTES OF HEALTH**
Patent Branch OGC NIH Building 31, 2B 9000
Rockville Pike
Bethesda, Maryland 20892(US)

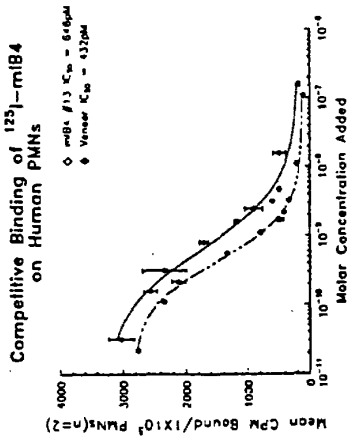
(72) Inventor: **Padlan, Eduardo A.**
4006 Simms Drive
Kensington, Maryland 20895(US)
Inventor: **Mark, George E., III**
4 Richmond Court
Princeton Junction, NJ 085550(US)
Inventor: **Daugherty, Bruce L.**
204 South Ridgewood Road
S. Orange, NJ 07079(US)

(74) Representative: **Thompson, John Dr. et al**
Merck & Co., Inc. European Patent
Department Terlings Park Eastwick Road
Harlow, Essex CM20 2QR(GB)

(54) **A method for reducing the immunogenicity of antibody variable domains.**

(57) A unique method is disclosed for identifying and replacing immunoglobulin surface amino acid residues which converts the antigenicity of a first mammalian species to that of a second mammalian species. The method will simultaneously change immunogenicity and strictly preserve ligand binding properties. The judicious replacement of exterior amino acid residues has no effect on the ligand binding properties but greatly alters immunogenicity.

Fig. 13



EP 0 519 596 A1

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Solvent exposure of sidechains of framework residues in KOL and J539 Fvs and the residues which occur most frequently at these positions in the various human VH subgroups.

Figure 2. Solvent exposure of sidechains of framework residues in KOL VL and the residues which occur most frequently at these positions in the various human V-lambda subgroups.

Figure 3. Solvent exposure of sidechains of framework residues in J539 VL and the residues which occur most frequently at these positions in the various human V-kappa subgroups.

Figure 4. Primers used to isolate DNA encoding murine kappa light chain variable region and murine IgG2a heavy chain variable region using PCR. Oligodeoxynucleotides used as PCR primers to generate a shortened IgG4 heavy chain. Oligodeoxynucleotides used in PCR to re-engineer the thymidine kinase (TK) promoter to facilitate the expression of the neomycin resistance gene. Oligodeoxynucleotide primers used in PCR to clone the IgH enhancer sequence. Oligodeoxynucleotides used as PCR primers to generate a human kappa light chain constant region.

Figure 5. Oligodeoxynucleotides used in the construction of the "veneered" 1B4 heavy and light chain variable regions plus those necessary to fuse the human signal and intronic sequenceds onto these variable regions.

Figure 6. PCR-recombination strategy used in the veneering of the 1B4 kappa light chain variable region.

Figure 7. Outline of the insertion of the "veneered" kappa light chain variable region and kappa constant region into the light chain expression vector.

Figure 8. PCR-recombination strategy used in the veneering of the 1B4 heavy chain variable region.

Figure 9. Outline of the insertion of the "veneered" heavy chain variable region into the heavy chain expression vector.

Figure 10. Outline of the construction of neomycin selectable expression vector.

Figure 11. Outline of the construction of the hygromycin selectable expression vector.

Figure 12. Amino acid sequence compleition of the "veneered"-1B4, murine 1B4 and human Gal heavy chain variable regions and the "veneered" 1B4, murine 1B4 and human Len kappa light chain variable regions. Check marks idicate the individual amino acid residues converted.

Figure 13. Competitive binding assay of native murine 1B4 (open diamonds) and recombinant "veneered" 1B4 (closed diamonds).

BACKGROUND OF THE INVENTION

The identification and production of murine monoclonal antibodies has lead to numerous therapeutic applications of these exquisitely specifc molecules in human disease. The technologies of molecular biology have further expanded the utility of many antibodies by allowing for the creation of class switched molecules whose functionality has been improved by the acquisition or loss of complement fixation. The size of the bioactive molecule may also be reduced so as to increase the tissue target availability of the antibody by either changing the class from an IgM to an IgG, removing most of the heavy chain constant region in the creation of a $F(ab)_2$ or both heavy and light chain constant regions may be dispensed with in the formation of a Fv antibody. Common to all of these potentially therapeutic forms of antibody are the requiste CDRs (complelmentary determining regions) which guide the molecule to its ligand and the framework residues (FRs) which support these latter structures and dictate the disposition of the CDRs relative to one another. Winter European Patent Application, Publication No. 239,400; Riechmann et al., Nature 332: 323-327 (1988). Crystallographic analyses of numerous antibody structures reveal that the combining site is composed almost entirely of the CDR residues arranged in a limited number of loop motifs, Padlan and Sheriff, 1990. The necessity of the CDRs to form these structures combined with the appreciated hypervariablity of their primary sequence leads to a great diversity in the antigen combining site, but one which has a finite number of possibilities. Thus, hypermutability and a limited primary sequence repetoire for each CDR would suggest that the CDRs derived for a given antigen from one species of animal would be the same derived from another species. Hence, they should be poorly immunogenic, if at all, when presented to a recipient organism in a non- foreign context.

Monoclonal antibody producing hybridomas have been most readily obtained from immunized rodents. Development of similar reagents from human sources has been frustrated by the current inability to maintain long term cultures of cells which produce sufficient quantities of antibody. Additional problems arise from the regulatory standpoint when cells of human origin are employed for the production of agents to be used in man. These considerations have lead to the widespread use of rodent mono- clonal antibodies for the imaging and treatment of malignancy, prophyllactic administration to guard against toxic shock, modification of graft rejection episodes, and to temper acute inflammatory reactions. In all scenarios where completely rodent or partially rodent (ie,rodent - human

chimeras) antibodies have been used for therapy the recipients have often illicited an immune response directed toward the antibody. These reactions have limited the duration and effectiveness of the therapy.

Various attempts have been made to minimize or eliminate the immunogenicity of non-human antibodies while perserving their antigen- binding properties. Initially, chimeric antibodies were constructed containing the rodent varible regions and their associated CDRs fused to human constant domains. The following references generally describe chimaeric antibody technology: Lobuglio et al., Proc. Natl. Acad. Sci. USA 86: 4220-4224 (1989); United States Patent 4,816,567; PCT International Publication No. WO 87/02671, published May 7,1987; European Patent Publication No. 255,694, published February 10 1988; European Patent Publication No. 274,394, published July 13, 1988; European Patent Publication No. 323,806, published July 12, 1989; PCT International Publication No. WO/89/00999, published February 9, 1989; European Patent Publication No. 327,000, published August 9, 1989; European Patent Publication No. 328,404, published August 16,1989; and European Patent Publication No. 332,424, published September 13, 1989. These proved to be less immunogenic but still approximately half of the recipients mounted an immune response to the rodent variable region framework residues. Further reduction of the "foreign" nature of the chimeric antibodies has been achieved by grafting only the CDRs from the rodent monoclonal into a human supporting framework prior to its subsequent fusion with an appropriate constant domain, Winter European Patent Application, Publication No. 239,400; Riechmann et al., Nature 332: 323-327 (1988). The procedures employed to accomplish CDR-grafting often result in imperfectly "humanized" antibodies. That is to say, the resultant antibody has either lost avidity (usually 2-3 fold, at best) or in an attempt to retain its original avidity a significant number of the murine framework residues have replaced the corresponding ones of the chosen human fromework. In this later case, the immunogenicity of the modified "humanized" antibody is difficult to anticipate a priori.

The ligand binding characteristics of an antibody combining site are determined primarily by the structure and relative disposition of the CDRs, although some neighboring residues also have been found to be involved in antigen binding (Davies et al., Ann. Rev. Biochem. 59: 439-473 [1990]). Fine specificity can be perserved in a "humanized" antibody only if the CDR structures, their interaction with each other, and their interaction with the rest of the variable domains are strictly maintained. One may anticipate that the key residues represent "interior" and interdomain contact residues, hence those surface exposed residues which are immediately available for immune surveillance should be non- inclusive of the structural residues.

OBJECTS OF THE INVENTION

It is, accordingly, an objective of the present invention to provide a means of converting a monoclonal antibody of one mammalian species to a monoclonal antibody of another mammalian species. Another object is to identify the amino acid residues responsible for species specificity or immunogenicity on the exterior of the monoclonal antibody. Another object is judiciously replace or veneer the exterior amino acid residues of one species with those of a second species so that the antibodies of the first species will not be immunogenic in the second species. A further object is to make replacements only in framework regions of the heavy and light chains of the antibody molecule and not in the complementarity-determining regions. Another object of the invention is to provide novel DNA sequences incorporating the replacement amino acid residues. Another object is to provide a vector containing the DNA sequences for the altered antibody. Another object is to provide a eukaryotic or procaryotic host transformed with a vector containing the DNA sequence for the veneered antibody.

SUMMARY OF THE INVENTION

A unique method is disclosed for identifying and replacing immunoglobulin surface amino acid residues which converts the antigenicity of a first mammalian species to that of a second mammalian species. The method will simultaneously change immunogenicity and strictly preserve ligind binding properties. The judicious replacement of exterior amino acid residues has no effect on the ligind binding properties but greatly alters immunogenicity.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a "humanization" procedure which simultaneously reduces the immunogenicity of the rodent monoclonal antibody while perserving its ligand binding properties in their entirety. Since the antigenicity of a protein is primarily dependent on the nature of its surface, the immunogenicity of an xenogenic or allogenic antibody could be reduced by replacing the exposed residues which differ from those usually found in another mammalian species antibodies. This judicious replacement of exterior resi-

dues should have little, or no, effect on the interior domains, or on the interdomain contacts. Thus, ligand binding properties should be unaffected as a consequence of alterations which are limited to the variable region framework residues. The process is refered to as "veneering" since only the outer surface or skin of the antibody is altered, the supporting residues remain undisturbed.

The procedure for "veneering" makes use of the available sequence data for human antibody variable domains complied by Kabat et al., "Sequences of Proteins of Immunological Interest", 4th ed., Bethesda, Maryland: National Institutes of Health, 1987, updates to this database, and other accessible U.S. and foreign databases (both nucleic acid and protein). The subgroups into which the various sequences have been combined are presented in Figures 1 - 3, indicating the most frequently occurring amino acid at each framework position. Also presented are the sequences of the various J-minigenes. The solvent accessibilities of the amino acids, as deduced from the known three-dimensional structure for human and mouse antibody fragments, are included in these figures.

High resolution X-ray crystallography of the variable domains of the antibodies KOL and J539 have been subjected to extensive refinement beginning with the structures available from the Protein Data Bank (Bernstein et al., J. Mol. Biol. 112: 535-542 1977; file 2FB4 for KOL and file 2FBJ for J529). The solvent accessibilities were computed as described by Padlan Proteins: Struct. Funct. Genet. 7: (1990).

There are two steps in the process of "veneering". First, the framework of a first animal species, i.e. the mouse, variable domains are compared with those corresponding frameworks of a second animal species, i.e. human. It is intended that this invention will allow the antigenic alteration of any animal species antibody. The present invention is illustrates the conversion of murine antibody to human antibody, but this is for illustrative purposes only. The most homologous human variable regions are then compared residue for residue to the corresponding murine regions. This will also define the human subgroup to which each mouse sequence most closely resembles. Second, those residues in the mouse framework which differ from its human counterpart are replaced by the residues present in the human counterpart. This switching occurrs only with those residues which are at least partially exposed (mE and Ex; Figures 1-3). One retains in the "veneered" mouse antibody: its CDRs, the residues neighboring the CDRs, those residues defined as buried or mostly buried (mB and Bu; Figures 1-3), and those residues believed to be involved with interdomain contacts (boldface, Figures 1-3).

Human and murine sequences frequently differ at the N-terminus of both heavy and light chains. The N-termini are contiguous with the CDR surface and are in position to be involved in ligand binding. Thus, wisdom would dictate that these murine termini be retained in its "veneered" version.

Finally, repalcement of some amino acid types could have a significant effect on the tertiary structure or electrostatic interactions of the variable region domains. Hence, care should be exercised in the replacement of proline, glycine, and charged amno acids.

These criteria and the following procedures are used to prepare recombinant DNA sequences which incorporate the CDRs of a first mammalian species, animal, mMAb, both light and heavy chains, into a second mammalian species, human, appearing frame works that can be used to transfect mammalian cells for the expression of recombinant human antibody with the antigen specificity of the animal monoclonal antibody. The present invention further comprises a method for constructing and expressing the altered antibody comprising: (i) mutagenesis and assembly of variable region domains including CDRs and mutagenesis and assembly of variable region domains including CDRs and FRs regions; (ii) preparation of an expression vector including at least one variable region which upon transfection into cells results in the secretion of protein sufficient for avidity and specificity determinations; and (iii) co-amplification of heavy and light chain expression vectors in appropriate cell lines. The present invention provides recombinant methods for incorporating CDRs from animal monoclonal antibodies into frameworks which appear to be human immunoglobulin in nature so that the resulting recombinant antibody will be either weakly immunogenic or non-immunogenic when administered to humans. Preferrably the recombinant immunoglobulins will be recognized as self proteins when administered for threapeutic purposes. This method of "veneering" will render the recombinant antibodies useful as therapeutic agents because they will be either weakly immunogenic or non-immunogenic when administered to humans. The invention is further contemplated to include the recombinant conversion of any animal monoclonal antibody into a recombinant "human-appearing" monoclonal antibody providing that a suitable framework region can be identified (as described below). The animal monoclonals may include, but are not limited to, those murine monoclonal antibodies described by Van Voorhis et al., J. Exp. Med. 158: 126-145 (1983) which bind to human leukocytes and the appropriate mMAbs produced by hybridomas deposited in the Hybridoma Cell Bank maintained by the American Type Culture Collection (ATCC) and

described in the ATCC Catalog of Cell Lines 8 Hybridomas, No. 6, 1988.

The CDR sequences from the animal monoclonal antibody are derived as follows. Total RNA is extracted from the murine hybridomas, for exampe the 1B4 myeloma cells described by Wright et al., Proc. Natl. Acad. Sci. USA 80: 5699-5703 (1983), the 60.3 cells described by Beatty et al., J. Immunol. 131:2913-2918 (1983), the TS1/18 cells described by Sanchez-Madrid et al., J. Exp. Med. 158: 1785-1803 (1983), and other anti-CD18 or CD11 monoclonal antibodies and hybridomas as described in Leukocyte Typing III, Springer-Verlag, New York (1988), using standard methods involving cellular solubilization with guanidinium isothiocyanate (Chirgwin et al., Biochem. 18: 5294-5299 [1979]).The murine 1B4 mMAb will be used as the primary example of animal MAb that can be "veneered" by the unique process being disclosed. The invention is intended to include the conversion of any animal immunoglobulin to a "human-appearing" immunoglobulin. It is further intended that "human-appearing" immunoglobulin (Ig) can contain either kappa or lambda light chains or be one of any of the following heavy chain isotypes (alpha, delta, epsilon, gamma and mu).

Pairs of degenerate oligodeoxynucleotide primers (Figure 4) representing sequences within framework 1 of the murine kappa light chain variable region and light chain constant domain, or those within framework 1 of the murine IgG2a heavy chain variable region and heavy chain constant CH1 domain are synthesized on an Applied Biosystem 381A DNA synthesizer, removed from the resin by treatment with concentrated NH_4OH and desalted on a NAP-5 column eluted with H_2O. Total RNA, about 2 µg, is reverse transcribed for 30 min at 42°C using Moloney MLV reverse transcriptase, about 200 units (BRL), and about 10 pmoles of the constant region complementary strand primers for either the heavy or light chain. The reverse transcriptase is heat inactivated, about 95° C for about 5 min, and the reactions are made to contain in about 100 µl of PCR buffer about 50 pmoles of each of the paired primers and and 2.5 units of Taq polymerase. About 45 cycles of amplification (2', 94° C; 2', 55° C; 2' 72° C) are followed by gel purification of the anticipated 400 + base pair (bp) DNA fragments. Prior to subcloning those DNAs into a blunt-ended intermediate plasmid such as pSP72 (Promega) they are terminally phosphorylated using T4 polynucleotide kinase. Multiple clones representing these PCR amplified sequences are grown and submitted to DNA sequence determinations using Sequenase® and T7 and SP6 specific sequencing primers. A unique DNA sequence representing a murine IgG2a heavy chain variable region is obtained by analysis of the derived amino acid sequences. Replacement of the "murine-appearing" framework residues with those residues compatible with a human variable region is accomplished utilizing the following unique processes. An appropriate human framework is determined utilizing the criteria discussed below. The light chain variable region framework with sufficient homology to the the m1B4 framework was determined to be the human LEN framework (FR). The Len FR shows a similarity of 90% and an identity of 81% when compared to murine 1B4. This sequence, with its leader, 3' intronic sequences and engrafted m1B4 CDRs had been subcloned into the intermediate vector pGEM3Z (Promega), as described in Daugherty et al. Nucleic Acids Res. 19: (1991). About eight oligodeoxynucleotide primers (Figure 5) are synthesized representing the primers necessary to generate by polymerase chain reaction (PCR) amplification four DNA fragments. Incorporated into all but the terminal oligodeoxynucleotide primers were those sequences corresponding to the veneered MAb 1B4 light chain, with its unaltered CDRs, and at least 15 bases of 5'-terminal complementarity to allow for the subsequent PCR- directed recombination of these four fragments. For the purposes of exemplifying the "veneering" process the LEN light chain variable region already containing an engrafted set of CDRs representing those within the light chain of murine 1B4 was used as the template into which mutations were placed so as to easily create the "veneered" framework sequence. The appropriate primer pair (S1 & V9, V10 & V11, etc.), about 50 pmole each, was combined with about 10 ng of plasmid DNA representing the LEN CDR-grafted framework, about 2.5 units of Taq DNA polymerase and about twenty-five (25) cycles of PCR amplification ensued (cycle periods: 1', 94° C; 1', 55° C; 2' 72° C). The products of the four reactions, purified by agarose gel electrophoresis, are combined, about 10 ng of each DNA fragment, along with terminal oligodeoxynucleotide primers (A1 &A2, Figure 6) and Taq DNA polymerase. The combined fragments were PCR amplified (25 cycles of: 2', 94° C; 2', 55° C; 2' 72° C). Following restriction endonuclease digestion with Hind III and Xba I the amplified DNA is purified by agarose gel electrophoresis and subcloned into compatible sites of an intermediate vector pSP72 (Promega) which contains the human kappa light chain constant region (see Figure 7). Genomic DNA, about 1 µg, purified from a human B cell line (GM0108A: NIGMS Human Genetic Mutant Cell Repository, Institute for Medical Research, Camden, NJ) is used as a template for PCR amplification (Figure 4) of about a 920 base pair fragment containing the splice acceptor for the kappa light chain constant domain, the exon and a portion of its 3'-untran-

slated region. The PCR product is purified by agarose gel electrophoresis, digested with Bam H1 endonuclease, and subcloned into pSP72 previously linearized with Bam H1. The individual clones representing the pSP72 intermediate vector containing both the 1B4 "veneered" light chain variable region and the human kappa constant region derived by PCR amplification of human DNA are used to determine the DNA sequence of the "veneered" light chain varisble region.

The "veneered" heavy chain portion of the recombinant antibody is derived from the mutated version of the marine 1B4 heavy chain variable region fused to the human constant region of a gamma 4 subtype obtained from a lambda library constructed by Flanagan and Rabbits, Nature 300: 709-713 (1982). The variable region of the "veneered" heavy chain is constructed from five DNA fragments representing a signal sequence, portions of the mutated murine heavy chain variable region, and an intronic sequence (Figure 8). Oligodeoxynucleotide primer pairs (Figure 5) are synthesized representing the primers necessary to generate by PCR amplification these five DNA fragments from about 10 ng of plasmid DNA template obtained from a pSP72 intermediate vector containing the heavy chain variable region previously used to determine the murine 1B4 CDR sequence. Amplification of the signal fragment, variable region fragments, and intron-containing fragment was as described above. The agarose gel purified products are combined, about 10 ng of each product, with terminal oligodeoxynucleotide primer pairs (Figure 8) and the PCR-generated in vitro recombined template is amplified using the standard procedures described above. Prior to subcloning into a Hind III and Bam HI digested expression vector containing the human heavy chain gamma 4 constant region (Figure 9), this recombined product is similarly digested and agarose gel purified. Individual clones are submitted to DNA sequence determination using Sequenase® and T7 and SP6 specific sequencing primers and one is chosen for subsequent expression. The gamma 4 heavy chain constant region is subcloned as about a 6.7 Kb Hind 111 fragment derived from the plasmid pAT84 into the Kind III site of the intermediate vector pSP72. This plasmid is then used as the template DNA from which a shortened version of the gamma 4 constant region is subcloned using PCR amplification and the primer pairs indicated in Figure 4. Eukaryotic expression vectors are constructed as described below.

Expression vectors are defined herein as DNA sequences that are required for the transcription of cloned copies of genes and the translation of their mRNAs in an appropriate host. Such vectors can be used to express eukaryotic genes in a variety of hosts such as bacteria, blue-green algae, plant cells, yeast cells, insect cells and animal cells. The immunoglobulins may also be expressed in a number of virus systems. Specifically designed vectors allow the shuttling of DNA between host such as bacteria-yeast or bacteria-animal cells. An appropriately constructed expression vector should contain: an origin of replication for autonomous replication in host cells, selectable markers, a limited number of useful restriction enzyme sites, a potential for high copy number, and strong promoters. A promoter is defined as a DNA sequence that directs RNA polymerase to bind to DNA and initiate RNA synthesis. A strong promoter is one which causes mRNAs to be initiated at high frequency. Expression vectors may include, but are not limited to, cloning vectors, modified cloning vectors, specifically designed plasmids or viruses. The heavy chain immunoglobulin molecule is transcribed from a plasmid carrying the neomycin (G418) resistance marker while the light chain immunoglobulin is transcribed from a plasmid carrying the hygromycin B resistance marker. With the exception of the drug resistance portion of these plasmids they are identical. The preferred progenitor of the immunoglobulin expression vectors is the pD5 (Berkner and Sharp, Nucl. Acids Res. 13: 841-857 [1985]) eukaryotic expression vector which contains the origin of adenovirus replication, the SV40 enhancer domain, the adenovirus major late promoter, the adenovirus 2 tripartite leader, a 5' splice donor from the adenovirus third leader and a 3' splice acceptor derived from an immunoglobulin locus, a multiple cloning site placed in the Bam H1 site subsequent to receipt of the vector, and the SV40 late polyadenylation signal (Figure 10). The origin of replication is removed by digestion with Eco R1 and Kpn I and replaced by two fragments representing the neo selectable marker gene (derived from plasmid pCMVIE-AK1-DHFR as an Eco R1/Bam H1 about 1.8 Kb fragment) and the Ig heavy chain enhancer (obtained as a PCR amplified fragment using human DNA as the template, and the oligodeoxynucleotides listed in Figure 4 as the primer pair, following its digestion with Bgl II and Kpn 1). The resultant expression vector is found to lack a small portion of the TK promoter responsible for the transcription of the neomycin gene. This is replaced by insertion into the Eco RI site about a 0.14 Kb PCR amplified fragment derived from the CMVIE-AK1-DHFR DNA using the primer pair listed in Figure 4. The resultant heavy chain expression vector (p8941) is modified by removal of the indicated Hind III and Xba I sites using standard procedures. To convert this vector into one expressing the hyg romycin B selectable marker the neomycin-resistance cassette is removed by digestion first with Eco R1 followed by

DNA polymerase-directed fill in of the 5' overhand, then subsequent Sal I digestion. The about 1.9 Kb hygromycin B expression cassette, TK promoter and TK polyadenylation signal flanking the hygromycin B gene, (obtained as a 1.8 kb Bam H1 fragment in plasmid pL690, Gritz and Davies, Gene 25: 179-188 [1981]) is removed from the plasmid pAL-2 by Bam H1 digestion and subcloned into the Bam H1 site of the intermediate vector pSP72. The hygromycin B cassette is removed from this vector by digestion with Sma I and Sal I and cloned into the expression vector linearized as described above to create a blunt end and Sal I end DNA fragment (Figure 11).

Expression of the 1B4 "veneered" kappa light chain is accomplished by transferring this cistron from the pSP72-based intermediate cloning vector (p8952), containing the human kappa constant region, to the hygromycin B selectable eukaryotic expression vector (Figure 7). An about 1.5 kb DNA fragment resulting from the endonuclease digestion of p8952 with Spe I and Cla I is purified by agarose gel electrophoresis and ligated into the expression vector which has previously been linearized, following digestion with the same two restriction enzymes, and agarose gel purified. The heavy chain eukaryotic expression vector is constructed in two steps. First, the p8950 vector containing the modified heavy chain variable region of murine 1B4 fragment is digested with Bgl II and Bam H1. The agarose gel purified 0.75 kb fragment is ligated into the Bam H1 site of the p8941 vector and recombinant clones containing this fragment in the proper orientation are identified. Plasmid DNA from one such clone is linearized by Bam H1 digestion and ligated with a 1.78 Kb Bam H1 fragment representing a short version of the human gamma 4 constant region, derived from plasmid pAT84 by PCR amplification. Following the identification of clones containing these inserts in the appropriate orientation, plasmid DNAs (one which is referred to as p8953) are grown and purified for transfection into recipient mammalian cells. Equal amounts, about 10 µg, of the plasmids encoding the 1B4 "veneered" IgG4 heavy chain and the 1B4 "veneered" kappa light chain are transfected by standard calcium phosphate precipitation procedures into the monkey kidney cell line CV-1P or the human embryonic kidney cell line 293. The culture supernants, assayed by a trapping ELISA (described below), were found to contain a human kappa light chain / human IgG4 immuno-globulin. Immulon-2 (Dynatech Labs.) 96-well plates are coated overnight with about a 5 µg/ml solution of mouse anti-human kappa chain constant domain monoclonal antibody (cat. #MC009, The Binding Site, Inc., San Diego, CA) in about 0.1 M $NaHCO_3$ buffer (pH 8.2) at about 4° C, and blocked with about 1% bovine serum (BSA) in about 0.1 M $NaHCO_3$ for about 1 hour at about 25° C. After this and all subsequent steps, washing was performed with phosphate buffered saline (PBS). The wells are then challenged with conditioned medium containing recombinant anti-CD18 antibody, or with predetermined quantities of human IgG4/kappa purified by protein A Sepharose (Pharmacia Fine Chemicals) chromatography from human IgG4 myeloma serum (cat. # BP026, The Binding Site, Inc.) All samples are diluted in PBS containing about 0.05% Tween-20. About 100 µl aliquots are incubated for about 1 hour at about 37° C in triplicate, and standard calibration curves are constructed using IgG4 concentrations ranging from about 10 ng/ml to about 100 ng/ml. Bound and fully assembled human IgG4 (either native or the recombinant 1B4 human "veneered" IgG4 constructs) are detected with about 100 µl aliquots of a 1:500 dilution of mouse anti-human IgG4 Fc monoclonal antibody conjugated to alkaline phosphatase (cat #05-3822, Zymed Laboratories, Inc.) in phosphate buffered saline (PBS) containing about 1 % BSA. After incubation for about 1 hour at about 37° C and subsequent washing, the quantities of bound conjugate are detected by incubating all samples with a 1 mg/ml solution of p-nitrophenyl phosphate in 0.1 M 2,2' amino methylpropanediol buffer, pH 10.3, for about 30 minutes at about 25° C. The adsorbance of the wells is determined with a UV Max ELISA plate reader (Molecular Devices) set at 405 nm. The antibody secreted by the transfected human 293 cells or monkey kidney CV1 P cells, either following transient expression or subsequent to stable clone isolation, is isolated by protein A chromatography, the concentration of recombinant human anti-CD18 antibodies determined by the trapping Elisa described above, and used to compete with the binding of radiolabeled murine 1B4 to the CD18 ligand on the surface of activated human PMNs. Affinities of r-anti-CD18 antibody constructs are determined using a competitive ^{125}I-1 B4 soluble binding assay with stimulated human polymorpho-nuclear leukocytes (PMNs). Purified murine anti-CD18 monoclonal antibody (50 ug) is iodinated using chloramine-T (Hunter, W.M. and Greenwood, F.C., Nature 194: 495-496, 1962), and the radiolabeled antibody purified using a Bio-Sil TSK250 (Biorad, Richmond, CA) gel filtration HPLC column (which fractionates proteins in the range of 1-300 x 103 daltons) equilibrated in 0.1 M phosphate buffer, pH 7.0. Effluent radioactivity is monitored with an in-line detector (Beckman Model 170; Beckman, Fullerton,CA) and total protein measured at OD_{280} with a Kratos Spectroflow 757 detector (Kratos, Mawah, N.J.). A single ^{125}I-1B4 peak composed of coincident OD_{280} and radioactivity tracings char-

acteristically elutes at about 6 minutes, 30 seconds following sample injection. Specific activity of the product is generally about 10 µCi/µg protein, and 97-99% of the counts are precipitable with 10% trichloroacetic acid. The binding of this radiolabeled antibody is assessed on human PMNs purified on a discontinuous Ficoll/Hypaque gradient (English, D. and Anderson, B.R., J. Immunol. Methods 5: 249-255, 1974) and activated with about 100 ng/ml phorbol myristate acetate for about 20 minutes at about 37°C (Lo et al., J.Exp. Med. 169: 1779-1793, 1989). To determine the avidity of antibodies for CD18 molecules on the PMN surface, about 1 x 10^5 activated PMNs are incubated in a buffer such as Hanks balanced salt solution containing about 20 mM Hepes (pH 7.2), about 0.14 units aprotinin (Sigma Chemical Co.) and about 2% human serum albumin (binding buffer) containing about 1.3 ng ^{125}I-1B4 (2.8×10^{-11} M) in the presence of increasing concentrations of unlabeled 1B4 antibody (about 10^{-7} to 10^{-15} M) in about a 300 µl reaction volume for about 1 hour at about 4°C with constant agitation. Cell bound 1B4 is separated from the unbound antibody by centrifugation through a 0.5 M sucorse cushion (4,800 x g, 3 minutes); the tubes are frozen on dry ice, and the tips cut off and counted with an LKB gamma counter. The IC_{50} of the anti-CD18 antibody for the inhibition of ^{125}I-1B4 antibody binding is calculated using a four-parameter fitter program (Rodbard et al., In, "Radioimmunoassay and Related Procedures in Medicine", International Atomic Energy Agency, Vienna, vol 1,469 - 504, 1978). The affinity of the "veneered" r-anti-CD18 antibody for the CD18 ligand is determined in a similar manner using murine ^{125}I-1B4 antibody and increasing quantities, as determined by the trapping Elisa, of unlabeled r-anti-CD18. The results of the binding assays are shown in Figure 13 and indicate that the avidity of the "veneered" recombinant 1B4 antibody is equal to that of the murine 1B4 monoclonal antibody. This result shows that an antibody with presumptive human isotype may be recombinantly constructed from the murine parent antibody by the introduction of numerous point mutations in its framework residues and expressed fused to human kappa and gamma 4 constant domains without loss in avidity for the antigen. It can be inferred from this result that the point mutations within the framework regions do not alter the presentation of the murine 1B4 light chain and heavy chain CDRs. Many of the examples of construction of recombinant human antibodies containing complementarity regions replaced by those found within murine monoclonal antibodies have resulted in loss of avidity for the ligand or antigen. Thus, although these latter transmutations are possible, the successful maintenance of avidity is not assured. This procedure described above demonstrates that when strict attention is payed to the framework regions, and the nature of the amino acids within each framework, "humanization" may potentially be achieved without the loss of avidity which accompanies the transfer of CDRs to the "generic" human frameworks ("humanization") employed by Winter, European Patent Publication No. 239,400, published September 30, 1987.

To identify human framework sequences compatible with the CDRs of, say, murine 1B4, human frameworks with a high degree of sequence similarity to those of murine 1B4 are identified. Sequence similarity is measured using identical residues as well as evolutionarily conservative amino acid substitutions. Similarity searches are performed using the murine 1B4 framework sequence from which the CDR sequences had been removed. This sequence is used to query a database of human immunoglobulin sequences that had been derived from multiple sources. Sequences with a high degree of sequence similarity are examined individually for their potential as humanizing framework sequences. In this way, the human homologue providing the murine CDRs with the structure most similar to their native murine framework is selected as the template for the construction of the "veneered" variable regions (Figure 12). Should human frameworks of sufficient similarity not be identifiable from compiled sequences, it is possible to isolate from human genomic DNA a group of closely related variable regions using recombinant technology. Thus, a degenerate 5' upstream oligodeoxynucleotide primer may be designed from the conserved sequences within the amino-terminus of each of the various human FR1 regions and paired with a degenerate 3' downstream oligodeoxynucleotide primers fashioned from the FR sequence determined from the murine monoclonal whose CDRs one wishes to transfer into a human context. These primer pairs are then used to PCR amplify from a human genomic template those DNA sequences which are flanked by the primer pair. The resulting DNAs may then be cloned and the DNA sequence derived from individual members will describe various murine-related human variable regions. The paucity of somatic mutations in framework residues and the conservation of amino acid sequence between mouse and man make this approach possible.

The construction of a complete recombinant human IgG4 antibody, whose heavy and light chain variable domains contain the CDR residues of the murine monoclonal antibody, with complete retention of the specificity and avidity of the parent murine monoclonal antibody is disclosed. The construction of the "veneered" light chain framework derived from the human sequence of LEN fused

with a human kappa light chain constant region is described above. The murine variable region framework sequence, devoid of CDR sequences, is used to query a database of complete human variable region sequences. The human sequences that are most similar to the murine framework region are then analyzed individually to determine both their sequence identity and similarity to the murine framework region. In the case of murine 1B4 these sequences include, but are not limited to, "Gal", chosen because of its high degree of both similarity and identity with the 1B4 heavy chain sequence. The Gal FR has been found to be 85% similar and 79% identical to murine 1B4. These values are based upon the Dayhoff similarity matrix of evolutionarily conserved amino acid substitutions (R. M. Schwartz, M. O. Dayhoff, in Atlas of Protein sequence and structure M. O. Dayhoff, Eds. (National Biomedical Research Foundation, Washington, DC [1979]) (Figure 12). To prepare a recombinant DNA encoding the murine heavy chain CDRs in the context of a human-appearing framework the following procedures are performed. A set of ten short oligodeoxynucleotides are synthesized. Each pair is combined in a separate PCR reaction with the DNA template representing the murine 1B4 heavy chain variable region, amplified and isolated following PCR of the RNA of the murine hybridoma 1B4 as described above. Thus, about 50 pmole of each primer pair was combined with about 10 ng of plasmid DNA representing the murine 1B4 heavy chain variable region, about 2.5 units of Taq DNA polymerase and about twenty-five (25) cycles of PCR amplification ensued (cycle periods: 1', 94°C; 1', 55°C; 2' 72°C). The products of the five reactions (Figure 8) encoded portions of the 1B4 heavy chain variable region, beginning with the signal peptide encoding region and ending with the 3' intronic sequence which resides between the variable region coding domain and the IgG4 constant region sequence, with the desired point mutations to create a "veneered" variable region framework. These five fragments are purified by agarose gel electrophoresis, combined, about 10 ng of each DNA fragment, along with terminal oligodeoxynucleotide primers (A1 &A2, Figure 5) and Taq DNA polymerase. The combined fragments were PCR amplified (25 cycles of: 2', 94°C; 2', 55°C; 2' 72°C). By virtue of the complementary ends of the five fragments, the polymerization/denaturation/polymerization cycles of the polymerase chain reaction result in the formation, and subsequent amplification, of the combined sequences. Following 25 cycles of amplification the combined 0.8 Kb fragment is electrophoretically purified from an agarose gel and was digested with restriction enzymes Spe I and Bam H1. Following agarose gel electrophoresis, the purified DNA fragment is ligated into the heavy chain expression vector, p8958 (see Figure 9), in place of the chimaeric variable region existing in this vector. Each "veneered" variable region, with its associated human constant region, residing within a pD5-based expression vector plasmid was co-transfected into 293 cells and CV1 P cells and recombinant human antibody is found to be present in the conditioned medium 48 hours post transfection. The "veneered" recombinant antibody is isolated by protein A chromatography. The avidity of this antibody for the CD18 ligand displayed on the surface of activated human PMNs is compared with that of the murine 1B4 monoclonal antibody parent. Figure 13 shows that although each antibody contains the same set of six CDRs within different framework domains, they exhibitidentical avidity for the ligand. Thus, the avidity of an antibody molecule does not rely upon the variable region framework residues which are surface exposed, rather the proper structure in which the CDRs are presented must be significantly influenced by the buried and inter/intra active residues. The parent murine monoclonal antibody demonstrates an IC_{50} of about 1.0 to about 0.7 nM, the "veneered" molecule has a similar IC_{50}.

This invention further relates to a method of inhibiting the influx or migration of leukocytes capable of expressing CD18 antigen (leukocyte integrin, beta subunit) on their surface into a site of inflammation or a tissue area or organ that will become inflamed following an influx of the cells. The inflammation which is the target of the method of the present invention may result from an infection with pathogenic microorganisms such as gram-positive and gram-negative bacteria, parasites and fungi. The response may also be induced by viruses and non-infectious means such as trauma or reprefusion following myocardial infarction or stroke, immune responses to foreign antigen and autoimmune responses. The recombinant human anti-CD18 antibodies are useful in the treatment of inflammation in lung, central nervous system, kidney, joints, endocardium, eyes, ears, skin, gastrointestinal tract and urogenital system. Disease states in which the recombinant human anti-CD18 antibodies are useful as therapeutic agents include, but are not limited to: infectious diseases where active infection exists at anybody site, such as meningitis; conditions such as chronic or acute secondary inflammations caused by antigen deposition; and other conditions such as, encephalitis; arthritis; uveitis; colitis; glomerulonephritis; dermatitis; psoriasis; and respiratory distress syndrome associated with sepsis and/or trama. Other inflammatory diseases which may be responsive to recombinant human anti-CD18 antibody include, but are not limited to, immune disorders and conditions involv-

ing T-cell and/or macrophage attachment/recognition, such as acute and delayed hypersensitivity, graft vs. host disease; primary autoimmune conditions such as pernicious anemia; infection related autoimmune conditions such as Type I diabetes mellitis; flares during rheumatoid arthritis; diseases that involve leukocyte diapedesis, such as multiple sclerosis; antigen-antibody complex mediated diseases including certain of the secondary infection states listed above; immunosuppression; and transplant rejection. Inflammatory conditions due to toxic shock or trauma such as adult respiratory distress syndrome and reperfusion injury; and disease states due to leukocyte dyscrasias and metastasis, are included within the scope of this invention. The present invention is also applicable to the inhibition of leukocyteendothelial attachment for diagnostic and therapeutic purposes; such as the iatrogenic opening of the endothelium to prevent the ingress of leukocytes during the ingress of a therapeutic drug in the instance of chemotherapy; or to enhance the harvesting of leukocytes from patients.

Recombinant human anti-CD18 antibodies or an active fragment thereof can be used to treat the above mentioned diseases. An active fragment will include the F(ab')2, the Fab and any other fragment that can bind to the CD18 antigen. Recombinant human anti-CD18 antibodies can be administered alone for non-infectious disease states or combined with antibiotics or other anti-infective agents for the treatment of infectious diseases for reasons discussed above. Administration will generally include the antibodies and possibly other substances in a physiologically acceptable medium or pharmaceutical carrier. Such physiologically acceptable media or phamaceutical carriers include, but are not limited to, physiological saline, phosphate buffered saline, phosphate buffered saline glucose, buffered saline and the like. The antibodies and any anti-infective agent will be administered by parenteral routes which include intravenous, intramuscular, subcutaneous and intraperitoneal injection or delivery. The amount of the antibodies and the mixture in the dosage form is dependent upon the particular disease state being treated. The amount of the recombinant human anti-CD18 antibody utilized in a dosage form can range from about 1 to about 1,000 mg, with a range of from about 10 mg to about 100 mg being preferred. The antibodies can be administered daily or less than daily as determined by the treating physician. The following examples illustrate the present invention without, however, limiting the same thereto.

EXAMPLE

Preparation of a "Veneered" Recombinant Antibody

An antibody was produced in which the variable domain of the light chain comprises the framework region of a murine light chain modified to contain surface exposed amino acids of human derivation. The variable domain of the heavy chain is similarly derived from the murine heavy chain with point mutations which replace murine exposed residues with human-appearing residues. The light chain human framework region was derived from human myeloma protein LEN. The CDR and framework sequences from the murine monoclonal antibody 1B4 which binds to CD18 (the beta subunit of the leukocyte integrin B-2 family which includes: LFA-1, Mac-1, and p150.95) were derived as follows. The hybridoma designated 1B4 which produces 1B4 monoclonal antibody was deposited under the Budapest Treaty at the International Depository Authority: American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD, 20852. Viability was determined on June 6,1989 and the hybridoma was designated HB 10164. Previous experiments had determined this antibody to be an IgG 2a with a kappa light chain (Wright et al., Proc. Natl. Acal. Sci. USA 80: 5699-5703 [1983]).

Total RNA was extracted from the 1B4 myeloma cells using standard methods involving cellular solubilization with guanidinium isothiocyanate (Chirgwin et al., Biochem. 18:5294-5299 [1979]). Sets of degenerate oligodeoxynucleotide primers (Figure 4) representing sequences within framework 1 of the murine kappa light chain variable region and kappa light chain constant domain, or those within framework 1 of the murine IgG2a heavy chain variable region and heavy chain constant CH1 domain were synthesized by standard phosphoramidite procedures on an Applied Biosystem 381A DNA synthesizer. Removal of the oligodeoxy-nucleotides (oligos) from the resin was accomplished by treatment with concentrated NH_4OH followed by desalting on a NAP-5 column (Pharmacia) with H_2O elution (when the oligos were <45 bases in length), or by use of an OPC column (Applied Biosystems Inc) with 20% acetonitrile elution (when the oligos were >45 bases in length), as recommended by the manufacturers. Total RNA (2µg) was reversed transcribed for 30' at 42°C using Moloney MLV reverse transcriptase (200 units, BRL) and 10 pmoles of the constant region complementary strand primers representing either heavy or light chain in a buffer (final volume of 20 µl) containing 50 mM Tris HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, and 20 units of RNAsin (Pharmacia). The reverse transcriptase was heat inactivated (95°C, 5') and the reactions were made to contain in 100µl of

PCR buffer (10 mM Tris HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 µM each dNTP), 50 pmoles of each of the paired primers, and 2.5 units of Taq polymerase (Perkin Elmer/Cetus). Polymerase chain reaction (PCR) amplification was carried out essentially as described by Saiki et al., Science 230: 1350-1354 (1985) and others (Mullis et al., Cold Spring Harbor Symp. Quant. Biol.51: 263-273 [1986], Dawasaki and Wang, PCR Technology, Princples and Applications for DNA Amplification, Erlich, Ed., Stockton Press, NY, pp. 89-97 [1989], Tung et al., ibid. pp. 99-104 [1989]). Forty five cycles of amplification by a DNA Thermal Cycler (Perkin Elmer Cetus Instruments)(2', 94°C; 2', 55°C; 2' 72°C) were followed by gel purification of the anticipated 400+ base pair (bp) DNA fragments. Prior to subcloning the DNAs into a blunt-ended intermediate plasmid (pSP72, Promega) they were terminally phosphorylated using T4 polynucleotide kinase (Boehringer Mannheim). Multiple clones representing these PCR amplified sequences were isolated form DH5 transformed E.coli plated on LB agar plates containing 50 µg/ml ampicillin, grown by described procedures (Maniatis et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), plasmid DNAs were extracted from the bacteria using the DNA preparation procedures of Birnboin and Doly, Nucleic Acid Res.7: 1515 (1979), and the double-stranded plasmid DNAs were submitted to DNA sequence determinations using Sequenase® (United States Biochemicals) and T7 and SP6 specific sequencing primers (Boehringer Mannheim) using the protocols recommended by the manufacturer. A unique DNA sequence representing a murine IgG2a heavy chain variable region was obtained, as was a kappa light chain variable region sequence.

To give the final appearance of a "veneered" murine light chain, several residues within a template composed of the human LEN framework, into which had been grafted the CDRs described for 1B4, were replaced by corresponding residues found in the murine 1B4 light chain framework. Replacement of the human LEN variable region residues with those unique to MAb 1B4 took place as follows. Eight oligodeoxynucleotides (Figure 5) were synthesized representing the primers necessary to generate by PCR amplification four DNA fragments. Incorporated into all but the terminal oligodeoxynucleotides were those sequences corresponding to the MAb 1B4 light chain variable region framework residues to be point mutated and at least 15 bases of 5'-terminal complementarity (see Figure 6). The appropriate primer pair (50 pmole each) was combined with 10 ng of a 1B4 CDR-grafted LEN framework-containing plasmid DNA, 2.5 units of Taq DNA polymerase, PCR reaction components and buffer, and twenty-five (25) cycles of PCR amplification ensued (cycle periods: 1', 94°C; 1', 55°C; 2' 72°C). The products of the four reactions, purified by agarose gel electrophoresis, were combined (10 ng of each DNA fragment) along with a terminal oligodeoxynucleotide primer pair (amplifier) (Figures 5 & 6), Taq DNA polymerase, PCR reaction components and buffer, and the subsequent recombined fragments were amplified, as described above, for twenty-five (see Figure 6). Following restriction endonuclease digestion with HindIII and XbaI the amplified DNA was purified from an agarose gel and subcloned into these same sites of an intermediate vector pSP72 (Promega) which contained the human kappa light chain constant region, obtained as follows. DNA (1µg) purified from a human B cell line (GM01018A; NIGMS Human Genetic Mutant Cell Repository, Institute for Medical Research, Camden, N.J. 08103) was used as a template for the oligodeoxynucleotide primers described in Figure 4 to PCR amplify a 920 base pair fragment containing the splice acceptor for the human kappa light chain constant domain, the exon and a portion of its 3'-untranslated region (PCR primer pair choice was selected based on the kappa constant region sequence described by Hieter et al., Cell 22: 197-207 (1980). The PCR product was purified by agarose gel electrophoresis, digested with Bam H1 endonuclease, and subcloned into pSP72 (Promega) previously linearized with Bam H1.

The individual clones representing the pSP72 intermediate vector containing both the 1B4 "veneered" light chain variable region derived as described above, and the human kappa constant region, derived by PCR amplification of human DNA, were used to verify the DNA sequence of the "veneered" light chain variable region. The "veneered" heavy chain portion of the recombinant antibody was derived from a point mutated murine 1B4 heavy chain variable region fused to the human constant region of gamma 4 subtype obtained from a lambda library constructed by Flanagan and Rebbitts, Nature 300: 709-713 (1982).

The variable region of the "veneered" heavy chain was constructed from five DNA fragments representing a signal sequence, mutated portions of the murine 1B4 heavy chain variable region, and an intronic sequence (Figure 8). Oligodeoxynucleotide primer pairs (Figure 5) were synthesized representing the primers necessary to generate by PCR amplification these five DNA fragments from 10 ng of plasmid DNA template containing the murine 1B4 heavy chain variable region previously used to determine the murine 1B4 CDR and framework sequences. Amplification of the five fragments was performed as described above for the four

light chain variable region fragments. The agarose gel purified products were combined (10 ng of each product) with terminal primer pairs (Figure 5) and the PCR-generated *in vitro* recombined template was amplified using the standard procedure also described above for recombining the fragments comprising the "veneered" light chain variable region. Prior to subcloning into a Hind III and Bam HI digested expression vector this recombined product was similarly digested and agarose gel purified. DNA was obtained following growth of individual bacterial clones and submitted to DNA sequence determination using Sequenase® and T7 and SP6 specific sequencing primers in order to verify the sequence of the reconstructed variable region and its flanking domains.

The gamma 4 heavy chain constant region had been subcloned as a 6.7 Kb Hind III fragment derived from the plasmid pAT84 (Flanagan and Rebbitts, supra)into the Hind III site of the intermediate vector pSP72 (Promega). This plasmid was then used as the template DNA from which a shortened version of the gamma 4 constant region was obtained using the standard PCR amplification procedures described above and the primer pairs indicated in Figure 4. Eukaryotic expression vectors were constructed as described below such that the heavy chain immunoglobulin molecule was transcribed from a plasmid carrying the neomycin (G418) (Rothstein and Reznikoff, Cell 23: 191-199 [1981]) resistance marker, while the light chain immunoglobulin was transcribed from a plasmid carrying the hygromycin B resistance marker (Gritz and Davies, Gene 25: 179-188 [1983]). With the exception of the drug resistance portion of these plasmids they are identical.

The progenitor of the immunoglobulin expression vectors was the pD5 eukaryotic expression vector (Berkner and Sharp, Nucl. Acids Res. 13: 841-857 [1985]) which contained the origin of adenovirus replication, the SV40 enhancer domain, the adenovirus major late promoter, the adenovirus 2 tripartite leader, a 5' splice donor from the adenovirus third leader and a 3' splice acceptor derived from an immunoglobulin locus, a multiple cloning site, and the SV40 late polyadenylation signal (Figure 10). The origin of replication was removed by digestion with Eco R1 and Kpn I and replaced by two fragments representing the neo selectable marker gene (derived from plasmid pCMVIE-AK1-DHFR (Silberklang et al., Modern Approaches to Animal Cell Technology. Ed. Spier et al., Butterworth, U.K., [1987]) as an Eco R1/Bam HI 1.8 Kb fragment) and the Ig heavy chain enhancer (obtained as a PCR amplified fragment using standard procedures described above and human DNA as the template; the oligodeoxynucleotide primer pair is listed in Figure 4) following its digestion with Bgl II and Kpn I. The resultant expression vector was found to lack a small portion of the TK promoter responsible for the transcription of the neomycin gene. This was replaced by insertion into the Eco RI site of a 0.14 kb PCR amplified fragment derived from the CMVIE-AK1-DHFR DNA using the primer pair also listed in Figure 4. The resultant heavy chain expression vector was subsequently modified by removal of the indicated Hind III and Xba I sites. To convert this neomycin selectable vector into one expressing the hygromycin B selectable marker (Figure 11) the neomycin-resistance cassette was removed by digestion first with Eco R1 followed by DNA polymerase-directed fill in of the 5' overhang, then subsequent Sal I digestion. The 1.9 kb hygromycin B expression cassette [TK promoter and TK polyadenylation signal flanking the hygromycin B gene obtained from Gritz and Davies, Gene 25: 179-188(1983), as the 1.9 kb Bam H1 fragment in plasmid (pLG90)] was removed from the plasmid pAL-2 by Bam H1 digestion and subcloned into the Bam H1 site of the intermediate vector pSP72 (Promega). The hygromycin B cassette was removed from this vector by digestion with Sma I and Sal I and cloned into the expression vector linearized as described above to create a blunt end and Sal I end DNA fragment.

Expression of the 1B4 "veneered" kappa light chain was accomplished by transferring this cistron from its position within the pSP72 intermediate vector to the hygromycin B selectable eukaryotic expression vector (Figure 7). A 1.5 kb DNA fragment resulting from the endonuclease digestion of v1B4 VK/pSP72 intermediate vector with Spe I and ClaI was purified by agarose gel electro-phoresis and ligated into the expression vector which had previously been linearized, by digestion with the same two restriction enzymes and agarose gel purified.

The 1B4 "veneered" heavy chain eukaryotic expression vector was constructed in one step (Figure 9) from an existing vector previously constructed to express a chimaeric form of the 1B4 heavy chain. The "veneered" heavy chain variable region created by PCR amplifcation (Figure 8) was digested with Hind III and Bam H1. The agarose gel purified 0.8 kb fragment was ligated into the Hind III and Bam H1 sites of the pD5/IgH-Enhancer/Neo/1B4 VH-Short Human C-Gamma 4 expression vector following its endonuclease digestion with these two enzymes and subsequent purification by agarose gel electrophoresesis (Figure 9). Transformants containing both variable and constant regions were identified. Plasmid DNAs were grown (Maniatis et al., supra) and purified for transfection into recipient mammalian cells (Maniatis et al., supra; Birbion and Doly, supra.).

Equal amounts (10µg) of the plasmids encoding the "veneered" IgG4 heavy chain and the "veneered" kappa light chain were transfected by standard calcium phosphate precipitation procedures into human 293 cells and african green monkey kidney CV-1P cells. The culture supernatant fluids were assayed by a trapping Elisa (described below) for the secretion of a human kappa light chain containing IgG4 immunoglobulin.

An Elisa was developed for the quantitation of the amounts of a 1B4 recombinant antibody expressed in conditioned mammalian cell growth medium. Immulon-2 (Dynatech Labs.) 96-well plates are coated overnight with a 5µg/ml solution of mouse anti-human k chain constant domain monoclonal antibody (cat. #MC009, The Binding Site, Inc., San Diego, CA) in 0.1M $NaHCO_3$ buffer (pH 8.2) at 4°C, and blocked with 1% bovine serum (BSA) in 0.1M $NaHCO_3$ for 1h at 25°C. After this and all subsequent steps, washing was performed with phosphate buffered saline (PBS). The wells are then challenged with conditioned medium containing recombinant anti-CD18 antibody, or with predetermined quantities of human IgG4 purified by protein A Sepharose (Pharmacia Fine Chemicals) chromatography from human IgG4 myeloma serum (cat. # BP026,The Binding Site, Inc.). All samples are diluted in PBS containing 0.05% Tween-20. 100µl aliquots are incubated for 1h at 37°C in triplicate, and standard calibration curves are constructed using IgG4 concentrations ranging from 10 ng/ml to 100 ng/ml. Bound and fully assembled human IgG4 (either native or recombinant "veneered"1B4 human IgG4 constructs) are detected with 100µl aliquots of a 1:500 dilution of mouse anti-human IgG4 Fc monoclonal antibody conjugated to alkaline phosphatase (cat #05-3822, Zymed Laboratories, Inc.) in phosphate buffered saline (PBS) containing 1% BSA. After incubation for 1h at 37°C and subsequent washing, the quantities of bound conjugate are detected by incubating all samples with a 1 mg/ml solution of p-nitrophenyl phosphate in 0.1M 2,2'amino-methyl-propanediol buffer, pH 10.3, for 30 min at 25°C. The adsorbance of the wells is determined with a UV Max ELISA plate reader (Molecular Devices) set at 405 nm. All supernatant fluids contain this immunoglobulin, though in various amounts. The antibody secreted by the transfected 293 cells is concentrated by protein A chromatography and the concentrations of the recombinant human "veneered" anti-CD18 antibody determined by the trapping Elisa described above, is used to compete with the binding of radiolabeled murine 1B4 to the CD18 ligand on the surface of activated human PMNs. Affinities of various anti-CD18 antibody constructs are determined using a competitive ^{125}I-m1B4 soluble binding assay with stimulated human polymorphonuclear leukocytes (PMNs). Purified murine anti-CD18 monoclonal antibody (50 ug; m1B4) is iodinated using chloramine-T (Hunter, W.M. and Greenwood, F.C., Nature 194: 495-496, 1962), and the radiolabeled antibody purified using a Bio-Sil TSK250 (Biorad, Richmond, CA) gel filtration HPLC column (which fractionates proteins in the range of 1-300 x 10^3 daltons) equilibrated in 0.1M phosphate buffer, pH 7.0. Effluent radioactivity is monitored with an in-line detector (Beckman Model 170; Beckman, Fullerton,CA) and total protein measured at OD_{280} with a Kratos Spectroflow 757 detector (Kratos, Mawah, N.J.). A single ^{125}I-m1B4 peak composed of coincident OD_{280} and radioactivity tracings characteristically elutes 6 minutes, 30 seconds following sample injection. Specific activity of the product is, generally about 10µCi/µg protein, and 97-99% of the counts are precipitable with 10% trichloroacetic acid. The binding of this radiolabeled antibody is assessed on human PMNs purified on a discontinuous Ficoll/Hypaque gradient (English, D. and Anderson, B.R., J. Immunol. Methods 5: 249-255, 1974) and activated with 100 ng/ml phorbol myristate for 20 minutes at 37°C (Lo, et al., J. Exp. Med. 169: 1779-1793, 1989). To determine the avidity of antibodies for CD18 molecules on the PMN surface, about 1 x 10^5 activated PMNs are incubated in a buffer such as Hanks balanced salt solution containing 20 mM Hepes (pH 7.2), 0.14 units aprotinin (Sigma Chemical Co.) and 2% human serum albumin (binding buffer) containing 1.3 ng ^{125}I-m1B4 (2.8 x 10^{-11}M) in the presence of increasing concentrations of unlabeled m1B4 antibody (10^{-7} to 10^{-15}M) in a 300 ul reaction volume for about 1 h at about 4°C with constant agitation. Cell bound 1B4 is separated from the unbound antibody by centrifugation through a 0.5M sucrose cushion (4,800 x g, 3 minutes); the tubes are frozen on dry ice, and the tips cut off and counted with an LKB gamma counter. The tC_{50} of the anti-CD18 antibody for the inhibition of ^{125}I-m1B4 antibody binding is calculated using a four parameter fitter program (Rodbard, D, Munson, P.J., and DeLean, In, "Radioimmunoassay and Related Procedures in Medicine", International Atomic Energy Agency, Vienna, vol I,469-504, 1978). The affinity of the "veneered" anti-CD18 antibody for the CD18 ligand is determined in a similar manner using murine ^{125}I-m1B4 antibody and increasing quantities, as determined by the trapping Elisa, of unlabeled "veneered" anti-CD18 antibody. The results of the binding assays are shown in Figure 13 and indicate that the avidity of the "veneered" heavy chain and light chain recombinant 1B4 antibody is equivalent to that of the murine 1B4 monoclonal antibody.

The "veneered" heavy and light chain expression vectors were co-transfected into CV1P mon-

key kidney cells using 20 µg of each plasmid to prepare 2 mL of the calcium phosphate precipitated solution. One mL was placed in the medium overlaying each 100 mm dish of CV1P cells. After 4 hr at 37°C the medium was replaced with 1 mL of 15% glycerol in 1 x HBS (Hepes buffered salt). Following the 3 min glycerol shock, 10 mL of PBS as added, the cell monolayers were aspirated, washed once with 10 mL of PBS, and re-fed with fresh medium (DMEM + 10% heat inactivated new born calf serum) containing 200 µg of hygromycin B and 800 µg of G418 per mL. Cloning cylinders (Fishney, In, Culture of Animal Cells, Alan R. Liss, Inc. New York, 1983) were used to isolate individual colonies prior to their expansion and subsequent assay for productivity. Two clones, #11 and #48, were found to express sufficient amounts of v1B4 to warrant their expansion and ultimate accessioning.

Claims

1. A method for identifying differences in mammalian species specific surface amino acid residues on an immunoglobulin comprising:
   a. comparing the framework amino acids of a variable domain of a first mammalian species with the variable domains of a second mammalian species;
   b. determining the subgroups of the second mammalian species to which the first mammalian species most closely corresponds;
   c. determining the second mammalian species sequence which is most similar to the first mammalian species sequence;
   d. identifying amino acid residues of the first mammalian species which differ from the amino acid residues of the second mammalian species, with said amino acids being mostly exposed or completely exposed on the immunoglobulin surface;
   e. identifying only those amino acid residues which are not within a complementarity-determining region or are not directly adjacent to a complementarity-determining region.

2. The method of claim 1 wherein the first mammalian species is mouse.

3. The method of claim 1 wherein the second mammalian species is human.

4. A method for converting an immunoglobulin having the immunogenicity of a first mammalian species to an antibody having the immunogenicity of a second mammalian species comprising:
   a. replacing the amino acid residues in a first mammalian species framework which differ from the amino acid residues of a second mammalian species with the corresponding amino acid residues from the most similar second mammalian species as identified by the method of claim 1.

5. The method of claim 2 wherein the first mammalian species is mouse.

6. The method of claim 2 wherein the second mammalian species is human.

7. A method comprising:
   a. preparing a DNA sequence encoding a veneered immunoglobulin having specificity for a known antigen wherein the surface amino acid residues of a first mammalian species which differ from the surface amino acid residues of a second mammalian species are replaced with the corresponding amino acids residues form the most similar second second mammalian species sequence as identified by the method of claim 1;
   b. inserting the sequence into a replicable expression vector operably linked to a suitable promoter compatible with a host cell;
   c. transforming the host cell with the vector of b;
   d. culturing the host cell; and
   e. recovering the veneered immunoglobulin from the host cell culture.

8. The method of claim 7 wherein the first mammalian species is mouse.

9. The method of claim 7 wherein the second mammalian species is human.

10. A composition comprising a veneered immunoglobulin having a specificity for a known antigen.

11. A DNA sequence encoding veneered 1B4 antibody.

12. A veneered murine 1B4 antibody exhibiting the antigenicity of human antibody of fragments thereof.

Fig. 1a

Position	KOL Residue	KOL Exposure	J539 Residue	J539 Exposure	Residues In Subgroup I	II	III
1	E	1.00 Ex	E	1.00 Ex	Q	Q	E
2	V	0.23 mB	V	0.37 mB	V	V	V M
3	Q	0.82 Ex	K	0.82 Ex	Q	T Q	Q
4	L	0.00 Bu	L	0.10 Bu	L	L	L
5	V	0.87 Ex	L	1.00 Ex	V	R Q K T	V L
6	Q	0.00 Bu	E	0.09 Bu	Q	E	E
7	S	0.94 Ex	S	0.94 Ex	S	S	S
8	G	1.00 Ex	G	1.00 Ex	G	G	G
9	G	0.00 Bu	G	0.00 Bu	A	P	G
10	G	1.00 Ex	G	1.00 Ex	E	A G T	G A
11	V	0.90 Ex	L	0.81 Ex	V	L	L F
12	V	0.25 mB	V	0.25 mB	K	V	V
13	Q	0.71 mE	Q	0.87 Ex	K	K	Q
14	P	0.59 PB	P	0.64 mE	P	P	P
15	G	1.00 Ex	G	1.00 Ex	G	T S	G
16	R	0.73 mE	G	1.00 Ex	S A	E Q	G
17	S	0.66 mE	S	0.75 mE	S	T	S
18	L	0.28 mB	L	0.26mB	V	L	L
19	R	0.66 mE	K	0.75 mE	R K	T S	R K
20	L	0.00 Bu	C	0.00 Bu	V	L	L
21	S	0.71 mE	S	0.82 Ex	S	T	S
22	C	0.00 Bu	C	0.00 Bu	C	C	C
23	S	1.00 Ex	A	1.00 Ex	K	T	A
24	S	0.00 Bu	A	0.00 Bu	A T V	F V	A
25	S	0.87 Ex	S	1.00 Ex	S	S	S
26	G	1.00 Ex	G	1.00 Ex	G	G	G
27	F	0.10 Bu	F	0.10 Bu	G Y D	F L G	F
28	I	0.85 Ex	D	0.72 mE	T	S	T N
29	F	0.00 Bu	F	0.00 Bu	F	L I	F
30	S	0.74 mE	S	0.83 Ex	S N V I	S	S
36	W	0.00 Bu	W	0.00 Bu	W	W	W
37	V	0.00 Bu	V	0.00 Bu	V	I	V
38	R	0.10 Bu	R	0.31 mB	R	R	R
39	Q	0.15 Bu	Q	0.28 mB	Q	Q	Q
40	A	0.95 Ex	A	0.75 mE	A	P	A
41	P	0.90 Ex	P	0.73 mE	P	P	P S
42	G	1.00 Ex	G	1.00 Ex	G	G	G
43	K	0.86 Ex	K	0.86 Ex	Q R K H	K R	K
44	G	1.00 Ex	G	1.00 Ex	G	A G	G S
45	L	0.00 Bu	L	0.00 Bu	L	L	L

Fig. 1b

Position	KOL Residue	KOL Exposure	J539 Residue	J539 Exposure	Residues In Subgroup I	II	III
46	E	0.75 mE	E	0.73 mE	E	E	E
47	W	0.10 Bu	W	0.04 Bu	W	W	W
48	V	0.00 Bu	I	0.00 Bu	MV	LI	V
49	A	0.00 Bu	G	0.00 Bu	G	AG	GSA
66	R	0.36 mB	K	0.51 pB	R	R	R
67	F	0.00 Bu	F	0.00 Bu	V	LV	F
68	T	0.87 Ex	I	0.88 Ex	T	T	T
69	I	0.00 Bu	I	0.00 Bu	VMI	IV	I
70	S	0.78 mE	S	0.79 mE	TS	ST	S
71	R	0.11 Bu	R	0.00 Bu	RLA	KV	R
72	N	0.61 mE	D	0.55 pB	DK	D	DN
73	D	0.44 pB	N	0.43 pB	PETAS	T	DN
74	S	0.65 Ex	A	0.97 Ex	S	S	S
75	K	0.88 Ex	K	0.77 mE	TF	KR	K
76	N	0.69 mE	N	0.68 mE	NST	N	N
77	T	0.41 pB	S	0.33 mB	TQ	Q	T
78	L	0.00 Bu	L	0.00 Bu	AV	VF	LA
79	F	0.45 pB	Y	0.35 mB	Y	VS	YF
80	L	0.00 Bu	L	0.00 Bu	M	L	L
81	Q	0.53 pB	Q	0.69 mE	E	TKSIN	Q
82	M	0.00 Bu	M	0.00 Bu	L	ML	M
82a	D	0.73 mE	S	0.58 pB	SVRT	TSNIR	ND
82b	S	0.98 Ex	K	0.96 Ex	S	NS	S
82c	L	0.00 Bu	V	0.00 Bu	L	VM	L
83	R	0.73 mE	R	0.83 Ex	RFI	DT	RE
84	P	0.75 mE	S	0.90 Ex	S	PA	PA
85	E	0.82 Ex	E	0.90 Ex	E	VA	ED
86	D	0.00 Bu	D	0.11 Bu	D	D	D
87	T	0.54 pB	T	0.47 pB	T	T	T
88	G	1.00 Ex	A	0.00 Bu	A	A	A
89	V	0.58 PB	L	0.63 mE	V	TV	VL
90	Y	0.00 Bu	Y	0.00 Bu	Y	Y	Y
91	F	0.00 Bu	Y	0.08 Bu	Y	Y	Y
92	C	0.00 Bu	C	0.00 Bu	C	C	C
93	A	0.00 Bu	A	0.00 Bu	A	A	AT
94	R	0.17 Bu	R	0.15 Bu	R	RH	RP

					JH1	JH2	JH3	JH4	JH5	JH6
103	W	0.09 Bu	W	0.07 Bu	W	W	W	W	W	W
104	G	0.00 Bu	G	1.00 Ex	G	G	G	G	G	G

Fig. 1c

Position	Residue	Fractional Accessibility KOL Exposure	Residue	J539 Exposure	Residues In Subgroup I	II		III		
					JH1	JH2	JH3	JH4	JH5	JH6
105	Q	0.93 Ex	Q	0.99 Ex	Q	R	Q	Q	Q	Q
106	G	0.00 Bu	G	0.00 Bu	G	G	G	G	G	G
107	T	0.22 mB	T	0.26 mB	T	T	T	T	T	T
108	P	0.99 Ex	L	0.67 mE	L	L	M	L	L	T
109	V	0.00 Bu	V	0.00 Bu	V	V	V	V	V	V
110	T	0.76 mE	T	0.69 mE	T	T	T	T	T	T
111	V	0.00 Bu	V	0.00 Bu	V	V	V	V	V	V
112	S	0.98 Ex	S	0.74 mE	S	S	S	S	S	S
113	S	0.94 Ex	A	0.84 Ex	S	S	S	S	S	S

Fig. 2a

Position	Residue	Exposure	Residues In Subgroup I	II	III	IV	V	VI
1	Q	1.00 Ex	Q	Q	SF	.	Q	ND
2	S	1.00 Ex	S	S	Y	S	S	F
3	V	0.77 mE	V	A	E	E	A	M
4	L	0.00 Bu	L	L	L	L	L	L
5	T	0.92 Ex	T	T	TK	T	T	T
6	Q	0.00 Bu	Q	Q	Q	Q	Q	Q
7	P	0.62 mE	P	P	P	D	P	P
8	P	1.00 Ex	P	ARP	P	P	P	H
9	S	1.00 Ex	S	S	S	A	S	S
10	-	-	-	-	-	-	-	-
11	A	0.34 mB	AV	V	V	V	A	V
12	S	0.71 mE	S	S	S	S	S	S
13	G	1.00 Ex	GA	G	VL	V	G	E
14	T	0.73 mE	TA	S	SA	A	S	S
15	P	0.75 mE	P	P	PA	L	PL	P
16	G	1.00 Ex	G	G	G	G	G	G
17	Q	0.69 mE	Q	Q	Q	Q	Q	K
18	R	0.79 mE	R	S	T	T	S	T
19	V	0.21 mB	V	IV	A	V	V	V
20	T	0.62 ME	T	T	RM	R	T	T
21	I	0.00 Bu	I	I	I	I	I	IFM
22	S	0.92 Ex	S	S	T	T	S	S
23	C	0.00 Bu	C	C	C	C	C	C
35	W	0.00 Bu	W	W	W	W	W	W
36	Y	0.00 Bu	Y	YF	Y	Y	Y	Y
37	Q	0.46 pB	Q	Q	Q	Q	Q	Q
38	Q	0.00 Bu	QH	Q	QE	Q	Q	Q
39	L	0.75 mE	LV	H	KR	K	H	R
40	P	0.91 Ex	P	P	PS	P	PA	P
41	G	1.00 EX	G	G	G	G	G	G
42	M	0.74 mE	T	K	QR	Q	RK	SRG
43	A	0.62 mE	A	A	A	A	A	A
44	P	0.00 Bu	P	P	P	P	P	P
45	K	0.95 Ex	K	K	V	L	K	T
46	L	0.23 mB	L	L	MLP	L	LV	T
47	L	0.15 Bu	L	MIL	V	V	VI	V
48	I	0.00 Bu	I	I	IV	I	I	I
49	Y	0.39 mB	Y	YF	Y	Y	FY	Y
57	G	1.00 Ex	G	G	GE	G	G	G
58	V	0.14 Bu	VI	VI	IV	I	V	V
59	P	0.70 mE	P	SP	P	P	P	P
60	D	0.95 Ex	D	DNL	EQA	D	D	D
61	R	0.31 mB	R	R	R	R	R	R

Fig. 2b

Position	Residue	Exposure	Residues In Subgroup I	II	III	IV	V	VI
62	F	0.12 Bu	F	F	F	F	F	F
63	S	0.85 Ex	S	S	S	S	S	S
64	G	0.00 Bu	G A	G	G S	G	G	G
65	S	1.00 Ex	S	S	S Y	S	S	S
66	K	0.41 pB	K	K	T S N	S	K	I F*
67	S	1.00 Ex	S	S	S	S	S	S
68	G	1.00 Ex	G	G	G	G	D G	S
69	A	0.71 mE	T	N	T N	H	N	N
70	S	1.00 Ex	S	T	T K S	T	T	S
71	A	0.00 Bu	A	A	A V	A	A	A
72	S	1.00 Ex	S T	S	T I	S	S	S
73	L	0.00 Bu	L	L	L	L	L	L
74	A	0.74 mE	A	T	T	T	T	T
75	I	0.00 Bu	I	I	I	I	V	I
76	G	1.00 Ex	S T	S	S N	T	S	S
77	G	1.00 Ex	G	G	G R	G	G	G
78	L	0.00 Bu	L	L	V A	A	L	L
79	Q	0.76 mE	Q R	Q	Q E	Q	R Q	K Q T
80	S	1.00 Ex	S T	A	A V	A	A	T
81	E	0.78 mE	E G	E	E G	E	E	E
82	D	0.09 Bu	D	D	D	D	D	D
83	E	0.64 mE	E	E	E	E	E	E
84	T	0.34 mB	A	A	A	A	A	A
85	D	0.30 mB	D	D	D	D	D	D
86	Y	0.00 Bu	Y	Y	Y	Y	Y	Y
87	Y	0.16 Bu	Y	Y	Y F	Y	Y	Y
88	C	0.00 Bu	C	C	C	C	C	C
			JL-1	JL-2	JL-3	JL-4	JL-5	
98	F	0.04 Bu	F	F	F	F	F	
99	G	0.00 Bu	G	G	G	G	G	
100	T	0.59 pB	T	G	G	S	S	
101	G	1.00 Ex	G	G	G	G	G	
102	T	0.00 Bu	T	T	T	T	T	
103	K	0.82 Ex	K	K	K	Q	Q	
104	V	0.00 Bu	V	L	L	L	L	
105	T	0.86 Ex	T	T	T	T	T	
106	V	0.19 Bu	V	V	V	V	V	
106a	L	0.70 mE	L	L	L	L	L	
107	G	1.00 Ex	G	G	G	S	G	

* additional residues after position 66:
66a D
66b S R D

Fig. 3a

Position	Residue	Exposure	Residues In Subgroup I	II	III	IV
1	E	0.99 Ex	D	D	E	D
2	I	0.16 Bu	I	I	I	I
3	V	0.87 Ex	Q	V	V	V
4	L	0.00 Bu	M	M	L	M
5	T	0.80 mE	T	T	T	T
6	Q	0.00 Bu	Q	Q	Q	Q
7	S	0.89 Ex	S	S	S	S
8	P	0.67 mE	P	P	P	P
9	A	1.00 Ex	S	L	G	D N
10	I	0.94 Ex	S	S	T	S
11	T	0.30 mB	L	L	L	L
12	A	0.59 pB	S	P	S	A
13	A	0.00 Bu	A	V	L	V
14	S	0.78 mE	S	T	S	S
15	L	0.79 mE	V	P	P	L
16	G	1.00 Ex	G	G	G	G
17	Q	0.64 mE	D	E	E	E
18	K	0.74 mE	R	P	R	R
19	V	0.22 mB	V	A	A	A
20	T	0.65 mE	T	S	T	T
21	I	0.00 Bu	I	I	L	I
22	T	0.69 mE	T	S	S	N
23	C	0.00 Bu	C	C	C	C
35	W	0.00 Bu	W	W	W	W
36	Y	0.00 Bu	Y	Y	Y	Y
37	Q	0.14 Bu	Q	L	Q	Q
38	Q	0.24 mB	Q	Q	Q	Q
39	K	0.69 mE	K	K	K	K
40	S	1.00 Ex	P	P	P	P
41	G	1.00 Ex	G	G	G	G
42	T	0.90 Ex	K	Q	Q	Q
43	S	0.30 mB	A	S	A	P
44	P	0.00 Bu	P	P	P	P
45	K	0.90 Ex	K	Q E R	R	K
46	P	0.43 pB	L	L	L	L
47	W	0.16 Bu	L	L	L	L
48	I	0.00 Bu	I	I	I	I
49	Y	0.42 pB	Y	Y	Y	Y
57	G	1.00 Ex	G	G	G	G
58	V	0.13 Bu	V	V	I	V
59	P	0.61 mE	P	P	P	P
60	A	1.00 Ex	S	D	D	D
61	R	0.36 mB	R	R	R	R
62	F	0.00 Bu	F	F	F	F
63	S	0.94 Ex	S	S	S	S

Fig. 3b

Position	Residue	Exposure	Residues In Subgroup I	II	III	IV
64	G	0.00 Bu	G	G	G	G
65	S	1.00 Ex	S	S	S	S
66	G	1.00 Ex	G	G	G	G
67	S	1.00 Ex	S	S	S	S
68	G	1.00 Ex	G	G	G	G
69	T	0.75 mE	T	T	T	T
70	S	0.98 Ex	D E Q	D	D	D
71	Y	0.09 Bu	F	F	F	F
72	S	0.70 mE	T	T	T	T
73	L	0.00 Bu	L	L	L	L
74	T	0.43 pB	T	K	T	T
75	I	0.00 Bu	I	I	I	I
76	N	0.83 Ex	S	S	S	S
77	T	0.83 Ex	S	R	R	S
78	M	0.00 Bu	L	V	L	L
79	E	0.63 mE	Q	E Q	E	Q
80	A	0.96 Ex	P	A	P	A
81	E	0.91 Ex	E D	E	E	E
82	D	0.13 Bu	D	D	D	D
83	A	0.55 pB	F I	V	F	V
84	A	0.00 Bu	A	G	A	A
85	I	0.58 pB	T	V	V	V
86	Y	0.00 Bu	Y	Y	Y	Y
87	Y	0.11 Bu	Y	Y	Y	Y
88	C	0.00 Bu	C	C	C	C

			JK-1	JK-2	JK-3	JL-4	JL-5
98	F	0.00 Bu	F	F	F	F	F
99	G	1.00 Ex	G	G	G	G	G
100	A	1.00 Ex	Q	Q	P	G	Q
101	G	0.00 Bu	G	G	G	G	G
102	T	0.00 Bu	T	T	T	T	T
103	K	0.79 mE	K	K	K	K	R
104	L	0.00 Bu	V	L	V	V	L
105	E	0.89 Ex	E	E	D	E	E
106	L	0.44 pB	I	I	I	I	I
106a	-	-	-	-	-	-	-
107	K	0.77 mE	K	K	K	K	K

Fig. 4

Mouse Light Chain Variable Region

5' upstream primer - FR1 of variable region

5'- TCT CGG ATC CGA (CT)AT (TC)GT G(AC)T (GC)AC CCA (GA) -3'

Bam H1

3' downstream primer - kappa constant region

5'- TCT CAA GCT TG GTG GCA AGA T(GA)G ATA CAG TTG GTG CAG C -3'

Hind III

Mouse Heavy Chain Variable Region

5' upstream primer - FR1 of variable region

i) 5'- TTC TGG ATC C(CG)A GGT (GCT)CA (AG)CT G(AC)A G(GC)A GTC (TA)GG -3'

Bam H1

ii) 5'- TTC TGG ATC C(CG)A GGT (GCT)AA GCT GGT G(GC)A GTC (TA)GG -3'

Bam H1

3' downstream primer - IgG2a CH1 region

5'- TCT CAA GCT TAC CGA TGG (GA)GC TGT TGT TTT GGC -3'

Hind III

SHORTEN VERSION OF THE IgG4 HEAVY CHAIN CONSTANT REGION

5'- ATT TGG ATC C TC TAG A CA TCG CGG ATA GAC AAG AAC -3'

Bam H1 Xba I

5'- AAT AAT GCG GCC GC A TCG AT G AGC TCA AGT ATG TAG ACG GGG TAC G -3'

Not I Cla I Sac I

TK PROMOTER FRAGMENT

5'- TAT AGA ATT C GG TAC CCT TCA TCC CCG TGG CCC G -3'

Eco R1 Kpn I

5'- TGC GTG TTC GAA TTC GCC -3'

Eco R1

Ig H ENHANCER

5'- TTT TAG ATC T GT CGA CAG ATG GCC GAT CAG AAC CAG -3'

Bgl II Sal I

5'- TTG GTC GAC GGT ACC AAT ACA TTT TAG AAG TCG AT -3'

Sal I Kpn I

HUMAN KAPPA CONSTANT REGION

5'- TCT CGG ATC CTC TAG AAG AAT GGC TGC AAA GAG C -3'

5'- TCT CGC TAG CGG ATC CTT GCA GAG GAT GAT AGG G -3'

Fig. 5

Oligodeoxynucleotides for PCR Amplification of the LEN Light Chain Variable Region

S1 5'-CAT TCG CTT ACC AGA TCT AAG CTT ACT AGT GAG ATC ACA GTT CTC TCT AC-3'
V9 5'-TGG CTC TGC AGC TGA TGG TG-3'

V10 5'-CAC CAT CAG CTG CAG AGC CA-3'
V11 5'-CTG TCT GGG ATC CCA GAT TC-3'

V12 5'-GAA TCT GGG ATC CCA GAC AG-3'
V13 5'-GTT GCA ACA TCT TCA GCC TCC ACG CTG CTG ATG-3'

V14 5'-GTG GAG GCT GAA GAT GTT GCA ACT TAT TAC TG-3'
I3 5'-GAA TGT GCC TAC TTT CTA GAG GAT CCA ACT GAG GAA GCA AAG-3'

A1 5'-CAT TCG CTT ACC AGA TCT-3'
A2 5'-GAA TGT GCC TAC TTT CTA G-3'

Oligodeoxynucleotides for PCR Amplification of the m1B4 Heavy Chain Variable Region

V1 5'-CCC TCC AGG CTT CAC TAA GTC TCC CCC-3'

V2 5'-TTA GTG AAG CCT GGA GGG TCC CTG AAA CTC-3'
V3 5'-GCC CCT TCC CAG GAG CTT GGC GAA CCC AAG ACA TG-3'

V4 5'-AAG CTC CTG GGA AGG GGC TGG AGT TGG TCG CAG CC-3'
V5 5'-TGT TCA TTT GTA GGT ACA GGG TGT TCT TGG AAT TGT CTC TGG AGA TGG TG-3'

V6 5'-TGT ACC TAC AAA TGA ACA GTC TGA GGG CTG AGG ACA CAG CCT TGT ATT-3'
V7 5'-CTG TGA GAA GGG TGC CTT GGC CCC AGT AG-3'

V8 5'-AAG GCA CCC TTC TCA CAG TCT CCT CAG GTG-3'
I2 5'-GAA TGT GCC TAC TTA AGC TTT CTA GAG GAT CCT ATA AAT CTC TGG CCA TG-3'

S1, A1, and A2, as above

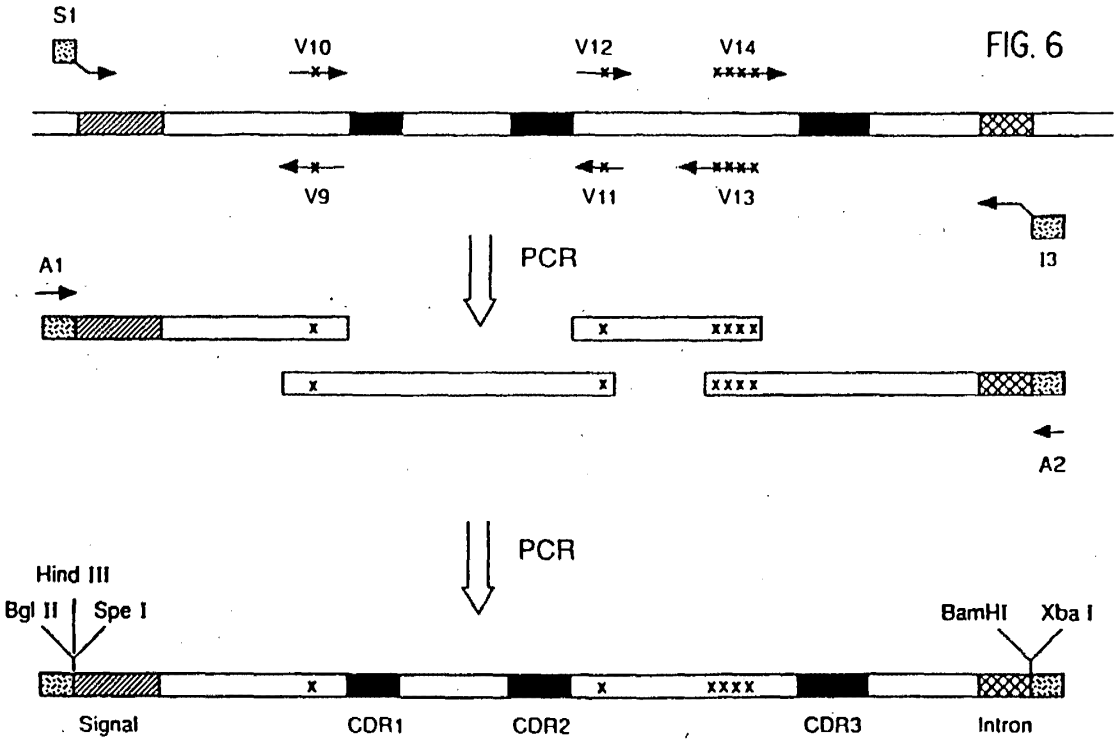


FIG. 6

EP 0 519 596 A1

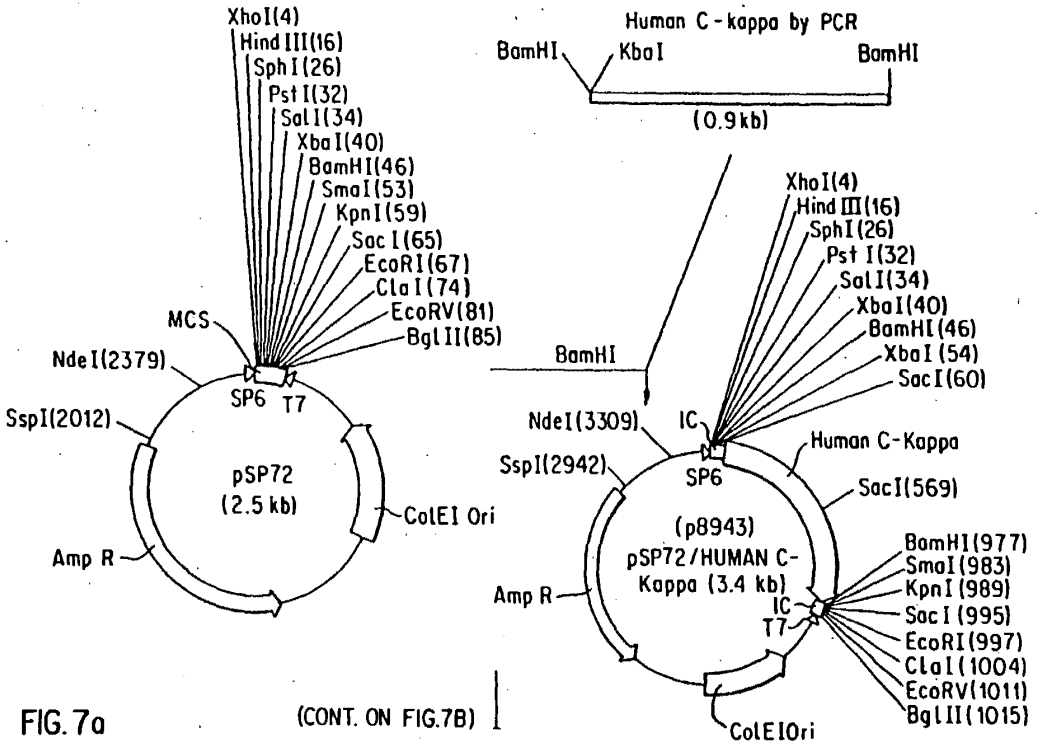


FIG. 7a

(CONT. ON FIG.7B)

EP 0 519 596 A1

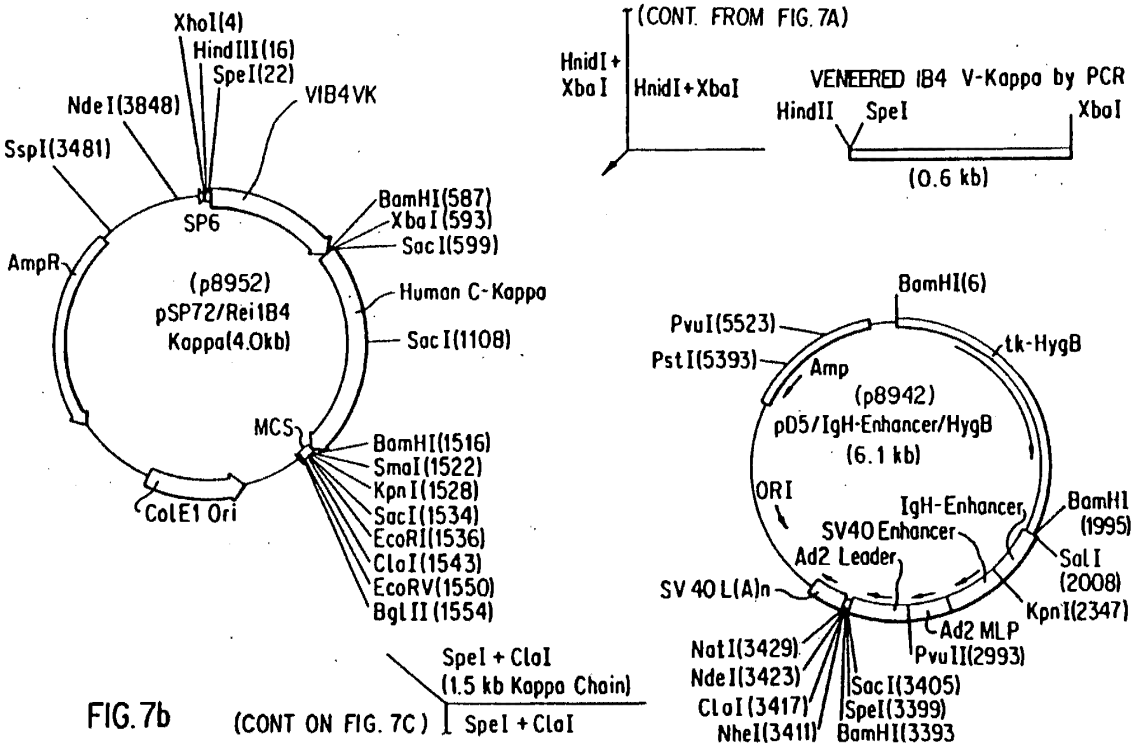


FIG. 7b

EP 0 519 596 A1

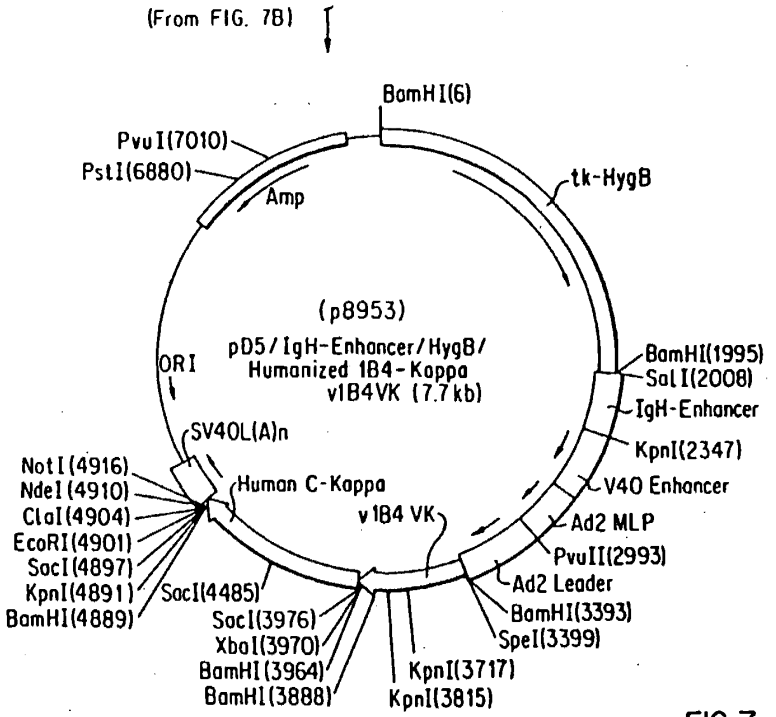
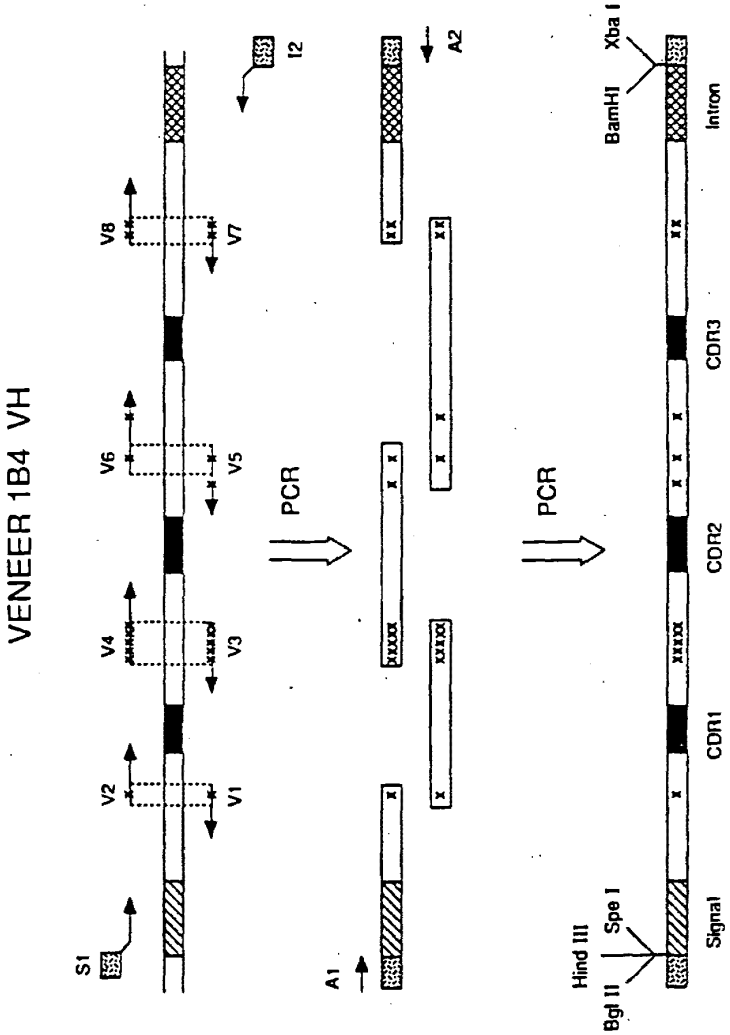


FIG. 7c

Fig. 8



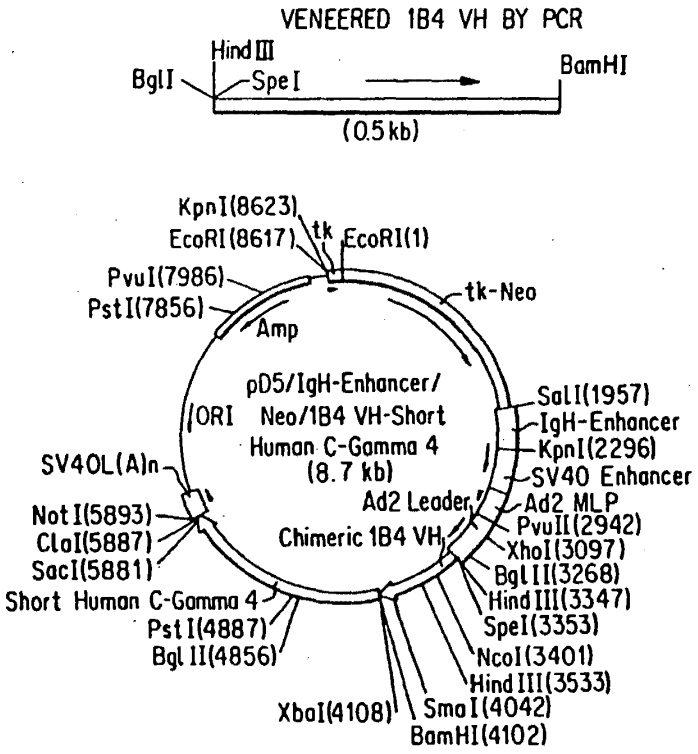


FIG. 9a

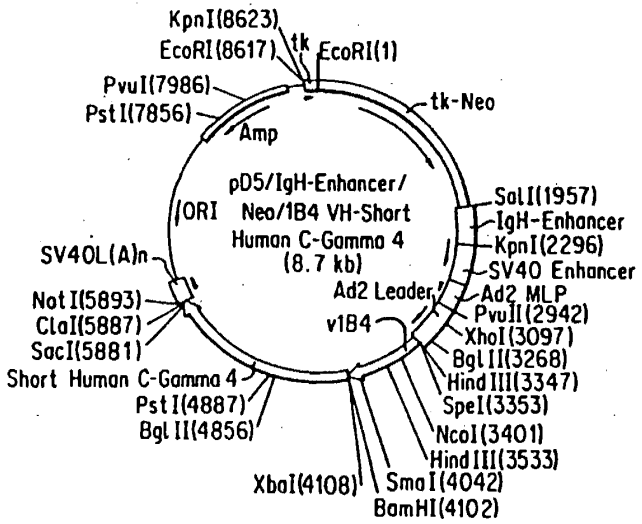


FIG. 9b

EP 0 519 596 A1

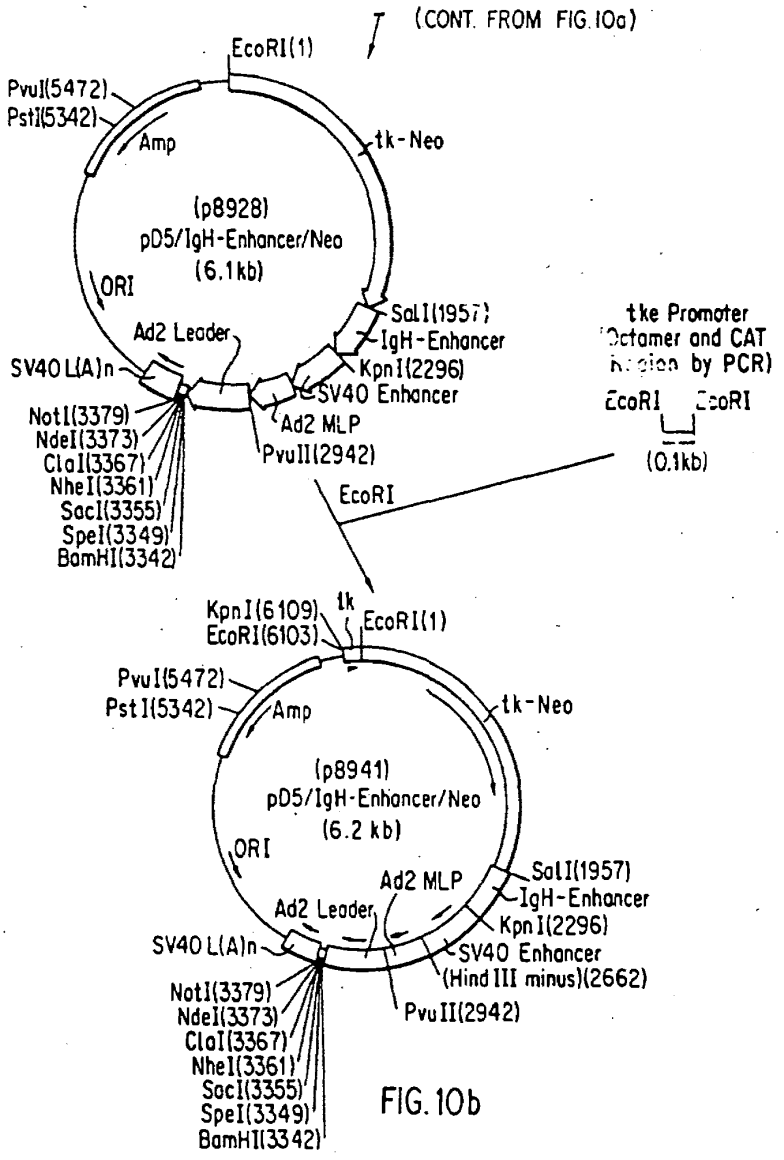


FIG. 10b

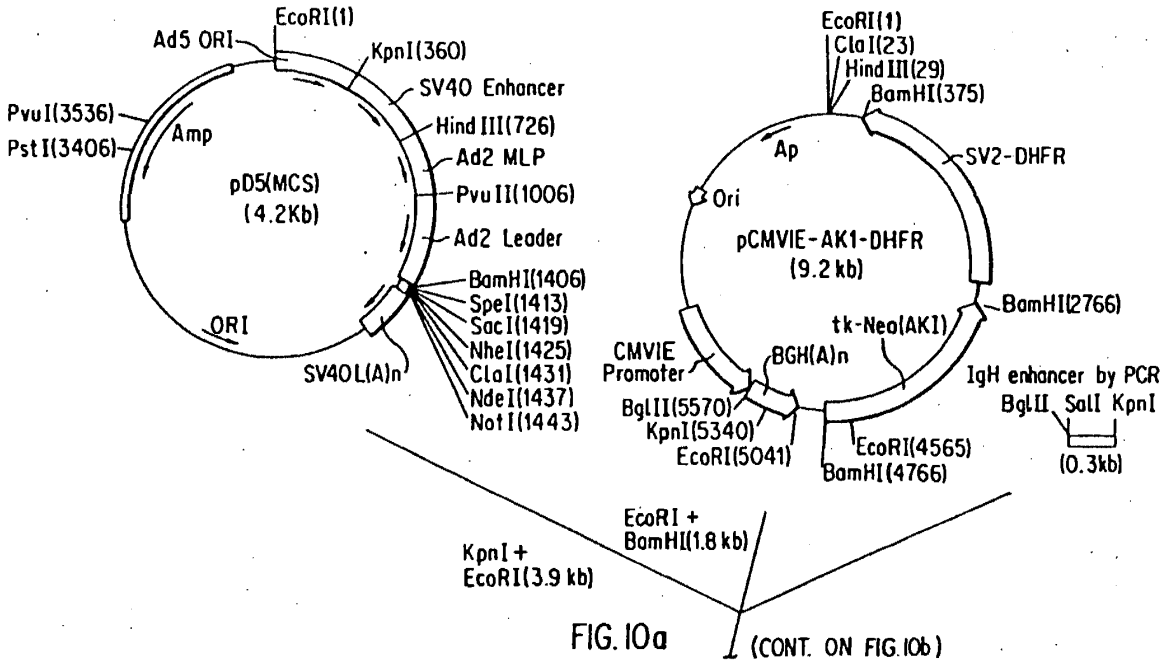


FIG. 10a (CONT. ON FIG. 10b)

EP 0 519 596 A1

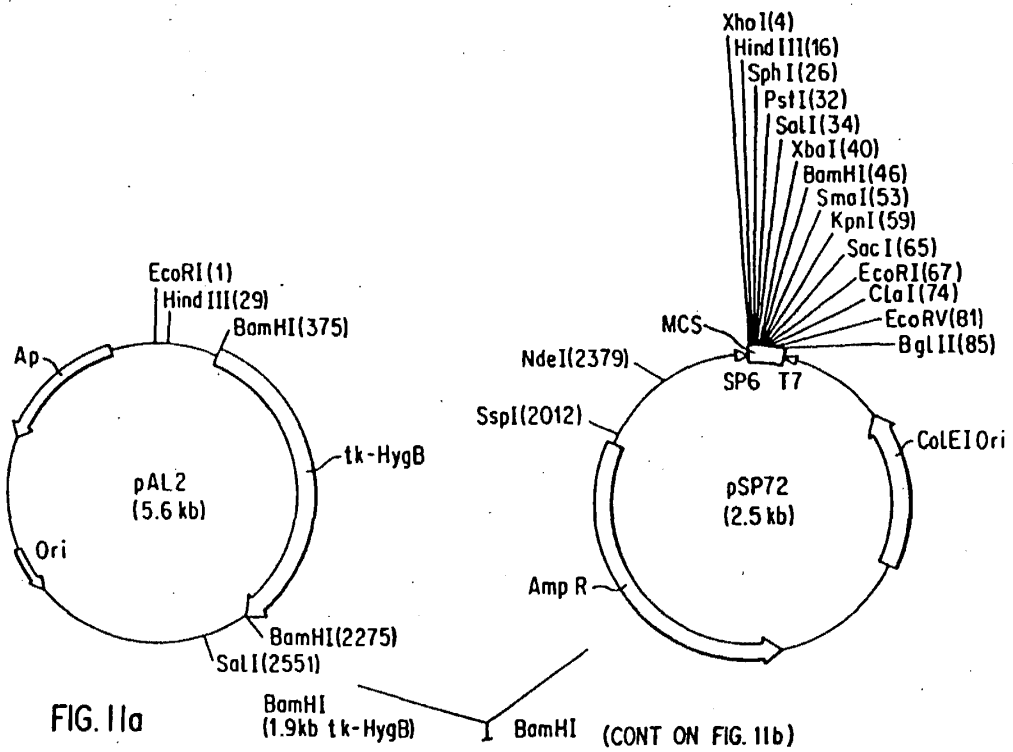


FIG. 11a

EP 0 519 596 A1

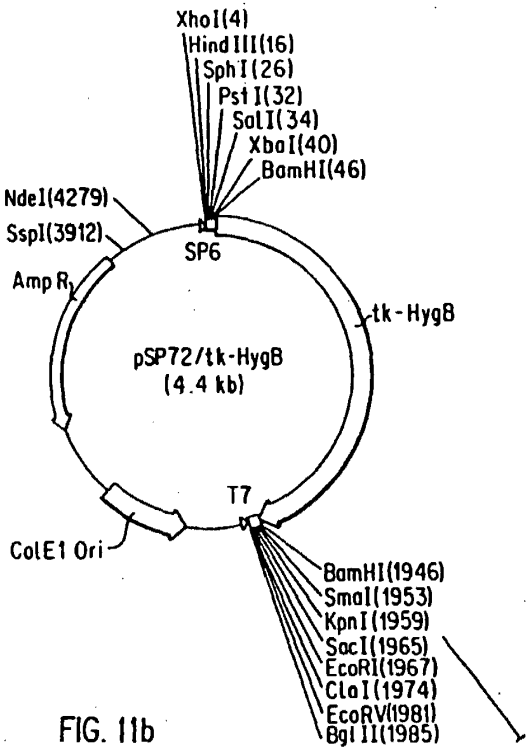
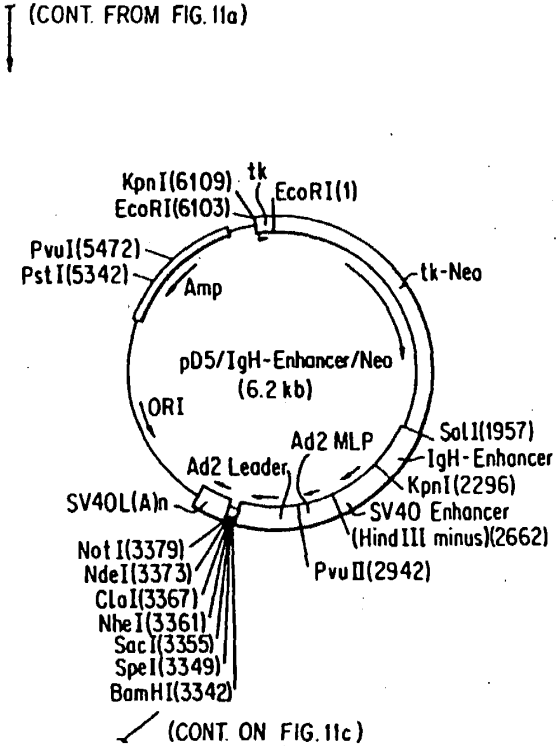


FIG. 11b



EP 0 519 596 A1

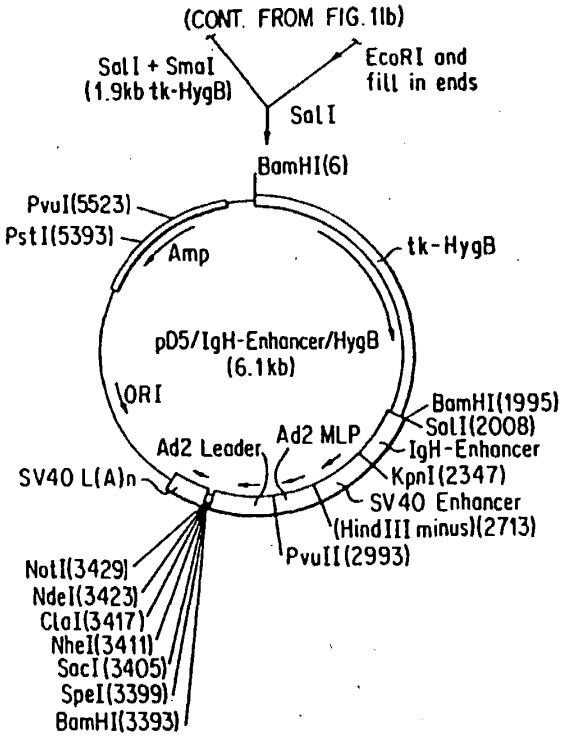
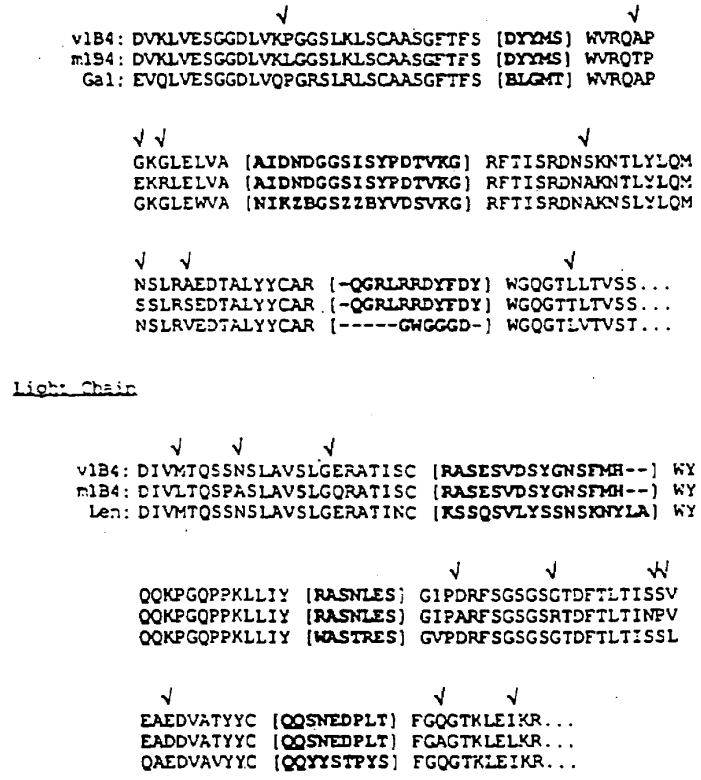


FIG. IIc

EP 0 519 596 A1

Fig. 12



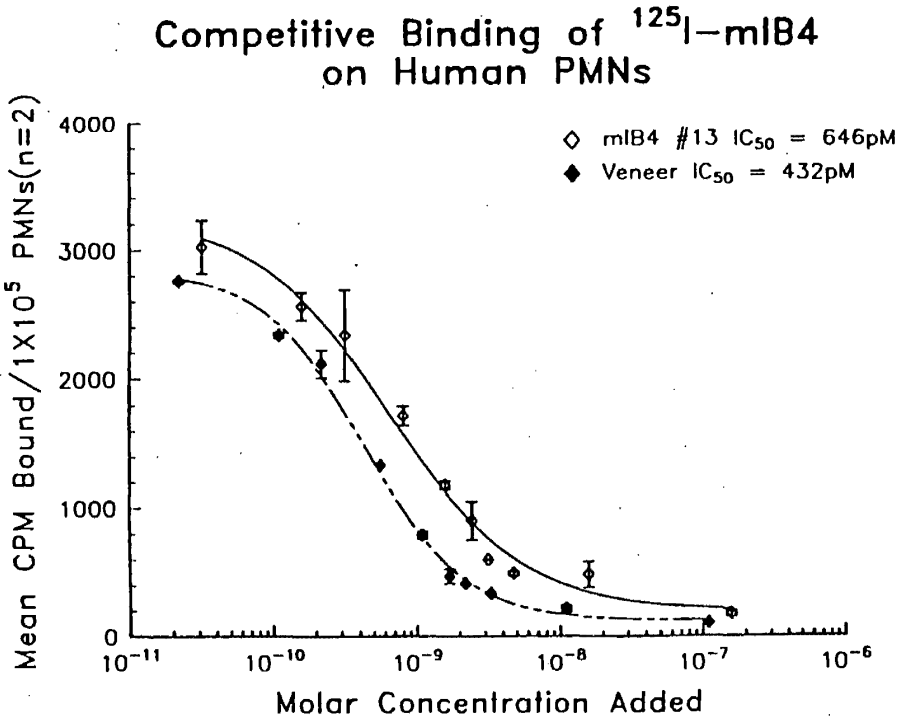


Fig. 13

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 92 30 4225

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA. vol. 88, no. 10, 15 May 1991, WASHINGTON DC, US pages 4181 - 4185 S. GORMAN ET AL. 'Reshaping a therapeutic CD4 antibody.' * the whole document * ---	1-12	C12N15/13 C12P21/08
D,Y	NUCLEIC ACIDS RESEARCH vol. 19, no. 9, 11 May 1991, LONDON, GB pages 2471 - 2476 B. DAUGHERTY ET AL. 'Polymerase chain reaction facilitates the cloning, CDR-grafting, and rapid expression of a murine monoclonal antibody directed against the CD18 component of leukocyte integrins.' * the whole document * ---	1-12	
D,Y	NATURE. vol. 332, 24 March 1988, LONDON, GB pages 323 - 327 L. RIECHMANN ET AL. 'Reshaping human antibodies for therapy.' * the whole document * ---	1-12	TECHNICAL FIELDS SEARCHED (Int. Cl.5) C07K
P,X	EP-A-0 438 312 (MERCK & CO., INC.) * claims * -----	1-12	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	30 SEPTEMBER 1992	NOOIJ F.J.M.

EPO FORM 1503 03.82 (P0401)

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number : **0 592 106 A1**

(12) EUROPEAN PATENT APPLICATION

(21) Application number : 93307051.8

(22) Date of filing : 07.09.93

(51) Int. Cl.[5] : C12N 15/13, C12N 15/62, C07K 15/00, C12P 21/08

(30) Priority : 09.09.92 US 942245

(43) Date of publication of application : 13.04.94 Bulletin 94/15

(84) Designated Contracting States : BE CH DE DK ES FR GB IE IT LI LU NL SE

(71) Applicant : IMMUNOGEN INC
148 Sidney Street
Cambridge, Massachusetts 02139 (US)

(72) Inventor : Pedersen, Jan T.
72 Great Pultney Street
Bath, Avon BA2 4DC (GB)

Inventor : Searle, Stephen M.J.
30 Eastbourne Avenue
Bath, Avon BA1 6NE (GB)
Inventor : Rees, Anthony R.
94 sydney Place
Bath, Avon BA2 6NE (GB)
Inventor : Roguska, Michael A.
16 Hilldale Road
Ashland, Massachusetts 01721 (US)
Inventor : Gulld, Braydon C.
109 Riverdale Road
Concord, Massachusetts 01742 (US)

(74) Representative : BROOKES & MARTIN
High Holborn House 52/54 High Holborn
London, WC1V 6SE (GB)

(54) Resurfacing of rodent antibodies.

(57) A method for determining how to humanize a rodent antibody or fragment thereof by resurfacing, said method comprising :

(a) determining the conformational structure of the variable region of said rodent antibody or fragment thereof by constructing a three-dimensional model of said rodent antibody variable region ;

(b) generating sequence alignments from relative accessibility distributions from x-ray crystallographic structures of a sufficient number of rodent antibody variable region heavy and light chains to give a set of heavy and light chain framework positions wherein said set is identical in 98% of said sufficient number of rodent antibody heavy and light chains ;

(c) defining for said rodent antibody or fragment thereof to be humanized a set of heavy and light chain surface exposed amino acid residues using said set of framework positions generated in said step (b) ;

(d) identifying from human antibody amino acid sequences a set of heavy and light chain surface exposed amino acid residues that is most closely identical to said set of surface exposed amino acid residues defined in said step (c), wherein said heavy and light chain from said human antibody are or are not naturally paired ;

(e) substituting, in the amino acid sequence of said rodent antibody or fragment thereof to be humanized said set of heavy and light chain surface exposed amino acid residues defined in said step (c) with said set of heavy and light chain surface exposed amino acid residues identified in said step (d) ;

(f) constructing a three-dimensional model of said variable region of said rodent antibody or fragment thereof resulting from the substituting specified in said step (e) ;

(g) identifying, by comparing said three-dimensional models constructed in said steps (a) and (f), any amino acid residues from said set identified in said step (d), that are within 5 Angstroms of any atom of any residue of the complementarity determining regions of said rodent antibody or fragment thereof to be humanized ; and

(h) changing any residues identified in said step (g) from the human to the original rodent amino acid residue to thereby define a rodent antibody humanizing set of surface exposed amino acid residues ; with the proviso that said step (a) need not be conducted first, but must be conducted prior to said step (g).

EP 0 592 106 A1

FIELD OF THE INVENTION

The present invention relates to the development of prediction rules that can be used to accurately model the variable regions (V-regions) of antibodies. The development of these rules and their application in the predictive molecular restructuring of the surfaces of variable domains of non-human monoclonal antibodies enables changing of the surface, resurfacing, of these monoclonal antibody V-regions to replicate the surface characteristics found on human antibody V-regions. This method of resurfacing non-human monoclonal antibody V-regions to resemble human antibody V-regions is expected to permit the production of functional altered antibodies, which retain the binding parameters of the original non-human monoclonal antibody, with improved therapeutic efficacy in patients due to the presentation of a human surface on the V-region.

BACKGROUND OF THE INVENTION

General Background of Antibodies

Murine monoclonal antibodies are widely used as diagnostic and therapeutic agents in the treatment of human disease. Mice can be readily immunized with foreign antigens to produce a broad spectrum of high affinity antibodies. Invariably, the introduction of murine or other rodent antibodies into humans results in the production of a human anti-mouse antibody (HAMA) response due to the presentation of a foreign protein in the body. The production of HAMA in patients can result from the introduction of foreign antibody in a single dose or from extended use in therapy, for example, for the treatment of cancer. Extended use of murine antibody is generally limited to a term of days or weeks in patients before concerns of anaphylaxis arise. Moreover, once HAMA has developed in a patient, future use of murine antibodies for diagnostic or therapeutic purposes is often precluded for the same reasons.

Beyond ethical considerations, attempts to produce human monoclonal antibodies have not been highly successful for a number of reasons. The production *in vitro* of human monoclonals rarely results in high affinity antibodies. *In vitro* cultures of human lymphocytes yield a restricted range of antibody responses relative to the broad spectrum of reactive antibodies produced *in vivo* through direct immunization of mice. Additionally, in humans, immune tolerance prevents the successful generation of antibodies to self-antigens. All of these factors have contributed to the search for ways to modify the structures of murine monoclonal antibodies to improve their use in patients. Many investigators have attempted to alter, reshape or humanize murine monoclonal antibodies in an effort to improve the therapeutic application of these molecules in patients.

Strategies of Antibody Humanization

The earliest reports of the controlled rearrangement of antibody domains to create novel proteins was demonstrated using rabbit and human antibodies as described by Bobrzecka, K. et al. (Bobrzecka, K., Konieczny, L., Laidler, P. and Rybarska, J. (1980), Immunology Letters **2**, pp. 151-155) and by Konieczny et al. (Konieczny, L., Bobrzecka, K., Laidler, P. and Rybarska, J. (1981), Haematologia **14** (I), pp. 95-99). In those reports, the protein subunits of antibodies, rabbit Fab fragments and human Fc fragments, were joined through protein disulfide bonds to form new, artificial protein molecules or chimeric antibodies.

Recombinant DNA technology was used to construct gene fusions between DNA sequences encoding mouse antibody variable light and heavy chain domains and human antibody light chain (LC) and heavy chain (HC) constant domains to permit expression of the first recombinant "near-human" antibody (chimeric antibody) product (Morrison, S.L., Johnson, M.J., Herzenberg, L.A. and Oi, V.T. (1984), Proc. Natl. Acad. Sci. U.S.A. **81**, pp. 6851-6855).

The kinetics and immune response in man to chimeric antibodies has been examined (LoBuglio, A.F., Wheeler, R.H., Trang, J., Haynes, A., Rogers, K., Harvey, E.B., Sun, L., Ghrayeb, J. and Khazaeli, M.B. (1989), Proc. Natl. Acad. Sci. **86**, pp. 4220-4224).

Chimeric antibodies contain a large number of non-human amino acid sequences and are immunogenic in man. The result is the production of human anti-chimera antibodies (HACA) in patients. HACA is directed against the murine V-region and can also be directed against the novel V-region/C-region (constant region) junctions present in recombinant chimeric antibodies.

To overcome some of the limitations presented by the immunogenicity of chimeric antibodies, the DNA sequences encoding the antigen binding portions or complementarity determining regions (CDR's) of murine monoclonal antibodies have been grafted by molecular means in the DNA sequences encoding the frameworks of human antibody heavy and light chains (Jones, P.T., Dear, P.H., Foote, J., Neuberger, M.S. and Winter, G. (1986), Nature **321**, pp. 522-525; Riechmann, L., Clark, M., Waldmann, H. and Winter, G. (1988), Nature **332**,

pp. 323-327). The expressed recombinant products called reshaped or humanized antibodies are comprised of the framework of a human antibody light or heavy chain and the antigen recognition portions, CDR's, of a murine monoclonal antibody. Several patent applications have been filed in this area including, for example, European Patent Application, Publication No. 0239400; European Patent Application, Publication Nos. 0438310A1 and 0438310A2; International Patent Publication No. WO 91/09967; and International Patent Publication No. WO 90/07861.

However, it is questionable whether European Patent Application (EP), Publication No. 0239400 is truly enabling. It is not assured in this patent that the best fit is made to assure proper presentation of the CDR loops at the antibody combining site.

EP Publication Nos. 0438310A1 and 0438310A2 go a step beyond EP Publication No. 0239400 by protecting the importance of uniquely selected human frameworks for the human light chain (LC) and heavy chain (HC) V-regions. These V-region frameworks should show a high degree of sequence similarity with the frameworks of the murine monoclonal antibody and present the CDR's in the appropriate configuration. However, the criteria for sequence matching are no more sophisticated than simple homology searching of the antibody protein or DNA databases.

International Patent Publication No. WO 91/09967 attempts a further variation of the method disclosed in EP Publication No. 0239400. In International Patent Publication No. WO 91/09967, homology of the donor sequences and the acceptor framework is not important, rather it discloses that a selected set of residues in the LC and HC are critically important to humanization. The ability to make changes at these positions is the basis of International Patent Publication No. WO 91/09967.

International Patent Publication No. WO 90/07861 proposes four important criteria for designing humanized antibodies. 1) Homology between human acceptor and non-human donor sequences. 2) Use donor rather than acceptor amino acids where the acceptor amino acid is unusual at that position. 3) Use donor framework amino acids at positions adjacent to the CDR. 4) Use donor amino acids at framework positions where the sidechain atom is within 3 Angstroms of the CDR in a 3-D model. The first antibody humanized by this method retained less than 1/3 the affinity of the original monoclonal antibody.

None of the above methods for designing a humanized antibody are predictable due to the questions that surround CDR framework interactions. By replacement of murine framework with human framework, there is no guarantee of identical conformations for CDR's because i) the V_L-V_H interaction is not identical in all V-regions and ii) accurate prediction of the CDR-framework interactions are key to faithful reproduction of the antigen binding contacts.

The above methods do not offer a general solution to solving the issues surrounding antibody humanization, rather the methods as outlined in each reference above involve a substantial amount of trial and error searching to obtain the desired affinity in the final humanized product. More importantly, there is no guarantee that corrective changes in framework amino acids will leave the reshaped V-regions resembling the surface character of a truly human antibody. Therefore, it can be argued that antibodies humanized by the above methods may be immunogenic in man.

Antigenicity of Antibodies

The antigenicity/immunogenicity of an antibody, including recombinant reshaped antibody products, introduced into humans can be viewed as a surface phenomenon. In general one can view the immune system as scanning the surface of a protein introduced to the body. If the F_V portion of a humanized antibody 'opens-up' in the circulation then internal residues can be presented to the immune system. On the other hand, if the F_V portion is stable and tightly packed then only the surface residues presented by the V-regions and the interface between the V_L and V_H regions will be 'scanned'.

Surface Reshaping or Resurfacing of Antibobies

The notion of surface presentation of proteins to the immune system raises the prospect of redesigning murine monoclonal antibodies to resemble human antibodies by humanizing only those amino acids that are accessible at the surface of the V-regions of the recombinant F_V. The resurfacing of murine monoclonal antibodies to reduce their immunogenicity could be beneficial in maintaining the avidity of the original monoclonal antibody in the reshaped version, because the natural framework-CDR interactions are retained. The value of maintaining the integrity of the framework-CDR interactions has been illustrated as summarized below.

In a recent research report, two different reshaped versions of the rat monoclonal antibody, Campath-9 (anti-human CD4), were generated (Gorman, S.D., Clark, M.R., Routledge, E.G., Cobbold, S.P. and Waldmann, H. (1991), Proc. Natl. Acad. Sci. U.S.A. 88, pp. 4181-4185). In one version, pV_HNEW/C_{G1}, the acceptor V_H fra-

mework was from the human NEW-based heavy chain, which has 47% identical residues to the Campath-9 V_H. While in the second version, pV_HKOL/C_{G1}, the acceptor V_H framework was from the human KOL antibody, which has 72% identical residues to Campath-9 V_H. Each reshaped antibody contained the identical V_L domain from the human REI antibody sequence. However, the recombinant product of pV_HKOL/C_{G1} had an avidity for CD4 that was substantially greater than the product of pV_HNEW/C_{G1}. The authors proposed a reshaping strategy where human sequences, that are highly homologous to the rodent antibody of interest, are transferred, by in vitro mutagenesis, into the rodent V-region to create a "bestfit" reshaped antibody. This strategy uses the term "bestfit" to describe the modeling process, however, there is no quantitative formula employed to assess "bestfit", and so in effect, the process is subjective. Additionally, there is no resurfacing concept presented in that paper.

The concept of reducing rodent-derived antibody immunogenicity through the replacement of exposed residues in the antibody framework regions which differ from those of human origin is discussed in a recent paper (Padlan, E.A. (1991), Molecular Immunology **28**, pp. 489-498). In that paper, the variable domains of two antibody structures, KOL (human) and J539 (mouse), are examined. The crystal structures of the Fab fragments of these two antibodies have been elucidated to high resolution. The solvent accessibility of the exposed framework residues in the variable domains of these two antibodies were compared to a sequence database of human and murine antibody V-region subgroups. On the basis of his findings, Padlan proposed to reduce the antigenicity of allogeneic variable domains [murine V-regions], through replacement of the exposed residues in the framework regions with residues usually found in human antibodies. In murine sequences with the highest similarity to a given human sequence, the number of changes necessary to "humanize" a murine V-region surface would range from 6-15 amino acid changes per V-region. This reference suggests how to convert one antibody surface into another but no general method is developed. Application of the procedure is provided by two examples, a worst-case and a best-case.

Worst Case:

Among the representative murine kappa V_L sequences examined for which its autologous V_H has been sequenced, $S107V_L$ has the most residues that need to be replaced to humanize it. $S107V_L$ is most similar to the members of the human subgroup VKIV and JK2. The exposed or partially exposed residues that need to be replaced are those at positions 9, 10, 14, 15, 16, 17, 18, 22, 41, 63, 80, 83, 85, 100 and 106. Murine V-region $S107V_H$ is most similar in its framework to the members of the human subgroup VHIII and JH6. The exposed or partially exposed residues in $S107V_H$ that need to be replaced are those at positions 3, 40, 68, 73, 75, 76, 82b and 89. A total of 23 residues need to be replaced to humanize the variable domains of S107.

Best Case:

Among the murine V_H sequences examined for which the autologous V_L has also been sequenced, $MOPC21V_H$ has the least number of residues that need to be replaced to humanize it. $MOPC21V_H$ is most similar in its framework to the members of the human subgroup HIII and JH6. The exposed or partially exposed residues that need to be replaced are those at positions 1, 42, 74, 82a, 84, 89 and 108. $MOPC21V_L$ is most similar in its framework to human subgroup VKIV and JK4. The exposed or partially exposed residues that need to be replaced are those at positions 1, 9, 12, 15, 22, 41, 63, 68, 83 and 85. A total of 17 amino acids need to be replaced to humanize the variable domains of MOPC21.

Of the light chains in the Best- and Worst-Case examples cited above, $S107V_L$ required changes at 15 positions and $MOPC21V_L$ required changes at 10 positions. Only seven of the changes are common to both of these light chain sequences (see underlined residues). Moreover, of the heavy chain residues that need to be replaced to humanize the respective V-regions, $S107V_H$ required changes at 8 positions and $MOPC21V_H$ required changes at 7 positions. In this instance, only one position is common to both of these heavy chain sequences (see residues in boldface).

An analysis of S107 V-regions alone would not have led to the prediction of which residues to change in MOPC21. The reason for this is that the surface residues in Padlan's analysis are only determined by reference to the crystal structure analysis of one antibody. In addition, the basis for defining the surface exposure of an amino acid at a particular position on that crystal structure is a continuous gradient of change, e.g., the fractional solvent accessibility values (Padlan, E.A. (1990), Molecular Immunology **28**, pp. 489-498) were computed, where: 0 to 0.2 = completely buried, 0.2 to 0.4 = mostly buried, 0.4 to 0.6 = partly buried/partly exposed, 0.6 to 0.8 = mostly exposed, and 0.8 or above = completely exposed. By limiting the analysis of exposed surface residues to a single crystal structure and by superimposing a broad range of solvent accessibility ratios on exposed residues, such a modeling strategy could be expected to have a wide margin of error in its calculations.

This model fails to take into account the great majority of structural information available in the database for other antibody crystal structures.

SUMMARY OF THE INVENTION

Accordingly, it is an object of this invention to provide humanized rodent antibodies or fragments thereof, and in particular, humanized rodent monoclonal antibodies that have improved therapeutic efficacy in patients due to the presentation of a human surface on the V-region. This and other objects have been attained by providing a method for determining how to humanize a rodent antibody or fragment thereof by resurfacing the method comprising:

(a) determining the conformational structure of the variable region of the rodent antibody or fragment thereof by constructing a three-dimensional model of the rodent antibody variable region;

(b) generating sequence alignments from relative accessibility distributions from x-ray crystallographic structures of a sufficient number of rodent antibody variable region heavy and light chains to give a set of heavy and light chain framework positions wherein the set is identical in 98% of the sufficient number of rodent antibody heavy and light chains;

(c) defining for the rodent antibody or fragment thereof to be humanized a set of heavy and light chain surface exposed amino acid residues using the set of framework positions generated in step (b);

(d) identifying from human antibody amino acid sequences a set of heavy and light chain surface exposed amino acid residues that is most closely identical to the set of surface exposed amino acid residues defined in step (c), wherein the heavy and light chain from the human antibody are or are not naturally paired;

(e) substituting, in the amino acid sequence of the rodent antibody or fragment thereof to be humanized the set of heavy and light chain surface exposed amino acid residues defined in step (c) with the set of heavy and light chain surface exposed amino acid residues identified in step (d);

(f) constructing a three-dimensional model of the variable region of the rodent antibody or fragment thereof resulting from the substituting specified in step (e);

(g) identifying, by comparing the three-dimensional models constructed in steps (a) and (f), any amino acid residues from the set identified in step (d), that are within 5 Angstroms of any atom of any residue of the complementarity determining regions of the rodent antibody or fragment thereof to be humanized; and

(h) changing any residues identified in step (g) from the human to the original rodent amino acid residue to thereby define a rodent antibody humanizing set of surface exposed amino acid residues; with the proviso that step (a) need not be conducted first, but must be conducted prior to step (g).

Also provided is a method for producing a humanized rodent antibody or fragment thereof from a rodent antibody or fragment thereof, the method comprising:

(I) carrying out the above-described method for determining how to humanize a rodent antibody or fragment thereof by resurfacing; and

(II) modifying the rodent antibody or fragment thereof by replacing the set of rodent antibody surface exposed amino acid residues with the rodent antibody humanizing set of surface exposed amino acid residues defined in step (h) of the above-described method.

In a preferred embodiment, the rodent antibody or fragment thereof is a murine antibody, and most preferably murine antibody N901.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows an algorithm that can be used for constructing a three-dimensional model of the rodent antibody variable region.

Figure 2 is a diagram showing the approach to determine how to humanize a rodent antibody or fragment thereof according to the present invention.

Figures 3A and 3B are plots of relative accessibility of amino acid residues for twelve antibody F_V structures, mapped onto the sequence alignment of these structures. Structures Glb2 (Jeffrey, P.D., Doctor of Philosophy Thesis, University of Oxford, United Kingdom, 1991), D1.3 (Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J. (1986), Science **233**, pp. 747-753), 3D6 (Grunow, R., Jahn, S., Porstman, T., Kiessig, T., Steinkeller, H., Steindl, F., Mattanovich, D., Gurtler, L., Deinhardt, F., Katinger, H. and von R., B. (1988), J. Immunol. Meth. **106**, pp. 257-265) and 36-71 (5fab) (Rose, D.R., Strong, R.K., Margolis, M.N., Gefter, M.L. and Petsko, G.A. (1990), Proc. Natl. Acad. Sci. U.S.A. **87**, pp. 338-342) are not yet present in the Brookhaven database. The other structures used were: 2hfl (Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C. and Davies, D.R. (1987), Proc. Natl. Acad. Sci. U.S.A. **84**, pp. 8075-8079), 3hfm (Padlan, E., Silverton, E., Sheriff, S., Cohen, G., Smith-Gill, S. and Davies, D. (1989), Proc. Natl. Acad. Sci. U.S.A. **86**, pp.

5938-5942), 2fbj (Mainhart, C.R., Potter, M. and Feldmann, R.J. (1984), Mol. Immunol. 21, pp. 469-478), 3fab (Saul, F.A., Amzel, L.M. and Poljak, R.J. (1978), J. Biol. Chem. 253, pp. 585-597), 4fab (Herron, J., He, X., Mason, M., Voss, E. and Edmunson, A. (1989), Proteins: Struct., Funct., Genet. 5, pp. 271-280), 2mcp (Segal, D., Padlan, E., Cohen, G., Rudikoff, S., Potter, M. and Davies, D. (1974), Proc. Natl. Acad. Sci. U.S.A. 71, pp. 4298-[??]), 2fb4 (Marquart, M. Deisenhofer, J. and Huber, R. (1980), J. Mol. Biol. 141, pp. 369-391), and 1f19 (Lascombe, M. Alzari, P., Boulot, G., Salujian, P., Tougard, P., Berek, C., Haba, S., Rosen, E., Nisonof, A. and Poljak, R. (1989), Proc. Natl. Acad. Sci. U.S.A. 86, p. 607). These structures are designated by their Brookhaven entry code. The sequence numbering used here is described in Figures 4A and 4B. Figure 3A graphically shows the relative accessibility for the heavy chain and Figure 3B graphically shows the relative accessibility for the light chain.

Figures 4A and 4B show alignments of sequences generated using the three methods of humanization. Sequences are: 1) Original rodent N901. 2+3) KOL (Marquart, M. Deisenhofer, J. and Huber, R. (1980), J. Mol. Biol. 141, pp. 369-391) and reshaped N901 using KOL surface. 4+5) Most homologous sequences, L(KV2F) (Klobeck, H., Meindl, A., Combriato, G., Solomon, A. and Zachau, H. (1985), Nucleic Acids Res. pp. 6499-6513) and H(G36005) (Schroeder, H. and Wang, J. (1990), Proc. Natl. Acad. Sci. U.S.A. 87), and reshaped N901 using these sequences. 6+7) Most homologous with respect to surface residues, L(KV4B) (Klobeck, H., Bronkamp, G., Combriato, G., Mocikat, R., Pohelnz, H. and Zachau, H. (1985), Nucleic Acids Res. 3, pp. 6515-6529) and H(PLO123) (Bird, J., Galili, N., Link, M., Sites, D. and Sklar, J. (1988), J. Exp. Med. 168, pp. 229-245), and reshaped N901 using these sequences. The numbering is the same as used in the antibody modelling program ABM (trademark for commercial software, Oxford Molecular Ltd., Oxford, U.K.), which is based on structural conservation and not sequence homology as used by Padlan et al. (Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M. and Gottesman, K.S. (1987), Sequences of Proteins of Immunological Interest. U.S. Department of Health and Human Services, Fourth Edition). The sequence changes which have to be introduced in order to resurface N901 with a given sequence are marked with bars, back-mutations as determined from F_V models are marked with stars. The sequence homology of given sequences to N901 are shown in brackets after each sequence.

Figure 5 is a stereo plot of mean antibody β-barrel, coordinates determined by iterative multiple fitting of eight antibody structures. Strands 7 and 8 comprise the 'take off' positions for CDR H3 and are not included in the fitting of V_L and V_H regions.

Figure 6 is a plot of RMS deviation from the mean of the eight β-sheet strands comprising the framework. The RMS was calculated from structures F19.9, 4-4-20, NEW, FBJ, KOL, HyHEL-5, HyHEL-10 and McPC603. N,Cα,C atoms are included in the plot. The residues used are shown in the alignment (Table 2). The most disordered residues are all the residues of strand HFR4, the last residue of LFR1, and the first and last residue of HFR2. The nomenclature of the strands is explained in the alignment in Table 2. LFR1 - #1, LFR2 - #2, LFR3 - #3, LFR4 - #4, HFR1 - #5, HFR2 - #6, HFR3 - #7, HFRS4 - #8.

Figure 7 is a flowchart of the overall modelling protocol known as CAMAL.

Figure 8 is a plot of superimposed loop backbones for models and x-ray structures discussed in Example 2. The loops are positioned after global framework fit. This does not represent the best local least squares fit, but shows how the loops are positioned globally onto the framework.

Figures 9A to 9D are stereo (N,C-α,C,O) representations of crystal structures and models of D1.3, 3671 and Gloop-2 variable domain and β-barrel strands described in Example 2. Crystal structures are shown with open bonds, model with solid bonds. The difference between the 3D6-H3 in the model and the crystal structure is due to a 5-7° twist in the extended β-sheet conformation of this loop, Figure 9A: D1.3, Figure 9B: 36-71, Figure 9C: Gloop-2, Figure 9D: 3D6.

Figure 10 is a histogram showing the distribution of loop length for CDR H3 loops, data from Kabat et al. (Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M. and Gottesman, K.S. (1987), Sequences of Proteins of Immunological Interest. U.S. Department of Health and Human Services, Fourth Edition).

DETAILED DESCRIPTION OF THE INVENTION

The existence of specific, yet different, surface patches in murine and human antibodies may be the origin of the inherited immunogenicity of murine antibodies in humans. Statistical analysis of a database of unique human and murine antibody F_V fragments has revealed that certain combinations of residues in exposed surface positions are specific for human and murine sequences. The combinations are not the same in human and murine F_V domains. However, it is possible to define families of surface residues for the two species of antibodies. These families reveal a novel method for the "humanization" or reshaping of murine antibodies. Humanization is the modification of the solvent accessible surface of a non-human antibody or fragment thereof to resemble the surface of a chosen human antibody or fragment thereof such that the modified non-human antibody or fragment thereof exhibits lower immunogenicity when administered to humans. Such a process

applies in the present application to antibody variable regions but could equally well apply to any other antibody fragment. The method is considered to be generally applicable to humanization of rodent antibodies.

According to the present invention, a statistical analysis is presented which is based on accessibility calculated for a range of antibody crystal structures. When this information is applied to an antibody sequence database, it is possible to discriminate between human and murine antibodies at the sequence level purely on the basis of their surface residue profiles.

Rational Resurfacing Approach

There are several key features of the resurfacing approach of the present invention.

1) This method uses as a starting point construction of a three-dimensional model of a rodent variable region by known methods;

2) A large number (e.g., twelve) of antibody F_V or Fab fragment x-ray crystallographic structures are analyzed to produce an unambiguous set of surface exposed amino acid residues that will be positionally identical for a majority (98%) of antibodies. The set is produced by identifying all those residues whose solvent accessibility is above a given cut-off (typically 30%), calculated using a modification of the method of Kabsch and Sander (Kabsch, W. and Sander, C. (1983), Biopolymers **22**, pp. 2257-2637) in which explicit atomic radii are used for each atom type to predict sidechain positions as is described below in more detail;

3) Using a complete human antibody database, the best set of human heavy and light chain surface exposed amino acid residues is selected on the basis of their closest identity to the set of surface amino acid residues of the murine antibody;

4) In order to retain the conformational structure- of the CDRs of the rodent antibody, replacement of any human surface exposed amino acid with the original rodent surface exposed amino acid residue is carried out whenever a surface residue is calculated from the three-dimensional model to be within 5 Angstroms of a CDR residues.

The general resurfacing approach of the present invention is illustrated in Figure 2. The approach can be divided into two stages. In the first, the rodent framework (white) is retained and only the surface residues changed from rodent (dark grey circles) to the closest human pattern (light grey circles). This should remove the antigenicity of the rodent antibody. In the second stage, surface residues within 5 Angstroms of the CDRs are replaced with the rodent equivalents in an attempt to retain antigen binding and CDR conformation.

The method of the present invention is applicable to whole antibodies as well as antibody fragments. Suitable antibody fragments that can be used can readily be determined by the skilled artisan. Examples of some suitable fragments include a single chain antibody (SCA), an antibody F_V fragment, Fab fragment, Fab_2 fragment, Fab' fragment, or other portion of an antibody comprising the binding site thereof.

According to the present invention, an important step in the method for determining how to modify a rodent antibody or fragment thereof by resurfacing is to determine the conformational structure of the variable region of the rodent antibody or fragment thereof to be humanized by constructing a three-dimensional model of the rodent antibody variable region. This can be done by known methods such as those described, for example, in Martin et al. (Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1989), Proc. Natl. Acad. Sci. U.S.A. **86**, pp. 9268-9272; Methods in Enzymology (1991), **203**, pp. 121-152) and as described in detail in Example 2.

Martin et al. describe an algorithm which is depicted in Figure 1. The algorithm applies to murine and human antibodies equally well. The present inventors therefore expect that, based on sequence similarity between antibodies of different species (Kabat, E.A. Segments of Proteins of Immunological Interest, National Institutes of Health, U.S.A. 1991), the algorithm will work equally well for rat and other rodent antibodies.

Briefly, the algorithm depicted in Figure 1 can be summarized as follows. The framework region of an antibody to be modelled is selected on the basis of sequence homology and constructed by a least squares fit onto the six conserved strands of the variable region β-barrel. Light and heavy chain complementarity determining regions are constructed using a combination of canonical structures (Chothia, C. and Lesk, A.M. (1987), J. Molec. Bio. **196**, pp. 901-917), database searching and conformational searching. Detailed descriptions of these methods are described in Example 2 herein and in the above two references (Martin et al. 1989 and 1991).

According to the present invention, another three-dimensional model is also constructed. The other three-dimensional model is of the rodent antibody variable region having human antibody surface amino acid residues substituted therein at particular rodent antibody surface residue positions.

This other three-dimensional model is constructed by carrying out the series of steps described next.

The first of the steps is to generate sequence alignments from relative accessibility distributions from x-ray crystallographic structures of a sufficient number of rodent antibody variable region heavy and light chains to give a set of framework positions of surface exposed amino acid residues which is identical in a majority

(98%) of the variable regions.

As used herein, the term "framework" means the antibody variable region from which the complementarity determining regions have been excluded.

"Complementarity determining regions" means those amino acid sequences corresponding to the following numbering system as defined by Kabat, E.A. (In Sequences of Immunological Interest, N.I.H., U.S.A., 1991).

Light Chain	L1	residues	24-34
Light Chain	L2	residues	50-56
Light Chain	L3	residues	89-97
Heavy Chain	H1	residues	31-358
Heavy Chain	H2	residues	50-58
Heavy Chain	H3	residues	95-102

A sufficient number of rodent antibody fragments that need to be analyzed in order to produce the set of framework positions of surface exposed amino acid residues can readily be determined by the skilled artisan through routine experimentation using a database of antibody sequences. Thus, this step can be conducted using suitable databases now in existence or later compiled.

The x-ray crystallographic structures are used to determine relative accessibility distributions of surface exposed amino acid residues. The relative accessibility distributions identify all those residues whose solvent accessibility is above a given cut-off (typically 30%), calculated using a modification of the method of Kabsch and Sander (Kabsch, W. and Sander C. (1983), Biopolymers **22**, pp. 2257-2637) in which explicit atomic radii are used for each atom type.

The relative accessibility distributions determined from the x-ray crystallographic structures can then be used to generate sequence alignments which give a set of framework positions of surface exposed amino acid residues which is identical in a majority (98%) of the variable regions.

The set of framework positions of surface exposed amino acid residues for the variable regions of murine antibodies is shown in Table 1, set forth in Example 1, and was produced using the sequence alignments and accessibility distributions shown in Figures 3A and 3B.

Once a set of framework positions of surface exposed amino acid residues for the variable regions of the rodent antibodies have been generated, the surface exposed residues of the heavy and light chain pair of the rodent antibody, or fragment thereof, to be humanized can be identified using an alignment procedure such as that described in Example 1 and shown in Figures 3A and 3B. This defines a set of surface exposed amino acid residues of a heavy and light chain pair of a rodent antibody or antibody fragment to be humanized.

Next, a complete human antibody sequence database is used to identify a set of surface exposed amino acid residues from a human antibody variable region that have the closest positional identity to the set of surface exposed amino acid residues of the variable region of the rodent antibody that is to be humanized. The set of surface exposed amino acid residues from the human antibodies can be separately identified for a heavy chain and for a light chain that are not naturally paired and/or a set can be identified from a natural human heavy and light chain pair, that is, a pair originating from a single B cell or hybridoma clone. Preferably, the set is one from a natural human heavy and light chain pair.

A humanized rodent antibody that gives the appearance of a human antibody is then predicted by substituting the set of surface exposed amino acid residues from the rodent antibody or fragment thereof to be humanized with the set of surface exposed amino acid residues from the human antibody.

A three-dimensional model can then be constructed from the resulting, fully substituted variable region of the rodent antibody or fragment thereof. The three-dimensional model is constructed using the same known methods mentioned above for constructing a 3-D model of the original rodent antibody or fragment thereof.

While the antigenicity of this fully "resurfaced" or humanized antibody should be removed, an additional factor to be addressed is the binding affinity or the binding strength of the resurfaced antibody. Changes in the framework of the variable domain introduced through resurfacing can influence the conformation of the CDR loops and therefore antigen binding of the antibody. According to the present invention, this problem is removed by the next step which is to identify, by means of a comparison of both of the above-described three-dimensional models of the rodent antibody variable region, any residues from the set of surface exposed amino acid residues of the variable region heavy and light chain pair of the human antibody identified that are within 5 Angstroms of any atom of any residue of the rodent antibody or antibody fragment complementarity deter-

mining regions (CDRs).

Any residue(s) so identified is then changed back from the human to the original rodent amino acid residue(s).

The results of this method can then be applied to a particular rodent antibody by well known methods. Briefly, genes for the humanized variable heavy and light chain regions are constructed using standard recombinant DNA methods (Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989), Molecular Cloning, Second Edition). For example, a PCR method can be used (Daugherty et al. (1991), Nucleic Acids Research 19, pp. 2471-2476).

Variable heavy chain or variable light chain gene constructs are subcloned into appropriate expression vectors. Suitable expression vectors contain either a human gamma or human kappa constant region gene, a suitable promoter, a sequence coding for a human immunoglobulin leader peptide (for example: met-gly-trp-ser-cys-ile-ile-leu-phe-leu-val-ala-thr-ala-thr (SEQ ID NO:39), Olandi et al. (1989), PNAS 86, pp. 3833-3837), and a drug selectable marker.

Heavy and light chain expression plasmids can be co-transfected, for example, by electroporation into suitable cells, for example, SP2/0 cells, and selected with an appropriate drug, G418, for example. Screening for intact antibody can be accomplished by ELISA assay. 96-well plates are coated with, for example, goat anti-human kappa chain antibody, and light chains are detected with, for example, goat anti-human antibody conjugated to alkaline phosphatase.

As another approach, light chain constructs are transfected, for example, by electroporation into suitable cells, for example, SP2/0 cells and selected, for example, in hygromycin. Screening for light chain expression can be accomplished by ELISA assay. 96-well plates are coated with, for example, goat anti-human kappa chain antibody, and light chains are detected with, for example, goat anti-human antibody conjugated to alkaline phosphatase.

A light chain producing line is then used as a host to electroporate in the heavy chain construct. The heavy chain plasmid is co-transfected with a plasmid containing the gene coding for another drug marker, for example, neomycin resistance and selected in the presence of the drug G418. Screening for intact antibody is accomplished by ELISA assay. 96-well plates are coated with, for example, goat anti-human Fc and detected with, for example, goat anti-human light chain conjugated to alkaline phosphatase.

EXAMPLE 1 AND COMPARATIVE EXAMPLES

The superiority of the presently claimed method for determining how to modify a rodent antibody or fragment thereof by resurfacing in order to produce a humanized rodent antibody will now be described by reference to the following example and comparative examples which are illustrative and are not meant to limit the present invention.

A) Analysis for Murine Antibodies

In order to determine the positions which are usually accessible on the surface of the F_V domain of murine antibodies, the accessibility was calculated for twelve Fab x-ray crystallographic structures obtained from the Brookhaven database (Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977), J. Mol. Biol. 112, pp. 535-542). The relative accessibility was calculated using the program MC (Pedersen, J. (1991)), which implements a modified version of the DSSP (Kabsch, W. and Sander, C. (1983), Biopolymers 22, pp. 2257-2637) accessibility calculation routine in which explicit atomic radii are specified for every atom. A residue was defined as being surface accessible when the relative accessibility was greater than 30%. The alignment positions of these residues were conserved in all twelve structures (98% identity). Surface accessible framework positions constitute 40% of the F_V surface area. The remaining surface accessible residues are in the CDRs and in the interdomain C-terminal region. Figures 3A and 3B show a sequence alignment of the twelve crystal structures, the average relative accessibility, and the 30% accessibility cutoff. Figure 3A shows the alignments relative accessibility for the twelve murine antibody light chains and Figure 3B shows the alignments and relative accessibility for the murine antibody heavy chains.

The surface accessible framework positions were mapped onto a database of unique human and mouse F_V sequences (see lists at the end of this Example). The frequency of particular residues in each of these positions is shown in Table 1. Only residue frequencies higher than 5% are listed.

Light chain		
Position	Human	Mouse
1	D 51 E 34 A 5 S 5	D 76 Q 9 E 6
3	V 38 Q 24 S 24 Y 6	V 63 Q 22 L 5
5	T 61 L 37	T 87
9	P 26 S 26 G 17 A 14 L 7	S 36 A 29 L 17 P 5
15	P 62 V 25 L 12	L 47 P 30 V 8 A 7
18	R 57 S 18 T 13 P 6	R 38 K 22 S 13 Q 12 T 9
46	P 94	P 82 S 9
47	G 89	G 71 D 18
51	K 43 R 31	K 70 Q 13 R 8 T 5
63	G 91	G 98
66	D 43 S 25 A 9	D 38 A 26 S 26
73	S 96	S 90 I 5
76	D 43 T 18 S 16 E 15	D 67 S 15 A 5 K 5
86	P 44 A 27 S 17 T 8	A 50 P 11 T 8 E 7 Q 6
87	E 71 D 11 G 7	E 91 D 6
111	K 74 R 12 N 6	K 93
115	K 54 L 40	K 87 L 5
116	R 60 G 33 S 5	R 89 G 9
117	Q 50 T 37 E 6 P 6	A 74 Q 14 P 5 R 5

Heavy chain		
Position	Human	Mouse
118	E 47 Q 46	E 59 Q 29 D 10
120	Q 83 T 7	Q 68 K 26
122	V 59 L 15 Q 13	Q 57 V 27 L 5 K 5
126	G 54 A 23 P 18	G 36 P 30 A 29
127	G 53 E 22 A 14 D 7	E 45 G 43 S 6
128	L 61 V 31 F 7	L 96
130	K 46 Q 41 E 5	K 52 Q 27 R 17
131	P 95	P 91 A 5
132	G 74 S 16 T 7	G 82 S 17
136	R 53 K 23 S 17 T 7	K 66 S 17 R 13
143	G 96	G 98
145	T 46 S 32 N 9 I 7	T 63 S 19 N 7 A 5 D 5
160	P 84 S 10	P 89 H 7
161	G 93	G 71 E 24
162	K 76 Q 10 R 8	K 50 Q 30 N 10 H 5
183	D 26 P 25 A 17 Q 10 T 7	E 31 P 22 D 17 A 12 Q 11
184	S 70 K 9 P 8	K 42 S 37 T 6
186	K 53 Q 22 R 7 N 5	K 83 Q 7
187	G 66 S 21 T 5	G 62 S 18 D 10
195	T 30 D 26 N 19 K 7	T 36 K 30 N 26 D 6
196	S 91	S 76 A 16
197	K 65 I 8 T 8 R 5	S 46 K 34 Q 11
208	R 46 T 18 K 17 D 6	T 55 R 26 K 8
209	A 50 P 21 S 13 T 8	S 67 A 14 T 11
210	E 46 A 18 D 13 S 9 Z 8 V 5	E 88 D 7
212	T 91	T 53 S 43
222	G 17 D 11 P 10 Y 9 V N 8	D 67 A 18

Table 1: Distribution of accessible residues in human V_H and V_L chain sequences. All of the positions appear to be conserved, which leads to the hypothesis that immunogenecity arises from a specific combination of these surface residues. The sequence numbering is explained in Figures 3A and 3B.

None of the entire combinations of surface residues in the human sequences are found in the murine sequences and *vice versa* (see lists at the end of this Example). However, the residues in individual positions appear to be conserved (see Table 1). There are few residues which differ significantly between the species;

these are at positions 54 and 91 of the L chain and 168 and 216 of the H chain. Of these positions only position 216 is a non conservative (V to Y) mutation. Differences between human and murine antigenicities are therefore believed to arise from the combinations of residues in these positions.

In order to determine whether the mouse sequences are more distantly related to human F_V sequences than to other mouse F_V sequences, the homology was calculated using a Dayhoff mutation matrix (Dayhoff, M., Barker, W. and Hunt, L. (1983), Meth. Enz. **91**, pp. 524-545). The homology was calculated between all the sequences in a pool of both human and mouse sequence patches made up of the surface accessible residues. The data was then represented as a density map (not shown) in which the sequences are plotted against each other. The density map can be used to discriminate "murine surfaces" from "human surfaces".

B) Reshaping of Antibody N901

In order to test the resurfacing approach suggested by the above analysis, three humanization experiments were set up. 1) Traditional loop grafting (Verhoeyen, M.E., Saunders, J.A., Broderick, E.L., Eida, S.J. and Badley, R.A. (1991), Disease markers **9**, pp. 3-4) onto a human F_V framework of known structure (KOL). 2) Resurfacing approach using most similar chain. 3) Resurfacing approach using human sequences with most similar surface residues.

The antibody used was the murine anti-N901 antibody (Griffin et al. (1983), J. Imm. **130**, pp. 2947-2951). The anti-N901 antibody (also referred to herein as the "N901 antibody") is available commercially from Coulter Corporation under the name NKH-1.

The alignment of the light chain sequences and heavy chain sequences in Figures 4A and 4B, respectively, show the original N901 antibody and the sequences used in each of the three approaches outlined here.

Figures 4A and 4B show alignments of sequences generated using the three methods of humanization. Sequences are: 1) Original rodent N901. 2+3) KOL (Marquart, M. Deisenhofer, J. and Huber, R. (1980), J. Mol. Biol. **141**, pp. 369-391) and reshaped N901 using KOL surface. 4+5) Most homologous sequences, L(KV2F) (Klobeck, H., Meindl, A., Combriato, G., Solomon, A. and Zachau, H. (1985), Nucleic Acids Res., pp. 6499-6513) and H(G36005) (Schroeder, H. and Wang, J. (1990), Proc. Natl. Acad. Sci. U.S.A. **87**) and reshaped N901 using these sequences. 6+7) Most homologous with respect to surface residues, L(KV4B) (Klobeck, H., Bronkamp, G., Combriato, G., Mocikat, R., Pohelnz, H. and Zachau, H. (1985), Nucleic Acids Res. 3, pp. 6515-6529) and H(PLO123) (Bird, J., Galili, N., Link, M., Sites, D. and Sklar, J. (1988), J. Exp. Med. **168**, pp. 229-245), and reshaped N901 using these sequences. The numbering is the same as used in the antibody modelling program ABM (ABM is a trademark for commercial software, Oxford Molecular Ltd., Oxford, U.K.), which is based on structural conservation and not sequence homology as used by Padlan et al. (Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M. and Gottesman, K.S. (1987), Sequences of Proteins of Immunological Interest. U.S. Department of Health and Human Services, Fourth Edition). The sequence changes which have to be introduced in order to reshape N901 with a given sequence are marked with bars, and back-mutations as determined from F_V models are marked with stars. The sequence homology of a given sequence to N901 is shown in brackets after each sequence.

(1) Classical Humanization

In classical humanization the rationale is to graft the rodent CDR's onto a framework of known structure, such that CDR-framework interactions can be accurately monitored by homology modelling. The model of the humanized antibody is compared to that of the original rodent antibody, and possible CDR interacting framework residues are back mutated (marked with '•' in alignment) in order to retain the three-dimensional shape of the CDR's. In this example the antibody KOL was used, giving a low homology score of only 77 and 46 in the heavy and light chains respectively.

(2) Most Similar Chain Resurfacing

A database of nonredundant human antibody sequences was compiled from available protein and nucleotide sequences. A total of 164 H and 129 L chains were sampled.

Each of the rodent chains, L and H, were then matched and the most similar human sequence found independently (G36005/KV2F) (Schroeder, H. and Wang, J. (1990), Proc. Natl. Acad. Sci. U.S.A. **87**); Klobeck, H., Meindl, A., Combriato, G., Solomon, A. and Zachau, H. (1985), Nucleic Acids Res., pp. 6499-6513). Surface residues, as outlined in Table 1, were then changed in the rodent sequences to match those of the human sequences. Subsequently a model was built of the resurfaced antibody and compared to the model of the original rodent antibody and back mutation of any CDR interacting residues was performed.

(3) Most Similar Surface Replacement According to the Present Invention

This method is identical to the above method, except that the similarity is calculated only over the surface residues outlined in Table 1 above.

The same procedure of surface mutation and subsequent back mutation was performed as in the previous methods. In this case the chosen sequences were PLO123/KV4B (Bird, J., Galili, N., Link, M., Sites, D. and Sklar, J. (1988), J. Exp. Med. **168**, pp. 229-245); Klobeck, H., Bronkamp, G., Combriato, G., Mocikat, R., Pohelnz, H. and Zachau, H. (1985), Nucleic Acids Res. **3**, pp. 6515-6529).

The following lists show the surface residue patterns in mouse and human light and heavy chain antibody variable regions. The sequences are ordered on similarity to one another. There are no pattern matches between mouse and human sequences although there are matches within a species.

MOUSE LIGHT CHAIN SURFACE PATCHES

1	KV5E$MOUSE	:KTSLRPGKGSSDYEKK*	(SEQ ID NO: 40)
2	PL0101	:KTSLRPGKGSSEYEKK*	(SEQ ID NO: 41)
3	N$1F19L	:QTSLRPDKGSSDHEKK*	(SEQ ID NO: 42)
4	KV5U$MOUSE	:QTSLRPDKGSSDQEKK*	(SEQ ID NO: 43)
5	MUSIGLDD	:QSSLRPDKGSSDQEKK*	(SEQ ID NO: 44)
6	PL0220	:QTSLRPDKGSSDPEKK*	(SEQ ID NO: 45)
7	KV5J$MOUSE	:QTSLRPDKGSSDPZKK*	(SEQ ID NO: 46)
8	MUSIGKABB	:QTSLRPDKGSSDPEXT*	(SEQ ID NO: 47)
9	MUSIGKCLG	:QTSLRADKGSSDQEKK*	(SEQ ID NO: 48)
10	MUSIGGVJ2	:QTSLRPDKGKSDSEKK*	(SEQ ID NO: 49)
11	MUSIGKCRN	:QTSLRPARGSSDQEKK*	(SEQ ID NO: 50)
12	MUSIGKCLF	:QTSLKPGRGSSDPEKK*	(SEQ ID NO: 51)
13	MUSIGKACM	:QTSLRPGRGSSDTEKK*	(SEQ ID NO: 52)
14	MUSIGKABE	:QISLRPGKGSSDSEKK*	(SEQ ID NO: 53)
15	KV5P$MOUSE	:QTSLRPGKGDSDEDKK*	(SEQ ID NO: 54)
16	MUSIGKCMK	:ETALRPGKGASDADKK*	(SEQ ID NO: 55)
17	KVJD$MOUSE	:VTALRPGKGASDEDKK*	(SEQ ID NO: 56)
18	MUSIGKAAW	:VTALRPGKGASDEEKK*	(SEQ ID NO: 57)
19	KVJG$MOUSE	:VTALRPGKGASBABKK*	(SEQ ID NO: 58)
20	KVJE$MOUSE	:VTALRPGKGASDEDDE*	(SEQ ID NO: 59)
21	MUSIGKAAZ	:QTSLRPDKGSSDQETT*	(SEQ ID NO: 60)
22	MUSIGKCNE	:QNSLTPGKGSSSPEKK*	(SEQ ID NO: 61)
23	MUSIGKBA	:VTKVRPGKGDSDSDKK*	(SEQ ID NO: 62)
24	KV5A$MOUSE	:VTKVRPGKGDSDAEKK*	(SEQ ID NO: 63)
25	MUSIGKV	:VTRVRPGKGDSDAEKK*	(SEQ ID NO: 64)
26	MUSIGKCNM	:LTKVRPGKGDSDSEKK*	(SEQ ID NO: 65)
27	MUSIGKCL1	:VTKVRPGKGDSDSEQK*	(SEQ ID NO: 66)
28	KV5B$MOUSE	:VTKVRPEKGDSDAEKK*	(SEQ ID NO: 67)
29	MUSIGKCSA	:VTKVRPEKGDSDSEKK*	(SEQ ID NO: 68)
30	MUSIGKCSR	:VTKVSPGKGDSDAEKK*	(SEQ ID NO: 69)
31	MUSIGKCST	:VTKVRSGKGESDAEKK*	(SEQ ID NO: 70)
32	MUSIGKAB	:VTSVKPGKGDSDAEKK*	(SEQ ID NO: 71)
33	PL0014	:VSSVKPGKGDSDAEKK*	(SEQ ID NO: 72)
34	MUSIGKACU	:VTSAKPGKGDSDAEKK*	(SEQ ID NO: 73)
35	PS0023	:VSSAKPGKGDSDAEKK*	(SEQ ID NO: 74)
36	N$2MCPL	:VTSARPGKGDSDAEKK*	(SEQ ID NO: 75)
37	MUSIGKADV	:VSPAKPGKGDSDAEKK*	(SEQ ID NO: 76)
38	MUSIGKCPF	:VTKARPGKGDSDVEKK*	(SEQ ID NO: 77)
39	MUSIGLDB	:VTLIPPGKGDSDAEKK*	(SEQ ID NO: 78)
40	MUSIGKCNZ	:VTLLQPGKGDSDAEKK*	(SEQ ID NO: 79)
41	B27887	:VTLLQPGKGDSDADKK*	(SEQ ID NO: 80)
42	H28840	:VTLLQPGKGDSDAERK*	(SEQ ID NO: 81)
43	KV2G$MOUSE	:VTLLQAGKGDSDAEKK*	(SEQ ID NO: 82)
44	C27887	:VTLLQPGEGDSDAEKK*	(SEQ ID NO: 83)
45	JL0029	:LTLLQPGNGDSDAEKK*	(SEQ ID NO: 84)
46	MUSIGKAZH	:VTLLQPGKGDSDAEKI*	(SEQ ID NO: 85)
47	PS0074	:VTLPQPGQGDSDPEKK*	(SEQ ID NO: 86)
48	MUSIGKCNY	:VTLPQPGKGDSDAEKK*	(SEQ ID NO: 87)
49	MUSIGKCNX	:VTLPQPGKGDWDAEKK*	(SEQ ID NO: 88)
50	KV2D$MOUSE	:VTFLSPGQGDSDAEKK*	(SEQ ID NO: 89)

51	MUSIGKADW	:ESSARPGKGDSDAEKK*	(SEQ ID NO: 90)
52	KV2A$MOUSE	:VTLSSPGQGDSDAEKK*	(SEQ ID NO: 91)
53	KV1A$MOUSE	:VTTAKPEKGDSDVEKK*	(SEQ ID NO: 92)
54	F30534	:VTTPKPDKGDSDVEKK*	(SEQ ID NO: 93)
55	MUSIGKCLO	:VTAPRPGKGASSAEKK*	(SEQ ID NO: 94)
56	G27887	:VTAPKPGKGTSSAEKK*	(SEQ ID NO: 95)
57	MUSIGVKV3	:VTTPKPGKGASSAEKK*	(SEQ ID NO: 96)
58	MUSIGKCNA	:VSAPKPGKGASSAEKK*	(SEQ ID NO: 97)
59	S03410	:VTAPRSGKGASSAEKK*	(SEQ ID NO: 98)
60	B32456	:VTAPKSGKGASSAEKK*	(SEQ ID NO: 99)
61	PL0013	:VTAPKPDKGVSSAEKK*	(SEQ ID NO: 100)
62	MUSIGLAET	:VTAPKSEKGVSSAEKK*	(SEQ ID NO: 101)
63	MUSIGVKV1	:FTAPKPGKGASSAEKK*	(SEQ ID NO: 102)
64	KV6K$MOUSE	:LTAPKPGRGVSSAEKK*	(SEQ ID NO: 103)
65	G30560	:VTAPKSGKGASSAEKR*	(SEQ ID NO: 104)
66	MUSIGKBO	:VSAPKPGKEGSSAEKK*	(SEQ ID NO: 105)
67	MUSIGKCNB	:VTAPKPRKGASSAEKK*	(SEQ ID NO: 106)
68	H33730	:VTPLSPGQGNSDAELP*	(SEQ ID NO: 107)
69	MUSIGKCPC	:VTPLSPGQGNSDEDLP*	(SEQ ID NO: 108)
70	KV2C$MOUSE	:VTLSSPQRGDSDAEKK*	(SEQ ID NO: 109)
71	MUSIGLAV	:VTAPKSSKGGSSAEKK*	(SEQ ID NO: 110)
72	MUSIGKCNH	:QTSPTPGKGSSDPEKK*	(SEQ ID NO: 111)
73	KV5R$MOUSE	:QISLIPGKGSYDDEKK*	(SEQ ID NO: 112)
74	KV6E$MOUSE	:VTALKSGKGASSAEKK*	(SEQ ID NO: 113)
75	MUSIGKCNI	:VTALKSDKGASSGEKK*	(SEQ ID NO: 114)
76	MUSIGLDA	:VTPPSPGQGDSAAEKK*	(SEQ ID NO: 115)
77	C26317	:VTPPSPGQGDSAREKK*	(SEQ ID NO: 116)
78	PS0073	:VTVRKPGKGDSSDEKK*	(SEQ ID NO: 117)
79	A23986	:QTSVRLGQGSSDPEKK*	(SEQ ID NO: 118)
80	MUSIGKABW	:KTSLRPWKGSSDSDKK*	(SEQ ID NO: 119)
81	KV5D$MOUSE	:QTDVTQGQGSSQPEKK*	(SEQ ID NO: 120)
82	MUSIGE6L	:QTAVSQGQGSSQSEKK*	(SEQ ID NO: 121)
83	MUSIGKCOE	:LTAPRTWRGSSDSEKK*	(SEQ ID NO: 122)
84	MUSIGKCKH	:VTAPSSHRGSSDTEKK*	(SEQ ID NO: 123)
85	MUSIGLVD	:LLSLSPLKGDSDPEKV*	(SEQ ID NO: 124)
86	S06822	:VTAPTPDTGAIKTEKL*	(SEQ ID NO: 125)
87	S06821	:VTIPTPDTGAIKTEKL*	(SEQ ID NO: 126)
88	MUSIGLAS	:AVSPTPDTGAIKTEKL*	(SEQ ID NO: 127)
89	MUSIGLAR	:AVSPTPDTGAIKTEKL*	(SEQ ID NO: 128)
90	KV2B$MOUSE	:AVSPTPDTGVIKTEKL*	(SEQ ID NO: 129)
91	MUSIGLAM	:AVSPTPDTGAIKTEPS*	(SEQ ID NO: 130)

HUMAN LIGHT CHAIN SURFACE PATCHES

1	LV4A$HUMAN	:YLPPTPGVIRSTAMKL*	(SEQ ID NO: 131)
2	LV4B$HUMAN	:YLPPTPGVIRSTAMRL*	(SEQ ID NO: 132)
3	LV4E$HUMAN	:YLPPTPGLIRSTSMKL*	(SEQ ID NO: 133)
4	LV4D$HUMAN	:YLPPTPGLIRSTSVKL*	(SEQ ID NO: 134)
5	LV4C$HUMAN	:YLPPTPGVIRSTAEKL*	(SEQ ID NO: 135)
6	LV5A$HUMAN	:YLPPTPGVIRSTAGKL*	(SEQ ID NO: 136)
7	LV7A$HUMAN	:YLPATPGVVRSSAGML*	(SEQ ID NO: 137)
8	LV2G$HUMAN	:SLPPSPGKVRSTAEKL*	(SEQ ID NO: 138)
9	LV2I$HUMAN	:SLPPSPGKVRSTANKL*	(SEQ ID NO: 139)
10	N$2RHE	:SLPPRPGKVRSSSEKL*	(SEQ ID NO: 140)
11	HUMIGLAN	:SLPPRPGKVRSSSDKL*	(SEQ ID NO: 141)
12	LV1A$HUMAN	:SLPPRPGRVRSSSEKL*	(SEQ ID NO: 142)
13	LV1B$HUMAN	:SLPPRPGKVRSSSEQL*	(SEQ ID NO: 143)
14	LV1F$HUMAN	:SLPPRPGKVRSSSETL*	(SEQ ID NO: 144)
15	LV1C$HUMAN	:SLPPKPGKIRSSTGKL*	(SEQ ID NO: 145)
16	A29700	:SLPPKPGRIRSSTGKL*	(SEQ ID NO: 146)
17	HUMIGLAM4	:SLPPKPGKIRSSTGQL*	(SEQ ID NO: 147)
18	LV1D$HUMAN	:SLPPEPGKIRSSTGRL*	(SEQ ID NO: 148)
19	LV2K$HUMAN	:SLAPSPGKIRSTAEKL*	(SEQ ID NO: 149)
20	LV1I$HUMAN	:SLPPRPGKIRSSTGNV*	(SEQ ID NO: 150)
21	LV2E$HUMAN	:SLRPSPGKVRSTAEKL*	(SEQ ID NO: 151)
22	LV2D$HUMAN	:SLRPSPGKVRSTADKL*	(SEQ ID NO: 152)
23	LV2C$HUMAN	:SLRPSPGKVRSTAEML*	(SEQ ID NO: 153)
24	LV2J$HUMAN	:SLRPSPGKVRSAVEKL*	(SEQ ID NO: 154)
25	LV1E$HUMAN	:SLPPRPGK-RSSAEKL*	(SEQ ID NO: 155)
26	LV2B$HUMAN	:SLAPSPGKVRSTVERL*	(SEQ ID NO: 156)
27	N$1MCMW	:SLAPSPDKIRSTPDKL*	(SEQ ID NO: 157)
28	LV2H$HUMAN	:SLALSPGKVRSTAEKL*	(SEQ ID NO: 158)
29	N$3MCG2	:SLPLSAGKVRSTAEKL*	(SEQ ID NO: 159)
30	LV2A$HUMAN	:SLAPSPGKVRSTAEYL*	(SEQ ID NO: 160)
31	S02083	:SLPLTPGLIRSTAEKL*	(SEQ ID NO: 161)
32	HUMIGLAM2	:SLPLTPRVIRSTAEKL*	(SEQ ID NO: 162)
33	LV6C$HUMAN	:FLHPTPGTDSSSTEKL*	(SEQ ID NO: 163)
34	LV6D$HUMAN	:FLLPTPGTDSSSTERL*	(SEQ ID NO: 164)
35	LV6E$HUMAN	:FLHPTRVTDSSSTEKL*	(SEQ ID NO: 165)
36	LV6B$HUMAN	:LLPPTPGTNSSSNDKL*	(SEQ ID NO: 166)
37	HUMIGLK5G	:VLPLSPHRIRSESEML*	(SEQ ID NO: 167)
38	HUMIGLVC	:SLAPSPAKFRSTAERD*	(SEQ ID NO: 168)
39	HUMIGVLLS	:VTAPRPGRIRSDPEKK*	(SEQ ID NO: 169)
40	HUMIGKAX	:VTAPRPGRVRSDPEKK*	(SEQ ID NO: 170)
41	E30609	:VTGPRPGRIRSDPEKK*	(SEQ ID NO: 171)
42	KV3B$HUMAN	:VTGPRPGRIRSDPDKK*	(SEQ ID NO: 172)
43	G30607	:VTGPRPGRVRSDPEKK*	(SEQ ID NO: 173)
44	KV3M$HUMAN	:VTGPRPGRIRSDPXKK*	(SEQ ID NO: 174)
45	KV3H$HUMAN	:VTAPRPGRIRSESERK*	(SEQ ID NO: 175)
46	KV3K$HUMAN	:VTGPSRGRIRSDPEKK*	(SEQ ID NO: 176)
47	KV3F$HUMAN	:VTVPRPSRIRSESERK*	(SEQ ID NO: 177)
48	B26555	:VTAPGPGRIRSESERK*	(SEQ ID NO: 178)
49	KV1Q$HUMAN	:QTSVRPGRVRSDPERK*	(SEQ ID NO: 179)
50	KV1W$HUMAN	:QTSVRPGKVRSDPERK*	(SEQ ID NO: 180)

51	KV1M$HUMAN	:QTSVRPGKVRSDPEKK*	(SEQ ID NO: 181)
52	KV1R$HUMAN	:QTSVRPGKVRSEPEKK*	(SEQ ID NO: 182)
53	KV1F$HUMAN	:QTSVRPGKVRSEPDKK*	(SEQ ID NO: 183)
54	KV1G$HUMAN	:QTSVRPGKVRAEPEKK*	(SEQ ID NO: 184)
55	KV1K$HUMAN	:QTSVRPGKVRSBPZKK*	(SEQ ID NO: 185)
56	KV1D$HUMAN	:QTSVRPGKVRSDPBKK*	(SEQ ID NO: 186)
57	KV1H$HUMAN	:QTSVRPGQVRSDPERK*	(SEQ ID NO: 187)
58	KV1B$HUMAN	:QTSVRPGKVRSHPEKK*	(SEQ ID NO: 188)
59	B27585	:QTSVRPGNVRSDPDKK*	(SEQ ID NO: 189)
60	N$1REIA	:QTSVRPGKVRSDPEKT*	(SEQ ID NO: 190)
61	KV1X$HUMAN	:QTSVRPGTVRSEPEKK*	(SEQ ID NO: 191)
62	KV1L$HUMAN	:QTSVRPEKVRSEPDKK*	(SEQ ID NO: 192)
63	IMGL38	:QTSVRPGKVRSESDKK*	(SEQ ID NO: 193)
64	A27585	:QTSVRPGEVRSEPDKK*	(SEQ ID NO: 194)
65	KV1N$HUMAN	:QTSVRPGBVRSBPZRK*	(SEQ ID NO: 195)
66	KV1C$HUMAN	:QTSVSPGKVRSDPEKK*	(SEQ ID NO: 196)
67	KV1V$HUMAN	:QTSVRPGKVNSDPEKK*	(SEQ ID NO: 197)
68	KV1T$HUMAN	:QTSVRPGKVRSDPDTK*	(SEQ ID NO: 198)
69	KV1U$HUMAN	:QTSVRPKKVRSDPZKK*	(SEQ ID NO: 199)
70	KV1A$HUMAN	:QTSVRPKKVRPDPEKK*	(SEQ ID NO: 200)
71	KV1S$HUMAN	:QTSVRSGKVRSEPETK*	(SEQ ID NO: 201)
72	KV4A$HUMAN	:VTNLRPGKVRSDAEKK*	(SEQ ID NO: 202)
73	KV4C$HUMAN	:VTDLRPGKVRSDAEKK*	(SEQ ID NO: 203)
74	HUMIGK2A1	:QTSVSPGNIRSESDKK*	(SEQ ID NO: 204)
75	HUMIGKBA	:KTSVTPGKPRSEPEKK*	(SEQ ID NO: 205)
76	HUMIGKBC	:VTLLPPGRVRSDAEKK*	(SEQ ID NO: 206)
77	KV2B$HUMAN	:VTLLPPGEVRSDAEKK*	(SEQ ID NO: 207)
78	KV2D$HUMAN	:VTLPPPGZVRSDAERK*	(SEQ ID NO: 208)
79	KV2C$HUMAN	:VTLPPPGZVRSBAZNK*	(SEQ ID NO: 209)
80	KV2E$HUMAN	:VTLPPPQQVRSDAEKK*	(SEQ ID NO: 210)
81	S03876	:VTLPPPGQVTSDAEKK*	(SEQ ID NO: 211)
82	KV2A$HUMAN	:VTLPPAGQVRSDAEKR*	(SEQ ID NO: 212)
83	HUMIGLAM5	:ALSPSSGQSSSASERL*	(SEQ ID NO: 213)

MOUSE HEAVY CHAIN SURFACE PATCHES

1	MUSIGHIT	:EKVGGLQPGRGTPGKASRGDSQRPES•	(SEQ ID NO: 214)
2	MUSIGHIU	:EKVGGLQPGRGTPGKVSRGDSQRPES•	(SEQ ID NO: 215)
3	MUSIGHIV	:EKVGGLQPGTGAPGKASRGDSQRPES•	(SEQ ID NO: 216)
4	MUSIGHYM	:EKVGGLQPGRGTPGKASKGNSQRAES•	(SEQ ID NO: 217)
5	PU0001	:EKMGGLQPGRGTPGKASKGNSQRAES•	(SEQ ID NO: 218)
6	MUSIGHFO	:EKVGGLQPGRGTPGKASKGTSQRAES•	(SEQ ID NO: 219)
7	AJ0515	:EKVGGLQPGRGTPGKASKGTSQRAET•	(SEQ ID NO: 220)
8	PL0018	:EKVGGLKPGRGTPGKASKGTSQRAET•	(SEQ ID NO: 221)
9	MUSIGHFK	:ENVGGLQPGRGTPGKASKGTSQRAET•	(SEQ ID NO: 222)
10	MUSIGHFQ	:EKVGGLQSGRGTPGKASKGTSQRAET•	(SEQ ID NO: 223)
11	PU0001	:EKVGGLQSGRGTPGKASKGTSQRAES•	(SEQ ID NO: 224)
12	EJ0540	:EKVGGLQPGRGTPGKASKGISQRAER•	(SEQ ID NO: 225)
13	HV17$MOUSE	:EKVGGLQPGRGTPGKSAKGBSZRAQS•	(SEQ ID NO: 226)
14	MUSIGHLN	:EKVGGLQPGSGTPGKASKGNSQRAES•	(SEQ ID NO: 227)
15	MUSIGHKG	:EKVGGLQPGSGTPGKASKGSSQRAES•	(SEQ ID NO: 228)
16	PU0004	:EKVGGLQPGRGTPRKASKGNSQRAES•	(SEQ ID NO: 229)
17	MUSIGHKJ	:EKMGNLQPGSGTPGKASKGNSQRPDS•	(SEQ ID NO: 230)
18	HV56$MOUSE	:EKVGGLKPGKGTPEKDSKGNARRSET•	(SEQ ID NO: 231)
19	C27888	:EKVGGLKPGKGAPEKDSKGNARRSET•	(SEQ ID NO: 232)
20	MUSIGHAAP	:EKVGGLKPGKGTPERDSKGNARRSET•	(SEQ ID NO: 233)
21	PH0097	:DKVGGLKPGKGTPEKDSKGNAKRSET•	(SEQ ID NO: 234)
22	E27888	:DKVGGLKPGKGTPEKDSKGNAKKSET•	(SEQ ID NO: 235)
23	MUSIGHJB	:DKVGGLKPGKGTPDKDNKGNAKKSET•	(SEQ ID NO: 236)
24	MUSIGHADL	:EKVGGLTPGKGTPEKDSKGNGRRSET•	(SEQ ID NO: 237)
25	A27888	:ENVGGLKPGKGTPEKDSKGNDRRSET•	(SEQ ID NO: 238)
26	H27887	:ENVGGLKPGKGTPEKDSKGNDKRSET•	(SEQ ID NO: 239)
27	B27888	:ENVGGLKPGKGTPEKDSKGNAKRSET•	(SEQ ID NO: 240)
28	B27889	:EQVGGLKPGKGTPEKDSKGNAKKSET•	(SEQ ID NO: 241)
29	D27889	:EQVGGLKPGKGTPEKDTKGNAKKSET•	(SEQ ID NO: 242)
30	HV55$MOUSE	:EQVGGLKPGKGAPEKDTKGNAKKSET•	(SEQ ID NO: 243)
31	MUSIGHAGT	:EKVGGLQPGKGTPEKDSKGNAKKSET•	(SEQ ID NO: 244)
32	MUSIGVH50	:EKVGGLQPGKGTPEKDTKGNAKKSET•	(SEQ ID NO: 245)
33	MUSIGHIW	:EKVGGLQPGRGTPEKDTKGNAKKSET•	(SEQ ID NO: 246)
34	MUSIGHAGZ	:EKVGGLQPGKGSPEKDSKGNAKKSET•	(SEQ ID NO: 247)
35	PH0098	:DKMGGLKPGKGTPEKDSKGNAKQSET•	(SEQ ID NO: 248)
36	MUSIGHIB	:EQVGGLQPGKGTPDKDSKGNAKKSET•	(SEQ ID NO: 249)
37	MUSIGHAGY	:EKVGGLQPGKGTPEKDSKGNAEKSET•	(SEQ ID NO: 250)
38	MUSIGHMF	:EQVGDLKPGKGTPEKDTKGNARRSET•	(SEQ ID NO: 251)
39	D27888	:ENVGDLKPGKGAPEKDSKGNARRSET•	(SEQ ID NO: 252)
40	MUSIGHIP	:EQVGGLQPGKGTSDKDSKGNAKKSET•	(SEQ ID NO: 253)
41	MUSIGHAGS	:EQVGGLQPGKGTPEKDSKGNAKKSGT•	(SEQ ID NO: 254)
42	HV16$MOUSE	:DQVGGLQPGKGTPEKDTKGNPKRSET•	(SEQ ID NO: 255)
43	BJ4871	:DQVGGLQPGQGTPEKDTKGNPKRSDT•	(SEQ ID NO: 256)
44	PH0094	:EKVGGLQPGKGTSEKDIKGNAKKSET•	(SEQ ID NO: 257)
45	PH0096	:DKVGGLKPGKRTPEKDNKGNAKKSET•	(SEQ ID NO: 258)
46	MUSIGVH62	:DKVGGLKLGKGTPEKDTKGNAKKSET•	(SEQ ID NO: 259)
47	MUSIGHAGR	:EKVGGLQPGKGTPEKDSKGNANTSET•	(SEQ ID NO: 260)
48	HV58$MOUSE	:ENVGGLKPGKGTPEKDSKGNAGRSET•	(SEQ ID NO: 261)
49	H27888	:EQVGGLQPGNGTPEKDTTGNAKRSET•	(SEQ ID NO: 262)
50	HV34$MOUSE	:EKEGGLQPGKGTPEKESKGDSKRAET•	(SEQ ID NO: 263)

51	HV33$MOUSE	:EKEGGLQPGKGTPEKESKGDSKRPET•	(SEQ ID NO: 264)
52	MUSIGHZAB	:EKEGGLQPGKGSPEKESKGDSKRAET•	(SEQ ID NO: 265)
53	N$4FABH	:EKDGGLQPGKGTPEKDSKGDSKRVEM•	(SEQ ID NO: 266)
54	I27888	:EQVGGLKPGRGTPEKDTTGDAQRSET•	(SEQ ID NO: 267)
55	G27888	:EQVGGLKPGRGTPEKDTTGNAKGSET•	(SEQ ID NO: 268)
56	HV59$MOUSE	:EKVGGSKPGKGTPEKDSKGNAKTSET•	(SEQ ID NO: 269)
57	MUSIGHOE	:SDQGGLKPGKGTPEKDTKGNARRSES•	(SEQ ID NO: 270)
58	N$2FVWH	:EKIGGLQPGKGDPGKPSKDNAKRSET•	(SEQ ID NO: 271)
59	MUSIGHJT	:EKLGGLQPGKGDPGKPSKDNAKRSET•	(SEQ ID NO: 272)
60	MUSIGHLY	:EKLGGLQPGKGDPGKPFKDNAKRSET•	(SEQ ID NO: 273)
61	S06816	:EKLGGLQPGKGDPGKLMKENAKRSET•	(SEQ ID NO: 274)
62	S06817	:ENLGGLQPGKGDPGKLXXENAKRPET•	(SEQ ID NO: 275)
63	MUSIGHAAI	:EKLGGLQPGNGDLGKPSKDNAKRSET•	(SEQ ID NO: 276)
64	HV42$MOUSE	:EKLGPLQLGKGDPGKPSKDDAKRSET•	(SEQ ID NO: 277)
65	MUSIGHAAL	:EQLGGLQPGGGTPGKPSKDNDKRSET•	(SEQ ID NO: 278)
66	MUSIGHABO	:EQLGGLQPGGGTPGKASKDNDKRSET•	(SEQ ID NO: 279)
67	MUSIGHEG	:EQVGGLKARKGTPEKDTTGNAKRSET•	(SEQ ID NO: 280)
68	MUSIGHWN	:EKVGVLEPGKGTPEKRQEGNAKRSET•	(SEQ ID NO: 281)
69	MUSIGKCLT	:EQVGGLQPKKGSPGKDSKDDSQKTET•	(SEQ ID NO: 282)
70	MUSIGHZAE	:EQVGGLQPKKGSPGKDSKDDSQKTER•	(SEQ ID NO: 283)
71	MUSIGHAAD	:QQVPELKPGRGTPGKEDKGTSARNDT•	(SEQ ID NO: 284)
72	MUSIGHAAM	:QQVPELKPGKGTPGKDDKGTSAKNET•	(SEQ ID NO: 285)
73	MUSIGHAMA	:QQVPELKPGKGTPGKDDKGTSAKNEM•	(SEQ ID NO: 286)
74	MUSIGHXZ	:QQKPELKPGKGSPGQEKKGTSSTSET•	(SEQ ID NO: 287)
75	A30502	:EQQPELKPGKGTPGQEKKGKSSTSES•	(SEQ ID NO: 288)
76	MUSIGHAAG	:EQQPELRPGKGTPGQEKKGKSSTSES•	(SEQ ID NO: 289)
77	B30902	:EQQPELKPGKGTPGQEKKGKSSASES•	(SEQ ID NO: 290)
78	MUSIGHADG	:EQQPELKPGKGTPGKQKKGKSSTSES•	(SEQ ID NO: 291)
79	MUSIGHFV	:EQQPELKPGKGTHGKQKKGKSSTSES•	(SEQ ID NO: 292)
80	MUSIGHAAMA	:EQQPELKPGKGSHGKQKKGKSSTSES•	(SEQ ID NO: 293)
81	MUSIGHZR	:EQQPELKPGKGSHGKQKKGKSSASES•	(SEQ ID NO: 294)
82	MUSIGHAI	:EQQPELKPGKGTHGKQKKGKSSTTES•	(SEQ ID NO: 295)
83	MUSIGHALA	:EQQPELKPGKGTHGKQKQGKSSTTES•	(SEQ ID NO: 296)
84	PL0011	:EQQPELKPGKGTHGKEKKDKSSTSES•	(SEQ ID NO: 297)
85	MUSIGKCLS	:EQQAELKPGKGSHGKQKKGKSSTSES•	(SEQ ID NO: 298)
86	MUSIGHADY	:EQQPELKPGKGTHGKQKKSNSSTSES•	(SEQ ID NO: 299)
87	MUSIGHWVX	:QQQAELRPGKGAPGQEKKGKSSTSES•	(SEQ ID NO: 300)
88	MUSIGHADO	:QQQAELRPGKGAPGQEKKGKSSTSDS•	(SEQ ID NO: 301)
89	MUSIGHVBM	:QQQAELRPGKGVPGQEKKGKSSTSDS•	(SEQ ID NO: 302)
90	A24672	:QQQPELKPGKGAPGKGKKGKSSTSES•	(SEQ ID NO: 303)
91	MUSIGHJG	:QQQPELKPGKGAPGKGKKDKSSTSES•	(SEQ ID NO: 304)
92	JL0044	:EQQPEAKPGKGTHGKQKKGKSSTSDS•	(SEQ ID NO: 305)
93	MUSIGHBA	:QQQAELKPGKGTHGKEKKDKSSTSDS•	(SEQ ID NO: 306)
94	MUSIGHAGP	:QQQAELRPGKGAPGQGKKGKSSTSES•	(SEQ ID NO: 307)
95	MUSIGHVBK	:QQQAELKPGRGTPGQEKKGKSSTSES•	(SEQ ID NO: 308)
96	A36194	:EQQAELRAGKGTPGQEKKGKSSTSES•	(SEQ ID NO: 309)
97	MUSIGHVBJ	:EQQAELRPGKGTPGQEKKGTSSTSES•	(SEQ ID NO: 310)
98	MUSIGHADV	:QQQAELRPGKGTPGHEKKGTSSTSES•	(SEQ ID NO: 311)
99	MUSIGHAAT	:QQQAELKPGKGTPGHEKKGTSSTSES•	(SEQ ID NO: 312)
100	MUSIGHJL	:QQQAELRPGKGTPGHENKGTSSTSES•	(SEQ ID NO: 313)

101	MUSIGHABM	:QQQAEVRPGKGTPGHEKKGTSSTSES*	(SEQ ID NO: 314)
102	MUSIGHFU	:QQQAELKPGKGTPGHENKGTSSTSES*	(SEQ ID NO: 315)
103	MUSIGHZZB	:QQQAELRPGKGTPGQQKKGKSSASES*	(SEQ ID NO: 316)
104	HV06$MOUSE	:HQQAELKPGKGTPGQQKKGKSSTSES*	(SEQ ID NO: 317)
105	MUSIGHRD	:EQQVELRAGKGTPGQEKKGKSSTSES*	(SEQ ID NO: 318)
106	MUSIGHVBH	:EQQAELRPGKGTPGQEKQGTSSTSES*	(SEQ ID NO: 319)
107	HV01$MOUSE	:EQQAELRPGKGTPGHDMKGTSSTSES*	(SEQ ID NO: 320)
108	MUSIGHADN	:QQQAEVRPGKGTPGHEKKGRSSTSES*	(SEQ ID NO: 321)
109	HV05$MOUSE	:QQQAELRPGKGTPGQQKKDKSSTSES*	(SEQ ID NO: 322)
110	MUSIGHAEF	:QQQAELKPGKGTPGQQKKDKSSTSES*	(SEQ ID NO: 323)
111	MUSIGHAAN	:QQQAELKPGKGTPGQQKKDKSSTSDS*	(SEQ ID NO: 324)
112	MUSIGHAAB	:QQQAELKPGKGSPGQQKKDKSSTSES*	(SEQ ID NO: 325)
113	C30560	:QHQAELKPGKGTPGQQKKNKSSTSES*	(SEQ ID NO: 326)
114	PS0024	:QQQAELKPGKGTPGQQNKDKSSTSES*	(SEQ ID NO: 327)
115	MUSIGHRG	:EQQAELRAGKGIPGQEKKGKSSTSES*	(SEQ ID NO: 328)
116	MUSIGHAAE	:QQQAELKPGKGTPGQEKKSKSSTSES*	(SEQ ID NO: 329)
117	MUSIGHLX	:QQQSELKPGKGTPGQEKKSKSSTSES*	(SEQ ID NO: 330)
118	HV04$MOUSE	:QQQTELKPGKGTPGQEKKSKSSTSES*	(SEQ ID NO: 331)
119	MUSIGHVBG	:EQQAELRTGKGTPGQERKGKSSTSES*	(SEQ ID NO: 332)
120	MUSIGHMX	:QQQAELKPGKGTPGQQKKDKSSTPES*	(SEQ ID NO: 333)
121	MUSIGHAAR	:EQQAELRPGTGAPGQEKKGKSSTSES*	(SEQ ID NO: 334)
122	HV15$MOUSE	:QQQPEVRPGKGTHAKQKKGKSSTSES*	(SEQ ID NO: 335)
123	MUSIGHAAU	:QQQPEVRPGKDTHAKQKKGKSSTSES*	(SEQ ID NO: 336)
124	MUSIGHVBO	:QQQAELKPGKGTPEQEKKGKSSTSES*	(SEQ ID NO: 337)
125	A26405	:EQQTELRAGKGTPGQEKKGRSSTSEA*	(SEQ ID NO: 338)
126	HV10$MOUSE	:QQQAELKPGKGTPGREKKSKPSTSES*	(SEQ ID NO: 339)
127	MUSIGJB44	:QQQSELKPGKGTPGREKKSKPSTSES*	(SEQ ID NO: 340)
128	MUSIGJB62	:QQRAELKPGKDTPGREKKNKPSTSES*	(SEQ ID NO: 341)
129	HV09$MOUSE	:QQQAELKPGKGTPGREKKSTSSTSES*	(SEQ ID NO: 342)
130	MUSIGKCLP	:QQQAELKPGKGTPGQEKKSTSSTSDS*	(SEQ ID NO: 343)
131	MUSIGBH	:QQQAELRPGKGTPIQQKKDKSSTSES*	(SEQ ID NO: 344)
132	HV11$MOUSE	:QQQAEPKPGKGTPGREHRSKPSTSES*	(SEQ ID NO: 345)
133	MUSIGHMC	:QQQAELRPGKGALGQEKKGKSSTSDS*	(SEQ ID NO: 346)
134	MUSIGHAGW	:QQQPEVKPGKGAPGKGNTDKSSTSES*	(SEQ ID NO: 347)
135	MUSIGHRF	:EQQAEVRAGKGSPGQEKKGKSSTSES*	(SEQ ID NO: 348)
136	MUSIGHVAD	:QQLAELKPGKGTPGHEKKGISSTSES*	(SEQ ID NO: 349)
137	MUSIGHVAF	:QQQAELKPGKGKPEQEKKGTSSTSES*	(SEQ ID NO: 350)
138	PL0012	:QQQPELKPGKGRNGKEMKGKSSTSES*	(SEQ ID NO: 351)
139	MUSIGGVD2	:QQQTELRPGRGTTGQERKGKSSTSES*	(SEQ ID NO: 352)
140	S06824	:QHQAELKPGKGTPGHEMKVTSSTSES*	(SEQ ID NO: 353)
141	MUSIGHRB	:EQQAELRAGKGTPGQEQKAKSSTSES*	(SEQ ID NO: 354)
142	MUSIGHAAB	:QQQAELKPGKGTPGQQKTGTSSTTES*	(SEQ ID NO: 355)
143	MUSIGHHS	:QQQAELKPGKGNPGQEKKSTSSASES*	(SEQ ID NO: 356)
144	MUSIGHAXA	:EQQTVLRPGKGTPGQQKKGTSATNES*	(SEQ ID NO: 357)
145	HV50$MOUSE	:QQLTELKPGMGTPGQEKKSKSSTSES*	(SEQ ID NO: 358)
146	MUSIGHVBP	:QQQSVLRPGKGTPGQEKKGTSSTSKS*	(SEQ ID NO: 359)
147	PH0100	:LQQPVLKPGKGSHGKQKKDKSSTSES*	(SEQ ID NO: 360)
148	MUSIGHAYA	:EQQPETKPGKGTLGKQKKSKSSTSES*	(SEQ ID NO: 361)
149	MUSIGHCP2	:QQQAELKPGQGTPGQEKKNKSSTPEF*	(SEQ ID NO: 362)
150	MUSIGHDZ	:EQQAELRPGKGNPEQPKQGTSSTSET*	(SEQ ID NO: 363)

151	MUSIGHNPI	:EQQAELRPGKGNPEQPKQGTSTTSET*	(SEQ ID NO: 364)
152	S06823	:EQQAELKPGKGNPEQPKQGTSSTSET*	(SEQ ID NO: 365)
153	MUSIGHASA	:EQQAELKPGKGNPEQPKQDTSSTSET*	(SEQ ID NO: 366)
154	S03484	:EQQAELKPGKGNPEQPKQGTSSTSGT*	(SEQ ID NO: 367)
155	MUSIGHVAA	:EQQAEVKPGKGNPEQPKQGTSSTSET*	(SEQ ID NO: 368)
156	MUSIGHNPD	:EQQAELRPGKGNPEQPKQVTSSTSET*	(SEQ ID NO: 369)
157	MUSIGHNPB	:EQQAELRPGKGNPEQPKQITSSTSET*	(SEQ ID NO: 370)
158	MUSIGHEC	:EQQAELRPGRGNPEQPKQVTSSTSET*	(SEQ ID NO: 371)
159	MUSIGHNPC	:EQQAELRPGRGNPEQPKHVTSSTSET*	(SEQ ID NO: 372)
160	MUSIGHNPF	:EQQAELRPGKGNTEQPKQVTSSTSET*	(SEQ ID NO: 373)
161	MUSIGHNPE	:EQQAELKPGKGNTEQPKLITSSTSET*	(SEQ ID NO: 374)
162	A27635	:TGQAELRPGKGAPEQGKKGKSSTSDR*	(SEQ ID NO: 375)
163	MUSIGHXW	:QYQAELRPGKGTPRQQKKGKSSTSES*	(SEQ ID NO: 376)
164	MUSIGHIZA	:QQQAVLRHGKGTHGQEKKGKSSTSES*	(SEQ ID NO: 377)
165	MUSIGHEH	:QQQTKLGPGRGTPGQGRKGKSSTSGS*	(SEQ ID NO: 378)
166	MUSIGHRH	:EQQAELRAGKGTPGQEKKGKSSVYFA*	(SEQ ID NO: 379)
167	HV00$MOUSE	:EQQAELKAGKGTPGQQKQGESTRSET*	(SEQ ID NO: 380)
168	N$1F19H	:QQKAELAASKGTPGQEKKGRSSTSES*	(SEQ ID NO: 381)
169	MUSIGHZAD	:QQQTELRPGKGTPGQEKKGKSSNLRL*	(SEQ ID NO: 382)
170	BJ0515	:EKVGGLQGSSFDPGKASKGTSQRAET*	(SEQ ID NO: 383)
171	MUSIGHZB	:EQQADLKLGKGNPEQPKLATPSTSET*	(SEQ ID NO: 384)
172	E27889	:EQVGGLKPGKGTPDKSDVKDNAKSET*	(SEQ ID NO: 385)
173	MUSIGHAAC	:DQQPDLKPSSGSPGHPSKSTSKTTET*	(SEQ ID NO: 386)
174	HV61$MOUSE	:DQQPDLKPSSGSPGNPSKSTSKTTET*	(SEQ ID NO: 387)
175	MUSIGVHR2	:DQQPDLKPSSGSPGNPSKSTSKTAET*	(SEQ ID NO: 388)
176	PL0100	:DQQPGLKPSSGSPGNPSKSTSKTTET*	(SEQ ID NO: 389)
177	MUSIGHAAO	:DQQPGLKPSSGSPGNPSKNTSKTTET*	(SEQ ID NO: 390)
178	MUSIGHGA6	:DQQPGLKPSSGSPGDPSKTTSKTTET*	(SEQ ID NO: 391)
179	MUSIGHJY	:DQQPGLKPSSGSPGNPSKTTSKTTET*	(SEQ ID NO: 392)
180	MUSIGHGA1	:DHQPGLKPSSGSPGNPSKNTSKTTET*	(SEQ ID NO: 393)
181	MUSIGHDX	:DQQPGLKPSSGSPGNPSRSTSKTTET*	(SEQ ID NO: 394)
182	HV62$MOUSE	:DQQPGLKPSAGSPGNPSKSTSKTAET*	(SEQ ID NO: 395)
183	MUSIGHAAGA	:EQQPGLKPSSGSPGNPSKSTSKTSET*	(SEQ ID NO: 396)
184	MUSIGHGA5	:DQQPGLKPSSGSPGNPSKNTSKTIET*	(SEQ ID NO: 397)
185	MUSIGHGA4	:DQQPGLKPSSGSPGDPSKNTSKTPET*	(SEQ ID NO: 398)
186	MUSIGHAGI	:EQQPSLKPSSGSPGNPSKSTSKTTET*	(SEQ ID NO: 399)
187	PL0102	:DQQPGLKPSSGSPGNPSKNTSETTET*	(SEQ ID NO: 400)
188	HV46$MOUSE	:DQQPGLKPSSGSPGNPSKNTSKTTZT*	(SEQ ID NO: 401)
189	MUSIGHZT	:EQQPSLKPSSGSPGNPSKSTSKTSET*	(SEQ ID NO: 402)
190	MUSIGHAGD	:EQQPSLKPSSGSPGNPSKSTSRTTET*	(SEQ ID NO: 403)
191	MUSIGHAGQ	:EQQPSLKPSSGSPGNPSKSTSKTAET*	(SEQ ID NO: 404)
192	MUSIGAMJ2	:DQQPDLKPSSGTPGNPSKSTSKTTET*	(SEQ ID NO: 405)
193	MUSIGHAFX	:EQQPSLKPSSGSPGKPSKSTSKTNET*	(SEQ ID NO: 406)
194	MUSIGHAGE	:EQQPSLKPSSGSPGNPSKSTYKTSET*	(SEQ ID NO: 407)
195	MUSIGHAGB	:EQQPSLKPSSGSPGNPSKSTSTTSET*	(SEQ ID NO: 408)
196	MUSIGHAGC	:EQQLSLKPSSGSPGNPSKSTSKTTET*	(SEQ ID NO: 409)
197	MUSIGHAAM	:QQQPGLKPSPGDPGKPSQSTSKTTET*	(SEQ ID NO: 410)
198	HV43$MOUSE	:QQKPGLAPSSGSPGKSTKSNSKQTDT*	(SEQ ID NO: 411)
199	MUSIGMUV1	:QQKPGLAPSSGSPGKSAKSNSKQTDT*	(SEQ ID NO: 412)
200	MUSIGHAEI	:QQKPGLAPSSGSPGKSAMSNSKQTDT*	(SEQ ID NO: 413)
201	MUSIGHBP	:QQKPGLAPSSGSPGKSAISNSKQTDT*	(SEQ ID NO: 414)
202	MUSIGHZZA	:QQKPGLQPSSGSPGKAAISNSKQSNT*	(SEQ ID NO: 415)
203	MUSIGMUV2	:QQKPGLQPSSGSPGKAAISNSKQANT*	(SEQ ID NO: 416)
204	AJ2456	:QQKPVLAPSSGSPGKSAMSNSKQIDT*	(SEQ ID NO: 417)
205	MUSIGHDB	:QQKPSLQPSSDSPGKAAMSNSKQADT*	(SEQ ID NO: 418)

HUMAN HEAVY CHAIN SURFACE PATCHES

1	HUMIGHVS	:ERVGDLEPGRGIPGKAPKGDSKKIET•	(SEQ ID NO: 419)
2	HUMIGHVR	:ERVGDLEPERGIPGKAPKGDSKKIET•	(SEQ ID NO: 420)
3	HJ6005	:EQVGGLKPGRGTPGKAPKGDSKKTET•	(SEQ ID NO: 421)
4	PL0122	:EQVGGLQPGKGTSGKASKGDSKKTET•	(SEQ ID NO: 422)
5	HVJDSHUMAN	:EQLGGLQPGRGTPGKBSKGDSKRAET•	(SEQ ID NO: 423)
6	HUMIGHAT	:EQLGGLQPGRGTPGKDSKGNSKRAET•	(SEQ ID NO: 424)
7	BJ4964	:EQLGGLQPGRGTPGKDSRGNSKRAET•	(SEQ ID NO: 425)
8	AJ4964	:EQVGGLQPGRGTPGKDSKGNSKRAET•	(SEQ ID NO: 426)
9	PL0123	:EQVGGLQPGRGTPGKDSKGNAKRAET•	(SEQ ID NO: 427)
10	HVJFSHUMAN	:EQVGGLQPGRGTPGKDSKGDSRRAET•	(SEQ ID NO: 428)
11	JL0048	:EQVGGLQPGRGTPGKDSKGNSRRAET•	(SEQ ID NO: 429)
12	HVJBSHUMAN	:QQVGGLEPGRGTPGKDSKGBSKRAET•	(SEQ ID NO: 430)
13	HUMIGHBV	:EQLGDLQPGRGTPGKASKGNSKRAET•	(SEQ ID NO: 431)
14	HVJESHUMAN	:EQVGGLQPGRGTTGKDSKGDSKRAET•	(SEQ ID NO: 432)
15	PL0116	:QQVGGVQPGRGTPGKDSKGNSKRAET•	(SEQ ID NO: 433)
16	HVJKSHUMAN	:QQVGGVQPGRGIPGKDSKGNSKRPET•	(SEQ ID NO: 434)
17	N92FB4H	:EQVGGVQPGRGIPGKDSKGDSKRPET•	(SEQ ID NO: 435)
18	HVJISHUMAN	:QQVGGVQPGRGTPGKDSNGDSKRPET•	(SEQ ID NO: 436)
19	HVJJSHUMAN	:QKVGGVQPGRGTPGKDSKGNSKRTET•	(SEQ ID NO: 437)
20	HVJGSHUMAN	:QEVGGVIPGRGTPGKBSKGBSKRAET•	(SEQ ID NO: 438)
21	HVJMSHUMAN	:EQLGGLQPGRGTPGKDSNGDSKQAIT•	(SEQ ID NO: 439)
22	HVJOSHUMAN	:EQLGGLQPGRGSPGKDTNGDSKEAIT•	(SEQ ID NO: 440)
23	HVJNSHUMAN	:AQLGGLQPGRGTPGKDSNGDSKQAIS•	(SEQ ID NO: 441)
24	HVJRSHUMAN	:EQLGGLQPGRGTPGKVSQGDSKQAIT•	(SEQ ID NO: 442)
25	HVJPSHUMAN	:EQVGGLQPGRGTPGKVSQGDSKEPIT•	(SEQ ID NO: 443)
26	HUMIGHCV	:EQLGGLQPERGTPGKESKGNSKRAET•	(SEQ ID NO: 444)
27	HVJTSHUMAN	:EQVGDLQPGRGBPGKDSKGNAKRVET•	(SEQ ID NO: 445)
28	HVJUSHUMAN	:EQVGDLQPGRGNPGKDSKGNAQRPET•	(SEQ ID NO: 446)
29	PL0098	:QQVGGVQPGRGTLGKDSKGNSKRAET•	(SEQ ID NO: 447)
30	HVJHSHUMAN	:QIVGGAIPGRGSPGKASKGBSKRAET•	(SEQ ID NO: 448)
31	HVJASHUMAN	:QQVGGLKPGRGSPGKDSKGNAQRTIT•	(SEQ ID NO: 449)
32	HVJSSHUMAN	:EQVGGLKPGRGTPGKNSNGDSKTPIT•	(SEQ ID NO: 450)
33	HUMIGHAM	:EQLGGLQPGRGTSREDSKGNSKRAFT•	(SEQ ID NO: 451)
34	HVJQSHUMAN	:EQVGALQPGRGTPGKDSQADSKEAIT•	(SEQ ID NO: 452)
35	AJ6040	:EQLGGLQPGRGTPGK----VEGSVFT•	(SEQ ID NO: 453)
36	HUMIGHAB	:EQVGAPQPGRGNSGKASKGDSKRPDT•	(SEQ ID NO: 454)
37	HUMIGHAO	:EQVGAPQPGKGNSGKASKGDSKRPDT•	(SEQ ID NO: 455)
38	HUMIGHAR	:EQVGAPQPGKGNSGKASKGDSNRPDT•	(SEQ ID NO: 456)
39	HVJLSHUMAN	:QQVGGVQAGRANPGKDSRGISKRTET•	(SEQ ID NO: 457)
40	HV1ASHUMAN	:QQVAEVKPGKGTPGQQKQGESTRSET•	(SEQ ID NO: 458)
41	AJ2483	:QQVAEVKPGKGTPGQQKQGTSTRSET•	(SEQ ID NO: 459)
42	HUMIGHAY	:QQVAEVKPGKGTPGQQKQGTSARSET•	(SEQ ID NO: 460)
43	HUMIGHCU	:QQVAEVKPGKGTPGQQKQGTSIRSDT•	(SEQ ID NO: 461)
44	HUMIGHBS	:QQVAEVKPGKGTPGQEKQGTSIRSDT•	(SEQ ID NO: 462)
45	HUMIGVHLS	:QQVAEVKPGKGTPGQQNQGTSTRSDT•	(SEQ ID NO: 463)
46	HUMIGHBX	:QQVGEVKPGRGTPGQQKQGTSTRSDT•	(SEQ ID NO: 464)
47	HV1C$HUMAN	:QQVAEVKPGRGTPGHPRQGASTRSDS•	(SEQ ID NO: 465)
48	HJ4964	:QQVSELKPGKGTPGQQGTGTSVKAFT•	(SEQ ID NO: 466)
49	HUMIGHCY	:EQVAEVKPGKGSPGKPSQGKSIKAST•	(SEQ ID NO: 467)
50	PL0119	:EQVAEVKPGRGSPGKPSQGKSIKAST•	(SEQ ID NO: 468)

51	HV1FSHUMAN	:QQVAEVKPGRGDPGRPRQASSTISAT*	(SEQ ID NO: 469)
52	D34964	:EQVAEVPQGKGRPGKSLQGKSLKAST*	(SEQ ID NO: 470)
53	HV1DSHUMAN	:QQMAEVKPGRGTPGKPGVVPSFFSET*	(SEQ ID NO: 471)
54	HV1ESHUMAN	:QQVAEVKPGRGTPGRYIWEPSFFNEG*	(SEQ ID NO: 472)
55	JL0047	:QQQAGLKPSSGSPGKPSKSTSKTAAT*	(SEQ ID NO: 473)
56	HUMIGHBW	:QQQPGLKPSSGSPGKPSKSTSKTAAT*	(SEQ ID NO: 474)
57	E34964	:QQQPGLKPSSGSPGKPSKSTSNTAAT*	(SEQ ID NO: 475)
58	HUMIGHCW	:QQQPGLKPSSGSAGKPSKSTSKTAAT*	(SEQ ID NO: 476)
59	HV2FSHUMAN	:RQQPGLKPSSGPPGKPSRGTSRSAAT*	(SEQ ID NO: 477)
60	HV2ISHUMAN	:QQQAGLKPSSGSPGRTSKSTSKTAAT*	(SEQ ID NO: 478)
61	HV2GSHUMAN	:QQEPGLRPSSGTPGRTPRSTSKTAAT*	(SEQ ID NO: 479)
62	NSJFABH	:XQEPGLRPSSGSPGRTPRSTSKTAAT*	(SEQ ID NO: 480)
63	PS0091	:QQQPGLKPSSGSPSRVSKSTSKTPET*	(SEQ ID NO: 481)
64	HUMIGHDA	:QHQAGLKRSSGPPGKPSTSTSKTAAT*	(SEQ ID NO: 482)
65	A26555	:ZQESGLKPTSGSPGKPSKSRSKAADA*	(SEQ ID NO: 483)
66	HV2ESHUMAN	:QTKPTLKPTTGSPGRPSKSTSKDPVT*	(SEQ ID NO: 484)
67	HV2DSHUMAN	:QTKPTLKPTTGSPGKPSRSTSRDPVS*	(SEQ ID NO: 485)
68	AJ6005	:ETRPALKPTTGSPGKTSKTTSKDPVT*	(SEQ ID NO: 486)
69	HV2HSHUMAN	:QNRPALKATTGSPGKTSETTSKDPAT*	(SEQ ID NO: 487)
70	HV2ASHUMAN	:QTTPALKPKTGSPGKTSRTDSKNPVT*	(SEQ ID NO: 488)
71	HV2CSHUMAN	:QTRPALRPTTGSPGEASETTSKGPGT*	(SEQ ID NO: 489)
72	HV2BSHUMAN	:QTRPALKPTTGSPGKTSETTSRDTAY*	(SEQ ID NO: 490)
73	JL0049	:LEGVQLWGGRGISRKYAKGNGKRDES*	(SEQ ID NO: 491)

EXAMPLE 2

DETAILED DESCRIPTION OF METHOD FOR CONSTRUCTING THREE-DIMENSIONAL MODEL OF ANTIBODY VARIABLE REGION

The references cited in the text below are listed at the end of this Example.

The first antibody Fab structure was determined in 1972. Since then, no more than about twelve Fab structures have been published, a number that represents a very small fraction of the total antibody repertoire (>10^8 antibodies). To understand the molecular basis of this antibody diversity will require knowledge of either a large number of x-ray structures, or the rules by which combining site topography is governed. The development of such prediction rules has now reached the point where variable regions of antibodies can be modelled to an accuracy approaching that of the medium resolution x-ray structure.

The interaction of an antibody with its cognate antigen is one of the most widely accepted paradigms of molecular recognition. To understand the antibody-antigen interaction in atomic detail requires knowledge of the three-dimensional structure of antibodies and of their antigen complexes. Traditionally such information has come from x-ray crystallographic studies (see Davies et al. for review (Davies et al., 1988)).

The modelling of antibody combining sites was first attempted by Padlan & Davies (Padlan et al., 1976) at a time when very few antibody structures were known. Nonetheless, Padlan and colleagues recognized that the key lay in high structural homology that existed within the β-sheet framework regions of different antibody variable domains. The antigen combining site is formed by the juxtaposition of six interstrand loops, or CDRs (Complementarity Determining Regions) (Kabat et al., 1987), on this framework. If the framework could be modelled by homology then it might be possible to model the CDRs in the same way. Padlan and Davies (Padlan et al., 1976) reasoned that CDR length was the important determinant of backbone conformation though the number of antibody structures was insufficient to thoroughly test this maximum overlap procedure (MOP). This notion was not picked up again until the early 1980's when Pedersen and Rees proposed a similar approach to modelling antibody combining sites based on a more extensive analysis of antibody structures (de la Pas et al., 1986).

Those essentially knowledge-based procedures are best exemplified for antibodies by the work of Chothia & Lesk (Chothia et al., 1986) who, in 1986, extended and modified the MOP procedure by introducing the concept of "key" residues. These residues allow the further subdivision of CDRs of the same length into "canonical" structures which differ in having residues at specified positions that, through packing, hydrogen bonding or the ability to assume unusual values of the torsion angels ϕ, ψ and ω, determine the precise CDR conformation

(Chothia et al., 1989). Similar knowledge-based methods have been proposed for predicting loop conformations in general (Thornton et al., 1988; Tramontano et al., 1989). These methods rely on the crystallographic database of protein structures. However, none of the above knowledge-based methods has been totally successful. In particular, the MOP or canonical structure approaches have succeeded in modelling only five of the six CDRs. This stems from the fact that the third CDR of the heavy chain, H3, is more variable in sequence, length and structure than any of the other CDRs.

To deal with this problem several groups have attempted to use *ab initio* methods to model the combining site (Bruccoleri and Karplus, 1987). The requirement with such methods is that the total allowable conformational space accessible to a particular CDR is sampled. Typical of purely geometric approaches is that of Go & Sheraga (Go and Sheraga, 1970) and more recently Palmer & Sheraga (Palmer and Sheraga, 1991), where the problem is reduced to one in which the central region of the polypeptide backbone, having characteristic bond length and bond angles, is constructed between the end points of the loop (CDR if an antibody loop) by a "chain closure" algorithm. In a modification of this algorithm, Bruccoleri & Karplus (Bruccoleri and Karplus, 1987) introduced an energy minimization procedure which greatly expanded the domain of conformational space searched during the chain closure procedure. This modification is incorporated into the conformational search program CONGEN (Bruccoleri and Karplus, 1987), which also allows the user to choose any set of standard bond length and bond angels such as the CHARMM (Brooks et al., 1983) standard geometry parameter sets. Other approaches such as minimization (Moult and James, 1986), or molecular dynamics (Fine et al., 1986) either fail to saturate conformational space or are unable to deal with the problem of long CDRs. Whichever of the *ab initio* methods is employed however, the problem is one of defining the selection criteria in such a way as to allow the unambiguous identification of the *correct* structure (in this context *correct* is defined by reference to an appropriate X-ray structure) within the ensemble of candidates, for every CDR. To date this has not been possible.

Recently a more holistic approach has been taken to the modelling of CDRs which combines the advantages of knowledge-based and *ab initio* methods in a single algorithm known as CAMAL (Combined Algorithm for Modelling Antibody Loops) (Martin et al., 1989; Martin et al., 1991). Previously this algorithm has been used to model individual CDRs in the presence of the crystal structure conformations of the other five. As is demonstrated below, CAMAL is able to predict the backbone conformations of all six CDRs of the antibody combining site to an accuracy approaching that of medium resolution x-ray structures. In addition the algorithm includes a procedure for selecting and fitting together the light and heavy chain framework regions prior to generation of CDR conformations, thus making possible the prediction of the entire variable region. Furthermore a new Monte Carlo (MC) simulated annealing method has been developed for the determination of sidechain conformations.

The Framework Region

Antibody framework regions consist of conserved β-strands that form the β-barrel structure characteristic of immunoglobulin V-type regions. In the procedure described here each V-region is built from a database of known antibody structures, using sequence homology for selection of the light (L) and heavy (H) chain V-domains. The two domains are then paired by least squares fitting on the most conserved strands of the antibody β-barrel (Table 2 and Figures 5 & 6. The strand orientations were determined by analyzing the barrels of known antibody crystal structures. Eight antibodies were analyzed using a multiple structure fitting program as follows. Seven structures were fitted onto one of the set selected at random and mean coordinates were calculated. All eight structures were then fitted onto these mean coordinates and new mean coordinates determined. This procedure was iterated until the mean coordinate set converged (5-10 cycles). The variance for the mean coordinates at each barrel point (N,Cα,C) was calculated. In Figure 5 this variance is plotted against the projected positions of these points onto the conjugate axis of the barrel.

Strand 8 and all but two residues of strand 7 in both light and heavy chains were eliminated as they showed deviations greater than 3σ (standard deviation units) from the mean coordinates. These two strands comprised the takeoff points of CDR H3, and suggests that any knowledge-based prediction of CDR H3 would have to account not only for sequence and length variation in the CDR itself, but also for the position of the participating strands. The remaining mean coordinates were used as a scaffold onto which the L and H chains were fitted. Strands 7 and 8 in the final framework were obtained from the database structure used in the construction. The framework strands are marked + in the multialignment in Table 2.

The sidechains were then replaced using a 'maximum overlap' method, in which sidechain templates were fitted on backbone atoms with the sidechain torsion angles being adjusted to match those of equivalent torsions in the parent sidechain.

The Combining Site

The procedure for predicting the structure of combining sites combines a database search with a conformational search procedure. The architecture of the program suite to perform this task is outlined in Figure 7.

The database search utilizes distance constraints for each of the six CDR loops determined from known antibody structures. These constraints were determined by calculating Cα-Cα distances within known loops and using a search range of $\bar{x} + 3.5\sigma$ (the mean ± 3.5 standard deviation units). A database containing all the proteins in the Brookhaven Protein Databank (Bernstein et al., 1977) is then searched for fragments which satisfy the constraints for a loop of the required length. The middle section of the loop is then deleted and reconstructed using the conformational search program CONGEN (Bruccoleri and Karplus, 1987). For loops of six or seven residues, the structure database appears to saturate the conformational space available to the backbone adequately and only sidechains are built by conformational search. Loops shorter than six residues are built by conformational search alone since this is computationally feasible and the number of loops selected from the database becomes unacceptably large as loop length decreases.

When modelling a complete combining site, loops of 6 or more residues are modelled individually with the other loops absent. If the loops are built consecutively, small errors can accumulate leading to a poor result (Martin, 1990). All the loop conformations are then evaluated using a solvent modified potential, which excludes the attractive van der Waals and electrostatic terms of the non-bonded energy function contained within the GROMOS (Åqvist et al., 1985) potential. The lowest five energy conformations are selected and filtered using a "structurally determining residue" algorithm (FILTER), based on backbone torsion angles observed in the original database loops. Since the database search is not used for the shortest loops of 5 residues or fewer, the FILTER algorithm cannot be used. Energy is thus the only available selection criterion and the short loops are built last, in the presence of the longer loops.

Side Chains

The determination of sidechain positions was previously done using the iterative sidechain determination algorithm described by Bruccoleri et al. (Bruccoleri and Karplus, 1987). Unfortunately the CHARMM (Brooks et al., 1983) force field fails to select the correct conformations of exposed hydrophobic sidechains. There is no penalty for having an exposed uncharged atom, without solvent present. CONGEN is also unable to saturate the conformational space for a large number of sidechains (more than 6 residues).

Recently Lee et al. (Lee and Levitt, 1991; Lee and Subbiah, 1991) has proposed a method for searching conformational space for a large number of sidechains using MC simulated annealing. A simple energy function is used for the evaluation of conformations generated by a biased random walk:

$$E = \sum_{i=1}^{n} \epsilon_o \left(\left(\frac{r_o}{r}\right)^6 - 2\left(\frac{r_o}{r}\right)^{12} \right) + \kappa_o \cdot COS(3\omega)$$

Where the first term is a simple *Lennard-Jones* potential which evaluates the non-bonded contacts between the atoms in a given molecule, the second term is a simple torsional term which only applies to C-C bonds. The torsional term biases the function towards 60° rotamers. ϵ_o and κ_o are constants. The metropolis function:

$$P = C^{\frac{-\delta E}{T}}$$

is used to evaluate the energy function. Any move which results in a decrease in energy is accepted, and any move which results in a positive δE is only accepted with the probability P. This simple method can be used to search the large conformational space defined by a set of torsion angles in amino-acid sidechains, and find or define the global minimum which exist for a set of sidechains. T is the simulation temperature.

When searching sidechain conformations using this method the simulation system usually gets trapped in an energetic minima well before the global minimum is encountered, at a high temperature, without the solution space having been searched sufficiently. This problem can be solved by truncating the *Lennard-Jones* potential, thus allowing atoms to pass through each other. In reality this function would converge towards infinity when the distance r between the atoms approaches zero.

The evaluation of sidechain conformations generated is done solely on the basis of energy, for internal (core) residues, since good van der Waal's interactions are considered to be equal to a good packing of the sidechains. The situation becomes more complicated when trying to predict the conformation of surface residues. The lowest van der Waal's interaction is obtained by a combination of sidechain conformations which minimize the overlap of atoms, this means that the lowest energy is obtained with extended conformations of

sidechains, without considering good packing of sidechains.

Using the fact that hydrophobic, bulky residues will be shielded by the hydrophilic sidechains, and will be buried in the surface, it is possible to generate a simple function which will evaluate these macroscopic observations. These functions can either be implemented in the objective evaluation function of the Monte Carlo simulation, or as is done here, added as a post processing step. Including an accessibility/hydrophobicity term in the evaluation function would slow down the calculation considerably, hence the term has been added as a post processing function. The function used is a sum of the product of relative exposed surface area multiplied by the residual hydrophobicities. The hydrophobicities used are taken from Cornette et al. (Cornette et al., 1987).

$$f_{conformation} = \sum_{i=1}^{n} -A_{irel} \cdot H_{irel}$$

n is the number of sidechains reconstructed. The surface area is calculated using the tesselated icosahedron approach (Chau and Dean, 1987), which is not very precise (0.1 percent), but is able to evaluate a large number of conformations. The function is evaluated for the final 2,000 conformations and the lowest value conformation selected as the best.

Using this simple approach it is possible to integrate over a large phase space with many degrees of freedom, and get a complete sampling of the space.

Predicted Structures of an Anti-hapten, Anti-peptide and Two Anti-protein Antibodies

In the following section the predicted structures of four different antibody F_V regions are presented and analyzed. The antibodies are:

- Gloop-2 (Darsley and Rees, 1985), an anti-lysozyme antibody whose Fab structure was determined by Jeffrey et al., (Jeffrey et al., 1991) and which was used as a learning exercise during the development of CAMAL.
- D1.3 (Amit et al., 1986), an anti-lysozyme antibody whose uncomplexed F_V coordinates were supplied by R. Poljak et al. after the model coordinates had been deposited.
- 36-71 (Rose et al., 1990), an anti-phenylarsonate antibody whose Fab structure was carried out by D. R. Rose, et al., and whose coordinates were obtained after the model coordinates had been deposited.
- 3D6 (Grunow et al., 1988), an anti-protein (GP41 of HIV) antibody whose Fab structure was carried out by D. Carter et al. (Carter, 1991) and whose coordinates were obtained after the model coordinates had been deposited. For this antibody, the model was generated using the canonical loop method of Chothia & Lesk (Chothia et al., 1989; Chothia et al., 1986) for CDRs L1, L2, H1 and H2, while L3 and H3, which cannot be modelled using canonical structures, were constructed using CAMAL.

All four models were subjected to both restrained and unrestrained energy minimization using the DISCOVER (TM Biosym Technology) potential with 300 cycles of steepest descents, followed by conjugate gradient minimization until convergence to within 0.01 Kcal occurred.

The resolution and R-factors of the x-ray structures are given in Table 3 together with the parent frameworks selected in building the models. The structures and models were compared by global fits of the loops. The β-barrel strands 1 to 6, as described above, were least squares fitted and the RMS deviation was then calculated over the loops. The backbone (N,Cα,C) RMS values for fitting model and crystal structure frameworks were between 0.4 and 0.9 Å, illustrating the conservation of the core β-barrel. Using all eight strands RMS deviations between 0.6 and 1.2 Å were observed.

Global fits (Table 4) give a more realistic measure of the accuracy of the model than a local least-squares fit over the loops since they account for the overall positioning of the loops in the context of the F_V structure. Local fits, which give lower RMS deviations, are also shown in Table 4. Differences between local and global RMS deviations arise from differences in V_H/V_L domain packing and differences in loop 'take off' angles and positions.

Table 5 shows the canonical loops selected from modelling 3D6. Backbone structures of the modelled CDRs, superimposed on the x-ray structures after global fitting are shown in Figure 8. General features and points of interest for each of the six CDRs are discussed below.

Analysis of the CDR Regions

During the comparison of CDR conformations in the V-region models and the x-ray Fab structures it was observed that at certain positions in a CDR, the peptide backbone may adopt either of two conformations by undergoing a "peptide flip" (1,4 shift). This phenomenon is also seen in type 2 β-turns (Paul et al., 1990). Dynamics simulations of β-turns show that the transformation energy between $\phi1 = -00$, $\psi1 = -30$, $\phi2 = -90$, $\psi2 = 0$ and $\phi1 = -00$, $\psi1 = 120$, $\phi2 = 90$, $\psi2 = 0$ has a maximum value of 5 kcal (Paul et al., 1990). This is low enough to allow selection of either conformation. The peptide flip is observed within several canonical classes (as described by Chothia et al. (Chothia et al., 1989)) and the hydrogen bonding pattern used to determine the conformation of a canonical class does not disallow the peptide flip. Any modelling procedure should therefore take these, or any other multiple conformations, into consideration where the transformation energies are sufficiently low to permit population of the different conformational forms. Table 6 shows an example of the "peptide-flip" phenomenon from the crystallographic database of antibody structures. It should be noted that a single crystal structure will not show multiple conformations since the crystallization will 'freeze out' one of the conformations. During the modelling procedure the two populations of conformers are easily extracted from a set of *ab initio* generated loops, by using a torsional clustering algorithm.

CDR-L1

In Gloop-2 and D1.3, all five low energy conformations were very similar with RMS deviations differing by less than 0.25 Å (backbone) and 0.35 Å (all atoms). The FILTER algorithm was unable to distinguish between the conformations and the lowest energy structure was selected.

Although CDR-L1 of 3D6 was originally built using the canonical loop from HyHEL-10, the mid-section was rebuilt by conformational search, for the following reason. HyHEL-10 and REI CDR-L1 loops are placed in the same canonical ensemble (Chothia et al., 1989) although they contain a 1-4 shift (peptide flip) relative to one another between the fifth and eighth residues of the loop (residues 28-31) (see Table 6).

36-71 shows the same 1-4 shift between the model and crystal structure CDRs. Both crystal structure and model were compared with other loops of the same canonical class as defined by Chothia et al. (Chothia et al., 1989). It was found that the hydrogen bonding pattern which determines the conformation was conserved.

CDR-L2

CDR-L2 of D1.3 has two adjacent threonines (49, 50) which in the x-ray structure are packed against the tyrosine at the fourth position of CDR-H3, thus minimizing the exposed hydrophobic sidechains. In the unminimized model the threonine sidechains are exposed to the solvent, but after energy minimization, this packing is observed.

CDR-L3

In Gloop-2, D1.3 and 36-71 the proline at the seventh position in the loop is correctly predicted in the *cis* conformation. It has previously been suggested that the conformation of CDR-L3 is dictated by the presence of a proline in position 8 or 9 (Chothia et al., 1989) within the loop. 3D6 does not have a proline in either position. Only 7 out of 290 CDR-L3 sequences (Kabat et al., 1987) lack a proline at both positions and in all of the published x-ray structures this proline is present. This is an example of a situation where either a new canonical class may need to be defined or where the canonical rule breaks down altogether, and an alternative method must be employed.

The 3D6 L3 loop is 7 residues in length and was built using database loops alone where conformational space is saturated by means of fragments selected from the crystallographic database (Global RMS: 2.01 Å, N,Cα,C), and by using CAMAL (Construction: Q[Q(YNS)Y]S, Global RMS: 1.97 Å, N,Cα,C). The similarity of the structures generated by the two procedures illustrates the utility of the database search and suggests that, for shorter loops it is capable of saturating the available conformational space.

CDR-H1

Using the Kabat and Wu definition of CDR-H1 places this loop as an extension of the β-sheet. The extended nature of this stretch of peptide limits its conformational flexibility and CDR-H1 is generally modelled accurately (Martin et al., 1989; Chothia et al., 1989).

In Gloop-2 and D1.3, the Phe or Tyr sidechain at the second position in the loop is poorly placed and packs against Leu at the penultimate position in HFR1 (see Table 2). 36-71 has a well-placed Asn at this position, rather than the more common bulky hydrophobic sidechain.

CDR-H2

CDR-H2 of 36-71 is similar in sequence to F19.9 (Strong et al., 1991), (36-71: YNNPGNGYIA (SEQ ID NO:492); F19.9: YINPGKGYLS (SEQ ID NO:493)). While the structurally determining residues specified by Chothia and Lesk (Chothia et al., 1989) are conserved, the backbone conformations are different: F19.9 has a bulge at the -PGN- Gly, compared with 36-71, giving the loop a 'kink' in the middle. The model of 36-71 shows a 1-4 shift, though the sidechains are still well placed.

In Gloop-2, the all atom RMS deviation is poor (3.00 Å) (Jeffrey et al., 1991) when compared with the $P2_1$ crystal structure, owing to rotations of the Phe at position 3 in the loop and Tyr at position 10 by approximately 120° about the χ_2 torsion angle. Gloop-2 has been solved in two different crystal forms, $P2_1$ and P1 (Jeffrey et al., 1991; Jeffrey, 1989). When compared with the P1 structure, the sidechains are placed almost perfectly and the all atom RMS (global fit) drops to 2.23 Å.

This concerted sidechain motion between crystal forms illustrates the effects of crystallization conditions on surface sidechain placement. Even though surface sidechains may show low temperature factors indicating low mobility in the crystal, their mobility in solution may be high. In the Gloop-2 P1 structure, the mean sidechain temperature factor for the F_V domain is 13.46 (σ = 8.20) while the sidechains of these two residues of H2 show mean temperature factors of 5.56 (σ = 0.68) for the Phe at position 3 and 7.10 (σ = 1.73) for the Tyr at position 10.

CDR-H3

CDR-H3 is the most variable of the six CDR's with all lengths up to 21 residues being represented in Kabat et al., (Kabat et al., 1987). This extreme variability results from V-D-J splicing (Schilling et al., 1980) and has always been a problem when attempting to model antibodies. Such loops may be divided into short (up to 7 residues), medium (up to 14 residues) and long (15 or more residues). Using the CAMAL procedure, short and medium CDR-H3's can be modelled as accurately as other CDR's of similar lengths. Although long CDR-H3's are more difficult and cannot, at present, be built to the same accuracy, the chain trace is still correct.

It is unlikely that the longer loops consist of 'pure' loops (i.e., all random coil or turn). In crystal structures of antibodies with medium to long CDR-H3 loops (McPC603 (Rudikoff et al., 1981): 11 amino acids (aa); KOL (Marquart et al., 1980): 17 aa; F19.9 (Lascombe et al., 1989): 15 aa) the loops consist of a disordered β-sheet extension from the β-barrel core and a 5-8 residue random coil/turn connecting these two strands.

To determine the nature of medium to long loops (>8 residues) which satisfy the CDR-H3 constraints, a complete search of the Protein Databank for loops of length 8-20 residues, was performed using the inter-Cα distance constraints determined from known antibody crystal structures for CDR-H3. The resulting loops were then analyzed using the DSSP (Kabsch and Sander, 1983) program, which is able to assign secondary structure to polypeptide structures. The amount of secondary structure for each length of loop was calculated, and it was observed that for loops longer than 12 residues the amount of secondary structure within each of the classes described in DSSP was constant. The number of loops selected is also constant (approximately 150 loops) for loops longer than 12 residues. A closer inspection of each of the length ensembles shows indeed that the loops are the same between the groups.

This analysis shows that, like the long CDR-H3 crystal structures, the selected fragments consist of β-strands connected by 5-8 residue loops. For loops above 12-13 residues in length, the same loops are selected, but with extensions to the β-strands. This is called the "sliding-ladder" effect. In addition, the maximum size of a random coil or turn fragment in any of the structures contained in the Protein Databank tends not to exceed 8 residues, as determined by DSSP. This implies that the conformational space of longer loops is not saturated by the database and, although it is unlikely that long loops in antibodies will differ significantly from long loops in other structures, confidence in the prediction must be correspondingly reduced.

By how much is the usefulness of the CAMAL algorithm reduced by this observation ?

The frequency of occurrence of different CDR-H3 lengths in antibody sequences described by Kabat et al. (Kabat et al., 1987) was analyzed. Figure 10 shows that more than 85% of H3 loops have lengths between 4 and 14 residues which can be modelled accurately by the CAMAL algorithm.

CDR-H3 of D1.3 is of average length (8 residues), though no loops of this length are seen in the available antibody structures. The crystal structure coordinate set showed an RMS of 1.9 Å compared with the model.

The 36-71 loop is 12 residues long. The conformation is correctly predicted as a short loop connecting an

extension of the β-sheet.

The 3D6 H3 loop is 17 residues long. While KOL (Marquart et al., 1980) has the same length it has only one residue in common with 3D6 and only one conservative mutation. There is thus no reason to believe that the conformations would be similar. The final predicted conformation of 3D6 is an extended β-sheet, as in the crystal structure. The difference between the predicted and the crystal structure of 3D6-H3 is due to a twist of 5-7° in the extended β-sheet loop (see Figures 9A-9D). Such a twist has also been observed for complexed and uncomplexed antibodies by Wilson et al. (Wilson and others). This suggests that long CDR-H3 loops may be flexible and actively involved in antigen binding.

The Complete Variable Region

Prediction of the strand positions and V_L-V_H orientation in the framework β-barrel was exact for all of the four antibodies. The backbone (N,Cα,C) RMS deviations from the crystal structures were between 0.56 and 0.86 Å, despite the fact that, in all cases the V_L and V_H regions of a particular model were derived from different antibody structures. This suggests that this method will do well in procedures such as humanization (Gorman et al., 1991), where correct framework positioning is important. The backbones of all six CDRs in all four antibodies are essentially correctly predicted, as shown in Figure 8. There are two important points to make about these predictions. First, the position of each CDR on its framework barrel is correct. Thus, CDR-framework interactions can be confidently monitored. The only deviation from the x-ray structure is CDR-H3 of antibody 3D6 which has been previously discussed. Second, the all atom RMS deviation between models and x-ray structures is dominated by sidechain positions. In most instances this deviation is due to a small number of incorrectly positioned, exposed sidechains (for example, in D1.3 the only sidechains which are incorrectly predicted are Tyr 9 of L1, Trp 4 of L3, Tyr 2 of H1 and Tyr 4 of H3). Since each CDR is constructed in the absence of other CDRs, the force field may choose a rotamer which is 120° away from that found in the crystal structure. This effect has also been observed by Lee et al. (Lee and Levitt, 1991).

Conclusion

For antibodies having CDR H3 regions of 14 residues or less the complete variable domain can be modelled to an accuracy approaching that of medium resolution x-ray structures. For antibodies with longer H3 loops the CAMAL algorithm is likely to need an additional procedure in which molecular dynamics simulations are also incorporated.

The canonical approach of Chothia et al. appears to work well (at least in modelling backbones) where it may be applied and may be used successfully in combination with the CAMAL procedure.

One important observation that has emerged from these studies is that a given loop can exist in several conformations. In particular, this seems likely for CDR-L1 and, to a lesser extent, CDR-L3 and longer CDR-H3's. A simple combinatorial calculation shows that, if each of these three loops can exist in three separate conformations, a given combining site can have $3^3 = 27$ different topographies. Clearly, this would explain the origins of cross reactivity and would allow for induced fit of antigens.

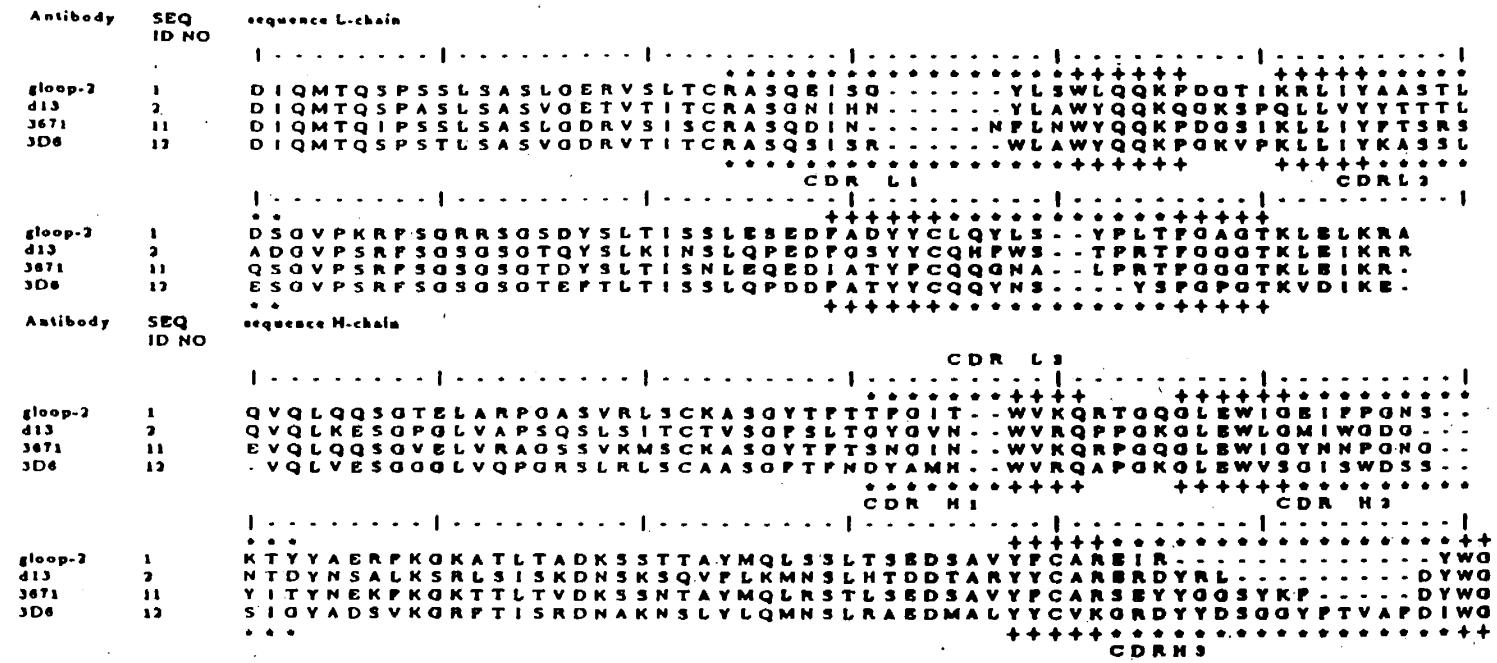


Table 2: Alignment of antibody sequences used in the modelling. '*' indicates CDR regions; '+' indicates β-strand regions used in the fitting for modelling frameworks. Nomenclature for β-barrel strands is (H or L - Chain) - FR(Framework region)-(Strand number), thus for example strand one of the heavy chain becomes HFR1.

Antibody	Resolution	R-factor	Framework Model	
			Light	Heavy
Gloop-2	2.80	21.2	REI	HyHEL-5
D1.3	-	-	REI	NEW
36-71	1.90	20.9	Gloop2	NEW
3D6	2.70	17.7	REI	KOL

Table 3: Details of the antibody crystal structures against which the models were compared and the parent frameworks used to build the models. Resolution data for D1.3 has not yet been published.

Antibody	CDR	sequence	SEQ ID NO	RMS local (Å)				RMS global (Å)			
				Cα	N,Cα,C	All CB	All MC	Cα	N,Cα,C	All CB	All MC
Gloop-2	L1	RAS[Q(EIS)G]YLS	496	0.73	0.71	2.08	1.96	0.66	0.87	2.09	2.12
D1.3		RAS[G(NIH)N]YLA	495	2.29	1.93	4.34	3.84	2.72	2.43	4.89	4.82
36-71		RAS[Q(DIN)N]FLN	496	2.71	2.43	4.99	4.59	3.81	3.31	5.19	5.47
3D6		RAS[Q(SIG)N]NLH	497	0.81	0.84	2.48	1.62	0.61	0.76	2.88	1.98
Gloop-2	L2	AASTLDS	498	0.26	0.23	0.80	1.00	0.66	0.68	1.10	1.10
D1.3		Y[T(TTL)A]D	499	0.67	0.73	1.88	1.48	0.99	1.02	2.01	1.98
36-71		F[T(SRS)Q]S	500	0.44	0.46	2.34	2.23	0.73	0.73	2.48	2.40
3D6		KASSLES	501	0.41	0.42	1.37	1.20	0.63	0.66	1.78	1.80
Gloop-2	L3	LQ[Y(LSY)P]LT	502	0.36	0.82	1.73	1.60	0.78	0.74	2.00	1.99
D1.3		QH[F(WST)P]RT	503	1.41	1.38	2.89	2.96	1.76	1.79	3.46	3.26
36-71		QQ[G(NAL)P]RT	504	1.08	1.08	3.26	3.18	1.48	1.36	3.37	3.26
3D6		Q[Q(YNS)Y]S	505	1.48	1.88	3.84	3.99	2.31	1.97	3.96	3.99
Gloop-2	H1	[T(FGI)T]	506	0.60	0.70	2.00	1.68	1.03	1.01	2.04	2.09
D1.3		[G(YGV)N]	507	0.44	0.62	2.38	2.00	0.83	0.90	2.24	2.08
36-71		[S(NGI)N]	508	0.90	0.83	2.22	1.96	1.04	0.97	2.81	2.23
3D6		DYAMH	509	0.67	0.77	1.62	1.11	0.81	0.72	1.69	1.20
Gloop-2	H2	EI[F(PGN)S]KTY	510	0.62	0.64	1.68	1.78	1.20	0.64	2.23	2.18
D1.3		MI[W(GDG)N]TD	511	0.62	0.62	1.56	1.48	0.87	0.88	1.68	1.68
36-71		YNN[P(GNG)Y]IA	512	0.84	0.78	2.01	2.20	1.47	1.41	1.79	1.98
3D6		ISWDSSSIG	513	0.68	0.82	2.38	2.08	0.98	0.89	2.88	2.10
Gloop-2	H3	[R(EIR)Y]	514	0.66	0.89	3.44	3.90	0.87	1.07	3.68	4.16
D1.3		ER[D(YRL)D]Y	515	0.36	0.83	1.66	1.28	1.26	0.81	1.96	1.83
36-71		SEYY[G(GSY)K]FDY	516	1.98	1.73	6.48	6.00	2.68	2.83	6.60	6.09
3D6		GRDYY[D(SGG)YF]TVAFDI	517	3.66	3.42	8.93	6.01	6.38	3.96	6.30	6.20

Table 4: Sequence and conformational search construction scheme for each of the 24 CDRs, [ ]=construction area, ( )= Chain closure, all sidechains are constructed. RMS(Root Mean Square) difference between model and crystal structure loop coordinates. The RMS values are a global fit calculated by least-squares fitting the conserved core of the two structures upon each other and calculating the RMS over the loops. The total RMS of the frameworks (N,Cα,C) is 0.81, 0.60, 0.86 and 0.56 respectivly

Loop	Canonical	Sequence	SEQ ID NO
L1	HyHEL-10	R A S Q S I S R W L A	518
	(3D6)	R A S Q S I G N N L H	497
L2	REI	E A S N D L A	519
	(3D6)	K A S S L E S	501
H1	McPC603	D F Y M E	520
	(3D6)	D Y A M H	509
H2	KOL	I I W D D G S D Q	521
	(3D6)	I S W D S S S I G	513

Table 5: Canonical loops selected for the model of 3D6(taken from Chothia *et al* (1989)).

	Residue Number	24	25	26	27	28*	29*
REI	Sequence	Q	A	S	Q	S	I
	ϕ/ψ	-/138	-103/157	-96/7	-158/142	-40/108	-112/9
HyHEL-10	Sequence	R	A	S	Q	S	I
	ϕ/ψ	-/108	-85/135	-88/64	172/160	-64/-38	9/63
	Residue Number	30*	31*	32	33	32	
REI	Sequence	I	K	Y	L	N	SEQ ID NO: 522
	ϕ/ψ	79/-77	-146/21	-104/89	-143/133	-144/-	
HyHEL-10	Sequence	G	N	N	L	H	SEQ ID NO: 518
	ϕ/ψ	-63/107	85/-15	-105/72	-129/118	-126/-	

Table 6: Backbone ϕ and ψ angles of residues in CDR-L1 from HyHEL-10 and REI classified in the same canonical group by Chothia *et al* (1989). The residues exhibiting a peptide flip are indicated by a *.

M.J. Darsley, P de al Paz, D.C. Phillips and A.R. Rees in Methodological Surveys in Biochemistry and Analysis, pages 63-68, Volume 15, 1985, Plenum Press (Eds. E. Reid, G.M.W. Cook and D.J. Morre), Presented at the Ninth International Subcellular Methodology Forum, September 3-6, 1984, Guildford, UK.

Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J. (1986). The Three-dimensional Structures of an Antigen-antibody Complex at 2.8 A Resolution. Science **233**, pp. 747-753.

Åqvist, J., van Gunsieren, W., Leifonmark, M. and Tapia, O. (1985), J. Mol. Biol. **183**, pp. 461-477.

Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977), J. Mol. Biol. **112**, pp. 535-542.

Brooks, B., Bruccoleri, R., Olaison, B., Statcs, D., Swaminathan, S. and Karplus, M. (1983), J. Comp. Chem. **4**, pp. 187-217.

Bruccoleri, R.E. and Karplus, M. (1987), Prediction of the Folding of Short Polypeptide Segments by Uniform Conformational Sampling. Biopolymers **26**, pp. 137-168.

Carter, D. et al. (1991). Protein Engineering, p. 9999.

Chau, P. and Dean P. (1987). Molecular Recognition: 3d Surface Structure Comparison by Gnomonic Projection. J. Mol. Graph. **5**, pp. 97-100.

Chothia, C., Lesk, A., Levitt, M., Amit, A., Mariuzza, R., Phillips, S. and Poljak, R. (1986). The Predicted Structure of Immunoglobulin D1.3 and its Comparison with the Crystal Structure. Science **233**, pp. 755-758.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D.R., Tulip, W.R., Colman, P.M., Alzri, P.M. and Poljak, R.J. (1989). Conformations of Immunoglobulin Hypervariable Regions. Nature (London) **342**, pp. 877-883.

Cornette, J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.A. and Delisi, C. (1987). Hydrophobicity Scales and Computational Techniques for Detecting Amphipatic Structures in Proteins. Journal of Molecular Biology **195.3**, pp. 659-685.

Darsley, M. and Rees, A. (1985), EMBO J. **4**, pp. 383-392.

Davies, D., Sheriff, S. and Padlan, E. (1988). Antibody Antigen Complexes. J. Biol. Chem. **263**, pp. 10541-10544.

de la Paz, P., Sutton, B., Darsly, M. and Rees, A. (1986). Modelling of the Combining Sites of Three Anti-lysozyme Monoclonal Antibodies and of the Complex Between One of the Antibodies and Its Epitope. EMBO J. **5**, pp. 415-425.

Fine, R., Wang, H., Shenkin, P., Yarmush, D. and Levinthal, C. (1986). Predicting Antibody Hypervariable Loop Conformations ii: Minimization and Molecular Dynamics Studies of McPC603 from Many Randomly Generated Loop Conformations. Proteins: Struct., Funct., Genet. **1**, pp. 342-362.

Go, N. and Sheraga, H. (1970). Ring Closure and Local Conformational Deformations of Chain Molecules. Macromolecules **3**, pp. 178-187.

Gorman, S., Clark, M., Rutledge, E., Cobbold, S. and Waldman, H. (1991). Reshaping a Therapeutic CD4 Antibody. Proc. Natl. Acad. Sci. U.S.A. **88**, pp. 4181-4185.

Grunow, R., Jahn, S., Porstman, T., Kiessig, T., Steinkeller, H., Steindl, F., Mattanovich, D., Gurtler, L., Deinhardt, F., Katinger, H. and von R., B. (1988). The High Efficiency, Human B Cell Immortalizing Heteromyeloma cb-f7. J. Immunol. Meth. **106**, pp. 257-265.

Jeffrey, P. (1989). The Structure and Specificity of Immunoglobulins. D. Phil. Thesis, University of Oxford.

Jeffrey, P.D., Gricst, R.E., Taylor, G.L. and Rees, A.R. (1991). Crystal Structure of the Fab Fragment of the Anti-peptide Antibody Gloop-2 and 2.8 Å. Manuscript in Preparation.

Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, B.M. and Gottesman, K.S. (1987). Sequences of Proteins of Immunological Interest. U.S. Department of Health and Human Services, Fourth Edition.

Kabsch, W. and Sander, C. (1983). Dictionary of Protein Secondary Structure. Biopolymers **22**, pp. 2577-2637.

Lascombe, M., Alzari, P., Boulot, G., Salujian, P., Tougard, P., Berek, C., Haba, S., Rosen, E., Nisonof, A. and Poljak, R. (1989). Three-dimensional Structure of Fab r19.9, A Monoclonal Murine Antibody Specific for the p-azobenzenearsonate Group. Proc. Natl. Acad. Sci. U.S.A. **86**, p. 607.

Lee, C. and Levitt, M. (1991). Accurate Prediction of the Stability and Activity Effects of Site-directed Mutagenesis on a Protein Core. Nature **352.6334**, pp. 448-451.

Lee, C. and Subbiah, S. (1991). Prediction of Protein Side-chain Conformation by Packing Optimization. Journal of Molecular Biology **217.2**, pp. 373-388.

Marquart, M., Deisenhofer, J. and Huber, R. (1980), Crystallographic Refinement and Atomic Models of the Intact Immunoglobulin Molecule KOL and Its Antigen-binding Fragment at 3.0 Å and and Resolution. J. Mol. Biol. **141**, pp. 369-391.

Martin, A.C.R. (1990). Molecular Modelling of Antibody Combining Sites. D. Phil. Thesis, University of Oxford.

Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1989). Modelling Antibody Hypervariable Loops: A Combined Algorithm. Proc. Natl. Acad. Sci. U.S.A. **86**, pp. 9268-9272.

Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1991). Modelling Antibody Hypervariable Loops using a 'Combined Algorithm'. Meth. Enz. In press.

Moult, J. and James, N. (1986). Proteins: Struct., Funct., Genet. **1**, p. 146.

Padlan, E., Davies, D., Pecht, I., Givol, D. and Wright, C. (1976). Model Building Studies of Antigen-binding Sites: The Hapten-Binding Site of MOPC-315. Cold Spring Harbor Quant. Symp. Biochem. **41**, pp. 627-637.

Palmer, K. and Sheraga, J. (1991). Standard-geometry Chains Fitted to X-ray Deviated Structures: Validation of the Rigid-geometry Approximation. I. Chain Closure through a Limited Search of Loop Conformations. J. Comp. Chem. **12**, pp. 505-526.

Paul, P., Burney, P., Campbell, M. and Odguthorpe, D. (1990). The Conformational Preferences of γ-lactam and Its Role in Constraining Peptide Structure. J. Comp.-aided. Mol. Des. **4**, pp. 239-253.

Rose, D.R., Strong, R.K., Margolis, M.N., Gefter, M.L. and Petsko, G.A. (1990). Crystal Structure of the Antigen-binding Fragment of the Murine Anti-arsonate Monoclonal Antibody 36-71 at 2.9 Å Resolution. Proc. Natl. Acad. Sci. U.S.A. **87**, pp. 338-342.

Rudikoff, S., Satow, Y., Padlan, E.A., Davies, D.R. and Potter, M. (1981). Kappa Chain Structure from a Crystallized Murine Fab': The Role of the Joining Segment in Hapten Binding. Mol. Immunol. **18**, pp. 705-711.

Schilling, J., Clevinger, B., Davie, J.M. and Hood, L. (1980). Amino Acid Sequence of Homogeneous Antibodies to Dextran and DNA Rearrangements in Heavy Chain V-region Gene Segments. Nature (London) **283**, pp. 35-40.

Strong, R., Campbell, R., Rose, D., Petsko, G., Sharon, J. and Margolies, M. (1991). Three-dimensional Structure of Murine Anti-p-azophenylarsonate Fab 36-71.1, X-ray Crystallography, Site-directed Mutagenesis, and Modeling of the Complex with Hapten. Biochemistry **30**, pp. 3739-3748.

Thornton, J., Sibanda, B., Edwards, M. and Barlow, D. (1988). Analysis, Design and Modification of Loop Regions in Proteins. BioEssays **8**, pp. 63-69.

Tramontano, A. Chothia, C. and Lesk, A. (1989). Structural Determinants of the Conformations of Medium-

sized Loops in Proteins. Proteins: Struct., Funct., Genet. 6, pp. 382-394.
Wilson, I. et al., Presented at Structure and Function Meeting in Honour of Sir David Phillips, 1-3 July, 1991, Oxford, UK.

SEQUENCE LISTING

GENERAL INFORMATION

(i) APPLICANT: PEDERSEN, Jan T.
SEARLE, Stephen M.J.
REES, Anthony R.
ROGUSKA, Michael A.
GUILD, Braydon C.

(ii) TITLE OF INVENTION: SURFACE RESIDUE VENEERING OF RODENT ANTIBODIES

(iii) NUMBER OF SEQUENCES: 522

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
(B) STREET: 2100 Pensylvania Avenue, N.W.
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: United States
(F) ZIP: 20037-3202

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: HP 9000/700 Workstation
(C) OPERATING SYSTEM: UNIX
(D) SOFTWARE: In house

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 07/942,245
(B) FILING DATE: 09-SEP-1992
(C) CLASSIFICATION:

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (202) 293-7060
(B) TELEFAX: (202) 293-7860
(C) TELEX: 6491103

(1) INFORMATION FOR SEQ ID NO:1

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Leu Gly
1 5 10 15

Glu Arg Val Ser Leu Thr Cys Arg Ala Ser Gln Glu Ile Ser Gly Tyr
20 25 30

Leu Ser Trp Leu Gln Gln Lys Pro Asp Gly Thr Ile Lys Arg Leu Ile
35 40 45

Tyr Ala Ala Ser Thr Leu Asp Ser Gly Val Pro Lys Arg Phe Ser Gly
50 55 60

Arg Arg Ser Gly Ser Asp Tyr Ser Leu Thr Ile Ser Ser Leu Glu Ser
65 70 75 80

Glu Asp Phe Ala Asp Tyr Tyr Cys Leu Gln Tyr Leu Ser Tyr Pro Leu
85 90 95

Thr Phe Gly Ala Gly Thr Lys Leu Glu Leu Lys Arg Ala
100 105

(2) INFORMATION FOR SEQ ID NO:2

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Asp Ile Gln Met Thr Gln Ser Pro Ala Ser Leu Ser Ala Ser Val Gly
1 5 10 15

Glu Thr Val Thr Ile Thr Cys Arg Ala Ser Gly Asn Ile His Asn Tyr
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Gln Gly Lys Ser Pro Gln Leu Leu Val
35 40 45

Tyr Tyr Thr Thr Thr Leu Ala Asp Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Gln Tyr Ser Leu Lys Ile Asn Ser Leu Gln Pro
65 70 75 80

Glu Asp Phe Gly Ser Tyr Tyr Cys Gln His Phe Trp Ser Thr Pro Arg
85 90 95

Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Arg Arg
100 105

(3) INFORMATION FOR SEQ ID NO:3

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Asp Ile Val Leu Thr Gln Ser Pro Ala Ile Met Ser Ala Ser Pro Gly
1 5 10 15

Glu Lys Val Thr Met Thr Cys Ser Ala Ser Ser Ser Val Asn Tyr Met
20 25 30

Tyr Trp Tyr Gln Gln Lys Ser Gly Thr Ser Pro Lys Arg Trp Ile Tyr
35 40 45

Asp Thr Ser Lys Leu Ala Ser Gly Val Pro Val Arg Phe Ser Gly Ser
50 55 60

Gly Ser Gly Thr Ser Tyr Ser Leu Thr Ile Ser Ser Met Glu Thr Glu
65 70 75 80

Asp Ala Ala Glu Tyr Tyr Cys Gln Gln Trp Gly Arg Asn Pro Thr Phe
85 90 95

Gly Gly Gly Thr Lys Leu Glu Ile Lys Arg Ala
100 105

(4) INFORMATION FOR SEQ ID NO:4

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asp Ile Val Leu Thr Gln Ser Pro Ala Thr Leu Ser Val Thr Pro Gly
1 5 10 15

Asn Ser Val Ser Leu Ser Cys Arg Ala Ser Gln Ser Ile Gly Asn Asn
20 25 30

Leu His Trp Tyr Gln Gln Lys Ser His Glu Ser Pro Arg Leu Leu Ile
35 40 45

Lys Tyr Ala Ser Gln Ser Ile Ser Gly Ile Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Asn Ser Val Glu Thr
65 70 75 80

Glu Asp Phe Gly Met Tyr Phe Cys Gln Gln Ser Asn Ser Trp Pro Tyr
85 90 95

Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Arg Ala
100 105

(5) INFORMATION FOR SEQ ID NO:5

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 108 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Glu Ile Val Leu Thr Gln Ser Pro Ala Ile Thr Ala Ala Ser Leu Gly
1 5 10 15

Gln Lys Val Thr Ile Thr Cys Ser Ala Ser Ser Ser Val Ser Ser Leu
20 25 30

His Trp Tyr Gln Gln Lys Ser Gly Thr Ser Pro Lys Pro Trp Ile Tyr
35 40 45

Glu Ile Ser Lys Leu Ala Ser Gly Val Pro Ala Arg Phe Ser Gly Ser
50 55 60

Gly Ser Gly Thr Ser Tyr Ser Leu Thr Ile Asn Thr Met Glu Ala Glu
65 70 75 80

Asp Ala Ala Ile Tyr Tyr Cys Gln Gln Trp Thr Tyr Pro Leu Ile Thr
85 90 95

Phe Gly Ala Gly Thr Lys Leu Glu Leu Lys Arg Ala
100 105

(6) INFORMATION FOR SEQ ID NO:6

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 112 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Glu Ser Val Leu Thr Gln Pro Pro Ser Ala Ser Gly Thr Pro Gly Gln
1 5 10 15

Arg Val Thr Ile Ser Cys Thr Gly Thr Ser Ser Asn Ile Gly Ser Ile
20 25 30

Thr Val Asn Trp Tyr Gln Gln Leu Pro Gly Met Ala Pro Lys Leu Leu
35 40 45

Ile Tyr Arg Asp Ala Met Arg Pro Ser Gly Val Pro Thr Arg Phe Ser
50 55 60

Gly Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile Ser Gly Leu Glu
65 70 75 80

Ala Glu Asp Glu Ser Asp Tyr Tyr Cys Ala Ser Trp Asn Ser Ser Asp
85 90 95

Asn Ser Tyr Val Phe Gly Thr Gly Thr Lys Val Thr Val Leu Gly Gln
100 105 110

(7) INFORMATION FOR SEQ ID NO:7

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 115 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu Ser Val Ser Ala Gly
1 5 10 15

Glu Arg Val Thr Met Ser Cys Lys Ser Ser Gln Ser Leu Leu Asn Ser
20 25 30

Gly Asn Gln Lys Asn Phe Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
35 40 45

Pro Pro Lys Leu Leu Ile Tyr Gly Ala Ser Thr Arg Glu Ser Gly Val
50 55 60

Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr
65 70 75 80

Ile Ser Ser Val Gln Ala Glu Asp Leu Ala Val Tyr Tyr Cys Gln Asn
85 90 95

Asp His Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Glu Ile
100 105 110

Lys Arg Ala
115

(8) INFORMATION FOR SEQ ID NO:8

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 103 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Ser Val Leu Thr Gln Pro Pro Ser Val Ser Gly Ala Pro Gly Gln Arg
1 5 10 15

Val Thr Ile Ser Cys Thr Gly Ser Ser Ser Asn Ile Gly Ala Gly Asn
20 25 30

His Val Lys Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu Leu
35 40 45

Ile Phe His Asn Asn Ala Arg Phe Ser Val Ser Lys Ser Gly Ser Ser
50 55 60

Ala Thr Leu Ala Ile Thr Gly Leu Gln Ala Glu Asp Glu Ala Asp Tyr
65 70 75 80

Tyr Cys Gln Ser Tyr Asp Arg Ser Leu Arg Val Phe Gly Gly Gly Thr
85 90 95

Lys Leu Thr Val Leu Arg Gln
100

(9) INFORMATION FOR SEQ ID NO:9

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly
1 5 10 15

Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser
20 25 30

Gln Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser
35 40 45

Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
50 55 60

Asn Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
65 70 75 80

Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser
85 90 95

Thr His Val Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys
100 105 110

Arg Ala

(10) INFORMATION FOR SEQ ID NO:10

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Asp Ile Gln Met Thr Gln Thr Thr Ser Ser Leu Ser Ala Ser Leu Gly
 1               5                  10                  15

Asp Arg Val Thr Ile Ser Cys Arg Ala Ser Gln Asp Ile Ser Asn Tyr
            20                  25                  30

Leu Asn Trp Tyr Gln Gln Lys Pro Asp Gly Thr Val Lys Leu Leu Val
        35                  40                  45

Tyr Tyr Thr Ser Arg Leu His Ser Gly Val Pro Ser Arg Phe Ser Gly
    50                  55                  60

Ser Gly Ser Gly Thr Asp Tyr Ser Leu Thr Ile Ser Asn Leu Glu His
65                  70                  75                  80

Glu Asp Ile Ala Thr Tyr Phe Cys Gln Gln Gly Ser Thr Thr Pro Arg
                85                  90                  95

Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Arg Arg
            100                 105
```

(11) INFORMATION FOR SEQ ID NO:11

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Asp Ile Gln Met Thr Gln Ile Pro Ser Ser Leu Ser Ala Ser Leu Gly
 1               5                  10                  15

Asp Arg Val Ser Ile Ser Cys Arg Ala Ser Gln Asp Ile Asn Asn Phe
            20                  25                  30

Leu Asn Trp Tyr Gln Gln Lys Pro Asp Gly Thr Ile Lys Leu Leu Ile
        35                  40                  45

Tyr Phe Thr Ser Arg Ser Gln Ser Gly Val Pro Ser Arg Phe Ser Gly
    50                  55                  60
```

Ser Gly Ser Gly Thr Asp Tyr Ser Leu Thr Ile Ser Asn Leu Glu Gln
65 70 75 80

Glu Asp Ile Ala Thr Tyr Phe Cys Gln Gln Gly Asn Ala Leu Pro Arg
85 90 95

Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Arg Ala
100 105

(12) INFORMATION FOR SEQ ID NO:12

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Asp Ile Gln Met Thr Gln Ser Pro Ser Thr Leu Ser Ala Ser Val Gly
1 5 10 15

Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Ser Ile Ser Arg Trp
20 25 30

Leu Ala Trp Tyr Gln Gln Lys Pro Gly Lys Val Pro Lys Leu Leu Ile
35 40 45

Tyr Lys Ala Ser Ser Leu Glu Ser Gly Val Pro Ser Arg Phe Ser Gly
50 55 60

Ser Gly Ser Gly Thr Glu Phe Thr Leu Thr Ile Ser Ser Leu Gln Pro
65 70 75 80

Asp Asp Phe Ala Thr Tyr Tyr Cys Gln Gln Tyr Asn Ser Tyr Ser Phe
85 90 95

Gly Pro Gly Thr Lys Val Asp Ile Lys Arg Thr
100 105

(13) INFORMATION FOR SEQ ID NO:13

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Gln Val Gln Leu Gln Gln Ser Gly Thr Glu Leu Ala Arg Pro Gly Ala
 1               5                  10                  15

Ser Val Arg Leu Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Thr Phe
            20                  25                  30

Gly Ile Thr Trp Val Lys Gln Arg Thr Gly Gln Gly Leu Glu Trp Ile
        35                  40                  45

Gly Glu Ile Phe Pro Gly Asn Ser Lys Thr Tyr Tyr Ala Glu Arg Phe
    50                  55                  60

Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Thr Thr Ala Tyr
65                  70                  75                  80

Met Gln Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
                85                  90                  95

Ala Arg Glu Ile Arg Tyr Trp Gly
            100
```

(14) INFORMATION FOR SEQ ID NO:14

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Gln Val Gln Leu Lys Glu Ser Gly Pro Gly Leu Val Ala Pro Ser Gln
 1               5                  10                  15

Ser Leu Ser Ile Thr Cys Thr Val Ser Gly Phe Ser Leu Thr Gly Tyr
            20                  25                  30

Gly Val Asn Trp Val Arg Gln Pro Pro Gly Lys Gly Leu Glu Trp Leu
        35                  40                  45

Gly Met Ile Trp Gly Asp Gly Asn Thr Asp Tyr Asn Ser Ala Leu Lys
    50                  55                  60
```

Ser Arg Leu Ser Ile Ser Lys Asp Asn Ser Lys Ser Gln Val Phe Leu
65 70 75 80

Lys Met Asn Ser Leu His Thr Asp Asp Thr Ala Arg Tyr Tyr Cys Ala
85 90 95

Arg Glu Arg Asp Tyr Arg Leu Asp Tyr Trp Gly
100 105

(15) INFORMATION FOR SEQ ID NO:15

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 106 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Val Gln Leu Gln Gln Ser Gly Ala Glu Leu Met Lys Pro Gly Ala Ser
1 5 10 15

Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Ser Asp Tyr Trp
20 25 30

Ile Glu Trp Val Lys Gln Arg Pro Gly His Gly Leu Glu Trp Ile Gly
35 40 45

Glu Ile Leu Pro Gly Ser Gly Ser Thr Asn Tyr His Glu Arg Phe Lys
50 55 60

Gly Lys Ala Thr Phe Thr Ala Asp Thr Ser Ser Ser Thr Ala Tyr Met
65 70 75 80

Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Gly Val Tyr Tyr Cys Leu
85 90 95

His Gly Asn Tyr Asp Phe Asp Gly Trp Gly
100 105

(16) INFORMATION FOR SEQ ID NO:16

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Asp Val Gln Leu Gln Glu Ser Gly Pro Ser Leu Val Lys Pro Ser Gln
 1               5                  10                  15

Thr Leu Ser Leu Thr Cys Ser Val Thr Gly Asp Ser Ile Thr Ser Asp
            20                  25                  30

Tyr Trp Ser Trp Ile Arg Lys Phe Pro Gly Asn Arg Leu Glu Tyr Met
        35                  40                  45

Gly Tyr Val Ser Tyr Ser Gly Ser Thr Tyr Tyr Asn Pro Ser Leu Lys
    50                  55                  60

Ser Arg Ile Ser Ile Thr Arg Asp Thr Ser Lys Asn Gln Tyr Tyr Leu
65                  70                  75                  80

Asp Leu Asn Ser Val Thr Thr Glu Asp Thr Ala Thr Tyr Tyr Cys Ala
                85                  90                  95

Asn Trp Asp Gly Asp Tyr Trp Gly
            100
```

(17) INFORMATION FOR SEQ ID NO:17

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
Glu Val Lys Leu Leu Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
 1               5                  10                  15

Ser Leu Lys Leu Ser Cys Ala Ala Ser Gly Phe Asp Phe Ser Lys Tyr
            20                  25                  30

Trp Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Ile
        35                  40                  45

Gly Glu Ile His Pro Asp Ser Gly Thr Ile Asn Tyr Thr Pro Ser Leu
    50                  55                  60
```

Lys Asp Lys Phe Ile Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Ser Lys Val Arg Ser Glu Asp Thr Ala Leu Tyr Tyr Cys
85 90 95

Ala Arg Leu His Tyr Tyr Gly Tyr Asn Ala Tyr Trp Gly
100 105

(18) INFORMATION FOR SEQ ID NO:18

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Glu Val Gln Leu Val Gln Ser Gly Gly Gly Val Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ser Ser Ser Gly Phe Ile Phe Ser Ser Tyr
20 25 30

Ala Met Tyr Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Ile Ile Trp Asp Asp Gly Ser Asp Gln His Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asn Asp Ser Lys Asn Thr Leu Phe
65 70 75 80

Leu Gln Met Asp Ser Leu Arg Pro Glu Asp Thr Gly Val Tyr Phe Cys
85 90 95

Ala Arg Asp Gly Gly His Gly Phe Cys Ser Ser Ala Ser Cys Phe Gly
100 105 110

Pro Asp Tyr Trp Gly
115

(19) INFORMATION FOR SEQ ID NO:19

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 113 amino acids
(B) TYPE: amino acid

(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
Glu Val Lys Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
 1               5                  10                  15

Ser Leu Arg Leu Ser Cys Ala Thr Ser Gly Phe Thr Phe Ser Asp Phe
            20                  25                  30

Tyr Met Glu Trp Val Arg Gln Pro Pro Gly Lys Arg Leu Glu Trp Ile
        35                  40                  45

Ala Ala Ser Arg Asn Lys Gly Asn Lys Tyr Thr Thr Glu Tyr Ser Ala
    50                  55                  60

Ser Val Lys Gly Arg Phe Ile Val Ser Arg Asp Thr Ser Gln Ser Ile
65                  70                  75                  80

Leu Tyr Leu Gln Met Asn Ala Leu Arg Ala Glu Asp Thr Ala Ile Tyr
                85                  90                  95

Tyr Cys Ala Arg Asn Tyr Tyr Gly Ser Thr Trp Tyr Phe Asp Val Trp
            100                 105                 110

Gly
```

(20) INFORMATION FOR SEQ ID NO:20

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Val Gln Leu Glu Gln Ser Gly Pro Gly Leu Val Arg Pro Ser Gln Thr
 1               5                  10                  15

Leu Ser Leu Thr Cys Thr Val Ser Gly Thr Ser Phe Asp Asp Tyr Tyr
            20                  25                  30
```

Ser Thr Trp Val Arg Gln Pro Pro Gly Arg Gly Leu Glu Trp Ile Gly
35 40 45

Tyr Val Phe Tyr His Gly Thr Ser Asp Thr Asp Thr Pro Leu Arg Ser
50 55 60

Arg Val Thr Met Leu Val Asn Thr Ser Lys Asn Gln Phe Ser Leu Arg
65 70 75 80

Leu Ser Ser Val Thr Ala Ala Asp Thr Ala Val Tyr Tyr Cys Ala Arg
85 90 95

Asn Leu Ile Ala Gly Cys Ile Asp Val Trp Gly
100 105

(21) INFORMATION FOR SEQ ID NO:21

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 109 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Glu Val Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Pro Met Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr
20 25 30

Trp Met Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val
35 40 45

Ala Gln Ile Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp
50 55 60

Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser
65 70 75 80

Val Tyr Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr
85 90 95

Tyr Cys Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly
100 105

(22) INFORMATION FOR SEQ ID NO:22

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 115 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Gln Val Gln Leu Lys Glu Ser Gly Ala Glu Leu Val Ala Ala Ser Ser
 1               5                  10                  15

Ser Val Lys Met Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Ser Tyr
            20                  25                  30

Gly Val Asn Trp Val Lys Gln Arg Pro Gly Gln Gly Leu Glu Trp Ile
        35                  40                  45

Gly Tyr Ile Asn Pro Gly Lys Gly Tyr Leu Ser Tyr Asn Glu Lys Phe
    50                  55                  60

Lys Gly Lys Thr Thr Leu Thr Val Asp Arg Ser Ser Ser Thr Ala Tyr
65                  70                  75                  80

Met Gln Leu Arg Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
                85                  90                  95

Ala Arg Ser Phe Tyr Gly Gly Ser Asp Leu Ala Val Tyr Tyr Phe Asp
            100                 105                 110

Ser Trp Gly
        115
```

(23) INFORMATION FOR SEQ ID NO:23

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 112 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
Glu Val Gln Leu Gln Gln Ser Gly Val Glu Leu Val Arg Ala Gly Ser
 1               5                  10                  15
```

Ser Val Lys Met Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Ser Asn
20 25 30

Gly Ile Asn Trp Val Lys Gln Arg Pro Gly Gln Gly Leu Glu Trp Ile
35 40 45

Gly Tyr Asn Asn Pro Gly Asn Gly Tyr Ile Ala Tyr Asn Glu Lys Phe
50 55 60

Lys Gly Lys Thr Thr Leu Thr Val Asp Lys Ser Ser Ser Thr Ala Tyr
65 70 75 80

Met Gln Leu Arg Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
85 90 95

Ala Arg Ser Glu Tyr Tyr Gly Gly Ser Tyr Lys Phe Asp Tyr Trp Gly
100 105 110

(24) INFORMATION FOR SEQ ID NO:24

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Asn Asp Tyr
20 25 30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ser Gly Ile Ser Trp Asp Ser Ser Ser Ile Gly Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65 70 75 80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Met Ala Leu Tyr Tyr Cys
85 90 95

Val Lys Gly Arg Asp Tyr Tyr Asp Ser Gly Gly Tyr Phe Thr Val Ala
100 105 110

Phe Asp Ile Trp Gly
115

(25) INFORMATION FOR SEQ ID NO:25

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 111 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Asp Val Leu Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly
1 5 10 15

Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ile Ile Ile His Ser
20 25 30

Asp Gly Asn Thr Tyr Leu Glu Trp Phe Leu Gln Lys Pro Gly Gln Ser
35 40 45

Pro Lys Leu Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
50 55 60

Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Met Ile
65 70 75 80

Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Tyr Cys Phe Gln Gly
85 90 95

Ser His Val Pro His Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
100 105 110

(26) INFORMATION FOR SEQ ID NO:26

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 110 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Gln Ser Val Leu Thr Gln Pro Pro Ser Ala Ser Gly Thr Pro Gly Gln
1 5 10 15

Arg Val Thr Ile Ser Cys Ser Gly Thr Ser Ser Asn Ile Gly Ser Ser
20 25 30

Thr Val Asn Trp Tyr Gln Gln Leu Pro Gly Met Ala Pro Lys Leu Leu
35 40 45

Ile Tyr Arg Asp Ala Met Arg Pro Ser Gly Val Pro Asp Arg Phe Ser
50 55 60

Gly Ser Lys Ser Gly Ala Ser Ala Ser Leu Ala Ile Gly Gly Leu Gln
65 70 75 80

Ser Glu Asp Glu Thr Asp Tyr Tyr Cys Ala Ala Trp Asp Val Ser Leu
85 90 95

Asn Ala Tyr Val Phe Gly Thr Gly Thr Lys Val Thr Val Leu
100 105 110

(27) INFORMATION FOR SEQ ID NO:27

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 111 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Gln Val Leu Met Thr Gln Thr Pro Ser Ser Leu Pro Val Thr Leu Gly
1 5 10 15

Gln Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ile Ile Ile His Ser
20 25 30

Asp Gly Asn Thr Tyr Leu Glu Trp Phe Leu Gln Lys Pro Gly Gln Ser
35 40 45

Pro Lys Leu Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
50 55 60

Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Ser Phe Thr Leu Ala Ile
65 70 75 80

Ser Arg Val Glu Ala Glu Asp Glu Gly Val Tyr Tyr Cys Phe Gln Gly
85 90 95

Ser His Val Pro His Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
100 105 110

(28) INFORMATION FOR SEQ ID NO:28

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 112 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Asp Val Val Met Thr Gln Ser Pro Leu Ser Leu Pro Val Thr Leu Gly
1 5 10 15

Gln Pro Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val Tyr Ser
20 25 30

Asp Gly Asn Thr Tyr Leu Asn Trp Phe Gln Gln Arg Pro Gly Gln Ser
35 40 45

Pro Arg Arg Leu Ile Tyr Lys Val Ser Asn Arg Asp Ser Gly Val Pro
50 55 60

Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
65 70 75 80

Ser Arg Val Glu Ala Glu Asp Val Gly Val Tyr Tyr Cys Met Gln Gly
85 90 95

Thr His Trp Ser Trp Thr Phe Gly Gln Gly Thr Lys Val Glu Ile Lys
100 105 110

(29) INFORMATION FOR SEQ ID NO:29

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 111 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Asp Val Leu Met Thr Gln Ser Pro Leu Ser Leu Pro Val Thr Leu Gly
1 5 10 15

Gln Pro Ala Ser Ile Ser Cys Arg Ser Ser Gln Ile Ile Ile His Ser
20 25 30

Asp Gly Asn Thr Tyr Leu Glu Trp Phe Gln Gln Arg Pro Gly Gln Ser
35 40 45

Pro Arg Leu Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
50 55 60

Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
65 70 75 80

Ser Arg Val Glu Ala Glu Asp Val Gly Val Tyr Tyr Cys Phe Gln Gly
85 90 95

Ser His Val Pro His Thr Phe Gly Gly Gly Thr Lys Val Glu Ile
100 105 110

(30) INFORMATION FOR SEQ ID NO:30

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 112 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Asp Ile Val Met Thr Gln Ser Pro Asp Ser Leu Ala Val Ser Leu Gly
1 5 10 15

Glu Arg Ala Thr Ile Asn Cys Lys Ser Ser Gln Ser Val Leu Tyr Ser
20 25 30

Ser Asn Asn Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
35 40 45

Pro Pro Lys Leu Leu Ile Tyr Trp Ala Ser Thr Arg Glu Ser Gly Val
50 55 60

Pro Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Thr
65 70 75 80

Ile Ser Ser Leu Gln Ala Glu Asp Val Ala Val Tyr Tyr Cys Gln Gln
85 90 95

Tyr Asp Thr Ile Pro Thr Phe Gly Gly Gly Thr Lys Val Glu Ile Lys
100 105 110

(31) INFORMATION FOR SEQ ID NO:31

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 111 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Asp Val Leu Met Thr Gln Thr Pro Asp Ser Leu Pro Val Ser Leu Gly
1 5 10 15

Asp Arg Ala Ser Ile Ser Cys Arg Ser Ser Gln Ile Ile Ile His Ser
20 25 30

Asp Gly Asn Thr Tyr Leu Glu Trp Phe Leu Gln Lys Pro Gly Gln Ser
35 40 45

Pro Lys Leu Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
50 55 60

Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Met Ile
65 70 75 80

Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Tyr Cys Phe Gln Gly
85 90 95

Ser His Val Pro His Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile
100 105 110

(32) INFORMATION FOR SEQ ID NO:32

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Asp Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Arg Lys Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Phe
20 25 30

Gly Met His Trp Val Arg Gln Ala Pro Glu Lys Gly Leu Glu Trp Val
35 40 45

Ala Tyr Ile Ser Ser Gly Ser Phe Thr Ile Tyr His Ala Asp Thr Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Pro Lys Asn Thr Leu Phe
65 70 75 80

Leu Gln Met Thr Ser Leu Arg Ser Glu Asp Thr Ala Met Tyr Tyr Cys
85 90 95

Ala Arg Met Arg Lys Gly Tyr Ala Met Asp Tyr Trp Gly Gln Gly Thr
100 105 110

Thr Val Thr Val Ser
115

(33) INFORMATION FOR SEQ ID NO:33

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Glu Val Gln Leu Val Gln Ser Gly Gly Gly Val Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ser Ser Ser Gly Phe Ile Phe Ser Ser Tyr
20 25 30

Ala Met Tyr Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Ile Ile Trp Asp Asp Gly Ser Asp Gln His Tyr Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asn Asp Ser Lys Asn Thr Leu Phe
65 70 75 80

Leu Gln Met Asp Ser Leu Arg Pro Glu Asp Thr Gly Val Tyr Phe Cys
85 90 95

Ala Arg Asp Gly Gly His Gly Phe Cys Ser Ser Ala Ser Cys Phe Gly
100 105 110

Pro Asp Tyr Trp Gly Gln Gly Thr Pro Val Thr Val Ser
115 120 125

(34) INFORMATION FOR SEQ ID NO:34

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Val Val Gln Pro Gly Arg
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Ile Phe Ser Ser Phe
20 25 30

Gly Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Tyr Ile Ser Ser Asp Gly Phe Thr Ile Tyr His Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Pro Lys Asn Thr Leu Phe
65 70 75 80

Leu Gln Met Thr Ser Leu Arg Ser Glu Asp Thr Ala Met Tyr Tyr Cys
85 90 95

Ala Arg Met Arg Lys Gly Tyr Ala Met Asp Tyr Trp Gly Gln Gly Thr
100 105 110

Thr Val Thr Val Ser
115

(35) INFORMATION FOR SEQ ID NO:35

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 120 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
Gln Val Gln Leu Val Glu Ser Gly Gly Gly Val Val Gln Pro Gly Arg
 1               5                  10                  15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
            20                  25                  30

Ala Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
        35                  40                  45

Ala Val Ile Ser Tyr Asp Gly Ser Asn Lys Tyr Tyr Ala Asp Ser Val
    50                  55                  60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
65                  70                  75                  80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
                85                  90                  95

Ala Arg Asp Arg Lys Asp Trp Gly Trp Ala Leu Phe Asp Tyr Trp Gly
            100                 105                 110

Gln Gly Thr Leu Val Thr Val Ser
        115                 120
```

(36) INFORMATION FOR SEQ ID NO:36

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
Gln Val Gln Leu Val Glu Ser Gly Gly Gly Val Val Gln Pro Gly Arg
 1               5                  10                  15
```

```
Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Phe
                20                  25                  30

Gly Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
            35                  40                  45

Ala Tyr Ile Ser Ser Gly Ser Phe Thr Ile Tyr Tyr Ala Asp Ser Val
        50                  55                  60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr
65                  70                  75                  80

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
                85                  90                  95

Ala Arg Met Arg Lys Gly Tyr Ala Met Asp Tyr Trp Gly Gln Gly Thr
            100                 105                 110

Leu Val Thr Val Ser
        115
```

(37) INFORMATION FOR SEQ ID NO:37

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 98 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```
Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1               5                   10                  15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr
            20                  25                  30

Trp Met Ser Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
        35                  40                  45

Ala Asn Ile Lys Gln Asp Gly Ser Glu Lys Tyr Tyr Val Asp Ser Val
    50                  55                  60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Ser Leu Tyr
65                  70                  75                  80
```

Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr Cys
85 90 95

Ala Arg

(38) INFORMATION FOR SEQ ID NO:38

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly
1 5 10 15

Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Phe
20 25 30

Gly Met His Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val
35 40 45

Ala Tyr Ile Ser Ser Gly Ser Phe Thr Ile Tyr His Ala Asp Ser Val
50 55 60

Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala Lys Asn Thr Leu Phe
65 70 75 80

Leu Gln Met Thr Ser Leu Arg Ala Glu Asp Thr Ala Met Tyr Tyr Cys
85 90 95

Ala Arg Met Arg Lys Gly Tyr Ala Met Asp Tyr Trp Gly Gln Gly Thr
100 105 110

Thr Val Thr Val Ser
115

(39) INFORMATION FOR SEQ ID NO:39

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr
1 5 10 15

(40) INFORMATION FOR SEQ ID NO:40

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Lys Thr Ser Leu Arg Pro Gly Lys Gly Ser Ser Asp Tyr Glu Lys Lys
1 5 10 15

(41) INFORMATION FOR SEQ ID NO:41

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Lys Thr Ser Leu Arg Pro Gly Lys Gly Ser Ser Glu Tyr Glu Lys Lys
1 5 10 15

(42) INFORMATION FOR SEQ ID NO:42

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Gln Thr Ser Leu Arg Pro Asp Lys Gly Ser Ser Asp His Glu Lys Lys
1 5 10 15

(43) INFORMATION FOR SEQ ID NO:43

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Gln Thr Ser Leu Arg Pro Asp Lys Gly Ser Ser Asp Gln Glu Lys Lys
1 5 10 15

(44) INFORMATION FOR SEQ ID NO:44

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Gln Ser Ser Leu Arg Pro Asp Lys Gly Ser Ser Asp Gln Glu Lys Lys
1 5 10 15

(45) INFORMATION FOR SEQ ID NO:45

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Gln Thr Ser Leu Arg Pro Asp Lys Gly Ser Ser Asp Pro Glu Lys Lys
1 5 10 15

(46) INFORMATION FOR SEQ ID NO:46

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Gln Thr Ser Leu Arg Pro Asp Lys Gly Ser Ser Asp Pro Glx Lys Lys
1 5 10 15

(47) INFORMATION FOR SEQ ID NO:47

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Gln Thr Ser Leu Arg Pro Asp Lys Gly Ser Ser Asp Pro Glu Lys Thr
1 5 10 15

(48) INFORMATION FOR SEQ ID NO:48

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Gln Thr Ser Leu Arg Ala Asp Lys Gly Ser Ser Asp Gln Glu Lys Lys
1 5 10 15

(49) INFORMATION FOR SEQ ID NO:49

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Gln Thr Ser Leu Arg Pro Asp Lys Gly Lys Ser Asp Ser Glu Lys Lys
1 5 10 15

(50) INFORMATION FOR SEQ ID NO:50

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Gln Thr Ser Leu Arg Pro Ala Arg Gly Ser Ser Asp Gln Glu Lys Lys
1 5 10 15

(51) INFORMATION FOR SEQ ID NO:51

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Gln Thr Ser Leu Lys Pro Gly Arg Gly Ser Ser Asp Pro Glu Lys Lys
1 5 10 15

(52) INFORMATION FOR SEQ ID NO:52

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Gln Thr Ser Leu Arg Pro Gly Arg Gly Ser Ser Asp Thr Glu Lys Lys
1 5 10 15

(53) INFORMATION FOR SEQ ID NO:53

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Gln Ile Ser Leu Arg Pro Gly Lys Gly Ser Ser Asp Ser Glu Lys Lys
1 5 10 15

(54) INFORMATION FOR SEQ ID NO:54

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Gln Thr Ser Leu Arg Pro Gly Lys Gly Asp Ser Asp Glu Asp Lys Lys
1 5 10 15

(55) INFORMATION FOR SEQ ID NO:55

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Glu Thr Ala Leu Arg Pro Gly Lys Gly Ala Ser Asp Ala Asp Lys Lys
1 5 10 15

(56) INFORMATION FOR SEQ ID NO:56

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Val Thr Ala Leu Arg Pro Gly Lys Gly Ala Ser Asp Glu Asp Lys Lys
1 5 10 15

(57) INFORMATION FOR SEQ ID NO:57

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Val Thr Ala Leu Arg Pro Gly Lys Gly Ala Ser Asp Glu Glu Lys Lys
1 5 10 15

(58) INFORMATION FOR SEQ ID NO:58

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Val Thr Ala Leu Arg Pro Gly Lys Gly Ala Ser Asx Ala Asx Lys Lys
1 5 10 15

(59) INFORMATION FOR SEQ ID NO:59

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Val Thr Ala Leu Arg Pro Gly Lys Gly Ala Ser Asp Glu Asp Asp Glu
1 5 10 15

(60) INFORMATION FOR SEQ ID NO:60

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Gln Thr Ser Leu Arg Pro Asp Lys Gly Ser Ser Asp Gln Glu Thr Thr
1 5 10 15

(61) INFORMATION FOR SEQ ID NO:61

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Gln Asn Ser Leu Thr Pro Gly Lys Gly Ser Ser Ser Pro Glu Lys Lys
1 5 10 15

(62) INFORMATION FOR SEQ ID NO:62

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Val Thr Lys Val Arg Pro Gly Lys Gly Asp Ser Asp Ser Asp Lys Lys
1 5 10 15

(63) INFORMATION FOR SEQ ID NO:63

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Val Thr Lys Val Arg Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(64) INFORMATION FOR SEQ ID NO:64

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Val Thr Arg Val Arg Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(65) INFORMATION FOR SEQ ID NO:65

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Leu Thr Lys Val Arg Pro Gly Lys Gly Asp Ser Asp Ser Glu Lys Lys
1 5 10 15

(66) INFORMATION FOR SEQ ID NO:66

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Val Thr Lys Val Arg Pro Gly Lys Gly Asp Ser Asp Ser Glu Gln Lys
1 5 10 15

(67) INFORMATION FOR SEQ ID NO:67

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

Val Thr Lys Val Arg Pro Glu Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(68) INFORMATION FOR SEQ ID NO:68

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Val Thr Lys Val Arg Pro Glu Lys Gly Asp Ser Asp Ser Glu Lys Lys
1 5 10 15

(69) INFORMATION FOR SEQ ID NO:69

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Val Thr Lys Val Ser Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(70) INFORMATION FOR SEQ ID NO:70

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Val Thr Lys Val Arg Ser Gly Lys Gly Glu Ser Asp Ala Glu Lys Lys
1 5 10 15

(71) INFORMATION FOR SEQ ID NO:71

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

Val Thr Ser Val Lys Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(72) INFORMATION FOR SEQ ID NO:72

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Val Ser Ser Val Lys Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(73) INFORMATION FOR SEQ ID NO:73

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Val Thr Ser Ala Lys Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(74) INFORMATION FOR SEQ ID NO:74

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Val Ser Ser Ala Lys Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(75) INFORMATION FOR SEQ ID NO:75

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Val Thr Ser Ala Arg Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(76) INFORMATION FOR SEQ ID NO:76

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Val Ser Pro Ala Lys Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(77) INFORMATION FOR SEQ ID NO:77

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Val Thr Lys Ala Arg Pro Gly Lys Gly Asp Ser Asp Val Glu Lys Asn
1 5 10 15

(78) INFORMATION FOR SEQ ID NO:78

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Val Thr Leu Ile Pro Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(79) INFORMATION FOR SEQ ID NO:79

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Val Thr Leu Leu Gln Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(80) INFORMATION FOR SEQ ID NO:80

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Val Thr Leu Leu Gln Pro Gly Lys Gly Asp Ser Asp Ala Asp Lys Lys
1 5 10 15

(81) INFORMATION FOR SEQ ID NO:81

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Val Thr Leu Leu Gln Pro Gly Lys Gly Asp Ser Asp Ala Glu Arg Lys
1 5 10 15

(82) INFORMATION FOR SEQ ID NO:82

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Val Thr Leu Leu Gln Ala Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(83) INFORMATION FOR SEQ ID NO:83

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Val Thr Leu Leu Gln Pro Gly Glu Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(84) INFORMATION FOR SEQ ID NO:84

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Leu Thr Leu Leu Gln Prc Gly Asn Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(85) INFORMATION FOR SEQ ID NO:85

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Val Thr Leu Leu Gln Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Ile
1 5 10 15

(86) INFORMATION FOR SEQ ID NO:86

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Val Thr Leu Phe Gln Pro Gly Gln Gly Asp Ser Asp Pro Glu Lys Lys
1 5 10 15

(87) INFORMATION FOR SEQ ID NO:87

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Val Thr Leu Pro Gln Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(88) INFORMATION FOR SEQ ID NO:88

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Val Thr Leu Pro Gln Pro Gly Lys Gly Asp Trp Asp Ala Glu Lys Lys
1 5 10 15

(89) INFORMATION FOR SEQ ID NO:89

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Val Thr Phe Leu Ser Pro Gly Gln Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(90) INFORMATION FOR SEQ ID NO:90

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Glu Ser Ser Ala Arg Pro Gly Lys Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(91) INFORMATION FOR SEQ ID NO:91

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Val Thr Leu Ser Ser Pro Gly Gln Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(92) INFORMATION FOR SEQ ID NO:92

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

Val Thr Thr Ala Lys Pro Glu Lys Gly Asp Ser Asp Val Glu Lys Lys
1 5 10 15

(93) INFORMATION FOR SEQ ID NO:93

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

Val Thr Thr Pro Lys Pro Asp Lys Gly Asp Ser Asp Val Glu Lys Lys
1 5 10 15

(94) INFORMATION FOR SEQ ID NO:94

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Val Thr Ala Pro Arg Pro Gly Lys Gly Ala Ser Ser Ala Glu Lys Lys
1 5 10 15

(95) INFORMATION FOR SEQ ID NO:95

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

Val Thr Ala Pro Lys Pro Gly Lys Gly Thr Ser Ser Ala Glu Lys Lys
1 5 10 15

(96) INFORMATION FOR SEQ ID NO:96

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Val Thr Thr Pro Lys Pro Gly Lys Gly Ala Ser Ser Ala Glu Lys Lys
1 5 10 15

(97) INFORMATION FOR SEQ ID NO:97

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Val Ser Ala Pro Lys Pro Gly Lys Gly Ala Ser Ser Ala Glu Lys Lys
1 5 10 15

(98) INFORMATION FOR SEQ ID NO:98

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

Val Thr Ala Pro Arg Ser Gly Lys Gly Ala Ser Ser Ala Glu Lys Lys
1 5 10 15

(99) INFORMATION FOR SEQ ID NO:99

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

Val Thr Ala Pro Lys Ser Gly Lys Gly Ala Ser Ser Ala Glu Lys Lys
1 5 10 15

(100) INFORMATION FOR SEQ ID NO:100

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Val Thr Ala Pro Lys Pro Asp Lys Gly Val Ser Ser Ala Glu Lys Lys
1 5 10 15

(101) INFORMATION FOR SEQ ID NO:101

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

Val Thr Ala Pro Lys Ser Glu Lys Gly Val Ser Ser Ala Glu Lys Lys
1 5 10 15

(102) INFORMATION FOR SEQ ID NO:102

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

Phe Thr Ala Pro Lys Pro Gly Lys Gly Ala Ser Ser Ala Glu Lys Lys
1 5 10 15

(103) INFORMATION FOR SEQ ID NO:103

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

Leu Thr Ala Pro Lys Pro Gly Arg Gly Val Ser Ser Ala Glu Lys Lys
1 5 10 15

(104) INFORMATION FOR SEQ ID NO:104

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

Val Thr Ala Pro Lys Ser Gly Lys Gly Ala Ser Ser Ala Glu Lys Arg
1 5 10 15

(105) INFORMATION FOR SEQ ID NO:105

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

Val Ser Ala Pro Lys Pro Gly Lys Glu Gly Ser Ser Ala Glu Lys Lys
1 5 10 15

(106) INFORMATION FOR SEQ ID NO:106

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

Val Thr Ala Pro Lys Pro Arg Lys Gly Ala Ser Ser Ala Glu Lys Lys
1 5 10 15

(107) INFORMATION FOR SEQ ID NO:107

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

Val Thr Phe Leu Ser Pro Gly Gln Gly Asn Ser Asp Ala Glu Leu Pro
1 5 10 15

(108) INFORMATION FOR SEQ ID NO:108

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

Val Thr Phe Leu Ser Pro Gly Gln Gly Asn Ser Asp Glu Asp Leu Pro
1 5 10 15

(109) INFORMATION FOR SEQ ID NO:109

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

Val Thr Leu Ser Ser Pro Gln Arg Gly Asp Ser Asp Ala Glu Lys Lys
1 5 10 15

(110) INFORMATION FOR SEQ ID NO:110

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

Val Thr Ala Pro Lys Ser Ser Lys Gly Gly Ser Ser Ala Glu Lys Lys
1 5 10 15

(111) INFORMATION FOR SEQ ID NO:111

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

Gln Thr Ser Pro Thr Pro Gly Lys Gly Ser Ser Asp Pro Glu Lys Lys
1 5 10 15

(112) INFORMATION FOR SEQ ID NO:112

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

Gln Ile Ser Leu Ile Pro Gly Lys Gly Ser Tyr Asp Asp Glu Lys Lys
1 5 10 15

(113) INFORMATION FOR SEQ ID NO:113

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

Val Thr Ala Leu Lys Ser Gly Lys Gly Ala Ser Ser Ala Glu Lys Lys
1 5 10 15

(114) INFORMATION FOR SEQ ID NO:114

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Val Thr Ala Leu Lys Ser Asp Lys Gly Ala Ser Ser Gly Glu Lys Lys
1 5 10 15

(115) INFORMATION FOR SEQ ID NO:115

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

Val Thr Pro Pro Ser Pro Gly Gln Gly Asp Ser Ala Ala Glu Lys Lys
1 5 10 15

(116) INFORMATION FOR SEQ ID NO:116

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

Val Thr Pro Pro Ser Pro Gly Gln Gly Asp Ser Ala Arg Glu Lys Lys
1 5 10 15

(117) INFORMATION FOR SEQ ID NO:117

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

Val Thr Val Arg Lys Pro Gly Lys Gly Asp Ser Ser Asp Glu Lys Lys
1 5 10 15

(118) INFORMATION FOR SEQ ID NO:118

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

Gln Thr Ser Val Arg Leu Gly Gln Gly Ser Ser Asp Pro Glu Lys Lys
1 5 10 15

(119) INFORMATION FOR SEQ ID NO:119

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

Lys Thr Ser Leu Arg Pro Trp Lys Gly Ser Ser Asp Ser Asp Lys Lys
1 5 10 15

(120) INFORMATION FOR SEQ ID NO:120

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Gln Thr Asp Val Thr Gln Gly Gln Gly Ser Ser Gln Pro Glu Lys Lys
1 5 10 15

(121) INFORMATION FOR SEQ ID NO:121

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Gln Thr Ala Val Ser Gln Gly Gln Gly Ser Ser Gln Ser Glu Lys Lys
1 5 10 15

(122) INFORMATION FOR SEQ ID NO:122

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Leu Thr Ala Pro Arg Thr Asn Arg Gly Ser Ser Asp Ser Glu Lys Lys
1 5 10 15

(123) INFORMATION FOR SEQ ID NO:123

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Val Thr Ala Pro Ser Ser His Arg Gly Ser Ser Asp Thr Glu Lys Lys
1 5 10 15

(124) INFORMATION FOR SEQ ID NO:124

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Leu Leu Ser Leu Ser Pro Leu Lys Gly Asp Ser Asp Pro Glu Lys Val
1 5 10 15

(125) INFORMATION FOR SEQ ID NO:125

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Val Thr Ala Pro Thr Pro Asp Thr Gly Ala Ile Lys Thr Glu Lys Leu
1 5 10 15

(126) INFORMATION FOR SEQ ID NO:126

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Val Thr Ile Pro Thr Pro Asp Thr Gly Ala Ile Lys Thr Glu Lys Leu
1 5 10 15

(127) INFORMATION FOR SEQ ID NO:127

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Ala Val Ser Pro Thr Pro Asp Thr Gly Ala Ile Lys Thr Glu Lys Leu
1 5 10 15

(128) INFORMATION FOR SEQ ID NO:128

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Ala Val Ser Pro Thr Pro Asp Thr Gly Ala Ile Lys Thr Glu Lys Leu
1 5 10 15

(129) INFORMATION FOR SEQ ID NO:129

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Ala Val Ser Pro Thr Pro Asp Thr Gly Val Ile Lys Thr Glu Lys Leu
1 5 10 15

(130) INFORMATION FOR SEQ ID NO:130

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Ala Val Ser Pro Thr Pro Asp Thr Gly Ala Ile Lys Thr Glu Pro Ser
1 5 10 15

(131) INFORMATION FOR SEQ ID NO:131

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Tyr Leu Pro Pro Thr Pro Gly Val Ile Arg Ser Thr Ala Met Lys Leu
1 5 10 15

(132) INFORMATION FOR SEQ ID NO:132

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Tyr Leu Pro Pro Thr Pro Gly Val Ile Arg Ser Thr Ala Met Arg Leu
1 5 10 15

(133) INFORMATION FOR SEQ ID NO:133

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

Tyr Leu Pro Pro Thr Pro Gly Leu Ile Arg Ser Thr Ser Met Lys Leu
1 5 10 15

(134) INFORMATION FOR SEQ ID NO:134

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

Tyr Leu Pro Pro Thr Pro Gly Leu Ile Arg Ser Thr Ser Val Lys Leu
1 5 10 15

(135) INFORMATION FOR SEQ ID NO:135

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Tyr Leu Pro Pro Thr Pro Gly Val Ile Arg Ser Thr Ala Glu Lys Leu
1 5 10 15

(136) INFORMATION FOR SEQ ID NO:136

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Tyr Leu Pro Pro Thr Pro Gly Val Ile Arg Ser Thr Ala Gly Lys Leu
1 5 10 15

(137) INFORMATION FOR SEQ ID NO:137

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Tyr Leu Pro Ala Thr Pro Gly Val Val Arg Ser Ser Ala Gly Met Leu
1 5 10 15

(138) INFORMATION FOR SEQ ID NO:138

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

Ser Leu Pro Pro Ser Pro Gly Lys Val Arg Ser Thr Ala Glu Lys Leu
1 5 10 15

(139) INFORMATION FOR SEQ ID NO:139

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Ser Leu Pro Pro Ser Pro Gly Lys Val Arg Ser Thr Ala Asn Lys Leu
1 5 10 15

(140) INFORMATION FOR SEQ ID NO:140

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Ser Leu Pro Pro Arg Pro Gly Lys Val Arg Ser Ser Ser Glu Lys Leu
1 5 10 15

(141) INFORMATION FOR SEQ ID NO:141

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

Ser Leu Pro Pro Arg Pro Gly Lys Val Arg Ser Ser Ser Asp Lys Leu
1 5 10 15

(142) INFORMATION FOR SEQ ID NO:142

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Ser Leu Pro Pro Arg Pro Gly Arg Val Arg Ser Ser Ser Glu Lys Leu
1 5 10 15

(143) INFORMATION FOR SEQ ID NO:143

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

Ser Leu Pro Pro Arg Pro Gly Lys Val Arg Ser Ser Ser Glu Gln Leu
1 5 10 15

(144) INFORMATION FOR SEQ ID NO:144

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

Ser Leu Pro Pro Arg Pro Gly Lys Val Arg Ser Ser Ser Glu Thr Leu
1 5 10 15

(145) INFORMATION FOR SEQ ID NO:145

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

Ser Leu Pro Pro Lys Pro Gly Lys Ile Arg Ser Ser Thr Gly Lys Leu
1 5 10 15

(146) INFORMATION FOR SEQ ID NO:146

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

Ser Leu Pro Pro Lys Pro Gly Arg Ile Arg Ser Ser Thr Gly Lys Leu
1 5 10 15

(147) INFORMATION FOR SEQ ID NO:147

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

Ser Leu Pro Pro Lys Pro Gly Lys Ile Arg Ser Ser Thr Gly Gln Leu
1 5 10 15

(148) INFORMATION FOR SEQ ID NO:148

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

Ser Leu Pro Pro Glu Pro Gly Lys Ile Arg Ser Ser Thr Gly Arg Leu
1 5 10 15

(149) INFORMATION FOR SEQ ID NO:149

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:149:

Ser Leu Ala Pro Ser Pro Gly Lys Ile Arg Ser Thr Ala Glu Lys Leu
1 5 10 15

(150) INFORMATION FOR SEQ ID NO:150

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:150:

Ser Leu Pro Pro Arg Pro Gly Lys Ile Arg Ser Ser Thr Gly Asn Val
1 5 10 15

(151) INFORMATION FOR SEQ ID NO:151

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:151:

Ser Leu Arg Pro Ser Pro Gly Lys Val Arg Ser Thr Ala Glu Lys Leu
1 5 10 15

(152) INFORMATION FOR SEQ ID NO:152

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:152:

Ser Leu Arg Pro Ser Pro Gly Lys Val Arg Ser Thr Ala Asp Lys Leu
1 5 10 15

(153) INFORMATION FOR SEQ ID NO:153

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:153:

Ser Leu Arg Pro Ser Pro Gly Lys Val Arg Ser Thr Ala Glu Asn Leu
1 5 10 15

(154) INFORMATION FOR SEQ ID NO:154

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:154:

Ser Leu Arg Pro Ser Pro Gly Lys Val Arg Ser Ala Val Glu Lys Leu
1 5 10 15

(155) INFORMATION FOR SEQ ID NO:155

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:155:

Ser Leu Pro Pro Arg Pro Gly Lys Arg Ser Ser Ala Glu Lys Leu
1 5 10 15

(156) INFORMATION FOR SEQ ID NO:156

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:156:

Ser Leu Ala Pro Ser Pro Gly Lys Val Arg Ser Thr Val Glu Arg Leu
1 5 10 15

(157) INFORMATION FOR SEQ ID NO:157

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:157:

Ser Leu Ala Pro Ser Pro Asp Lys Ile Arg Ser Thr Pro Asp Lys Leu
1 5 10 15

(158) INFORMATION FOR SEQ ID NO:158

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:158:

Ser Leu Ala Leu Ser Pro Gly Lys Val Arg Ser Thr Ala Glu Lys Leu
1 5 10 15

(159) INFORMATION FOR SEQ ID NO:159

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:159:

Ser Leu Pro Leu Ser Ala Gly Lys Val Arg Ser Thr Ala Glu Lys Leu
1 5 10 15

(160) INFORMATION FOR SEQ ID NO:160

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:160:

Ser Leu Ala Pro Ser Pro Gly Lys Val Arg Ser Thr Ala Glu Tyr Leu
1 5 10 15

(161) INFORMATION FOR SEQ ID NO:161

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:161:

Ser Leu Pro Leu Thr Pro Gly Leu Ile Arg Ser Thr Ala Glu Lys Leu
1 5 10 15

(162) INFORMATION FOR SEQ ID NO:162

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:162:

Ser Leu Pro Leu Thr Pro Arg Val Ile Arg Ser Thr Ala Glu Lys Leu
1 5 10 15

(163) INFORMATION FOR SEQ ID NO:163

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:163:

Phe Leu His Pro Thr Pro Gly Thr Asp Ser Ser Ser Thr Glu Lys Leu
1 5 10 15

(164) INFORMATION FOR SEQ ID NO:164

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:164:

Phe Leu Leu Pro Thr Pro Gly Thr Asp Ser Ser Ser Thr Glu Arg Leu
1 5 10 15

(165) INFORMATION FOR SEQ ID NO:165

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:165:

Phe Leu His Pro Thr Arg Val Thr Asp Ser Ser Ser Thr Glu Lys Leu
1 5 10 15

(166) INFORMATION FOR SEQ ID NO:166

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:166:

Leu Leu Pro Pro Thr Pro Gly Thr Asn Ser Ser Ser Asn Asp Lys Leu
1 5 10 15

(167) INFORMATION FOR SEQ ID NO:167

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:167:

Val Leu Pro Leu Ser Pro His Arg Ile Arg Ser Glu Ser Glu Asn Leu
1 5 10 15

(168) INFORMATION FOR SEQ ID NO:168

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:168:

Ser Leu Ala Pro Ser Pro Ala Lys Phe Arg Ser Thr Ala Glu Arg Asp
1 5 10 15

(169) INFORMATION FOR SEQ ID NO:169

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:169:

Val Thr Ala Pro Arg Pro Gly Arg Ile Arg Ser Asp Pro Glu Lys Lys
1 5 10 15

(170) INFORMATION FOR SEQ ID NO:170

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:170:

Val Thr Ala Pro Arg Pro Gly Arg Val Arg Ser Asp Pro Glu Lys Lys
1 5 10 15

(171) INFORMATION FOR SEQ ID NO:171

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:171:

Val Thr Gly Pro Arg Pro Gly Arg Ile Arg Ser Asp Pro Glu Lys Lys
1 5 10 15

(172) INFORMATION FOR SEQ ID NO:172

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:172:

Val Thr Gly Pro Arg Pro Gly Arg Ile Arg Ser Asp Pro Asp Lys Lys
1 5 10 15

(173) INFORMATION FOR SEQ ID NO:173

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:173:

Val Thr Gly Pro Arg Pro Gly Arg Val Arg Ser Asp Pro Glu Lys Lys
1 5 10 15

(174) INFORMATION FOR SEQ ID NO:174

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:174:

Val Thr Gly Pro Arg Pro Gly Arg Ile Arg Ser Asp Pro Xaa Lys Lys
1 5 10 15

(175) INFORMATION FOR SEQ ID NO:175

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:175:

Val Thr Ala Pro Arg Pro Gly Arg Ile Arg Ser Glu Ser Glu Arg Lys
1 5 10 15

(176) INFORMATION FOR SEQ ID NO:176

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:176:

Val Thr Gly Pro Ser Arg Gly Arg Ile Arg Ser Asp Pro Glu Lys Lys
1 5 10 15

(177) INFORMATION FOR SEQ ID NO:177

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:177:

Val Thr Val Pro Arg Pro Ser Arg Ile Arg Ser Glu Ser Glu Arg Lys
1 5 10 15

(178) INFORMATION FOR SEQ ID NO:178

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:178:

Val Thr Ala Pro Gly Pro Gly Arg Ile Arg Ser Glu Ser Glu Arg Lys
1 5 10 15

(179) INFORMATION FOR SEQ ID NO:179

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:179:

Gln Thr Ser Val Arg Pro Gly Arg Val Arg Ser Asp Pro Glu Arg Lys
1 5 10 15

(180) INFORMATION FOR SEQ ID NO:180

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:180:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Asp Pro Glu Arg Lys
1 5 10 15

(181) INFORMATION FOR SEQ ID NO:181

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:181:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Asp Pro Glu Lys Lys
1 5 10 15

(182) INFORMATION FOR SEQ ID NO:182

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:182:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Glu Pro Glu Lys Lys
1 5 10 15

(183) INFORMATION FOR SEQ ID NO:183

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:183:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Glu Pro Asp Lys Lys
1 5 10 15

(184) INFORMATION FOR SEQ ID NO:184

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:184:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ala Glu Pro Glu Lys Lys
1 5 10 15

(185) INFORMATION FOR SEQ ID NO:185

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:185:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Asx Pro Glx Lys Lys
1 5 10 15

(186) INFORMATION FOR SEQ ID NO:186

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:186:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Asp Pro Asx Lys Lys
1 5 10 15

(187) INFORMATION FOR SEQ ID NO:187

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:187:

Gln Thr Ser Val Arg Pro Gly Gln Val Arg Ser Asp Pro Glu Arg Lys
1 5 10 15

(188) INFORMATION FOR SEQ ID NO:188

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:188:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser His Pro Glu Lys Lys
1 5 10 15

(189) INFORMATION FOR SEQ ID NO:189

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:189:

Gln Thr Ser Val Arg Pro Gly Asn Val Arg Ser Asp Pro Asp Lys Lys
1 5 10 15

(190) INFORMATION FOR SEQ ID NO:190

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Asp Pro Glu Lys Thr
1 5 10 15

(191) INFORMATION FOR SEQ ID NO:191

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:191:

Gln Thr Ser Val Arg Pro Gly Thr Val Arg Ser Glu Pro Glu Lys Lys
1 5 10 15

(192) INFORMATION FOR SEQ ID NO:192

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:192:

Gln Thr Ser Val Arg Pro Glu Lys Val Arg Ser Glu Pro Asp Lys Lys
1 5 10 15

(193) INFORMATION FOR SEQ ID NO:193

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:193:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Glu Ser Asp Lys Lys
1 5 10 15

(194) INFORMATION FOR SEQ ID NO:194

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:194:

Gln Thr Ser Val Arg Pro Gly Glu Val Arg Ser Glu Pro Asp Lys Lys
1 5 10 15

(195) INFORMATION FOR SEQ ID NO:195

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

Gln Thr Ser Val Arg Pro Gly Asx Val Arg Ser Asx Pro Glx Arg Lys
1 5 10 15

(196) INFORMATION FOR SEQ ID NO:196

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:196:

Gln Thr Ser Val Ser Pro Gly Lys Val Arg Ser Asp Pro Glu Lys Lys
1 5 10 15

(197) INFORMATION FOR SEQ ID NO:197

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

Gln Thr Ser Val Arg Pro Gly Lys Val Asn Ser Asp Pro Glu Lys Lys
1 5 10 15

(198) INFORMATION FOR SEQ ID NO:198

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

Gln Thr Ser Val Arg Pro Gly Lys Val Arg Ser Asp Pro Asp Thr Lys
1 5 10 15

(199) INFORMATION FOR SEQ ID NO:199

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:199:

Gln Thr Ser Val Arg Pro Lys Lys Val Arg Ser Asp Pro Glx Lys Lys
1 5 10 15

(200) INFORMATION FOR SEQ ID NO:200

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:200:

Gln Thr Ser Val Arg Pro Lys Lys Val Arg Phe Asp Pro Glu Lys Lys
1 5 10 15

(201) INFORMATION FOR SEQ ID NO:201

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:201:

Gln Thr Ser Val Arg Ser Gly Lys Val Arg Ser Glu Pro Glu Thr Lys
1 5 10 15

(202) INFORMATION FOR SEQ ID NO:202

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:202:

Val Thr Asn Leu Arg Pro Gly Lys Val Arg Ser Asp Ala Glu Lys Lys
1 5 10 15

(203) INFORMATION FOR SEQ ID NO:203

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:203:

Val Thr Asp Leu Arg Pro Gly Lys Val Arg Ser Asp Ala Glu Lys Lys
1 5 10 15

(204) INFORMATION FOR SEQ ID NO:204

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:204:

Gln Thr Ser Val Ser Pro Gly Asn Ile Arg Ser Glu Ser Asp Lys Lys
1 5 10 15

(205) INFORMATION FOR SEQ ID NO:205

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:205:

Lys Thr Ser Val Thr Pro Gly Lys Phe Arg Ser Glu Pro Glu Lys Lys
1 5 10 15

(206) INFORMATION FOR SEQ ID NO:206

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:206:

Val Thr Leu Leu Pro Pro Gly Arg Val Arg Ser Asp Ala Glu Lys Lys
1 5 10 15

(207) INFORMATION FOR SEQ ID NO:207

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

Val Thr Leu Leu Pro Pro Gly Glu Val Arg Ser Asp Ala Glu Lys Lys
1 5 10 15

(208) INFORMATION FOR SEQ ID NO:208

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

Val Thr Leu Pro Pro Pro Gly Glx Val Arg Ser Asp Ala Glu Arg Lys
1 5 10 15

(209) INFORMATION FOR SEQ ID NO:209

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

Val Thr Leu Pro Pro Pro Gly Glx Val Arg Ser Asx Ala Glx Asn Lys
1 5 10 15

(210) INFORMATION FOR SEQ ID NO:210

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

Val Thr Leu Pro Pro Pro Gln Gln Val Arg Ser Asp Ala Glu Lys Lys
1 5 10 15

(211) INFORMATION FOR SEQ ID NO:211

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

Val Thr Leu Pro Pro Pro Gly Gln Val Thr Ser Asp Ala Glu Lys Lys
1 5 10 15

(212) INFORMATION FOR SEQ ID NO:212

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

Val Thr Leu Pro Pro Ala Gly Gln Val Arg Ser Asp Ala Glu Lys Arg
1 5 10 15

(213) INFORMATION FOR SEQ ID NO:213

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

Ala Leu Ser Pro Ser Ser Gly Gln Ser Ser Ser Ala Ser Glu Arg Leu
1 5 10 15

(214) INFORMATION FOR SEQ ID NO:214

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Arg Gly Asp Ser Gln Arg Pro Glu Ser
20 25

(215) INFORMATION FOR SEQ ID NO:215

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Val
1 5 10 15

Ser Arg Gly Asp Ser Gln Arg Pro Glu Ser
20 25

(216) INFORMATION FOR SEQ ID NO:216

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

Glu Lys Val Gly Gly Leu Gln Pro Gly Thr Gly Ala Pro Gly Lys Ala
1 5 10 15

Ser Arg Gly Asp Ser Gln Arg Pro Glu Ser
20 25

(217) INFORMATION FOR SEQ ID NO:217

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Asn Ser Gln Arg Ala Glu Ser
20 25

(218) INFORMATION FOR SEQ ID NO:218

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:

Glu Lys Met Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Asn Ser Gln Arg Ala Glu Ser
20 25

(219) INFORMATION FOR SEQ ID NO:219

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Thr Ser Gln Arg Ala Glu Ser
20 25

(220) INFORMATION FOR SEQ ID NO:220

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Thr Ser Gln Arg Ala Glu Thr
20 25

(221) INFORMATION FOR SEQ ID NO:221

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:

Glu Lys Val Gly Gly Leu Lys Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Thr Ser Gln Arg Ala Glu Thr
20 25

(222) INFORMATION FOR SEQ ID NO:222

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:

Glu Asn Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Thr Ser Gln Arg Ala Glu Thr
20 25

(223) INFORMATION FOR SEQ ID NO:223

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:

Glu Lys Val Gly Gly Leu Gln Ser Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Thr Ser Gln Arg Ala Glu Thr
20 25

(224) INFORMATION FOR SEQ ID NO:224

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

Glu Lys Val Gly Gly Leu Gln Ser Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Thr Ser Gln Arg Ala Glu Ser
20 25

(225) INFORMATION FOR SEQ ID NO:225

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Ile Ser Gln Arg Ala Glu Arg
20 25

(226) INFORMATION FOR SEQ ID NO:226

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ser
1 5 10 15

Ala Lys Gly Asx Ser Glx Arg Ala Gln Ser
20 25

(227) INFORMATION FOR SEQ ID NO:227

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:

Glu Lys Val Gly Gly Leu Gln Pro Gly Ser Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Asn Ser Gln Arg Ala Glu Ser
20 25

(228) INFORMATION FOR SEQ ID NO:228

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

Glu Lys Val Gly Gly Leu Gln Pro Gly Ser Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Ser Ser Gln Arg Ala Glu Ser
20 25

(229) INFORMATION FOR SEQ ID NO:229

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Arg Lys Ala
1 5 10 15

Ser Lys Gly Asn Ser Gln Arg Ala Glu Ser
20 25

(230) INFORMATION FOR SEQ ID NO:230

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

Glu Lys Met Gly Asn Leu Gln Pro Gly Ser Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Asn Ser Gln Arg Pro Asp Ser
20 25

(231) INFORMATION FOR SEQ ID NO:231

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

Glu Lys Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Arg Arg Ser Glu Thr
20 25

(232) INFORMATION FOR SEQ ID NO:232

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

Glu Lys Val Gly Gly Leu Lys Pro Gly Lys Gly Ala Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Arg Arg Ser Glu Thr
20 25

(233) INFORMATION FOR SEQ ID NO:233

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

Glu Lys Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Arg Asp
1 5 10 15

Ser Lys Gly Asn Ala Arg Arg Ser Glu Thr
20 25

(234) INFORMATION FOR SEQ ID NO:234

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

Asp Lys Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Arg Ser Glu Thr
20 25

(235) INFORMATION FOR SEQ ID NO:235

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

Asp Lys Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(236) INFORMATION FOR SEQ ID NO:236

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

Asp Lys Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Asp Lys Asp
1 5 10 15

Asn Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(237) INFORMATION FOR SEQ ID NO:237

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

Glu Lys Val Gly Gly Leu Thr Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Gly Arg Arg Ser Glu Thr
20 25

(238) INFORMATION FOR SEQ ID NO:238

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

Glu Met Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Asp Arg Arg Ser Glu Thr
20 25

(239) INFORMATION FOR SEQ ID NO:239

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

Glu Met Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Asp Lys Arg Ser Glu Thr
20 25

(240) INFORMATION FOR SEQ ID NO:240

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

Glu Met Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Arg Ser Glu Thr
20 25

(241) INFORMATION FOR SEQ ID NO:241

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

Glu Gln Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(242) INFORMATION FOR SEQ ID NO:242

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

Glu Gln Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(243) INFORMATION FOR SEQ ID NO:243

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

Glu Gln Val Gly Gly Leu Lys Pro Gly Lys Gly Ala Pro Glu Lys Asp
1 5 10 15

Thr Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(244) INFORMATION FOR SEQ ID NO:244

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

Glu Lys Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(245) INFORMATION FOR SEQ ID NO:245

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

Glu Lys Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(246) INFORMATION FOR SEQ ID NO:246

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

Glu Lys Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(247) INFORMATION FOR SEQ ID NO:247

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

Glu Lys Val Gly Gly Leu Gln Pro Gly Lys Gly Ser Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(248) INFORMATION FOR SEQ ID NO:248

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

Asp Lys Met Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Gln Ser Glu Thr
20 25

(249) INFORMATION FOR SEQ ID NO:249

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

Glu Gln Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Asp Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(250) INFORMATION FOR SEQ ID NO:250

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

Glu Lys Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Glu Lys Ser Glu Thr
20 25

(251) INFORMATION FOR SEQ ID NO:251

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

Glu Gln Val Gly Asp Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Lys Gly Asn Ala Arg Arg Ser Glu Thr
20 25

(252) INFORMATION FOR SEQ ID NO:252

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

Glu Asn Val Gly Asp Leu Lys Pro Gly Lys Gly Ala Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Arg Arg Ser Glu Thr
20 25

(253) INFORMATION FOR SEQ ID NO:253

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

Glu Gln Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Ser Asp Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(254) INFORMATION FOR SEQ ID NO:254

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

Glu Gln Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Lys Ser Gly Thr
20 25

(255) INFORMATION FOR SEQ ID NO:255

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

Asp Gln Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Lys Gly Asn Pro Lys Arg Ser Glu Thr
20 25

(256) INFORMATION FOR SEQ ID NO:256

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

Asp Gln Val Gly Gly Leu Gln Pro Gly Gln Gly Thr Pro Glu Lys Asn
1 5 10 15

Thr Lys Gly Asn Pro Lys Arg Ser Asp Thr
20 25

(257) INFORMATION FOR SEQ ID NO:257

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

Glu Lys Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Ser Glu Lys Asp
1 5 10 15

Ile Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(258) INFORMATION FOR SEQ ID NO:258

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

Asp Lys Val Gly Gly Leu Lys Pro Gly Lys Arg Thr Pro Glu Lys Asp
1 5 10 15

Asn Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(259) INFORMATION FOR SEQ ID NO:259

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

Asp Lys Val Gly Gly Leu Lys Leu Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Lys Gly Asn Ala Lys Lys Ser Glu Thr
20 25

(260) INFORMATION FOR SEQ ID NO:260

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

Glu Lys Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Asn Thr Ser Glu Thr
20 25

(261) INFORMATION FOR SEQ ID NO:261

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

Glu His Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Gly Arg Ser Glu Thr
20 25

(262) INFORMATION FOR SEQ ID NO:262

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

Glu Gln Val Gly Gly Leu Gln Pro Gly Asn Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Thr Gly Asn Ala Lys Arg Ser Glu Thr
20 25

(263) INFORMATION FOR SEQ ID NO:263

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

Glu Lys Glu Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Glu
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Ala Glu Thr
20 25

(264) INFORMATION FOR SEQ ID NO:264

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:264:

Glu Lys Glu Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Glu
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Pro Glu Thr
20 25

(265) INFORMATION FOR SEQ ID NO:265

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:265:

Glu Lys Glu Gly Gly Leu Gln Pro Gly Lys Gly Ser Pro Glu Lys Glu
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Ala Glu Thr
20 25

(266) INFORMATION FOR SEQ ID NO:266

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

Glu Lys Asp Gly Gly Leu Gln Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Val Glu Met
20 25

(267) INFORMATION FOR SEQ ID NO:267

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

Glu Gln Val Gly Gly Leu Lys Pro Gly Arg Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Thr Gly Asp Ala Gln Arg Ser Glu Thr
20 25

(268) INFORMATION FOR SEQ ID NO:268

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

Glu Gln Val Gly Gly Leu Lys Pro Gly Arg Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Thr Gly Asn Ala Lys Gly Ser Glu Thr
20 25

(269) INFORMATION FOR SEQ ID NO:269

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:269:

Glu Lys Val Gly Gly Ser Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Thr Ser Glu Thr
20 25

(270) INFORMATION FOR SEQ ID NO:270

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:270:

Ser Asp Gln Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Lys Gly Asn Ala Arg Arg Ser Glu Ser
20 25

(271) INFORMATION FOR SEQ ID NO:271

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:271:

Glu Lys Ile Gly Gly Leu Gln Pro Gly Lys Gly Asp Pro Gly Lys Pro
1 5 10 15

Ser Lys Asp Asn Ala Lys Arg Ser Glu Thr
20 25

(272) INFORMATION FOR SEQ ID NO:272

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:272:

Glu Lys Leu Gly Gly Leu Gln Pro Gly Lys Gly Asp Pro Gly Lys Pro
1 5 10 15

Ser Lys Asp Asn Ala Lys Arg Ser Glu Thr
20 25

(273) INFORMATION FOR SEQ ID NO:273

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:273:

Glu Lys Leu Gly Gly Leu Gln Pro Gly Lys Gly Asp Pro Gly Lys Pro
1 5 10 15

Phe Lys Asp Asn Ala Lys Arg Ser Glu Thr
20 25

(274) INFORMATION FOR SEQ ID NO:274

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

Glu Lys Leu Gly Gly Leu Gln Pro Gly Lys Gly Asp Pro Gly Lys Leu
1 5 10 15

Met Lys Glu Asn Ala Lys Arg Ser Glu Thr
20 25

(275) INFORMATION FOR SEQ ID NO:275

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

Glu Asn Leu Gly Gly Leu Gln Pro Gly Lys Gly Asp Pro Gly Lys Leu
1 5 10 15

Lys Xaa Glu Asn Ala Lys Arg Pro Glu Thr
20 25

(276) INFORMATION FOR SEQ ID NO:276

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

Glu Lys Leu Gly Gly Leu Gln Pro Gly Asn Gly Asp Leu Gly Lys Pro
1 5 10 15

Ser Lys Asp Asn Ala Lys Arg Ser Glu Thr
20 25

(277) INFORMATION FOR SEQ ID NO:277

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

Glu Lys Leu Gly Pro Leu Gln Leu Gly Lys Gly Asp Pro Gly Lys Pro
1 5 10 15

Ser Lys Asp Asp Ala Lys Arg Ser Glu Thr
20 25

(278) INFORMATION FOR SEQ ID NO:278

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Gly Gly Thr Pro Gly Lys Pro
1 5 10 15

Ser Lys Asp Asn Asp Lys Arg Ser Glu Thr
20 25

(279) INFORMATION FOR SEQ ID NO:279

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:279:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Gly Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Asp Asn Asp Lys Arg Ser Glu Thr
20 25

(280) INFORMATION FOR SEQ ID NO:280

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:280:

Glu Gln Val Gly Gly Leu Lys Ala Arg Lys Gly Thr Pro Glu Lys Asp
1 5 10 15

Thr Thr Gly Asn Ala Lys Arg Ser Glu Thr
20 25

(281) INFORMATION FOR SEQ ID NO:281

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:281:

Glu Met Val Gly Val Leu Glu Pro Gly Lys Gly Thr Pro Glu Lys Arg
1 5 10 15

Gln Glu Gly Asn Ala Lys Arg Ser Glu Thr
20 25

(282) INFORMATION FOR SEQ ID NO:282

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

Glu Gln Val Gly Gly Leu Gln Pro Lys Lys Gly Ser Pro Gly Lys Asp
1 5 10 15

Ser Lys Asp Asp Ser Gln Lys Thr Glu Thr
20 25

(283) INFORMATION FOR SEQ ID NO:283

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

Glu Gln Val Gly Gly Leu Gln Pro Lys Lys Gly Ser Pro Gly Lys Asp
1 5 10 15

Ser Lys Asp Asp Ser Gln Lys Thr Glu Arg
20 25

(284) INFORMATION FOR SEQ ID NO:284

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:284:

Gln Gln Val Pro Glu Leu Lys Pro Gly Arg Gly Thr Pro Gly Lys Glu
1 5 10 15

Asp Lys Gly Thr Ser Ala Arg Asn Asp Thr
20 25

(285) INFORMATION FOR SEQ ID NO:285

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

Gln Gln Val Pro Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Lys Asp
1 5 10 15

Asp Lys Gly Thr Ser Ala Lys Asn Glu Thr
20 25

(286) INFORMATION FOR SEQ ID NO:286

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

Gln Gln Val Pro Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Lys Asp
1 5 10 15

Asp Lys Gly Thr Ser Ala Lys Asn Glu Met
20 25

(287) INFORMATION FOR SEQ ID NO:287

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

Gln Gln Lys Pro Glu Leu Lys Pro Gly Lys Gly Ser Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Thr Ser Ser Thr Ser Glu Thr
20 25

(288) INFORMATION FOR SEQ ID NO:288

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(289) INFORMATION FOR SEQ ID NO:289

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:289:

Glu Gln Gln Pro Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(290) INFORMATION FOR SEQ ID NO:290

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:290:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Ala Ser Glu Ser
20 25

(291) INFORMATION FOR SEQ ID NO:291

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:291:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(292) INFORMATION FOR SEQ ID NO:292

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Thr His Gly Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(293) INFORMATION FOR SEQ ID NO:293

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Ser His Gly Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(294) INFORMATION FOR SEQ ID NO:294

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Ser His Gly Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Ala Ser Glu Ser
20 25

(295) INFORMATION FOR SEQ ID NO:295

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:295:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Thr His Gly Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Phe Glu Ser
20 25

(296) INFORMATION FOR SEQ ID NO:296

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:296:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Thr His Gly Lys Gln
1 5 10 15

Lys Gln Gly Lys Ser Ser Thr Phe Glu Ser
20 25

(297) INFORMATION FOR SEQ ID NO:297

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Thr His Gly Lys Glu
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(298) INFORMATION FOR SEQ ID NO:298

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

Glu Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Ser His Gly Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(299) INFORMATION FOR SEQ ID NO:299

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

Glu Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Thr His Gly Lys Gln
1 5 10 15

Lys Lys Ser Asn Ser Ser Thr Ser Glu Ser
20 25

(300) INFORMATION FOR SEQ ID NO:300

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:300:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Ala Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(301) INFORMATION FOR SEQ ID NO:301

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Ala Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Asp Ser
20 25

(302) INFORMATION FOR SEQ ID NO:302

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:302:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Val Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Asp Ser
20 25

(303) INFORMATION FOR SEQ ID NO:303

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:303:

Gln Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Ala Pro Gly Lys Gly
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(304) INFORMATION FOR SEQ ID NO:304

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:304:

Gln Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Ala Pro Gly Lys Gly
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(305) INFORMATION FOR SEQ ID NO:305

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:305:

Glu Gln Gln Pro Glu Ala Lys Pro Gly Lys Gly Thr His Gly Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Asp Ser
20 25

(306) INFORMATION FOR SEQ ID NO:306

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:306:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr His Gly Lys Glu
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Asp Ser
20 25

(307) INFORMATION FOR SEQ ID NO:307

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:307:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Ala Pro Gly Gln Gly
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(308) INFORMATION FOR SEQ ID NO:308

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:308:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Arg Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(309) INFORMATION FOR SEQ ID NO:309

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:309:

Glu Gln Gln Ala Glu Leu Arg Ala Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(310) INFORMATION FOR SEQ ID NO:310

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(311) INFORMATION FOR SEQ ID NO:311

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:311:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly His Glu
1 5 10 15

Lys Lys Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(312) INFORMATION FOR SEQ ID NO:312

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:312:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly His Glu
1 5 10 15

Lys Lys Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(313) INFORMATION FOR SEQ ID NO:313

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly His Glu
1 5 10 15

Asn Lys Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(314) INFORMATION FOR SEQ ID NO:314

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

Gln Gln Gln Ala Glu Val Arg Pro Gly Lys Gly Thr Pro Gly His Glu
1 5 10 15

Lys Lys Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(315) INFORMATION FOR SEQ ID NO:315

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly His Glu
1 5 10 15

Asn Lys Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(316) INFORMATION FOR SEQ ID NO:316

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Ala Ser Glu Ser
20 25

(317) INFORMATION FOR SEQ ID NO:317

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:317:

His Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(318) INFORMATION FOR SEQ ID NO:318

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:318:

Glu Gln Gln Val Glu Leu Arg Ala Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(319) INFORMATION FOR SEQ ID NO:319

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:319:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Gln Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(320) INFORMATION FOR SEQ ID NO:320

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:320:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly His Asp
1 5 10 15

Asn Lys Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(321) INFORMATION FOR SEQ ID NO:321

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:321:

Gln Gln Gln Ala Glu Val Arg Pro Gly Lys Gly Thr Pro Gly His Glu
1 5 10 15

Lys Lys Gly Arg Ser Ser Thr Ser Glu Ser
20 25

(322) INFORMATION FOR SEQ ID NO:322

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:322:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(323) INFORMATION FOR SEQ ID NO:323

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:323:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(324) INFORMATION FOR SEQ ID NO:324

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:324:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Asp Ser
20 25

(325) INFORMATION FOR SEQ ID NO:325

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:325:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Ser Pro Gly Gln Gln
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(326) INFORMATION FOR SEQ ID NO:326

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:326:

Gln His Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Lys Asn Lys Ser Ser Thr Ser Glu Ser
20 25

(327) INFORMATION FOR SEQ ID NO:327

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:327:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Asn Lys Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(328) INFORMATION FOR SEQ ID NO:328

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:328:

Glu Gln Gln Ala Glu Leu Arg Ala Gly Lys Gly Ile Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(329) INFORMATION FOR SEQ ID NO:329

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:329:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Ser Lys Ser Ser Thr Ser Glu Ser
20 25

(330) INFORMATION FOR SEQ ID NO:330

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:330:

Gln Gln Gln Ser Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Ser Lys Ser Ser Thr Ser Glu Ser
20 25

(331) INFORMATION FOR SEQ ID NO:331

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:331:

Gln Gln Gln Thr Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Ser Lys Ser Ser Thr Ser Glu Ser
20 25

(332) INFORMATION FOR SEQ ID NO:332

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:332:

Glu Gln Gln Ala Glu Leu Arg Thr Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Arg Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(333) INFORMATION FOR SEQ ID NO:333

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:333:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Phe Glu Ser
20 25

(334) INFORMATION FOR SEQ ID NO:334

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:334:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Thr Gly Ala Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(335) INFORMATION FOR SEQ ID NO:335

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:335:

Gln Gln Gln Pro Glu Val Arg Pro Gly Lys Gly Thr His Ala Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(336) INFORMATION FOR SEQ ID NO:336

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:336:

Gln Gln Gln Pro Glu Val Arg Pro Gly Lys Asp Thr His Ala Lys Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(337) INFORMATION FOR SEQ ID NO:337

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:337:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Glu Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(338) INFORMATION FOR SEQ ID NO:338

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:338:

Glu Gln Gln Thr Glu Leu Arg Ala Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Arg Ser Ser Thr Ser Glu Ala
20 25

(339) INFORMATION FOR SEQ ID NO:339

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:339:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Arg Glu
1 5 10 15

Lys Lys Ser Lys Pro Ser Thr Ser Glu Ser
20 25

(340) INFORMATION FOR SEQ ID NO:340

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:340:

Gln Gln Gln Ser Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Arg Glu
1 5 10 15

Lys Lys Ser Lys Pro Ser Thr Ser Glu Ser
20 25

(341) INFORMATION FOR SEQ ID NO:341

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:341:

Gln Gln Arg Ala Glu Leu Lys Pro Gly Lys Asp Thr Pro Gly Arg Glu
1 5 10 15

Lys Lys Asn Lys Pro Ser Thr Ser Glu Ser
20 25

(342) INFORMATION FOR SEQ ID NO:342

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:342:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Arg Glu
1 5 10 15

Lys Lys Ser Thr Ser Ser Thr Ser Glu Ser
20 25

(343) INFORMATION FOR SEQ ID NO:343

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:343:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Ser Thr Ser Ser Thr Ser Asp Ser
20 25

(344) INFORMATION FOR SEQ ID NO:344

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:344:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Ile Gln Gln
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(345) INFORMATION FOR SEQ ID NO:345

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:345:

Gln Gln Gln Ala Glu Phe Lys Pro Gly Lys Gly Thr Pro Gly Arg Glu
1 5 10 15

His Arg Ser Lys Pro Ser Thr Ser Glu Ser
20 25

(346) INFORMATION FOR SEQ ID NO:346

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:346:

Gln Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Ala Leu Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Asp Ser
20 25

(347) INFORMATION FOR SEQ ID NO:347

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:347:

Gln Gln Gln Pro Glu Val Lys Pro Gly Lys Gly Ala Pro Gly Lys Gly
1 5 10 15

Asn Thr Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(348) INFORMATION FOR SEQ ID NO:348

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:348:

Glu Gln Gln Ala Glu Val Arg Ala Gly Lys Gly Ser Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(349) INFORMATION FOR SEQ ID NO:349

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:349:

Gln Gln Leu Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly His Glu
1 5 10 15

Lys Lys Gly Ile Ser Ser Thr Ser Glu Ser
20 25

(350) INFORMATION FOR SEQ ID NO:350

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:350:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Lys Pro Glu Gln Glu
1 5 10 15

Lys Lys Gly Thr Ser Ser Thr Ser Glu Ser
20 25

(351) INFORMATION FOR SEQ ID NO:351

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:351:

Gln Gln Gln Pro Glu Leu Lys Pro Gly Lys Gly Arg Asn Gly Lys Glu
1 5 10 15

Asn Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(352) INFORMATION FOR SEQ ID NO:352

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:352:

Gln Gln Gln Thr Glu Leu Arg Pro Gly Arg Gly Thr Thr Gly Gln Glu
1 5 10 15

Arg Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(353) INFORMATION FOR SEQ ID NO:353

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:353:

Gln His Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly His Glu
1 5 10 15

Asn Lys Val Thr Ser Ser Thr Ser Glu Ser
20 25

(354) INFORMATION FOR SEQ ID NO:354

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:354:

Glu Gln Gln Ala Glu Leu Arg Ala Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Gln Lys Ala Lys Ser Ser Thr Ser Glu Ser
20 25

(355) INFORMATION FOR SEQ ID NO:355

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:355:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Thr Gly Thr Ser Ser Thr Thr Glu Ser
20 25

(356) INFORMATION FOR SEQ ID NO:356

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:356:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Asn Pro Gly Gln Glu
1 5 10 15

Lys Lys Ser Thr Ser Ser Ala Ser Glu Ser
20 25

(357) INFORMATION FOR SEQ ID NO:357

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:357:

Glu Gln Gln Thr Val Leu Arg Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Lys Gly Thr Ser Ala Thr Asn Glu Ser
20 25

(358) INFORMATION FOR SEQ ID NO:358

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:358:

Gln Gln Leu Thr Glu Leu Lys Pro Gly Asn Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Ser Lys Ser Ser Thr Ser Glu Ser
20 25

(359) INFORMATION FOR SEQ ID NO:359

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:359:

Gln Gln Gln Ser Val Leu Arg Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Thr Ser Ser Thr Ser Lys Ser
20 25

(360) INFORMATION FOR SEQ ID NO:360

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:360:

Leu Gln Gln Pro Val Leu Lys Pro Gly Lys Gly Ser His Gly Lys Gln
1 5 10 15

Lys Lys Asp Lys Ser Ser Thr Ser Glu Ser
20 25

(361) INFORMATION FOR SEQ ID NO:361

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:361:

Glu Gln Gln Pro Glu Thr Lys Pro Gly Lys Gly Thr Leu Gly Lys Gln
1 5 10 15

Lys Lys Ser Lys Ser Ser Thr Ser Glu Ser
20 25

(362) INFORMATION FOR SEQ ID NO:362

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:362:

Gln Gln Gln Ala Glu Leu Lys Pro Gly Gln Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Asn Lys Ser Ser Thr Pro Glu Phe
20 25

(363) INFORMATION FOR SEQ ID NO:363

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:363:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Gly Thr Ser Ser Thr Ser Glu Thr
20 25

(364) INFORMATION FOR SEQ ID NO:364

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:364:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Gly Thr Ser Thr Thr Ser Glu Thr
20 25

(365) INFORMATION FOR SEQ ID NO:365

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:365:

Glu Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Gly Thr Ser Ser Thr Ser Glu Thr
20 25

(366) INFORMATION FOR SEQ ID NO:366

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:366:

Glu Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Asp Thr Ser Ser Thr Ser Glu Thr
20 25

(367) INFORMATION FOR SEQ ID NO:367

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:367:

Glu Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Gly Thr Ser Ser Thr Ser Gly Thr
20 25

(368) INFORMATION FOR SEQ ID NO:368

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:368:

Glu Gln Gln Ala Glu Val Lys Pro Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Gly Thr Ser Ser Thr Ser Glu Thr
20 25

(369) INFORMATION FOR SEQ ID NO:369

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:369:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Val Thr Ser Ser Thr Ser Glu Thr
20 25

(370) INFORMATION FOR SEQ ID NO:370

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:370:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Ile Thr Ser Ser Thr Ser Glu Thr
20 25

(371) INFORMATION FOR SEQ ID NO:371

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:371:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Arg Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Gln Val Thr Ser Ser Thr Ser Glu Thr
20 25

(372) INFORMATION FOR SEQ ID NO:372

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:372:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Arg Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys His Val Thr Ser Ser Thr Ser Glu Thr
20 25

(373) INFORMATION FOR SEQ ID NO:373

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:373:

Glu Gln Gln Ala Glu Leu Arg Pro Gly Lys Gly Asn Thr Glu Gln Pro
1 5 10 15

Lys Gln Val Thr Ser Ser Thr Ser Glu Thr
20 25

(374) INFORMATION FOR SEQ ID NO:374

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:374:

Glu Gln Gln Ala Glu Leu Lys Pro Gly Lys Gly Asn Thr Glu Gln Pro
1 5 10 15

Lys Leu Ile Thr Ser Ser Thr Ser Glu Thr
20 25

(375) INFORMATION FOR SEQ ID NO:375

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:375:

Thr Gly Gln Ala Glu Leu Arg Pro Gly Lys Gly Ala Pro Glu Gln Gly
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Asp Arg
20 25

(376) INFORMATION FOR SEQ ID NO:376

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:376:

Gln Tyr Gln Ala Glu Leu Arg Pro Gly Lys Gly Thr Pro Arg Gln Gln
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(377) INFORMATION FOR SEQ ID NO:377

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:377:

Gln Gln Gln Ala Val Leu Arg His Gly Lys Gly Thr His Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Thr Ser Glu Ser
20 25

(378) INFORMATION FOR SEQ ID NO:378

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:378:

Gln Gln Gln Thr Lys Leu Gly Pro Gly Arg Gly Thr Pro Gly Gln Gly
1 5 10 15

Arg Lys Gly Lys Ser Ser Thr Ser Gly Ser
20 25

(379) INFORMATION FOR SEQ ID NO:379

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:379:

Glu Gln Gln Ala Glu Leu Arg Ala Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Lys Ser Ser Val Tyr Phe Ala
20 25

(380) INFORMATION FOR SEQ ID NO:380

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:380:

Glu Gln Gln Ala Glu Leu Lys Ala Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Gln Gly Glu Ser Thr Arg Ser Glu Thr
20 25

(381) INFORMATION FOR SEQ ID NO:381

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:381:

Gln Gln Lys Ala Glu Leu Ala Ala Ser Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Lys Gly Arg Ser Ser Thr Ser Glu Ser
20 25

(382) INFORMATION FOR SEQ ID NO:382

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:382:

Gln Gln Gln Thr Glu Leu Arg Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Arg Gly Lys Ser Ser Asn Leu Arg Leu
20 25

(383) INFORMATION FOR SEQ ID NO:383

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:383:

Glu Lys Val Gly Gly Leu Gln Gly Ser Ser Phe Asp Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Thr Ser Gln Arg Ala Glu Thr
20 25

(384) INFORMATION FOR SEQ ID NO:384

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:384:

Glu Gln Gln Ala Asp Leu Lys Leu Gly Lys Gly Asn Pro Glu Gln Pro
1 5 10 15

Lys Leu Ala Thr Pro Ser Thr Ser Glu Thr
20 25

(385) INFORMATION FOR SEQ ID NO:385

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:385:

Glu Gln Val Gly Gly Leu Lys Pro Gly Lys Gly Thr Pro Asp Lys Ser
1 5 10 15

Asp Val Lys Asp Asn Ala Lys Ser Glu Thr
20 25

(386) INFORMATION FOR SEQ ID NO:386

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:386:

Asp Gln Gln Pro Asp Leu Lys Pro Ser Ser Gly Ser Pro Gly His Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Thr Glu Thr
20 25

(387) INFORMATION FOR SEQ ID NO:387

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:387:

Asp Gln Gln Pro Asp Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Thr Glu Thr
20 25

(388) INFORMATION FOR SEQ ID NO:388

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:388:

Asp Gln Gln Pro Asp Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ala Glu Thr
20 25

(389) INFORMATION FOR SEQ ID NO:389

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:389:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Thr Glu Thr
20 25

(390) INFORMATION FOR SEQ ID NO:390

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:390:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Asn Thr Ser Lys Thr Thr Glu Thr
20 25

(391) INFORMATION FOR SEQ ID NO:391

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:391:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asp Pro
1 5 10 15

Ser Lys Thr Thr Ser Lys Thr Thr Glu Thr
20 25

(392) INFORMATION FOR SEQ ID NO:392

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:392:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Thr Thr Ser Lys Thr Thr Glu Thr
20 25

(393) INFORMATION FOR SEQ ID NO:393

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:393:

Asp His Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Asn Thr Ser Lys Thr Thr Glu Thr
20 25

(394) INFORMATION FOR SEQ ID NO:394

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:394:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Arg Ser Thr Ser Lys Thr Thr Glu Thr
20 25

(395) INFORMATION FOR SEQ ID NO:395

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:395:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ala Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ala Glu Thr
20 25

(396) INFORMATION FOR SEQ ID NO:396

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:396:

Glu Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ser Glu Thr
20 25

(397) INFORMATION FOR SEQ ID NO:397

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:397:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Asn Thr Ser Lys Thr Ile Glu Thr
20 25

(398) INFORMATION FOR SEQ ID NO:398

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:398:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asp Pro
1 5 10 15

Ser Lys Asn Thr Ser Lys Thr Pro Glu Thr
20 25

(399) INFORMATION FOR SEQ ID NO:399

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:399:

Glu Gln Gln Pro Ser Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Thr Glu Thr
20 25

(400) INFORMATION FOR SEQ ID NO:400

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:400:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Asn Thr Ser Glu Thr Thr Glu Thr
20 25

(401) INFORMATION FOR SEQ ID NO:401

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:401:

Asp Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Asn Thr Ser Glu Thr Thr Glx Thr
20 25

(402) INFORMATION FOR SEQ ID NO:402

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:402:

Glu Gln Gln Pro Ser Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ser Glu Thr
20 25

(403) INFORMATION FOR SEQ ID NO:403

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:403:

Glu Gln Gln Pro Ser Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Arg Thr Thr Glu Thr
20 25

(404) INFORMATION FOR SEQ ID NO:404

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:404:

Glu Gln Gln Pro Ser Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ala Glu Thr
20 25

(405) INFORMATION FOR SEQ ID NO:405

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:405:

Asp Gln Gln Pro Asp Leu Lys Pro Ser Ser Gly Phe Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Thr Glu Thr
20 25

(406) INFORMATION FOR SEQ ID NO:406

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:406:

Glu Gln Gln Pro Ser Leu Lys Pro Ser Ser Gly Ser Pro Gly Lys Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Asn Glu Thr
20 25

(407) INFORMATION FOR SEQ ID NO:407

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:407:

Glu Gln Gln Pro Ser Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Phe Lys Thr Ser Glu Thr
20 25

(408) INFORMATION FOR SEQ ID NO:408

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:408:

Glu Gln Gln Pro Ser Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Thr Thr Ser Glu Thr
20 25

(409) INFORMATION FOR SEQ ID NO:409

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:409:

Glu Gln Gln Leu Ser Leu Lys Pro Ser Ser Gly Ser Pro Gly Asn Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Thr Glu Thr
20 25

(410) INFORMATION FOR SEQ ID NO:410

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:410:

Gln Gln Gln Pro Gly Leu Lys Pro Ser Phe Gly Pro Pro Gly Lys Pro
1 5 10 15

Ser Gln Ser Thr Ser Lys Thr Thr Glu Thr
20 25

(411) INFORMATION FOR SEQ ID NO:411

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:411:

Gln Gln Lys Pro Gly Leu Ala Pro Ser Ser Gly Ser Pro Gly Lys Ser
1 5 10 15

Thr Lys Ser Asn Ser Lys Gln Thr Asp Thr
20 25

(412) INFORMATION FOR SEQ ID NO:412

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:412:

Gln Gln Lys Pro Gly Leu Ala Pro Ser Ser Gly Ser Pro Gly Lys Ser
1 5 10 15

Ala Lys Ser Asn Ser Lys Gln Thr Asp Thr
20 25

(413) INFORMATION FOR SEQ ID NO:413

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:413:

Gln Gln Lys Pro Gly Leu Ala Pro Ser Ser Gly Ser Pro Gly Lys Ser
1 5 10 15

Ala Met Ser Asn Ser Lys Gln Thr Asp Thr
20 25

(414) INFORMATION FOR SEQ ID NO:414

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:414:

Gln Gln Lys Pro Gly Leu Ala Pro Ser Ser Gly Ser Pro Gly Lys Ser
1 5 10 15

Ala Ile Ser Asn Ser Lys Gln Thr Asp Thr
20 25

(415) INFORMATION FOR SEQ ID NO:415

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:415:

Gln Gln Lys Pro Gly Leu Gln Pro Ser Ser Gly Ser Pro Gly Lys Ala
1 5 10 15

Ala Ile Ser Asn Ser Lys Gln Ser Asn Thr
20 25

(416) INFORMATION FOR SEQ ID NO:416

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:416:

Gln Gln Lys Pro Gly Leu Gln Pro Ser Ser Gly Ser Pro Gly Lys Ala
1 5 10 15

Ala Ile Ser Asn Ser Lys Gln Ala Asn Thr
20 25

(417) INFORMATION FOR SEQ ID NO:417

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:417:

Gln Gln Lys Pro Val Leu Ala Pro Ser Ser Gly Ser Pro Gly Lys Ser
1 5 10 15

Ala Met Ser Asn Ser Lys Gln Ile Asp Thr
20 25

(418) INFORMATION FOR SEQ ID NO:418

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:418:

Gln Gln Lys Pro Ser Leu Gln Pro Ser Ser Asp Ser Pro Gly Lys Ala
1 5 10 15

Ala Met Ser Asn Ser Lys Gln Ala Asp Thr
20 25

(419) INFORMATION FOR SEQ ID NO:419

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:419:

Glu Arg Val Gly Asp Leu Glu Pro Gly Arg Gly Ile Pro Gly Lys Ala
1 5 10 15

Pro Lys Gly Asp Ser Lys Lys Ile Glu Thr
20 25

(420) INFORMATION FOR SEQ ID NO:420

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:420:

Glu Arg Val Gly Asp Leu Glu Pro Glu Arg Gly Ile Pro Gly Lys Ala
1 5 10 15

Pro Lys Gly Asp Ser Lys Lys Ile Glu Thr
20 25

(421) INFORMATION FOR SEQ ID NO:421

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:421:

Glu Gln Val Gly Gly Leu Lys Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Pro Lys Gly Asp Ser Lys Lys Thr Glu Thr
20 25

(422) INFORMATION FOR SEQ ID NO:422

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:422:

Glu Gln Val Gly Gly Leu Gln Pro Gly Lys Gly Thr Ser Gly Lys Ala
1 5 10 15

Ser Lys Gly Asp Ser Lys Lys Thr Glu Thr
20 25

(423) INFORMATION FOR SEQ ID NO:423

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids

(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:423:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asx
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Ala Glu Thr
20 25

(424) INFORMATION FOR SEQ ID NO:424

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:424:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ser Lys Arg Ala Glu Thr
20 25

(425) INFORMATION FOR SEQ ID NO:425

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:425:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Arg Gly Asn Ser Lys Arg Ala Glu Thr
20 25

(426) INFORMATION FOR SEQ ID NO:426

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:426:

Glu Gln Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ser Lys Arg Ala Glu Thr
20 25

(427) INFORMATION FOR SEQ ID NO:427

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:427:

Glu Gln Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Arg Ala Glu Thr
20 25

(428) INFORMATION FOR SEQ ID NO:428

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:428:

Glu Gln Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asp Ser Arg Arg Ala Glu Thr
20 25

(429) INFORMATION FOR SEQ ID NO:429

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:429:

Glu Gln Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ser Arg Arg Ala Glu Thr
20 25

(430) INFORMATION FOR SEQ ID NO:430

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:430:

Gln Gln Val Gly Gly Leu Glu Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asx Ser Lys Arg Ala Glu Thr
20 25

(431) INFORMATION FOR SEQ ID NO:431

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:431:

Glu Gln Leu Gly Asp Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Asn Ser Lys Arg Ala Glu Thr
20 25

(432) INFORMATION FOR SEQ ID NO:432

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:432:

Glu Gln Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Thr Gly Lys Asp
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Ala Glu Thr
20 25

(433) INFORMATION FOR SEQ ID NO:433

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:433:

Gln Gln Val Gly Gly Val Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ser Lys Arg Ala Glu Thr
20 25

(434) INFORMATION FOR SEQ ID NO:434

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:434:

Gln Gln Val Gly Gly Val Gln Pro Gly Arg Gly Ile Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ser Lys Arg Pro Glu Thr
20 25

(435) INFORMATION FOR SEQ ID NO:435

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:435:

Glu Gln Val Gly Gly Val Gln Pro Gly Arg Gly Ile Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Pro Glu Thr
20 25

(436) INFORMATION FOR SEQ ID NO:436

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:436:

Gln Gln Val Gly Gly Val Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Asn Gly Asp Ser Lys Arg Pro Glu Thr
20 25

(437) INFORMATION FOR SEQ ID NO:437

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:437:

Gln Lys Val Gly Gly Val Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ser Lys Arg Thr Glu Thr
20 25

(438) INFORMATION FOR SEQ ID NO:438

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:438:

Gln Glu Val Gly Gly Val Glx Pro Gly Arg Gly Thr Pro Gly Lys Asx
1 5 10 15

Ser Lys Gly Asx Ser Lys Arg Ala Glu Thr
20 25

(439) INFORMATION FOR SEQ ID NO:439

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:439:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Asn Gly Asp Ser Lys Gln Ala Glx Thr
20 25

(440) INFORMATION FOR SEQ ID NO:440

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:440:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Ser Pro Gly Lys Asp
1 5 10 15

Thr Asn Gly Asp Ser Lys Glu Ala Glx Thr
20 25

(441) INFORMATION FOR SEQ ID NO:441

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:441:

Ala Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Asn Gly Asp Ser Lys Gln Ala Glx Ser
20 25

(442) INFORMATION FOR SEQ ID NO:442

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:442:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Val
1 5 10 15

Ser Gln Gly Asp Ser Lys Gln Ala Glx Thr
20 25

(443) INFORMATION FOR SEQ ID NO:443

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:443:

Glu Gln Val Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Val
1 5 10 15

Ser Gln Gly Asp Ser Lys Glu Pro Glx Thr
20 25

(444) INFORMATION FOR SEQ ID NO:444

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:444:

Glu Gln Leu Gly Gly Leu Gln Pro Glu Arg Gly Thr Pro Gly Lys Glu
1 5 10 15

Ser Lys Gly Asn Ser Met Arg Ala Glu Thr
20 25

(445) INFORMATION FOR SEQ ID NO:445

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:445:

Glu Gln Val Gly Asp Leu Gln Pro Gly Arg Gly Asx Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Lys Arg Val Glu Thr
20 25

(446) INFORMATION FOR SEQ ID NO:446

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:446:

Glu Gln Val Gly Asp Leu Gln Pro Gly Arg Gly Asn Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Gln Arg Pro Glu Thr
20 25

(447) INFORMATION FOR SEQ ID NO:447

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:447:

Gln Gln Val Gly Gly Val Gln Pro Gly Arg Gly Thr Leu Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ser Lys Arg Ala Glu Thr
20 25

(448) INFORMATION FOR SEQ ID NO:448

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids

(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:448:

Gln Glx Val Gly Gly Ala Glx Pro Gly Arg Gly Ser Pro Gly Lys Ala
1 5 10 15

Ser Lys Gly Asx Ser Lys Arg Ala Glu Thr
20 25

(449) INFORMATION FOR SEQ ID NO:449

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:449:

Gln Gln Val Gly Gly Leu Lys Pro Gly Arg Gly Ser Pro Gly Lys Asp
1 5 10 15

Ser Lys Gly Asn Ala Gln Arg Thr Glx Thr
20 25

(450) INFORMATION FOR SEQ ID NO:450

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:450:

Asp Gln Val Gly Gly Leu Lys Pro Gly Arg Gly Thr Pro Gly Lys Asn
1 5 10 15

Ser Asn Gly Asp Ser Lys Thr Pro Glx Thr
20 25

(451) INFORMATION FOR SEQ ID NO:451

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:451:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Thr Ser Arg Glu Asp
1 5 10 15

Ser Lys Gly Asn Ser Lys Arg Ala Glu Thr
20 25

(452) INFORMATION FOR SEQ ID NO:452

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:452:

Glu Gln Val Gly Ala Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Asp
1 5 10 15

Ser Gln Ala Asp Ser Lys Glu Ala Glx Thr
20 25

(453) INFORMATION FOR SEQ ID NO:453

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:453:

Glu Gln Leu Gly Gly Leu Gln Pro Gly Arg Gly Thr Pro Gly Lys Val
1 5 10 15

Glu Gly Ser Val Glu Thr
20

(454) INFORMATION FOR SEQ ID NO:454

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:454:

Glu Gln Val Gly Ala Phe Gln Pro Gly Arg Gly Asn Ser Gly Lys Ala
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Pro Asp Thr
20 25

(455) INFORMATION FOR SEQ ID NO:455

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:455:

Glu Gln Val Gly Ala Phe Gln Pro Gly Lys Gly Asn Ser Gly Lys Ala
1 5 10 15

Ser Lys Gly Asp Ser Lys Arg Pro Asp Thr
20 25

(456) INFORMATION FOR SEQ ID NO:456

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:456:

Glu Gln Val Gly Ala Phe Gln Pro Gly Lys Gly Asn Ser Gly Lys Ala
1 5 10 15

Ser Lys Gly Asp Ser Asn Arg Pro Asp Thr
20 25

(457) INFORMATION FOR SEQ ID NO:457

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:457:

Gln Gln Val Gly Gly Val Gln Ala Gly Arg Ala Asn Pro Gly Lys Asp
1 5 10 15

Ser Arg Gly Ile Ser Lys Arg Thr Glu Thr
20 25

(458) INFORMATION FOR SEQ ID NO:458

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:458:

Gln Gln Val Ala Glu Val Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Gln Gly Glu Ser Thr Arg Ser Glu Thr
20 25

(459) INFORMATION FOR SEQ ID NO:459

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:459:

Gln Gln Val Ala Glu Val Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Gln Gly Thr Ser Thr Arg Ser Glu Thr
20 25

(460) INFORMATION FOR SEQ ID NO:460

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:460:

Gln Gln Val Ala Glu Val Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Gln Gly Thr Ser Ala Arg Ser Glu Thr
20 25

(461) INFORMATION FOR SEQ ID NO:461

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:461:

Gln Gln Val Ala Glu Val Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Gln Gly Thr Ser Ile Arg Ser Asp Thr
20 25

(462) INFORMATION FOR SEQ ID NO:462

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:462:

Gln Gln Val Ala Glu Val Lys Pro Gly Lys Gly Thr Pro Gly Gln Glu
1 5 10 15

Lys Gln Gly Thr Ser Ile Arg Ser Asp Thr
20 25

(463) INFORMATION FOR SEQ ID NO:463

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:463:

Gln Gln Val Ala Glu Val Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Asn Gln Gly Thr Ser Thr Arg Ser Asp Thr
20 25

(464) INFORMATION FOR SEQ ID NO:464

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:464:

Gln Gln Val Gly Glu Val Lys Pro Gly Arg Gly Thr Pro Gly Gln Gln
1 5 10 15

Lys Gln Asp Thr Ser Thr Arg Ser Asp Thr
20 25

(465) INFORMATION FOR SEQ ID NO:465

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:465:

Gln Gln Val Ala Glu Val Lys Pro Gly Arg Gly Thr Pro Gly His Pro
1 5 10 15

Arg Gln Gly Ala Ser Phe Arg Ser Asp Ser
20 25

(466) INFORMATION FOR SEQ ID NO:466

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:466:

Gln Gln Val Ser Glu Leu Lys Pro Gly Lys Gly Thr Pro Gly Gln Gln
1 5 10 15

Gly Thr Gly Thr Ser Val Lys Ala Glu Thr
20 25

(467) INFORMATION FOR SEQ ID NO:467

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:467:

Glu Gln Val Ala Glu Val Lys Pro Gly Lys Gly Ser Pro Gly Lys Pro
1 5 10 15

Ser Gln Gly Lys Ser Ile Lys Ala Ser Thr
20 25

(468) INFORMATION FOR SEQ ID NO:468

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:468:

Glu Gln Val Ala Glu Val Lys Pro Gly Arg Gly Ser Pro Gly Lys Pro
1 5 10 15

Ser Gln Gly Lys Ser Ile Lys Ala Ser Thr
20 25

(469) INFORMATION FOR SEQ ID NO:469

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:469:

Gln Gln Val Ala Glu Val Lys Pro Gly Arg Gly Asp Pro Gly Arg Pro
1 5 10 15

Arg Gln Ala Ser Ser Thr Ile Ser Ala Thr
20 25

(470) INFORMATION FOR SEQ ID NO:470

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:470:

Glu Gln Val Ala Glu Val Pro Gln Gly Lys Gly Arg Pro Gly Lys Ser
1 5 10 15

Leu Gln Gly Lys Ser Leu Lys Ala Ser Thr
20 25

(471) INFORMATION FOR SEQ ID NO:471

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:471:

Gln Gln Met Ala Glu Val Lys Pro Gly Arg Gly Thr Pro Gly Lys Pro
1 5 10 15

Gly Val Val Pro Ser Phe Phe Ser Glu Thr
20 25

(472) INFORMATION FOR SEQ ID NO:472

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:472:

Gln Gln Val Ala Glu Val Lys Pro Gly Arg Gly Thr Pro Gly Arg Tyr
1 5 10 15

Ile Trp Glu Pro Ser Phe Phe Asn Glu Gly
20 25

(473) INFORMATION FOR SEQ ID NO:473

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids

(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:473:

Gln Gln Gln Ala Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Lys Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ala Ala Thr
20 25

(474) INFORMATION FOR SEQ ID NO:474

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:474:

Gln Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Lys Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ala Ala Thr
20 25

(475) INFORMATION FOR SEQ ID NO:475

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:475:

Gln Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Lys Pro
1 5 10 15

Ser Lys Ser Thr Ser Asn Thr Ala Ala Thr
20 25

(476) INFORMATION FOR SEQ ID NO:476

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:476:

Gln Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Ala Gly Lys Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ala Ala Thr
20 25

(477) INFORMATION FOR SEQ ID NO:477

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:477:

Arg Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Pro Pro Gly Lys Pro
1 5 10 15

Ser Arg Gly Thr Ser Arg Ser Ala Ala Thr
20 25

(478) INFORMATION FOR SEQ ID NO:478

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:478:

Gln Gln Gln Ala Gly Leu Lys Pro Ser Ser Gly Ser Pro Gly Arg Thr
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Ala Ala Thr
20 25

(479) INFORMATION FOR SEQ ID NO:479

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:479:

Gln Gln Glu Pro Gly Leu Arg Pro Ser Ser Gly Thr Pro Gly Arg Thr
1 5 10 15

Pro Arg Ser Thr Ser Lys Thr Ala Ala Thr
20 25

(480) INFORMATION FOR SEQ ID NO:480

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:480:

Xaa Gln Glu Pro Gly Leu Arg Pro Ser Ser Gly Ser Pro Gly Arg Thr
1 5 10 15

Pro Arg Ser Thr Ser Lys Thr Ala Ala Thr
20 25

(481) INFORMATION FOR SEQ ID NO:481

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:481:

Gln Gln Gln Pro Gly Leu Lys Pro Ser Ser Gly Ser Pro Ser Arg Val
1 5 10 15

Ser Lys Ser Thr Ser Lys Thr Pro Glu Thr
20 25

(482) INFORMATION FOR SEQ ID NO:482

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:482:

Gln His Gln Ala Gly Leu Lys Arg Ser Ser Gly Pro Pro Gly Lys Pro
1 5 10 15

Ser Thr Ser Thr Ser Lys Thr Ala Ala Thr
20 25

(483) INFORMATION FOR SEQ ID NO:483

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:483:

Glx Gln Glu Ser Gly Leu Lys Pro Thr Ser Gly Ser Pro Gly Lys Pro
1 5 10 15

Ser Lys Ser Arg Ser Lys Ala Ala Asp Ala
20 25

(484) INFORMATION FOR SEQ ID NO:484

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:484:

Gln Thr Lys Pro Thr Leu Lys Pro Thr Thr Gly Ser Pro Gly Arg Pro
1 5 10 15

Ser Lys Ser Thr Ser Lys Asp Pro Val Thr
20 25

(485) INFORMATION FOR SEQ ID NO:485

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:485:

Gln Thr Lys Pro Thr Leu Lys Pro Thr Thr Gly Ser Pro Gly Lys Pro
1 5 10 15

Ser Arg Ser Thr Ser Arg Asp Pro Val Ser
20 25

(486) INFORMATION FOR SEQ ID NO:486

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:486:

Glu Thr Arg Pro Ala Leu Lys Pro Thr Thr Gly Ser Pro Gly Lys Thr
1 5 10 15

Ser Lys Thr Thr Ser Lys Asp Pro Val Thr
20 25

(487) INFORMATION FOR SEQ ID NO:487

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:487:

Gln Asn Arg Pro Ala Leu Lys Ala Thr Thr Gly Ser Pro Gly Lys Thr
1 5 10 15

Ser Glu Thr Thr Ser Lys Asp Pro Ala Thr
20 25

(488) INFORMATION FOR SEQ ID NO:488

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:488:

Gln Thr Thr Pro Ala Leu Lys Pro Lys Thr Gly Ser Pro Gly Lys Thr
1 5 10 15

Ser Arg Thr Asp Ser Lys Asn Pro Val Thr
20 25

(489) INFORMATION FOR SEQ ID NO:489

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:489:

Gln Thr Arg Pro Ala Leu Arg Pro Thr Thr Gly Ser Pro Gly Glu Ala
1 5 10 15

Ser Glu Thr Thr Ser Lys Gly Pro Gly Thr
20 25

(490) INFORMATION FOR SEQ ID NO:490

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:490:

Gln Thr Arg Pro Ala Leu Lys Pro Thr Thr Gly Ser Pro Gly Lys Thr
1 5 10 15

Ser Glu Thr Thr Ser Arg Asp Thr Ala Tyr
20 25

(491) INFORMATION FOR SEQ ID NO:491

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:491:

Leu Glu Gly Val Gln Leu Trp Gly Gly Arg Gly Ile Ser Arg Lys Tyr
1 5 10 15

Ala Lys Gly Asn Gly Lys Arg Glu Asp Ser
20 25

(492) INFORMATION FOR SEQ ID NO:492

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:492:

Tyr Asn Asn Pro Gly Asn Gly Tyr Ile Ala
1 5 10

(493) INFORMATION FOR SEQ ID NO:493

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:493:

Tyr Ile Asn Pro Gly Lys Gly Tyr Leu Ser
1 5 10

(494) INFORMATION FOR SEQ ID NO:494

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:494:

Arg Ala Ser Gln Glu Ile Ser Gly Tyr Leu Ser
1 5 10

(495) INFORMATION FOR SEQ ID NO:495

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:495:

Arg Ala Ser Gly Asn Ile His Asn Tyr Leu Ala
1 5 10

(496) INFORMATION FOR SEQ ID NO:496

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:496:

Arg Ala Ser Gln Asp Ile Asn Asn Phe Leu Asn
1 5 10

(497) INFORMATION FOR SEQ ID NO:497

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:497:

Arg Ala Ser Gln Ser Ile Gly Asn Asn Leu His
1 5 10

(498) INFORMATION FOR SEQ ID NO:498

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:498:

Ala Ala Ser Thr Leu Asp Ser
1 5

(499) INFORMATION FOR SEQ ID NO:499

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:499:

Tyr Thr Thr Thr Leu Ala Asp
1 5

(500) INFORMATION FOR SEQ ID NO:500

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:500:

Phe Thr Ser Arg Ser Gln Ser
1 5

(501) INFORMATION FOR SEQ ID NO:501

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:501:

Lys Ala Ser Ser Leu Glu Ser
1 5

(502) INFORMATION FOR SEQ ID NO:502

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:502:

Leu Gln Tyr Leu Ser Tyr Pro Leu Thr
1 5

(503) INFORMATION FOR SEQ ID NO:503

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:503:

Gln His Phe Trp Ser Thr Pro Arg Thr
1 5

(504) INFORMATION FOR SEQ ID NO:504

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:504:

Gln Gln Gly Asn Ala Leu Pro Arg Thr
1 5

(505) INFORMATION FOR SEQ ID NO:505

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:505:

Gln Gln Tyr Asn Ser Tyr Ser
1 5

(506) INFORMATION FOR SEQ ID NO:506

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:506:

Thr Phe Gly Ile Thr
1 5

(507) INFORMATION FOR SEQ ID NO:507

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:507:

Gly Tyr Gly Val Asn
1 5

(508) INFORMATION FOR SEQ ID NO:508

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:508:

Ser Asn Gly Ile Asn
1 5

(509) INFORMATION FOR SEQ ID NO:509

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:509:

Asp Tyr Ala Met His
1 5

(510) INFORMATION FOR SEQ ID NO:510

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:510:

Glu Ile Phe Pro Gly Asn Ser Lys Thr Tyr
1 5 10

(511) INFORMATION FOR SEQ ID NO:511

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:511:

Met Ile Trp Gly Asp Gly Asn Thr Asp
1 5

(512) INFORMATION FOR SEQ ID NO:512

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:512:

Tyr Asn Asn Pro Gly Asn Gly Tyr Ile Ala
1 5 10

(513) INFORMATION FOR SEQ ID NO:513

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:513:

Ile Ser Trp Asp Ser Ser Ser Ile Gly
1 5

(514) INFORMATION FOR SEQ ID NO:514

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:514:

Arg Glu Ile Arg Tyr
1 5

(515) INFORMATION FOR SEQ ID NO:515

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:515:

Glu Arg Asp Tyr Arg Leu Asp Tyr
1 5

(516) INFORMATION FOR SEQ ID NO:516

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:516:

Ser Glu Tyr Tyr Gly Gly Ser Tyr Lys Phe Asp Tyr
1 5 10

(517) INFORMATION FOR SEQ ID NO:517

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:517:

Gly Arg Asp Tyr Tyr Asp Ser Gly Gly Tyr Phe Thr Val Ala Phe Asp
1 5 10 15

Ile

(518) INFORMATION FOR SEQ ID NO:518

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:518:

Arg Ala Ser Gln Ser Ile Ser Arg Trp Leu Ala
1 5 10

(519) INFORMATION FOR SEQ ID NO:519

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:519:

Glu Ala Ser Asn Asp Leu Ala
1 5

(520) INFORMATION FOR SEQ ID NO:520

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:520:

Asp Phe Tyr Met Glu
1 5

(521) INFORMATION FOR SEQ ID NO:521

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:521:

Ile Ile Trp Asp Asp Gly Ser Asp Gln
1 5

(522) INFORMATION FOR SEQ ID NO:522

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:522:

Gln Ala Ser Gln Ser Ile Ile Lys Tyr Leu Asn
1 5 10

Claims

1. A method for determining how to humanize a rodent antibody or fragment thereof by resurfacing, said method comprising:
(a) determining the conformational structure of the variable region of said rodent antibody or fragment thereof by constructing a three-dimensional model of said rodent antibody variable region;
(b) generating sequence alignments from relative accessibility distributions from x-ray crystallographic structures of a sufficient number of rodent antibody variable region heavy and light chains to give a set of heavy and light chain framework positions wherein said set is identical in 98% of said sufficient number of rodent antibody heavy and light chains;

(c) defining for said rodent antibody or fragment thereof to be humanized a set of heavy and light chain surface exposed amino acid residues using said set of framework positions generated in said step (b);
(d) identifying from human antibody amino acid sequences a set of heavy and light chain surface exposed amino acid residues that is most closely identical to said set of surface exposed amino acid residues defined in said step (c), wherein said heavy and light chain from said human antibody are or are not naturally paired;
(e) substituting, in the amino acid sequence of said rodent antibody or fragment thereof to be humanized said set of heavy and light chain surface exposed amino acid residues defined in said step (c) with said set of heavy and light chain surface exposed amino acid residues identified in said step (d);
(f) constructing a three-dimensional model of said variable region of said rodent antibody or fragment thereof resulting from the substituting specified in said step (e);
(g) identifying, by comparing said three-dimensional models constructed in said steps (a) and (f), any amino acid residues from said set identified in said step (d), that are within 5 Angstroms of any atom of any residue of the complementarity determining regions of said rodent antibody or fragment thereof to be humanized; and
(h) changing any residues identified in said step (g) from the human to the original rodent amino acid residue to thereby define a rodent antibody humanizing set of surface exposed amino acid residues; with the proviso that said step (a) need not be conducted first, but must be conducted prior to said step (g).

2. The method of claim 1, wherein said rodent antibody is an antibody fragment.

3. The method of claim 2, wherein said rodent antibody fragment is a single chain antibody, a F_V fragment, a Fab fragment, a Fab_2 fragment or a Fab' fragment.

4. The method of claim 1 or 2, wherein said step (d) identifies a set of naturally paired heavy and light chain surface exposed amino acid residues that is most closely identical to said set of surface exposed amino acid residues defined in said step (c).

5. The method of claim 1 or 2, wherein said surface exposed amino acid residues are those residues whose solvent accessibility is above 30%.

6. The method of claim 1 or 2, wherein the rodent antibody or fragment thereof to be humanized is a murine antibody.

7. The method of claim 6, wherein said set of framework positions of surface exposed amino acid residues is defined by the set shown in Table 1 and the alignments set forth in Figures 3A and 3B.

Light Chain		
Position	Human	Mouse
1	D 51 E 34 A 5 S 5	D 76 Q 9 E 6
3	V 38 Q 24 S 24 Y 6	V 63 Q 22 L 5
5	T 61 L 37	T 87
9	P 26 S 26 G 17 A 14 L 7	S 36 A 29 L 17 P 5
15	P 62 V 25 L 12	L 47 P 30 V 8 A 7
18	R 57 S 18 T 13 P 6	R 38 K 22 S 13 Q 12 T 9
46	P 94	P 82 S 9
47	G 89	G 71 D 18
51	K 43 R 31	K 70 Q 13 R 8 T 5
63	G 91	G 98

66	D 43 S 25 A 9	D 38 A 26 S 26
73	S 96	S 90 I 5
76	D 43 T 18 S 16 E 15	D 67 S 15 A 5 K 5
86	P 44 A 27 S 17 T 8	A 50 P 11 T 8 E 7 Q 6
87	E 71 D 11 G 7	E 91 D 6
111	K 74 R 12 N 6	K 93
115	K 54 L 40	K 87 L 5
116	R 60 G 33 S 5	R 89 G 9
117	Q 50 T 37 E 6 P 6	A 74 Q 14 P 5 R 5
	Heavy Chain	
Position	Human	Mouse
118	E 47 Q 46	E 59 Q 29 D 10
120	Q 83 T 7	Q 68 K 26
122	V 59 L 15 Q 13	Q 57 V 27 L 5 K 5
126	G 54 A 23 P 18	G 36 P 30 A 29
127	G 53 E 22 A 14 D 7	E 45 G 43 S 6
128	L 61 V 31 F 7	L 96
130	K 46 Q 41 E 5	K 52 Q 27 R 17
131	P 95	P 91 A 5
132	G 74 S 16 T 7	G 82 S 17
136	R 53 K 23 S 17 T 7	K 66 S 17 R 13
143	G 96	G 98
145	T 46 S 32 N 9 I 7	T 63 S 19 N 7 A 5 D 5
160	P 84 S 10	P 89 H 7
161	G 93	G 71 E 24
162	K 76 Q 10 R 8	K 50 Q 30 N 10 H 5
183	D 26 P 25 A 17 Q 10 T 7	E 31 P 22 D 17 A 12 Q 11
184	S 70 K 9 P 8	K 42 S 37 T 6
186	K 53 Q 22 R 7 N 5	K 83 Q 7
187	G 66 S 21 T 5	G 62 S 18 D 10
195	T 30 D 26 N 19 K 7	T 36 K 30 N 26 D 6
196	S 91	S 76 A 16
197	K 65 I 8 T 8 R 5	S 46 K 34 Q 11

208	R 46 T 18 K 17 D 6	S 67 A 14 T 11
209	A 50 P 21 S 13 T 8	E 88 D 7
210	E 46 A 18 D 13 S 9 Z 8 V 5	T 53 S 43
212	T 91	

Table 1

8. The method of claim 1 or 2, wherein the rodent antibody or fragment thereof to be humanized is murine antibody anti-N901.

9. The method of claim 8, wherein said set of framework positions of surface exposed amino acid residues is defined by the set shown in Table 1 and the alignments set forth in Figures 3A and 3B.

Light Chain		
Position	Human	Mouse
1	D 51 E 34 A 5 S 5	D 76 Q 9 E 6
3	V 38 Q 24 S 24 Y 6	V 63 Q 22 L 5
5	T 61 L 37	T 87
9	P 26 S 26 G 17 A 14 L 7	S 36 A 29 L 17 P 5
15	P 62 V 25 L 12	L 47 P 30 V 8 A 7
18	R 57 S 18 T 13 P 6	R 38 K 22 S 13 Q 12 T 9
46	P 94	P 82 S 9
47	G 89	G 71 D 18
51	K 43 R 31	K 70 Q 13 R 8 T 5
63	G 91	G 98
66	D 43 S 25 A 9	D 38 A 26 S 26
73	S 96	S 90 I 5
76	D 43 T 18 S 16 E 15	D 67 S 15 A 5 K 5
86	P 44 A 27 S 17 T 8	A 50 P 11 T 8 E 7 Q 6
87	E 71 D 11 G 7	E 91 D 6
111	K 74 R 12 N 6	K 93
115	K 54 L 40	K 87 L 5
116	R 60 G 33 S 5	R 89 G 9
117	Q 50 T 37 E 6 P 6	A 74 Q 14 P 5 R 5
Heavy Chain		
Position	Human	Mouse
118	E 47 Q 46	E 59 Q 29 D 10

120	Q 83 T 7	Q 68 K 26
122	V 59 L 15 Q 13	Q 57 V 27 L 5 K 5
126	G 54 A 23 P 18	G 36 P 30 A 29
127	G 53 E 22 A 14 D 7	E 45 G 43 S 6
128	L 61 V 31 F 7	L 96
130	K 46 Q 41 E 5	K 52 Q 27 R 17
131	P 95	P 91 A 5
132	G 74 S 16 T 7	G 82 S 17
136	R 53 K 23 S 17 T 7	K 66 S 17 R 13
143	G 96	G 98
145	T 46 S 32 N 9 I 7	T 63 S 19 N 7 A 5 D 5
160	P 84 S 10	P 89 H 7
161	G 93	G 71 E 24
162	K 76 Q 10 R 8	K 50 Q 30 N 10 H 5
183	D 26 P 25 A 17 Q 10 T 7	E 31 P 22 D 17 A 12 Q 11
184	S 70 K 9 P 8	K 42 S 37 T 6
186	K 53 Q 22 R 7 N 5	K 83 Q 7
187	G 66 S 21 T 5	G 62 S 18 D 10
195	T 30 D 26 N 19 K 7	T 36 K 30 N 26 D 6
196	S 91	S 76 A 16
197	K 65 I 8 T 8 R 5	S 46 K 34 Q 11
208	R 46 T 18 K 17 D 6	S 67 A 14 T 11
209	A 50 P 21 S 13 T 8	E 88 D 7
210	E 46 A 18 D 13 S 9 Z 8 V 5	T 53 S 43
212	T 91	

Table 1

10. A method for producing a humanized rodent antibody or fragment thereof from a rodent antibody or fragment thereof by resurfacing, said method comprising.
(I) carrying out the method of claim 1; and
(II) modifying the rodent antibody or fragment thereof by replacing the set of rodent antibody surface exposed amino acid residues with the rodent antibody humanizing set of surface exposed amino acid residues defined in said step (h).

11. The method of claim 10, wherein said rodent antibody is an antibody fragment.

12. The method of claim 11, wherein said rodent antibody fragment is a single chain antibody, a F_V fragment,

a Fab fragment, a Fab_2 fragment or a Fab' fragment.

13. The method of claim 10 or 11, wherein said step (d) identifies a set of naturally paired heavy and light chain surface exposed amino acid residues that is most closely identical to said set of surface exposed amino acid residues defined in said step (c).

14. The method of claim 10 or 11, wherein said surface exposed amino acid residues are those residues whose solvent accessibility is above 30%.

15. The method of claim 10 or 11, wherein the rodent antibody or fragment thereof to be humanized is a murine antibody.

16. The method of claim 15, wherein said set of framework positions of surface exposed amino acid residues is defined by the set shown in Table 1 and the alignments set forth in Figures 3A and 3B.

Light Chain		
Position	Human	Mouse
1	D 51 E 34 A 5 S 5	D 76 Q 9 E 6
3	V 38 Q 24 S 24 Y 6	V 63 Q 22 L 5
5	T 61 L 37	T 87
9	P 26 S 26 G 17 A 14 L 7	S 36 A 29 L 17 P 5
15	P 62 V 25 L 12	L 47 P 30 V 8 A 7
18	R 57 S 18 T 13 P 6	R 38 K 22 S 13 Q 12 T 9
46	P 94	P 82 S 9
47	G 89	G 71 D 18
51	K 43 R 31	K 70 Q 13 R 8 T 5
63	G 91	G 98
66	D 43 S 25 A 9	D 38 A 26 S 26
73	S 96	S 90 I 5
76	D 43 T 18 S 16 E 15	D 67 S 15 A 5 K 5
86	P 44 A 27 S 17 T 8	A 50 P 11 T 8 E 7 Q 6
87	E 71 D 11 G 7	E 91 D 6
111	K 74 R 12 N 6	K 93
115	K 54 L 40	K 87 L 5
116	R 60 G 33 S 5	R 89 G 9
117	Q 50 T 37 E 6 P 6	A 74 Q 14 P 5 R 5
Heavy Chain		
Position	Human	Mouse
118	E 47 Q 46	E 59 Q 29 D 10
120	Q 83 T 7	Q 68 K 26
122	V 59 L 15 Q 13	Q 57 V 27 L 5 K 5

126	G 54 A 23 P 18	G 36 P 30 A 29
127	G 53 E 22 A 14 D 7	E 45 G 43 S 6
128	L 61 V 31 F 7	L 96
130	K 46 Q 41 E 5	K 52 Q 27 R 17
131	P 95	P 91 A 5
132	G 74 S 16 T 7	G 82 S 17
136	R 53 K 23 S 17 T 7	K 66 S 17 R 13
143	G 96	G 98
145	T 46 S 32 N 9 I 7	T 63 S 19 N 7 A 5 D 5
160	P 84 S 10	P 89 H 7
161	G 93	G 71 E 24
162	K 76 Q 10 R 8	K 50 Q 30 N 10 H 5
183	D 26 P 25 A 17 Q 10 T 7	E 31 P 22 D 17 A 12 Q 11
184	S 70 K 9 P 8	K 42 S 37 T 6
186	K 53 Q 22 R 7 N 5	K 83 Q 7
187	G 66 S 21 T 5	G 62 S 18 D 10
195	T 30 D 26 N 19 K 7	T 36 K 30 N 26 D 6
196	S 91	S 76 A 16
197	K 65 I 8 T 8 R 5	S 46 K 34 Q 11
208	R 46 T 18 K 17 D 6	S 67 A 14 T 11
209	A 50 P 21 S 13 T 8	E 88 D 7
210	E 46 A 18 D 13 S 9 Z 8 V 5	T 53 S 43
212	T 91	

Table 1

17. The method of claim 10 or 11, wherein the rodent antibody or fragment thereof to be humanized is murine antibody anti-N901.

18. The method of claim 17, wherein said set of framework positions of surface exposed amino acid residues is defined by the set shown in Table 1 and the alignments set forth in Figures 3A and 3B.

Light Chain		
Position	Human	Mouse
1	D 51 E 34 A 5 S 5	D 76 Q 9 E 6
3	V 38 Q 24 S 24 Y 6	V 63 Q 22 L 5

5	T 61 L 37	T 87
9	P 26 S 26 G 17 A 14 L 7	S 36 A 29 L 17 P 5
15	P 62 V 25 L 12	L 47 P 30 V 8 A 7
18	R 57 S 18 T 13 P 6	R 38 K 22 S 13 Q 12 T 9
46	P 94	P 82 S 9
47	G 89	G 71 D 18
51	K 43 R 31	K 70 Q 13 R 8 T 5
63	G 91	G 98
66	D 43 S 25 A 9	D 38 A 26 S 26
73	S 96	S 90 I 5
76	D 43 T 18 S 16 E 15	D 67 S 15 A 5 K 5
86	P 44 A 27 S 17 T 8	A 50 P 11 T 8 E 7 Q 6
87	E 71 D 11 G 7	E 91 D 6
111	K 74 R 12 N 6	K 93
115	K 54 L 40	K 87 L 5
116	R 60 G 33 S 5	R 89 G 9
117	Q 50 T 37 E 6 P 6	A 74 Q 14 P 5 R 5
	Heavy Chain	
Position	Human	Mouse
118	E 47 Q 46	E 59 Q 29 D 10
120	Q 83 T 7	Q 68 K 26
122	V 59 L 15 Q 13	Q 57 V 27 L 5 K 5
126	G 54 A 23 P 18	G 36 P 30 A 29
127	G 53 E 22 A 14 D 7	E 45 G 43 S 6
128	L 61 V 31 F 7	L 96
130	K 46 Q 41 E 5	K 52 Q 27 R 17
131	P 95	P 91 A 5
132	G 74 S 16 T 7	G 82 S 17
136	R 53 K 23 S 17 T 7	K 66 S 17 R 13
143	G 96	G 98
145	T 46 S 32 N 9 I 7	T 63 S 19 N 7 A 5 D 5
160	P 84 S 10	P 89 H 7
161	G 93	G 71 E 24

162	K 76 Q 10 R 8	K 50 Q 30 N 10 H 5
183	D 26 P 25 A 17 Q 10 T 7	E 31 P 22 D 17 A 12 Q 11
184	S 70 K 9 P 8	K 42 S 37 T 6
186	K 53 Q 22 R 7 N 5	K 83 Q 7
187	G 66 S 21 T 5	G 62 S 18 D 10
195	T 30 D 26 N 19 K 7	T 36 K 30 N 26 D 6
196	S 91	S 76 A 16
197	K 65 I 8 T 8 R 5	S 46 K 34 Q 11
208	R 46 T 18 K 17 D 6	S 67 A 14 T 11
209	A 50 P 21 S 13 T 8	E 88 D 7
210	E 46 A 18 D 13 S 9 Z 8 V 5	T 53 S 43
212	T 91	

Table 1

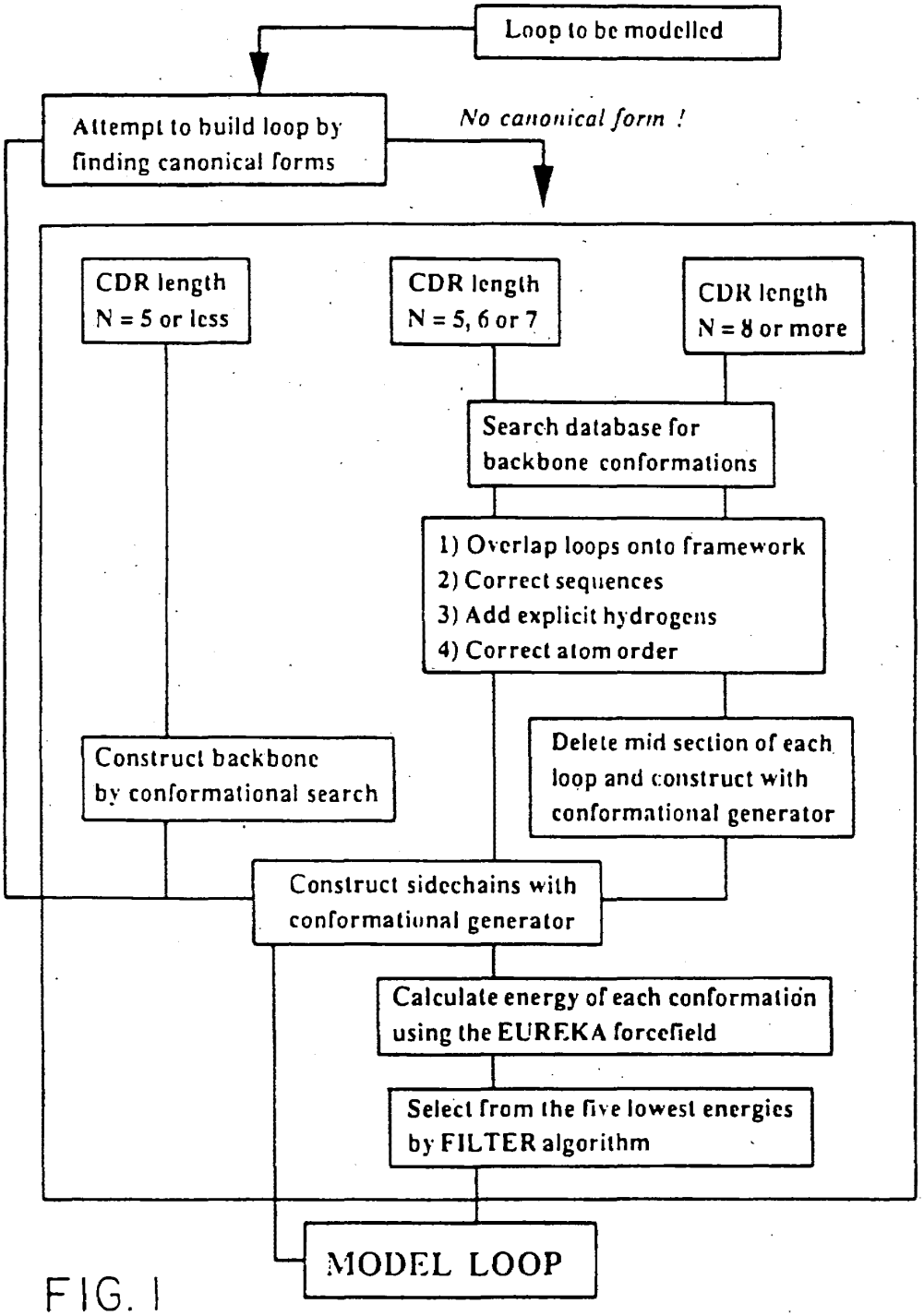


FIG. 1

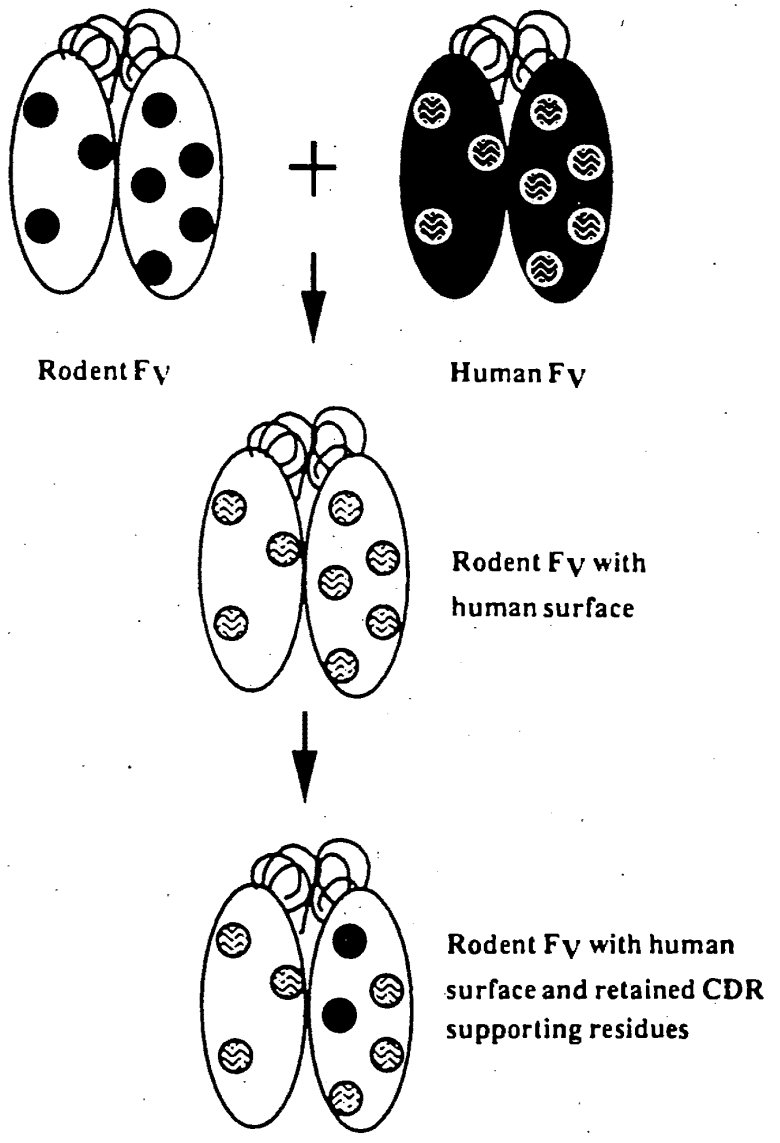
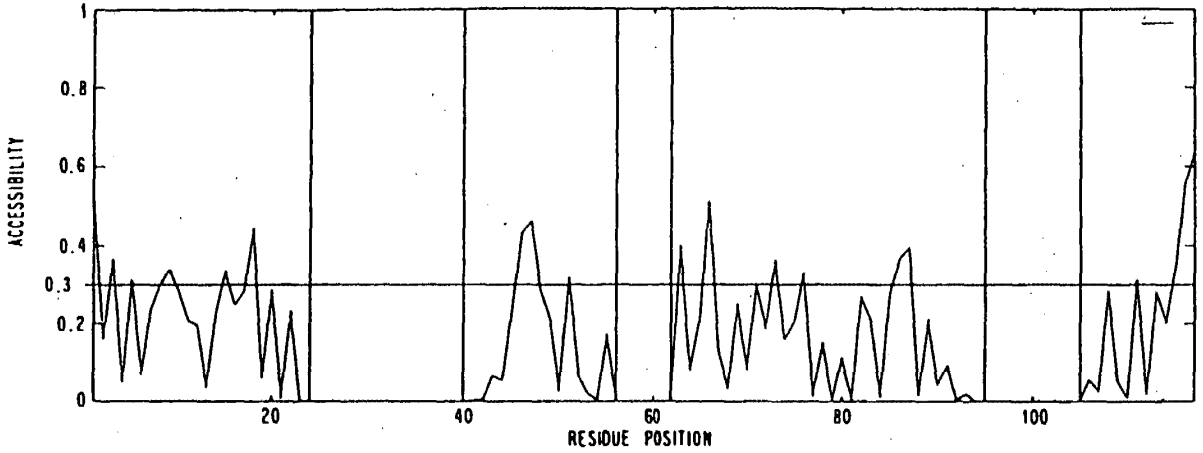


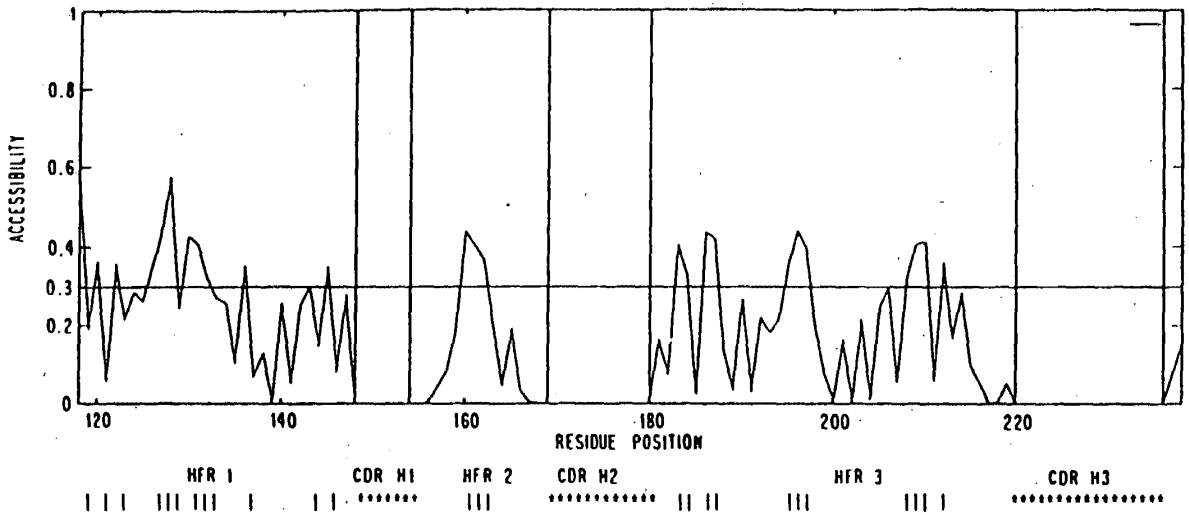
FIG. 2

FIG. 3A




```
                LFR 1          CDR L1          LFR 2    CDR L2          LFR 3          CDR L3    LFR 4
                | | |  |    |  |  **************** ||  |  *******|  |    |  |     ||  **********  |  |||
g1b2 DIQMTQSPSSLSASLGERVSLTCRASQEISG------YLSWLQQKPDGTIKRLIYAASTLDSGVPKRFSGRRSGSDYSLTISSLESEDFADYYCLQYLS--YPLTFGAGTKLELKRA  (SEQ ID NO: 1)
1fdl DIQMTQSPASLSASVGETVTITCRASGNIHN------YLAWYQQKQGKSPQLLVYYTTTLADGVPSRFSGSGSGTQYSLKINSLQPEDFGSYYCQHFWS--TPRTFGGGTKLEIKRR  (SEQ ID NO: 2)
2hfl DIVLTQSPAIMSASPGEKVTMTCSASSSVN-------YMYWYQQKSGTSPKRWIYDTSKLASGVPVRFSGSGSGTSYSLTISSMETEDAAEYYCQQWSR--NP-TFGGGTKLEIKRA  (SEQ ID NO: 3)
3hfm DIVLTQSPATLSVTPGNSVSLSCRASQSIGN------NLHWYQQKSHESPRLLIKYASQSISGIPSRFSGSGSGTDFTLSINSVETEDFGMYFCQQSNS--WPYTFGGGTKLEIKRA  (SEQ ID NO: 4)
2fbj EIVLTQSPAITAASLGQKVTITCSASSSVSS-------LHWYQQKSGTSPKPWIYEISKLASGVPARFSGSGSGTSYSLTINTMEAEDAAIYYCQQWTY--PLITFGAGTKLELKRA  (SEQ ID NO: 5)
2fb4 ESVLTQPPSASG-TPGQRVTISCTGTSSNIG----SITVNWYQQLPGMAPKLLIYRDAMRPSGVPTRFSGSKSGTSASLAISGLEAEDESDYYCASWNSSDNSYVFGTGTKVTVLGQ  (SEQ ID NO: 6)
2mcp DIVMTQSPSSLSVSAGERVTMSCKSSQSLLNSGNQKNFLAWYQQKPGQPPKLLIYGASTRESGVPDRFTGSGSGTDFTLTISSVQAEDLAVYYCQNDHS--YPLTFGAGTKLEIKRA  (SEQ ID NO: 7)
7fab -SVLTQPPSVSG-APGQRVTISCTGSSSNIG---AGNHVKWYQQLPGTAPKLLIFHNNA-------RFSVSKSGSSATLAITGLQAEDEADYYCQSYDR--SLRVFGGGTKLTVLRQ  (SEQ ID NO: 8)
4fab DVVMTQTPLSLPVSLGDQASISCRSSQSLVHS-QGNTYLRWYLQKPGQSPKVLIYKVSNRFSGVPNRFSGSGSGTDFTLKISRVEAEDLGVYFCSQSTH--VPWTFGGGTKLEIKRA  (SEQ ID NO: 9)
1f19 DIQMTQTTSSLSASLGDRVTISCRASQDISN------YLNWYQQKPDGTVKLLVYYTSRLESGVPSRFSGSGSGTDYSLTISNLEHEDIATYFCQQGST--TPRTFGGGTKLEIKRR  (SEQ ID NO: 10)
6fab DIQMTQTPSSLSASLGDRVSISCRASQDINN------FLNWYQQKPDGTIKLLIYFTSRSQSGVPSRFSGSGSGTDYSLTISNLEQEDIATYFCQQGNA--LPRTFGGGTKLEIKRA  (SEQ ID NO: 11)
1dfb DIQMTQSPSTLSASVGDRVTITCRASQSISR------WLAWYQQKPGKAPKLLIYKASSLESGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCQQYNS----YSFGPGTKVDIKRT  (SEQ ID NO: 12)
```

FIG. 3B




```
g1b2 QVQLQQSGTELARPGASVRLSCKASGYTFTTFGIT--WVKQRTGQGLEWIGEIFPGNS--KTYYAERFKGKATLTADKSSTTAYMQLSSLTSEDSAVYFCAREIR----------YWG  (SEQ ID NO: 13)
1fdl QVQLKESGPGLVAPSQSLSITCTVSGFSLTGYGVN--WVRQPPGKGLEWLGMIWGDG---NTDYNSALKSRLSISKDNSKSQVFLKMNSLHTDDTARYYCARERDYRL---------DYWG  (SEQ ID NO: 14)
2hfl -VQLQQSGAELMKPGASVKISCKASGYTFSDYWIE--WVKQRPGHGLEWIGEILPGSG--STNYHERFKGKATFTADTSSSTAYMQLNSLTSEDSGVYYCLHGNYDF----------DGWG  (SEQ ID NO: 15)
3hfm DVQLQESGPSLVKPSQTLSLTCSVTGDSITSDYWS--WIRKFPGNRLEYMGYVSYSG---STYYNPSLKSRISITRDTSKNQYYLDLNSVTTEDTATYYCANWDG----------DYWG  (SEQ ID NO: 16)
2fbj EVKLLESGGGLVQPGGSLKLSCAASGFDFSKYWMS--WVRQAPGKGLEWIGEIHPDSG--TINYTPSLKDKFIISRDNAKNSLYLQMSKVRSEDTALYYCARLHYYGYN--------AYWG  (SEQ ID NO: 17)
2fb4 EVQLVQSGGGVVQPGRSLRLSCSSSGFIFSSYAMY--WVRQAPGKGLEWVAIIWDDGS--DQHYADSVKGRFTISRNDSKNTLFLQMDSLRPEDTGVYFCARDGGHGFCSSASCFGPDYWG  (SEQ ID NO: 18)
2mcp EVKLVESGGGLVQPGGSLRLSCATSGFTFSDFYME--WVRQPPGKRLEWIAASRNKGNKYTTEYSASVKGRFIVSRDTSQSILYLQMNALRAEDTAIYYCARNYYGSTWYF------DVWG  (SEQ ID NO: 19)
7fab -VQLEQSGPGLVRPSQTLSLTCTVSGTSFDDYYST--WVRQPPGRGLEWIGYVFYHG---TSDTDTPLRSRVTMLVNTSKNQFSLRLSSVTAADTAVYYCARNLIAGCI--------DVWG  (SEQ ID NO: 20)
4fab EVKLDETGGGLVQPGRPMKLSCVASGFTFSDYWMN--WVRQSPEKGLEWVAQIRNKPYNYETYYSDSVKGRFTISRDDSKSSVYLQMNNLRVEDMGIYYCTGSYYGM----------DYWG  (SEQ ID NO: 21)
1f19 QVQLKESGAELVAASSSVKMSCKASGYTFTSYGVN--WVKQRPGQGLEWIGYINPGKG--YLSYNEKFKGKTTLTVDRSSSTAYMQLRSLTSEDSAVYFCARSFYGGSDLAVYYF--DSWG  (SEQ ID NO: 22)
6fab EVQLQQSGVELVRAGSSVKMSCKASGYTFTSNGIN--WVKQRPGQGLEWIGYINPGNG--YTAYNEKFKGKTTLTVDKSSSTAYMQLRSLTSEDSAVYFCARSEYYGGSYKF-----DYWG  (SEQ ID NO: 23)
1dfb EVQLVESGGGLVQPGRSLRLSCAASGFTFNDYAMH--WVRQAPGKGLEWVSGISWDSS--SIGYADSVKGRFTISRDNAKNSLYLQMNSLRAEDMALYYCVKGRDYYQSGGYYTVAFDIWG  (SEQ ID NO: 24)
```

FIG. 4A

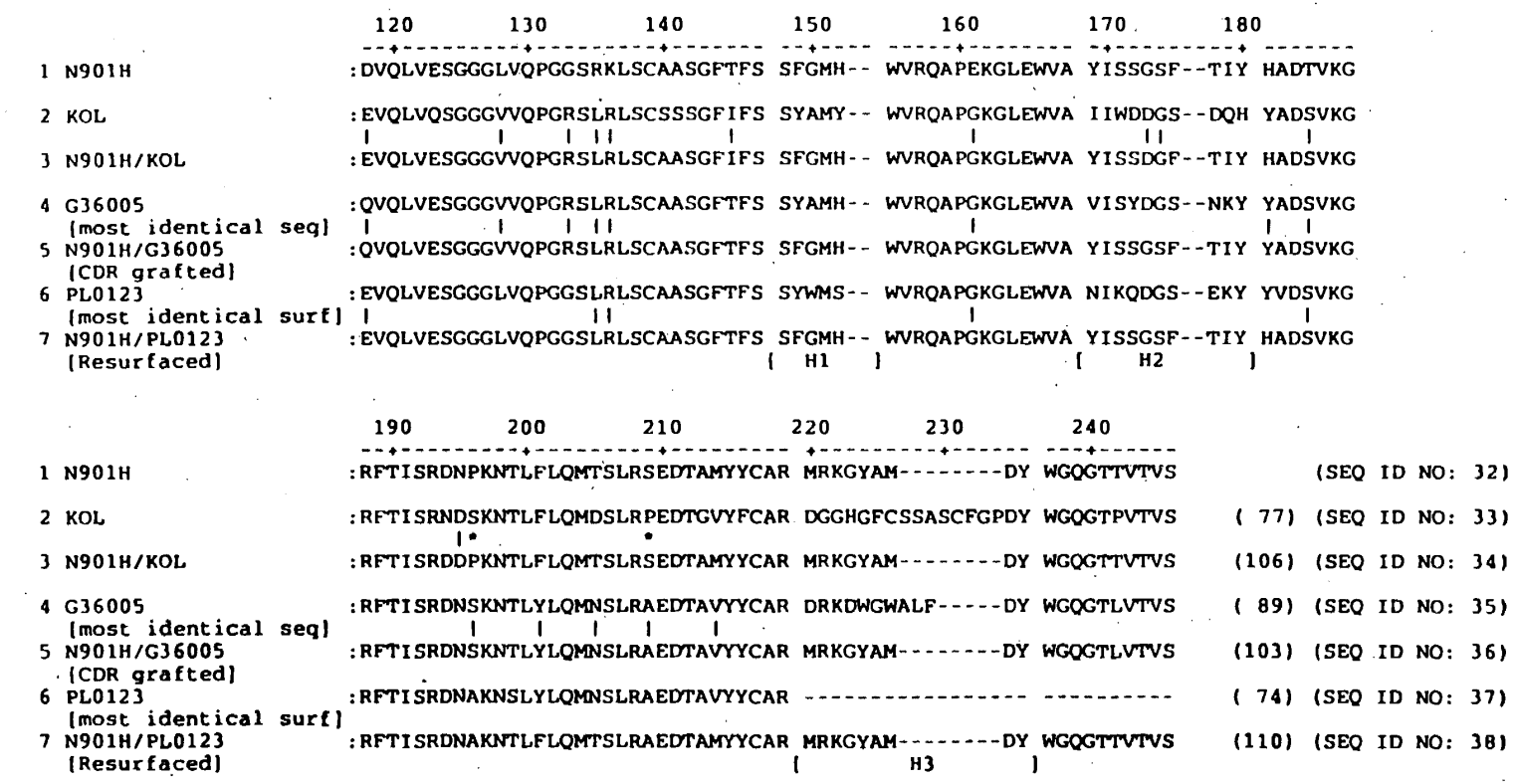
Light Chain Sequences

```
                                10        20        30        40        50        60        70
                                ---------+---------+---------+---------+---------+---------+---------+
1 N901L                        :DVLMTQTPLSLPVSLGDQASISC RSSQIIIHSDGNTY-LE WFLQKPGQSPKLLIY KVSNRFS GVPDRFSG

2 KOL                          :QSVLTQPPSASG-TPGQRVTISC SGTSSNIGS----STVN WYQQLPGMAPKLLIY RDAMRPS GVPDRFSG
                                | *     | .   |  |
3 N901L/KOL                    :QVLMTQTPSSLPVTLGQQASISC RSSQIIIHSDGNTY-LE WFLQKPGQSPKLLIY KVSNRFS GVPDRFSG

4 KV2F$HUMAN                   :DVVMTQSPLSLPVTLGQPASISC RSSQSLVYSDGNTY-LN WFQQRPGQSPRRLIY KVSNRDS GVPDRFSG
  [most identical seq]           *   |       |  ||                          *
5 N901L/KV2F                   :DVLMTQSPLSLPVTLGQPASISC RSSQIIIHSDGNTY-LE WFQQRPGQSPRLLIY KVSNRFS GVPDRFSG
  [CDR grafted]
6 KV4B$HUMAN                   :DIVMTQSPDSLAVSLGERATINC KSSQSVLYSSNNKNYLA WYQQKPGQPPKLLIY WASTRES GVPDRFSG
  [most identical surf]          *      |       |
7 N901L/KV4B                   :DVLMTQTPDSLPVSLGDRASISC RSSQIIIHSDGNTY-LE WFLQKPGQSPKLLIY KVSNRFS GVPDRFSG
  [Resurfaced]                                  [       L1       ]               [  L2   ]

                                80        90        100       110
                                ---------+---------+------ ---+------- --+------
1 N901L                        :SGSGTDFTLMISRVEAEDLGVYYC FQGSH--VPHT FGGGTKLEI-          (SEQ ID NO: 25)

2 KOL                          :SKSGASASLAIGGLQSEDETDYYC AAWDVSLNAYV FGTGTKVTVL    ( 44)  (SEQ ID NO: 26)
                                      |   |         |                      *
3 N901L/KOL                    :SGSGTSFTLAISRVEAEDEGVYYC FQGSH--VPHT FGGGTKLEI-    (104)  (SEQ ID NO: 27)

4 KV2F$HUMAN                   :SGSGTDFTLKISRVEAEDVGVYYC MQGTH--WSWT FGQGTKVEIK    ( 87)  (SEQ ID NO: 28)
  [most identical seq]                 |         |                    *
5 N901L/KV2F                   :SGSGTDFTLKISRVEAEDVGVYYC FQGSH--VPHT FGGGTKVEI-    (101)  (SEQ ID NO: 29)
  [CDR grafted]
6 KV4B$HUMAN                   :SGSGTDFTLTISSLQAEDVAVYYC QQYDT---IPT FGGGTKVEIK    ( 71)  (SEQ ID NO: 30)
  [most identical surf]
7 N901L/KV4B                   :SGSGTDFTLMISRVEAEDLGVYYC FQGSH--VPHT FGGGTKLEI-    (109)  (SEQ ID NO: 31)
  [Resurfaced]                                    [     L3    ]
```


EP 0 592 106 A1

FIG. 4B

Heavy Chain Sequences



220

EP 0 592 106 A1

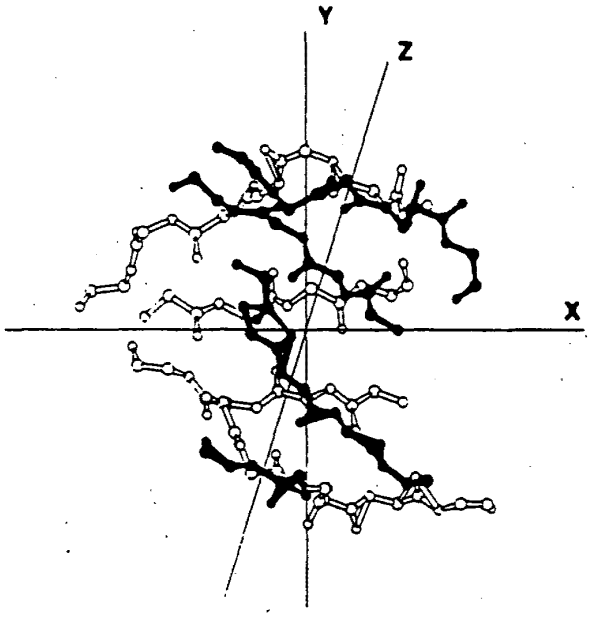
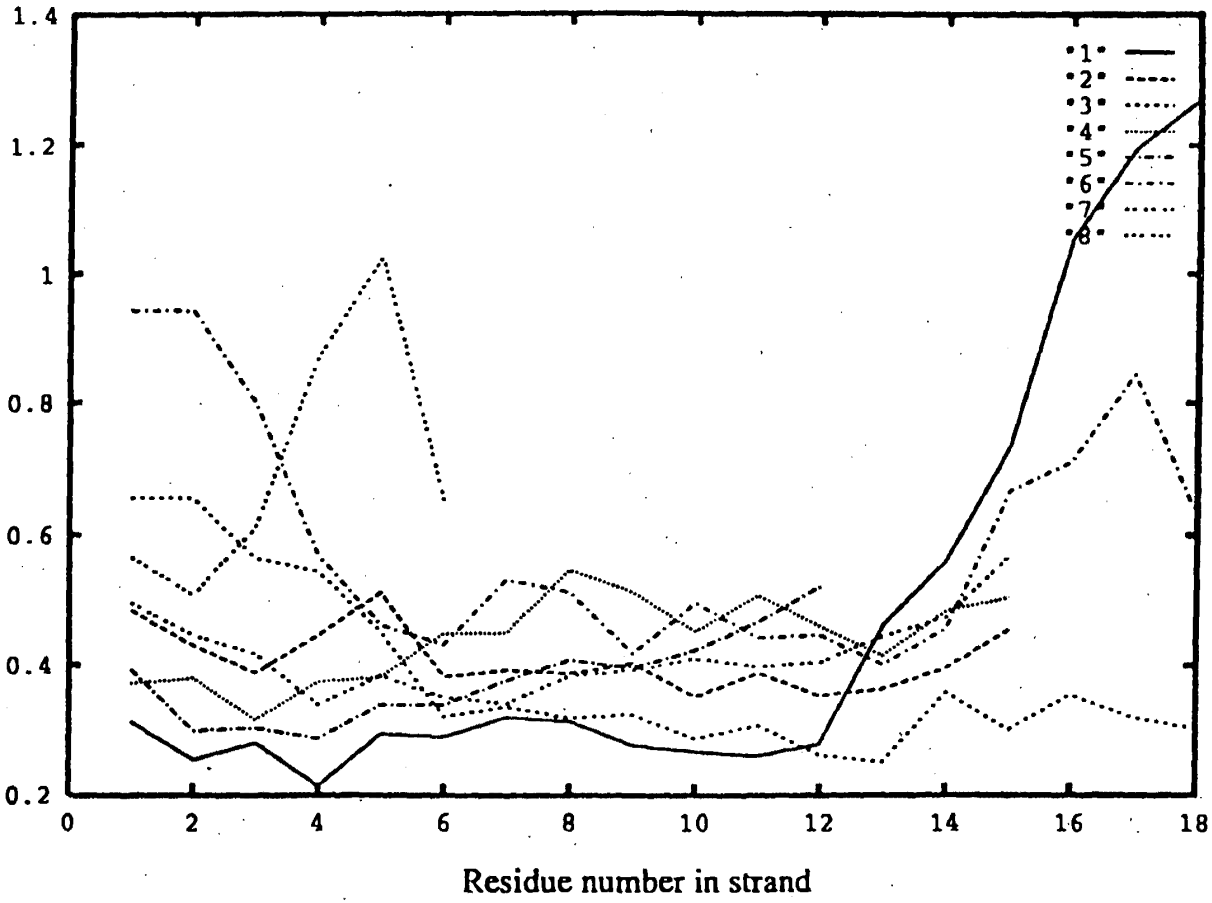


FIG. 5

FIG. 6



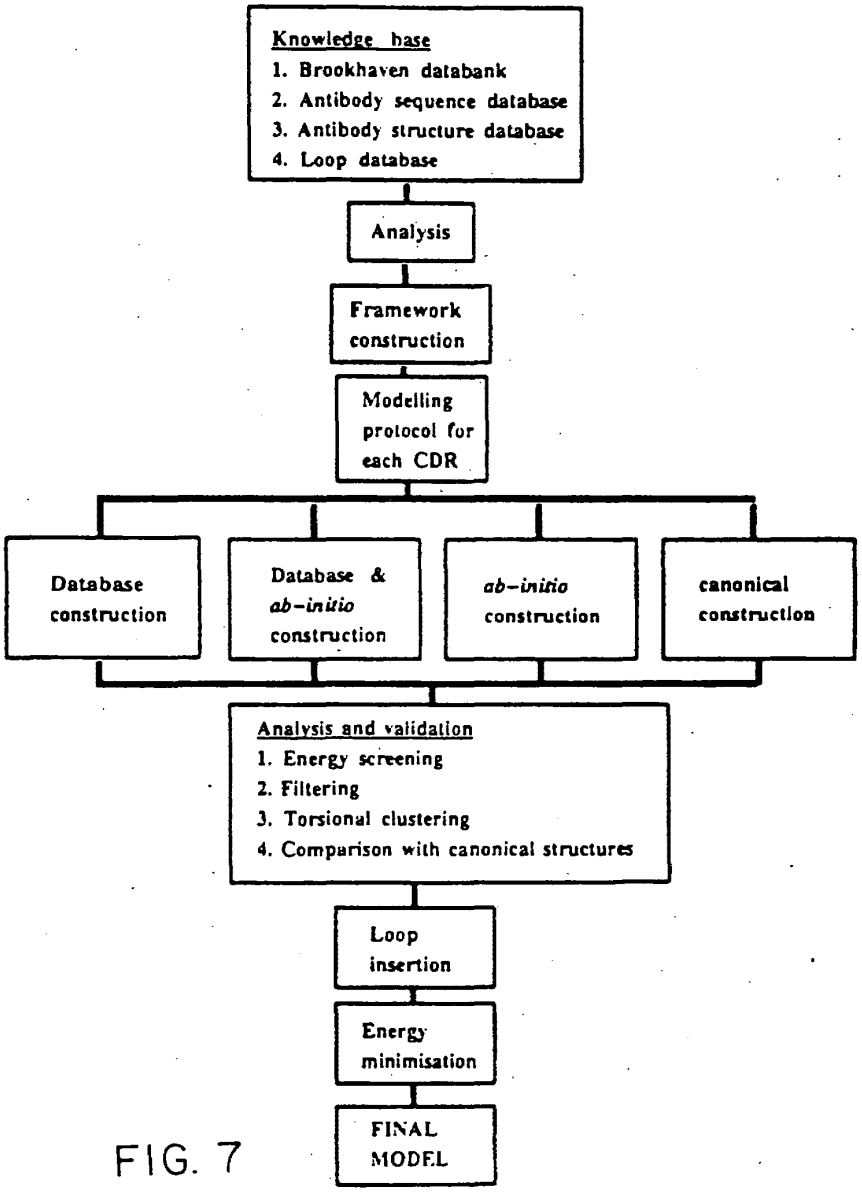


FIG. 7

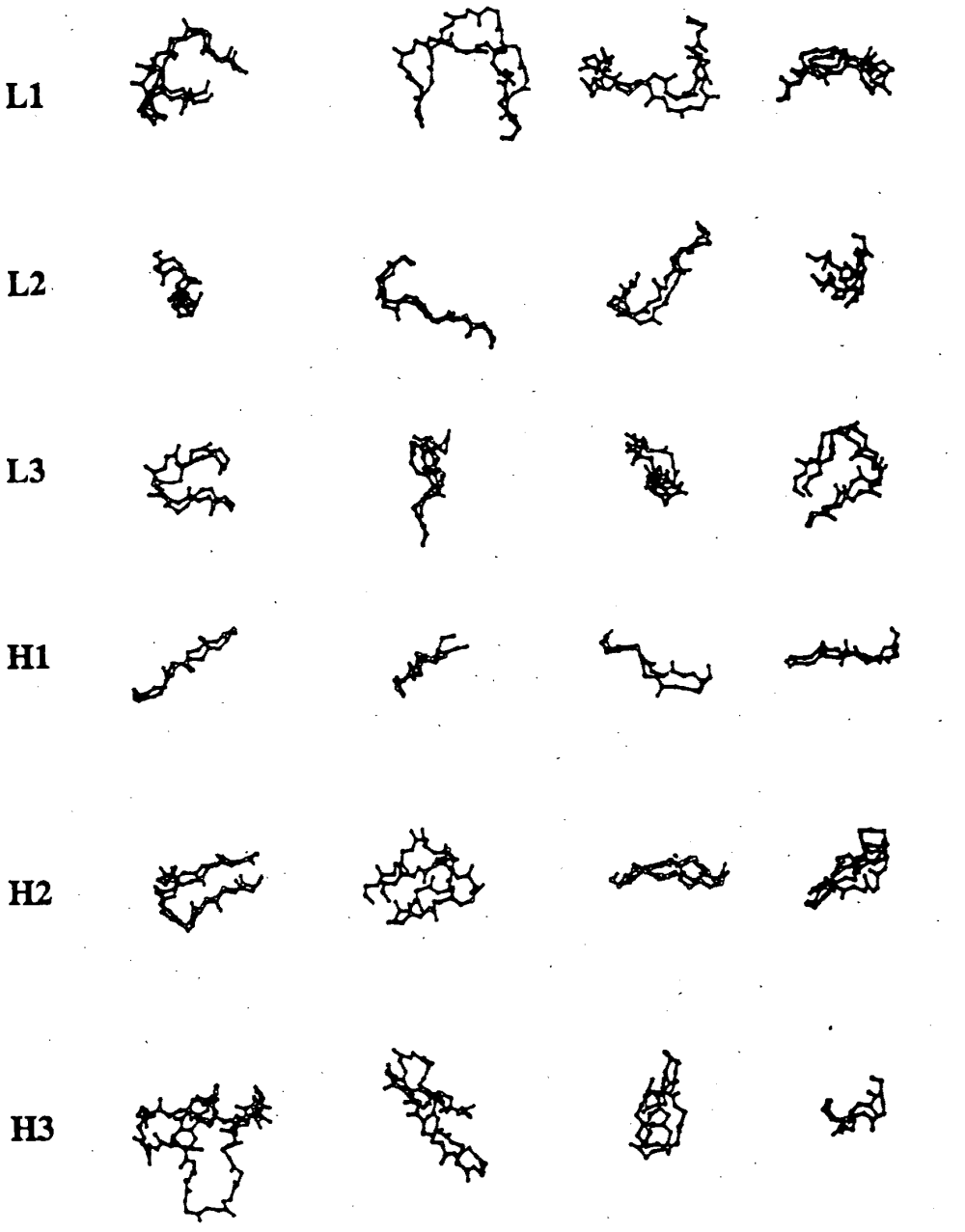


FIG. 8

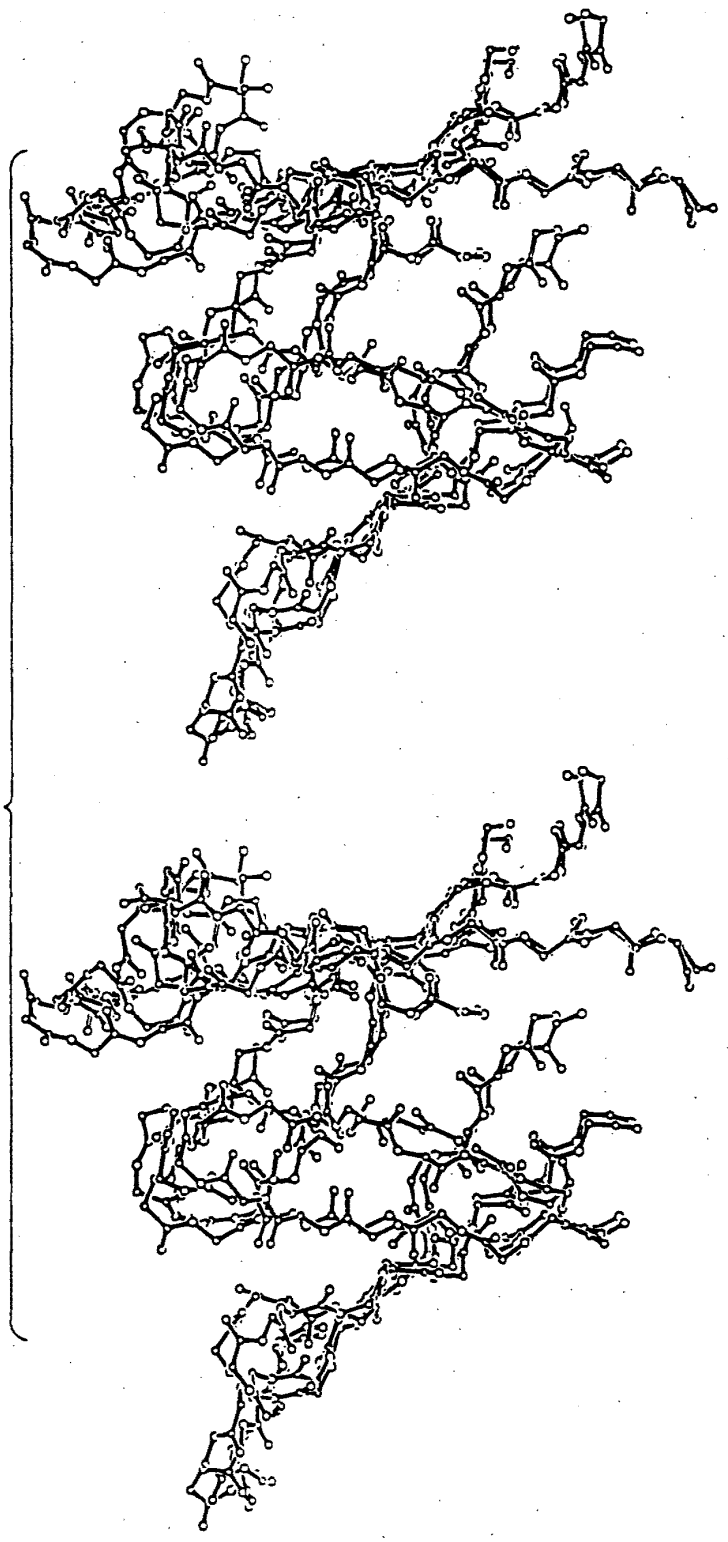


FIG. 9A

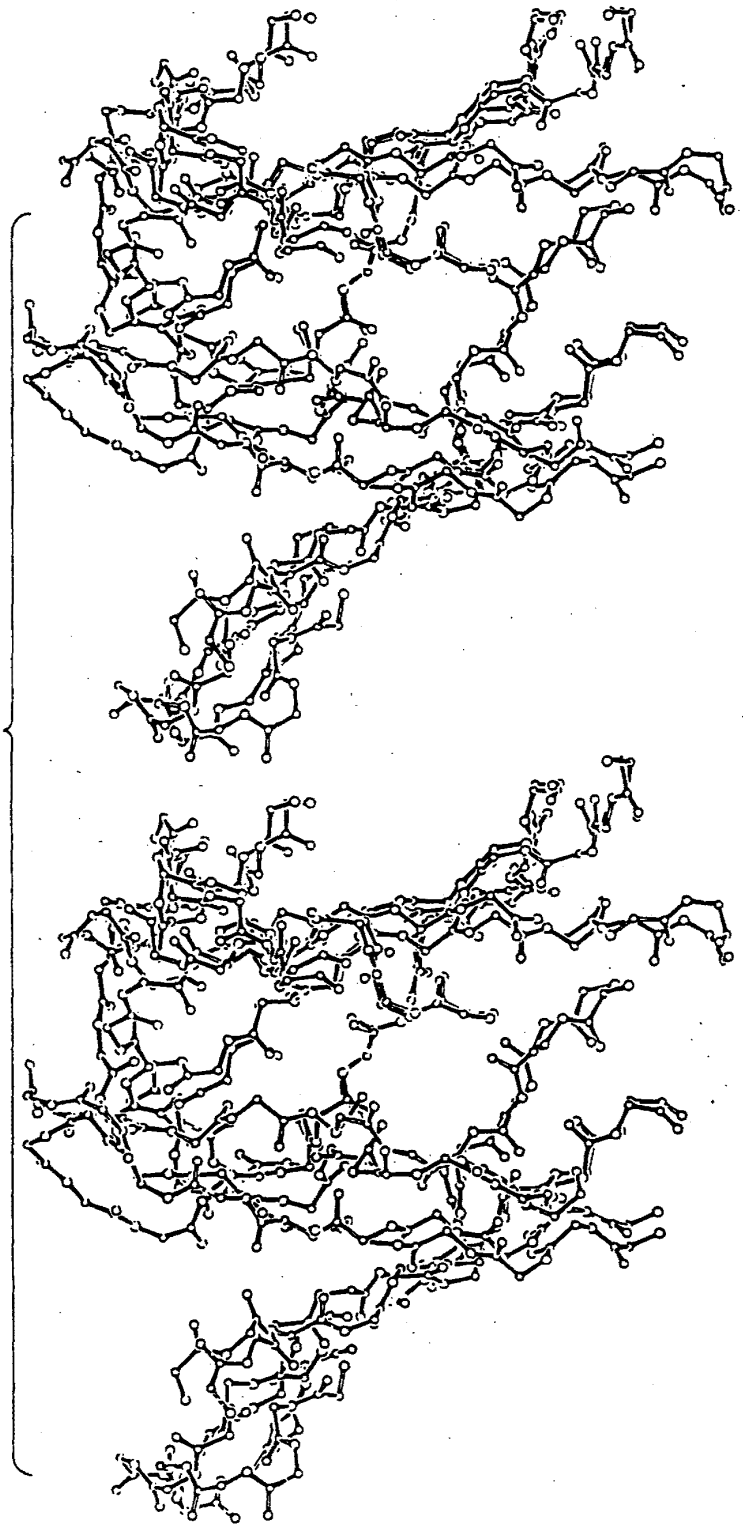


FIG. 9B

226

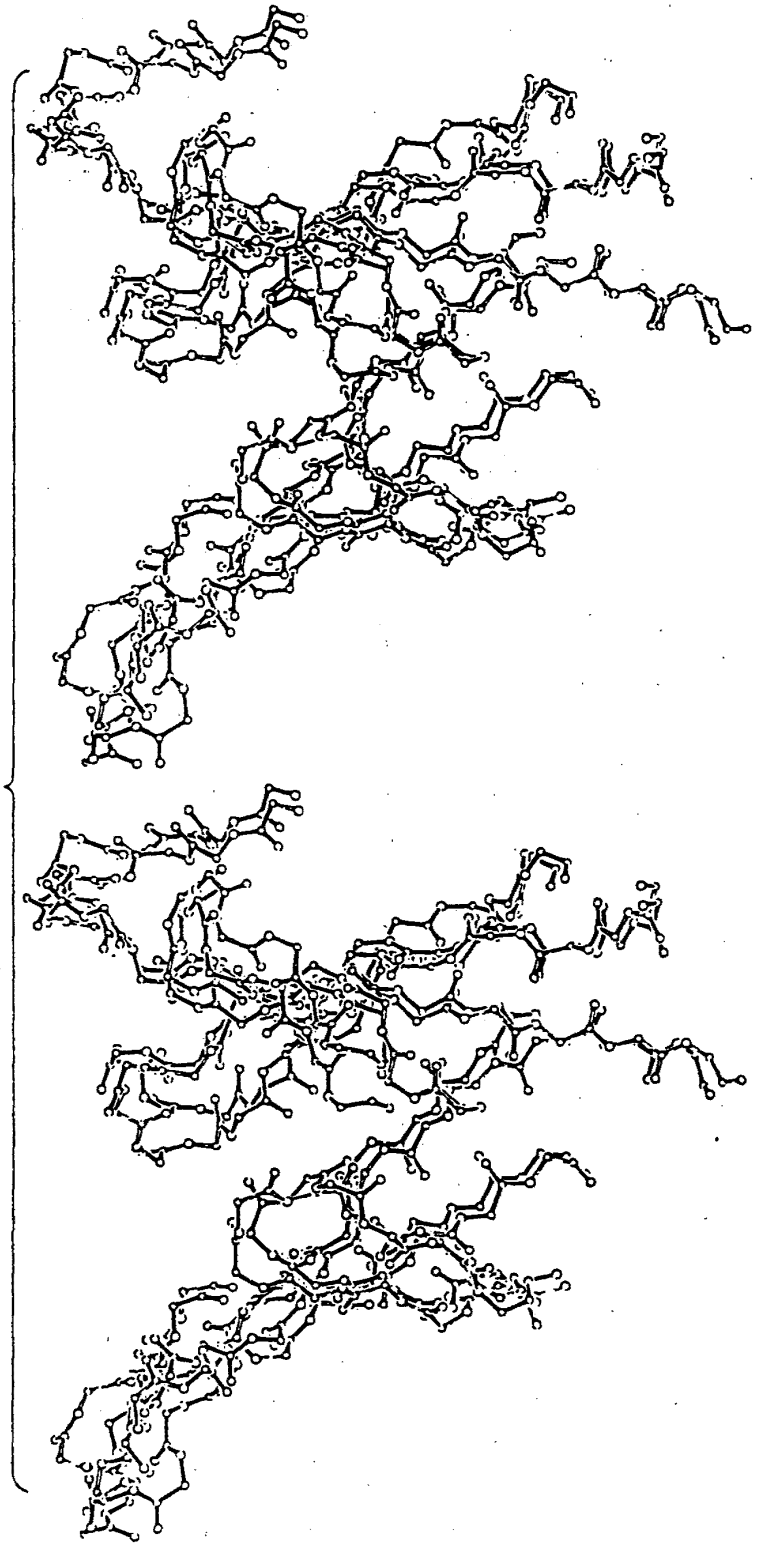


FIG. 9C

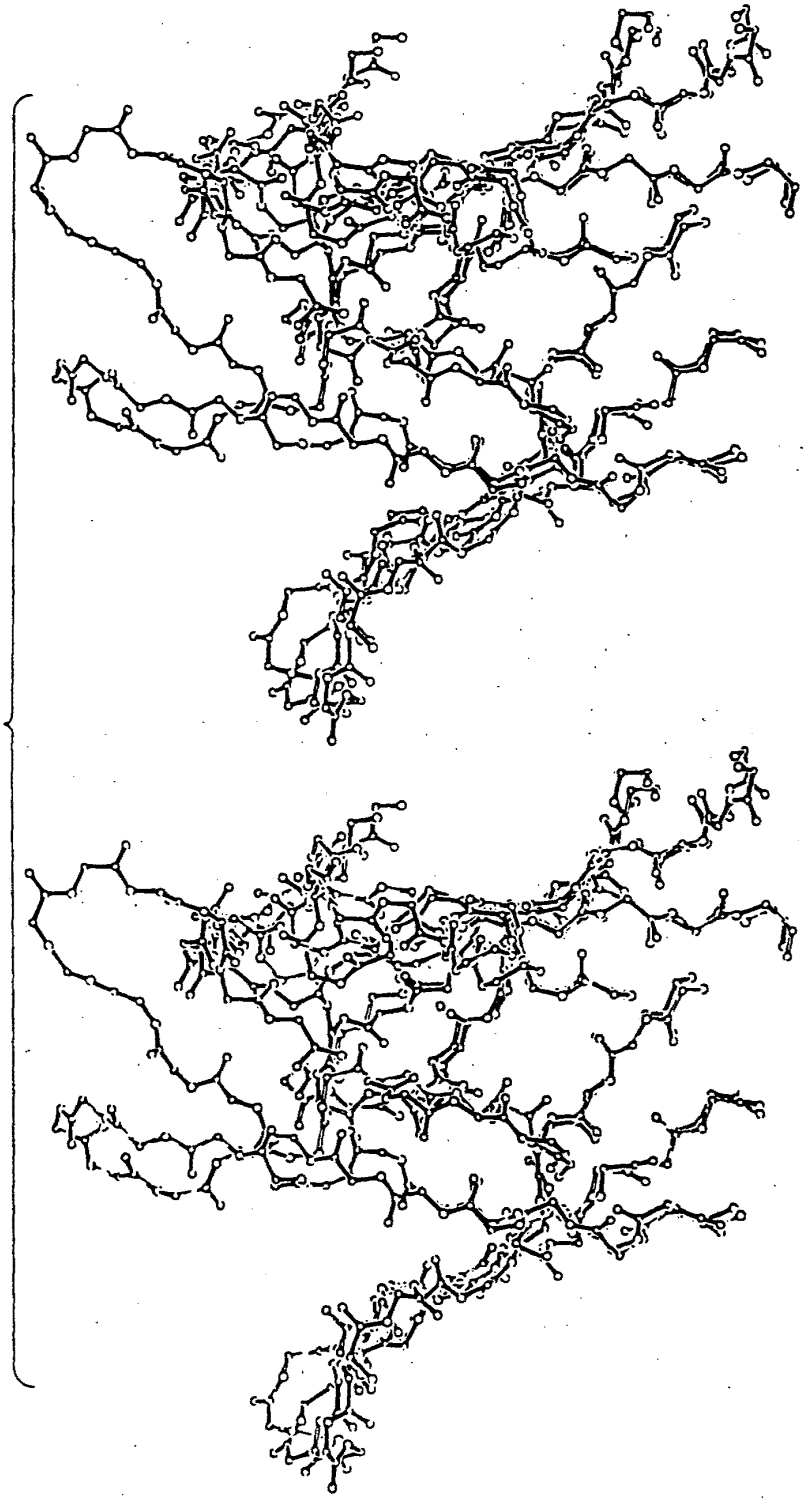


FIG. 9D

228

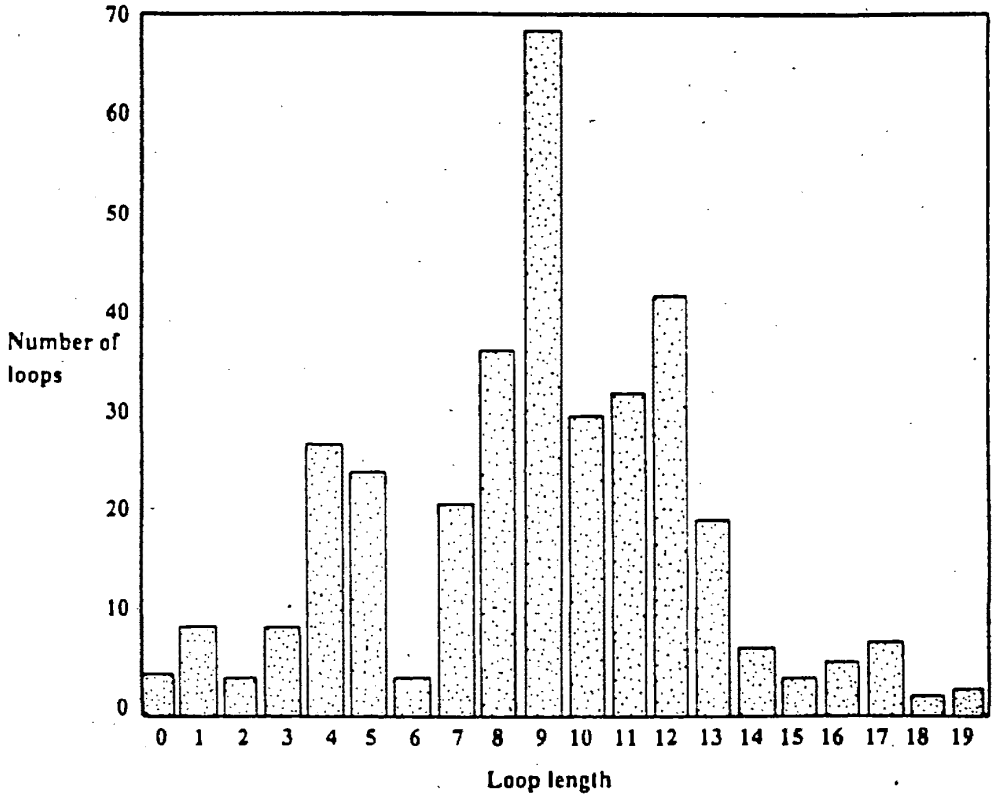


FIG. 10

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 93 30 7051

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
D,A	MOLECULAR IMMUNOLOGY vol. 28, no. 4/5, 1991, GB pages 489 - 498 PADLAN A E 'POSSIBLE PROCEDURE FOR REDUCING THE IMMUNOGENICITY OF ANTIBODY VARIABLE DOMAINS WHILE PRESERVING THEIR LIGAND-BINDING PROPERTIES' * Materials and Methods, p.490, 491, Tables 1-3 *		C12N15/13 C12N15/62 C07K15/00 C12P21/08
D,A	WO-A-9 109 967 (CELLTECH LIMITED) 11 July 1991 * p. 5, second paragraph, p. 6 second paragraph, "Rational" pp. 19-23 *		
P,A	EP-A-0 519 596 (MERCK & CO. INC.) 23 December 1992 * Claims *		
			TECHNICAL FIELDS SEARCHED (Int. Cl.5) C07K

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
MUNICH	12 JANUARY 1994	Germinario C.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P0401)

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 120 694**
A2

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 84301996.9

(22) Date of filing: 23.03.84

(51) Int. Cl.³: C 12 N 15/00
C 12 P 21/00, A 61 K 39/00
A 61 K 39/395, A 61 K 47/00
B 01 D 15/08, G 01 N 33/54
//C12R1/19, C12R1/125,
C12R1/865

(30) Priority: 25.03.83 GB 8308235

(43) Date of publication of application:
03.10.84 Bulletin 84/40

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: CELLTECH LIMITED
244-250 Bath Road
Slough Berkshire SL1 4DY(GB)

(72) Inventor: Boss, Michael Alan
20 Hunter Court Huntercombe Lane North
Slough Berkshire SL1 6DS(GB)

(72) Inventor: Kenten, John Henry
24 Avery Avenue
Downley High Wycombe Buckinghamshire(GB)

(72) Inventor: Emtage, John Spencer
12, Benjamin House Amersham Hill
High Wycombe Buckinghamshire(GB)

(72) Inventor: Wood, Clive Ross
Whych Wood Buddle Hill
North Gorley NEAR Fordingbridge Hants(GB)

(74) Representative: Votier, Sidney David et al,
CARPMAELS & RANSFORD 43, Bloomsbury Square
London WC1A 2RA(GB)

(54) Multichain polypeptides or proteins and processes for their production.

(57) This invention relates to multichain polypeptides or proteins and processes for their production in cells of host organisms which have been transformed by recombinant DNA techniques.

According to a first aspect of the present invention, there is provided a process for producing a heterologous multichain polypeptide or protein in a single host cell, which comprises transforming the host cell with DNA sequences coding for each of the polypeptide chains and expressing said polypeptide chains in said transformed host cell.

According to another aspect of the present invention there is provided as a product of recombinant DNA technology an Ig heavy or light chain or fragment thereof having an intact variable domain.

The invention also provides a process for increasing the level of protein expression in a transformed host cell and vectors and transformed host cells for use in the processes.

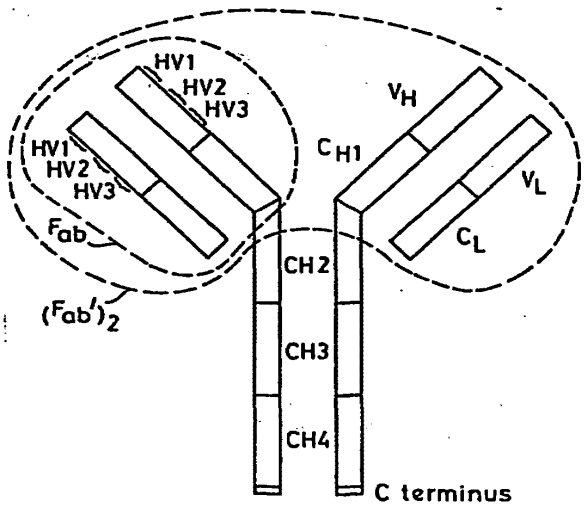


FIG. 1

EP 0 120 694 A2

Croydon Printing Company Ltd.

MULTICHAIN POLYPEPTIDES OR PROTEINS AND PROCESSES FOR THEIR PRODUCTION

This invention relates to multichain polypeptides or proteins and processes for their production in cells of host organisms which have been transformed by recombinant DNA techniques.

In recent years advances in molecular biology based on recombinant DNA techniques have provided processes for the production of heterologous (foreign) polypeptides or proteins in host cells which have been transformed with heterologous DNA sequences which code for the production of these products.

Theoretically, the recombinant DNA approach may be applied to the production of any heterologous polypeptide or protein in a suitable host cell, provided that appropriate DNA coding sequences can be identified and used to transform the host cell. In practice, when the recombinant DNA approach was first applied to the production of commercially useful products, its application for the production of any specified polypeptide or protein presented particular problems and difficulties, and the success of applying this approach to the production of any particular polypeptide or product was not readily predictable.

However, a large number of heterologous single chain polypeptides or proteins have now been produced by host cells transformed by recombinant DNA techniques. Examples of such heterologous single chain polypeptides or proteins include human interferons, the A and B chains of human insulin, human and bovine growth hormone, somatostatin, calf prochymosin and urokinase. Such

transformed host cells provide a reproducible supply of authentic heterologous polypeptide or protein which may be produced on an industrial scale using industrial fermentation technology.

It should be pointed out that some of these polypeptides, for instance urokinase, after secretion by a host cell appear as two chain molecules. However, in such cases, the molecule is synthesised by the host cell as a single chain polypeptide, coded for by a single DNA sequence, which is cleaved in the host cell subsequent to synthesis to form the two chain structure.

It is known that in both human and animal systems there are a number of polypeptides or proteins which have multichain structure in which the chains are not derived from the cleavage of a single chain polypeptide coded for by a single DNA sequence. In such cases, the gene for each of the chains may be located at different points on the same chromosome or even on different chromosomes. In these cases, the polypeptide chains are synthesised separately and then assembled into the complete molecule subsequent to synthesis. Heretofore, no such multichain polypeptide or protein has been produced by recombinant DNA techniques from a single host cell.

A particular example of a class of such multichain polypeptides or proteins is the immunoglobulins.

Immunoglobulins, commonly referred to as antibodies, are protein molecules produced in animals by B-lymphocyte cells in response to challenge with foreign antigenic agents, such as bacteria, viruses and foreign proteins. The immunoglobulins comprise a crucial part of the immune systems of humans and animals. The immunoglobulins

recognise specific parts of the foreign agents and bind onto them. The specific parts are usually known as antigenic determinants or antibody binding sites. A given foreign agent is likely to have a number of different antigenic determinants.

A typical immunoglobulin (Ig) molecule is shown in Figure 1 of the accompanying drawings, to which reference is now made. The Ig molecule comprises two identical polypeptide chains of about 600 amino acid residues (usually referred to as the heavy chains H), disulphide bonded to each other, and two identical shorter polypeptide chains of about 220 amino acid residues (usually referred to as the light chains L), each light chain being disulphide bonded to one end of each heavy chain as shown.

When the Ig molecule is correctly folded, each chain is formed into a number of distinct globular areas, usually known as domains, joined by a more linear polypeptide chain. The light chains have two such domains, one of which is of variable sequence V_L, and the other of which is of constant sequence C_L. The heavy chains have a single variable domain V_H adjacent the variable domain V_L of the light chain, and three or four constant domains $C_{H1-3 \text{ or } 4}$.

The variable domains on both the heavy and the light chains each contain three hypervariable regions (HV1-3) which, when the Ig molecule is correctly folded, are located adjacent one another and form the antigen binding site. It is this area which recognises and binds to the antigenic determinant for which the Ig molecule is specific.

The constant domains of the Ig molecule do not take part in the binding to the antigenic determinant, but mediate the actions triggered by the binding of the Ig molecule to the antigenic determinant. It is believed that this triggering is caused by an allosteric effect induced by the binding of the Ig molecule to the antigenic determinant. The constant domain may enable the Ig molecule to fix complement or may cause mast cells to release histamine.

Ig 's may be categorised by class or subclass, depending on which of a number of possible heavy chain constant domains they contain, there being eight such possible heavy chains in mice. Thus, for instance, an Ig molecule with a μ heavy chain belongs to the class IgM, and one with a γ_1 heavy chain to the class IgG_1.

Ig's may also contain one of two light chains, designated as κ and λ light chains, which have different constant domains and different sets of variable domains.

The structure of the Ig molecule and the location of the genes coding for the various domains thereof are discussed more fully by Early and Hood, in Genetic Engineering, Principles and Methods, Vol.3, pages 153-158 (edited by Setlow and Hollaender, Plenum Press).

It is known that Ig molecules on digestion with selected enzymes can produce a number of immunologically functional fragments. Two such fragments are known as the F_{ab} and $(F_{ab}')_2$ fragments. The F_{ab} fragment comprises one light chain linked to the V_H and C_{H1} domains of a heavy chain as shown in Figure 1. The $(F_{ab}')_2$ fragment consists essentially of two F_{ab} fragments linked together by a small additional portion of the heavy chains as shown

in Figure 1. These fragments and other similar fragments can be of use in various tests and diagnostic and medical methods.

The principle method employed for the production of Ig's involves the immunisation of susceptible animals with the antigenic agents to provide an immune reaction. Generally the animal is immunised a second time to improve the yield of Ig. The animal is then bled and the Ig is recovered from the serum.

However, the product of this method is not a homogeneous protein. The animal will produce Ig of different classes and also Ig specific for each of the different antigenic determinants on the antigenic agent, and its blood will therefore contain a heterogenous mixture of Ig's. Obtaining a specific Ig of particular class and desired specificity from such a mixture requires very difficult and tedious purification procedures.

Recently, it has become possible to produce a homogeneous Ig of a single class and a single specificity by a technique first described by Kohler and Milstein (Nature, 256, 495-479,1975). The technique involves the fusion of single Ig-producing parent cells with cancer cells to produce a monoclonal hybridoma cell which produces the Ig. Ig produced by this technique is usually known as monoclonal antibody. The nature of the monoclonal Ig is determined by the class and specificity of the Ig produced by the parent cell.

Recently, attempts have been made to use recombinant DNA techniques to produce fragments of Ig molecules. For instance, Amster et al. (Nucleic Acid Research, 8, No.9, 1980, pp 2055 to 2065) disclose the cloning of

double stranded cDNA sequences encoding for a mouse I_g light chain into a plasmid. An E. Coli strain transformed by the plasmid synthesised a protein thought to comprise the complete constant domain of the light chain and about 40 amino acid residues of its variable region.

Kemp and Cowman (Proc. Natl. Acad. Sci. USA, 78,, 1981, pp 4520 to 4524) disclose the cloning of cDNA sequences encoding for mouse heavy chain fragments and the transforming of an E. Coli strain which then synthesised heavy chain polypeptide fragments.

In both these cases, the polypeptides were produced as fusion proteins, in which the fragments of the Ig polypeptides were fused with additional non-Ig polypeptide sequences, and with incomplete variable domains. Thus, the polypeptide chains produced in these studies were not immunologically functional polypeptides as they were incapable of combining with complementary heavy or light chains to provide Ig molecules having intact antigen binding sites and immunological function.

Research studies have also been carried out in mammalian systems. For instance, Falkner and Zachau (Nature, 298, 1982, pp 286 to 288) report the cloning of cDNA sequences encoding for mouse light chains into a plasmid which was used to transfect genomic eukaryotic cells which could then transiently synthesise light chains.

Rice and Baltimore (Proc.Natl, Acad. Sci. USA, 79, 1982, pp 7862 to 7865) report on the transfection of a functionally rearranged K light chain Ig gene into a murine leukemia virus-transformed lymphoid cell line. The cell line is then able to express the gene continuously. In both these cases, the K genes used to transfect the

mammalian cells were obtained from myeloma cells and the K polypeptides produced were of indeterminate immunological function.

A further approach is exemplified in a series of papers by Valle et al. (Nature, 291, 1981, 338-340; Nature, 300, 1982, 71-74; and J. Mol. Biol., 160, 1982, 459-474), which describe the microinjection of mRNAs encoding for heavy or light chains of Ig isolated from a mouse myeloma line into oocytes of Xenopus laevis. Under certain conditions complete Ig molecules were formed. However, the mRNAs were obtained from myeloma cells and the Ig molecules were of indeterminate immunological function.

It can thus be seen that hitherto it has not been possible to produce functional Ig by recombinant DNA technology.

According to a first aspect of the present invention, there is provided a process for producing a heterologous multichain polypeptide or protein in a single host cell, which comprises transforming the host cell with DNA sequences coding for each of the polypeptide chains and expressing said polypeptide chains in said transformed host cell.

According to a second aspect of the present invention, there is provided a heterologous multichain polypeptide or protein produced by recombinant DNA technology from a single host cell.

The present invention is of particular, but not exclusive, application in the production of Ig molecules and immunologically functional

Ig fragments of the type referred to above. However, it will be appreciated that the present invention can be applied to the production of other multichain polypeptides or proteins.

In relation to the product of Ig molecules according to the invention it will be appreciated that, in order to produce a functional molecule, the DNA sequences used to transform the host cell will need to encode for at least the V_L and V_H domains of an Ig molecule. Moreover, these domains will need to be complementary so that when the two polypeptide chains fold together they form an antigen binding site of predetermined specificity.

Preferably, the Ig molecule or fragment includes a complete light chain and at least the C_{H1} domain in addition to the V_H domain of the heavy chain. Most preferably the Ig molecule is intact.

It has also been shown by the present applicants that it is now possible to produce individual heavy and light chains having intact variable domains. This has not previously been possible. Therefore, according to a third aspect of the present invention there is provided as a product of recombinant DNA technology an Ig heavy or light chain or fragment thereof having an intact variable domain.

Advantageously, the Ig molecule or functional fragment thereof according to the present invention has a variable region (formed by the V_L and V_H domains) which defines a binding site for an antigenic determinant of clinical or industrial importance. The DNA coding sequences necessary to produce such a molecule may be derived from naturally occurring or hybridoma (monoclonal)Ig-producing cells with the desired specificity.

The constant domains of the Ig molecule or fragment, if present, may be derived from the same cell line as the variable region. However, the constant domains may be specifically altered, partially or completely omitted, or derived from a cell line producing a different class of Ig to provide Ig molecules or fragments having desired properties.

For example, an Ig molecule may be produced having variable domains (V_H and V_L) idential with those from a monoclonal antibody having a desired specificity, and constant domain(s) from a different monoclonal antibody having desired properties, for instance to provide human compatibility or to provide a complement binding site.

Such alterations in the amino acid sequence of the constant domains may be achieved by suitable mutation or partial synthesis and replacement or partial or complete substitution of appropriate regions of the corresponding DNA coding sequences. Substitute constant domain portions may be obtained from compatible recombinant DNA sequences.

The invention may be utilised for the production of Ig molecules or fragments useful for immunopurification, immunoassays, cytochemical labelling and targetting methods, and methods of diagnosis or therapy. For example, the Ig molecule or fragment may bind to a therapeutically active protein such as interferon or a blood clotting factor, for example Factor VIII, and may therefore be used to produce an affinity chromatorgraphy medium for use in the immunopurification or assay of the protein.

It is also envisaged that the Ig molecule may be synthesised by a host cell with another peptide moiety attached to one of its constant domains. Such a further peptide moiety may be cytotoxic or enzymatic. Alternatively, the moiety may be useful in attaching the Ig molecule to a biological substrate, such as a cell or tissue, or to a non-biological substrate, such as a chromatography medium. Such a peptide moiety is herein referred to as a structural peptide moiety.

It is further envisaged that cytotoxic, enzymic or structural peptide moieties could be attached to the Ig molecule by normal peptide chemical methods, as are already known in the art, rather than by being synthesised with the Ig molecule.

The Ig molecule or fragment may also comprise a therapeutic agent in its own right. For instance, an Ig molecule or fragment specific for D blood group antigen may be useful for the prevention of haemolytic disease of the new born.

Any suitable recombinant DNA technique may be used in the production of the multichain polypeptides or proteins of the present invention. Typical expression vectors such as plasmids are constructed comprising DNA sequences coding for each of the chains of the polypeptide or protein.

It will be appreciated that a single vector may be constructed which contains the DNA sequences coding for more than one of the chains. For instance, the DNA sequences coding for Ig heavy and light chains may be inserted at different positions on the same plasmid.

Alternatively, the DNA sequence coding for each chain may be inserted individually into a plasmid, thus producing a number of constructed plasmids, each coding for a particular chain. Preferably the plasmids into which the sequences are inserted are compatible.

The or each constructed plasmid is used to transform a host cell so that each host cell contains DNA sequences coding for each of the chains in the polypeptide or protein.

Suitable expression vectors which may be used for cloning in bacterial systems include plasmids, such as Col El, pcRl, pBR322, pACYC 184 and RP4, phage DNA or derivatives of any of these.

For use in cloning in yeast systems, suitable expression vectors include plasmids based on a 2 micron origin.

Any plasmid containing an appropriate mammalian gene promoter sequence may be used for cloning in mammalian systems. Such vectors include plasmids derived from, for instance, pBR322, bovine papilloma virus, retroviruses, DNA viruses and vaccinia viruses.

Suitable host cells which may be used for expression of the heterologous multichain polypeptide or protein include bacteria, such as E.Coli and B. Subtilis, streptomyces, yeasts, such as S. cervisiae, and eukaryotic cells, such as insect or mammalian cell lines. Examples of suitable bacterial host cells include E. Coli HB 101, E. Coli X1776, E. Coli X2882, E. Coli PS 410, E. Coli MRC 1, E. Coli RV308, E. Coli E103S and E. Coli B.

The present invention also includes constructed expression vectors and transformed host cells for use in producing the multichain polypeptides or proteins of the present invention.

After expression of the individual chains in the same host cell, they may be recovered to provide the complete multichain polypeptide or protein in active form, for instance to provide an Ig molecule of predetermined immunological function.

It is envisaged that in preferred forms of the invention, the individual chains will be processed by the host cell to form the complete polypeptide or protein which advantageously is secreted therefrom.

However, it may be that the individual chains may be produced in insoluble or membrane-bound form. It may therefore be necessary to solubilise the individual chains and allow the chains to refold in solution to form the active multichain polypeptide or protein. A suitable procedure for solubilising polypeptide chains expressed in insoluble or membrane-bound form is disclosed in our copending application No. (Protein Recovery, Agent's Ref. GF 402120 and 402121).

It will be appreciated that the present application shows for the first time that it is possible to transform a host cell so that it can express two or more separate polypeptides which may be assembled to form a complete multichain polypeptide or protein. There is no disclosure or suggestion of the present invention in the prior art, which relates solely to the production of a single chain heterologous polypeptide or protein from each host cell.

The present invention will now be described, by way of example only, with reference to the accompanying drawings, in which:-

Figure 1 shows a diagrammatic representation of a typical intact Ig molecule;

Figure 2 shows the construction of plasmids for the direct synthesis of a λ light chain in E. Coli;

Figure 3 shows the construction of plasmids for the direct synthesis of a μ heavy chain in E. Coli;

Figure 4 is a diagrammatic representation of μ mRNA sequences around the initiation codon;

Figure 5 shows the construction of plasmids having altered secondary structure around the initiation codon;

Figure 6 is a polyacrylamide gel showing expression and distribution of μ protein from E. Coli B;

Figure 7 is a polyacrylamide gel showing pulse chase autoradiograms of μ protein in E. Coli B and in E. Coli HB101;

Figure 8 is a polyacrylamide gel showing the results of λ gene expression in E. Coli;

Figure 9 is a polyacrylamide gel showing the distribution of recombinant λ light chain polypeptide between the soluble and insoluble cell fractions;

Figure 10 is a polyacrylamide gel showing expression and distribution of λ protein from E. Coli E103S;

Figure 11 shows the results of the fracitonation of μ and λ protein expressed by E. Coli B on DEAE Sephacel;

Figure 12 shows the specific hapten binding of reconstituted Ig molecules; and

Figure 13 shows the heteroclitic nature of the hapten binding of the reconstituted Ig molecules.

In the following examples, there is described the production of Ig light and heavy chain polypeptides derived from monoclonal antibodies which recognise and bind to the antigenic determinant 4-hydroxy-3-nitrophenyl acetyl (NP), using E.coli and S. cerevisiae as the host cells Recombinant DNA techniques were used to enable the host cells to express both the polypeptide chains.

It will be appreciated that the invention is not limited to the specific methods and construction described hereafter.

Construction of Lambda Light Chain Expression Plasmid

Figure 2, to which reference is now made, shows schematically the method used to construct a λ light chain expression plasmid.

It was decided to express the lambda gene in E. coli by direct expression of the gene lacking the eucaryotic signal peptide but containing a methionine initiator residue at the amino-terminus (met-lambda). The approach used for bacterial synthesis of met-lambda was to reconstruct the gene in vitro from restriction fragments of a cDNA clone and to utilise synthetic DNA fragments for insertion into the bacterial plasmid pCT54 (Emtage et al., Proc. Natl. Acad. Sci. USA., 80, 3671 to 3675, 1983). This vector contains the E. coli trp promoter, operator and leader ribosome binding site; in addition 14 nucleotides downstream of the ribosome binding site is an initiator ATG followed immediately by EcoR1 and HindIII sites and the terminator for E. coli RNA polymerase from bacteriophage T7.

As a source of light chain we used a plasmid pABλ1-15 which contains a full-length λ_1 light chain cDNA cloned into the PstI site of pBR322. This λ_1 light chain is derived from a monoclonal antibody, designated S43, which binds to 4-hydroxy-3-nitrophenylacetyl (NP) haptens.

In order to create a HindIII site 3' to the end of the lambda gene for insertion into the HindIII site of pCT54, the cDNA was excised from pABλ1-15 using Pst1. The cohesive ends were blunt ended using the Klenow fragment of DNA polymerase and synthetic HindIII linker molecules of sequence 5'-CCAAGCTTGG-3' ligated. The DNA was digested with HindIII and the 850bp lambda gene isolated by gel electrophoresis and cloned into HindIII cut pAT153 to yield plasmid pATλ1-15. The 3' end of the lambda gene was isolated from pATλ1-15 by HindIII plus partial Sac1 digestion as a 630bp SacI-HindIII fragment (2 in Figure 2). The HindIII cohesive end was dephosphorylated by calf intestinal alkaline phosphatase during isolation of the fragment to prevent unwanted ligations at this end in subsequent reactions.

A HinfI restriction site is located between codons 7 and 8 and the lambda sequence. The 5' end of the lambda gene was isolated as a 148bp HinfI to SacI fragment (1 in Figure 2).

Two oligodeoxyribonucleotides were designed to restore codons 1-8, and to provide an initiator ATG as well as Bc1I and HinfI sticky ends. The two chemically synthesised oligonucleotides made to facilitate assembly of the gene had the sequences:

R45 5' -pGATCAATGCAGGCTGTTGTG 3'

R44 3' CCGACAACACTGAGTCCTTAp- 5'

pCT54 was cut with both Bcl1 and HindIII and the resulting linear molecules isolated, mixed together with the two oligodeoxyribonucleotide linkers R44 and R45 and both fragments 1 and 2, and ligated using T4 ligase (Figure 2). The mixture was used to transform E. coli DH1 to ampicillin resistance. Recombinant clones in pCT54 were identified by hybridisation of DNA from replica plated colonies on nitrocellulose to a nick-translated probe derived from the pATλ1-15 insert.

A clone was identified which hybridised to lambda cDNA and also showed the predicted restriction fragment pattern. This plasmid (designated pCT54 19-1) was sequenced from the ClaI site and shown to have the anticipated sequence except that there was a mutation of the fourth codon from CTG to ATG, changing the amino acid at this point from valine to methionine.

The sequence in this area was:

...GATTGATCA.ATG.CAG.GCT.GTT.ATG.ACT.CAG.GAA.TCT.GCA.CTC.ACC.ACA.TCA..

met gln ala val met thr gln glu ser ala leu thr thr ser

The restriction enzyme sites in pCT54 between the Shine and Dalgarno sequence (AAGG), which is important for ribosome binding, and the ATG allow for the adjustment of the SD-ATG distance, an important parameter in determining expression rates. The SD-ATG distance was reduced by cutting the plasmid with ClaI or BclI and creating blunt ended species by digestion with S_I nuclease. 2 µg of ClaI cut DNA was digested with 200 units of S_1 nuclease for 30 minutes at 30° using standard buffer conditions. The solution was deproteinised with phenol and the DNA recovered by ethanol precipitation. This DNA on religation with T4 DNA ligase and transformation into E.coli strain HB101 gave rise to a number of plasmids which had lost the ClaI or BclI site.

The plasmids which had lost their ClaI site were sequenced in the region surrounding the initiator ATG.

....AAGGGTATTGATCAATG CAG.... plasmid pNP3
SD met glu

....AAGGGTTTGATCAATG CAG plasmid pNP4
SD met glu

In order to achieve high level expression a number of other approaches were followed. Firstly, a series of constructs were obtained which had increasing amounts of the 3' untranslated region of the cDNA removed by Bal 31 exonuclease. Secondly, a high copy number plasmid containing λ cDNA was constructed. This plasmid contained a par function (Meacock, P.A.and Cohen, S.N.Cell, 20, 529-542, 1980) as well as being present in high copy number. Thirdly, the pNP3 plasmid was transformed into a number of protease-deficient strains or into HB101 in conjunction with a protease deficient dominant acting plasmid (Grossman, A.D.et al Cell, 32, 151-159, 1983).

Construction of μ Heavy Chain Expression Plasmid

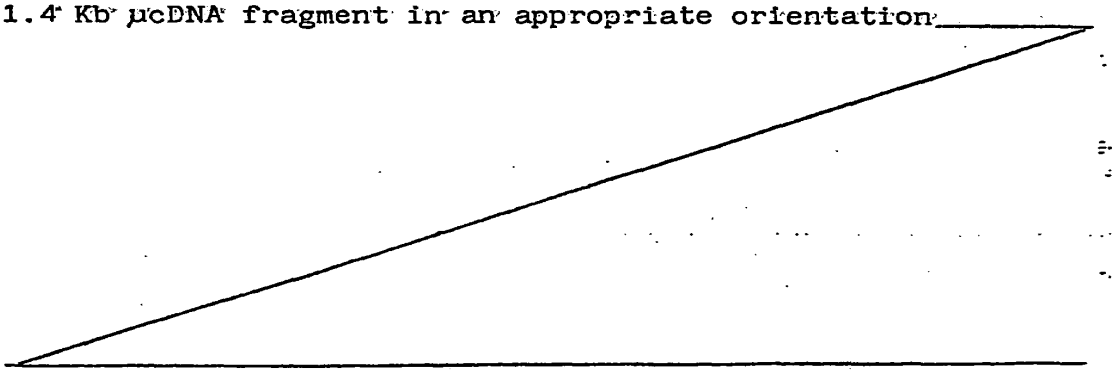
The full-length μ heavy chain cDNA derived from the NP binding monoclonal antibody B1-8 had been cloned into the PstI site of pBR322 yielding a plasmid designated pABμ-11 (Bothwell et al., Cell, 24, 625-637, 1981). In order to achieve high level expression, the μ cDNA minus the eukaryotic leader was reconstructed into pCT54. The construction of the μ heavy chain expression plasmid is shown diagramatically in Figure 3. Two chemically synthesised oligonucleotides were made to facilitate this. These have the following sequences:

R43 5' GATCAATGCAGGTTCAGCTGCA 3'

R46 3' TTACGTCCAAGTCG 5'

These were ligated into BclI cut pCT54 using T4 DNA ligase and the resulting plasmid designated pCT54 Pst. The linkers were designed to replace the sequence in pCT54 between the BclI site and the ATG, to provide an internal PstI site and to recreate the sequence from pABμ-11 5' to the PstI site up to codon +1. pCT54 Pst was cut briefly with PstI, treated with alkaline phosphatase and full length DNA glass isolated from a 1% agarose gel. Similarly pABμ-11 was briefly cut with PstI and the full length μ insert isolated following agarose gel electrophoresis. The μ cDNA was ligated with the full length pCT54Pst fragment using T4 DNA ligase under standard conditions and a plasmid designated pCT54μ was identified which by restriction enzyme analysis was shown to contain a full-length μ insert. The plasmid was sequenced around the 5' linker region and was found to have the anticipated sequence:

5' ....TGATCAATGCAGGTTCAGCTGCAGGGGGGGGATGGGATGGAG.... 3', demonstrating that it was, indeed, a full length clone. A complete PstI digest of pCT54μ liberated a 1.4 Kb fragment which was purified by 0.8% agrose gel electrophoresis and glass powder isolation. This was ligated with PstI cut pCT54 Pst (see above) using T4 DNA ligase under standard conditons, followed by transformation into HB101. A plasmid designated pNP1 was isolated which was shown by restriction endonuclease pattern analysis to contain the 1.4 Kb μcDNA fragment in an appropriate orientation



(Figure 3). pNP1 was a plasmid which consisted of an appropriate 5' end for expression, whilst pCT54 contained an appropriate 3' end. The full length gene was reconstructed into pCT54 by cutting both pNP1 and PCT54μ with AvaI which cuts once in the plasmid and once in the μ gene. Both digests were run on a 1% agrose gel and the 1.9 Kb fragment from pNP1 and the 3.65 Kb fragment from pCT54μ isolated. Following alkaline phosphatase treatment of the 3.65 Kb pCT54μ fragment, the two pieces of DNA were ligated to each other and transformed into HB101. A plasmid designated pNP2 was identified which demonstrated the correct restriction endonuclease pattern. It was sequenced in the area surrounding the initiator ATG and found to have the anticipated sequence:

5' ....TTGATCAATGCAGGTTCAGCTGCAGCAGCCTGGGGCTGAGCTTGTGAAG.... 3'

The vector pCT54 had been constructed to include two restriction sites (Bcl1 and Cla1) between the S-D sequence (AAGG) and the initiation codon so that the distance between these sequence elements could be varied. As most E.coli mRNAs have 6-11 nucleotides between the S-D sequence and the AUG the distance in pNP2 was reduced by modification at the Cla1 site. pNP2 was cut with Cla1 and incubated with S1. The amount of S1 nuclease was adjusted so that some DNA molecules would lose 1-2 extra base pairs as a result of 'nibbling' by the enzyme. This DNA on religation with T4 DNA ligase and transformation into E.coli strain HB101 gave rise to a number of colonies harbouring plasmids which had lost the Cla1 site. The sequences around the initiation codon of four plasmids pNP223, pNP261, pNP9 and pNP282 were determined and are given in Table 1, part A below.

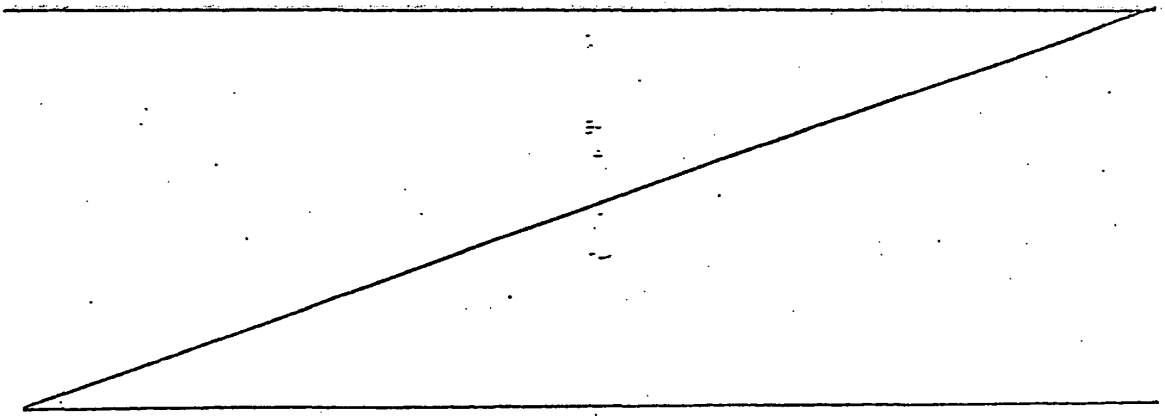


TABLE 1

		S-D to ATG distance (no. of residues)	Sequence (5' - 3') ClaI	Expression: [mu] expts. A	B	[ru] C
(A)	pNP2	14	-AAAAAGGGTATCGATTGATCAATGCAGGTTCAG-	0.7	0.6	0.8
S1 derivatives of pNP2	pNP223	12	-AAAAAGGGTATATTGATCAATGCAGGTTCAG-	-	0.7	0.6
	pNP261	10	-AAAAAGGGTATTGATCAATGCAGGTTCAG-	-	1.1	1.1
	pNP282	7	-AAAAAGGGTGATCAATGCAGGTTCAG-	-	0.7	0.6
	pNP9	9	-AAAAAGGGTATGATCAATGCAGGTTCAG-	1.0	1.0	1.0
(B)	pNP11	11	-AAAAAGGGTATTGCACATATGCAAGTGCAA-	5.3	6.3	9.1
oligo constructs	pNP12	11	-AAAAAGGGTATTGCACATATGCAGGTTCAG-	1.8	2.3	1.3
	pNP14	9	-AAAAAGGGTATGATCAATGCAAGTGCAA-	72.2	106.7	124.6
	pNP8	-	-	77.0	53.4	145.5
	pCT70	-	-	0	0	0

The expression of these constructs in E. coli B was examined in three experiments, and concentrations of μ protein were determined by ELISA as relative units (ru), a measure of μ concentration normalised to pNP9. The results are shown in Table 1.

Inductions were carried out by resuspending cells at 1:50 dilution from overnight cultures into induction medium, consisting of: KH_2PO_4 (3g/L), Na_2HPO_4 (6g/L), NaCl (0.5g/L), Difco vitamin assay casamino acids (30g/L), NH_4Cl (4g/L) glycerol (16g/L), proline (0.5g/L), Difco yeast extract (1g/L), $CaCl_2$. $6H_2O$ (0.022g/L), $MgSO_4.7H_2O$ (0.025g/L), thiamine (0.01g/L) and carbenicillin (0.1g/L). Cultures were shaker at 37°C and cells harvested by centrifugation.

ELISA assays were performed essentially as described by Patel et al. (Nucleic Acid Research, 10, 5605 to 5620, 1982), using affinity purified goat anti-mouse IgM (Tago) and affinity purified goat anti-mouse IgM (Tago) conjugated toperoxidase, with 3, 3', 5, 5' tetramethylbenzidine (Miles) as a substrate. Bacterial cell pellets were boiled for 2 minutes in 9M urea (BRL) and loaded directly into the ELISA, such that all wells contained 2.25M urea, 10mM Tris-HCl pH7.8, 150mM NaCl, 0.05% NP40 and 0.5% casein (Hammarsten, B.D.H.).

Altering the S-D to ATG distance with nuclease S1 was found to increase the level of μ expression, relative to the parental plasmid pNP2, but only to a small extent (compare pNP9). The optimal S-D to AUG distance was found to be 9-10 nucleotides, as found in pNP9 and pNP261 respectively.

Secondary Structure Analysis

Hairpin loops were identified by use of the computer programme HAIRGU developed by Dr. Roger Staden. The ΔG values were calculated as described by Trioco et al. (Nature NB, 246, 40-41, 1973).

Analysis of the potential secondary structure of μ mRNA encoded by pNP2 and pNP9 (one of the S1 derivatives of pNP2) revealed an extensive array of possible hairpin loops. Attention was focused on the region of this mRNA containing the ribosome binding site since the sequestering of S-D sequences and initiation codons into secondary structure may inhibit the initiation of translation. It was found that a hairpin loop including the U and G of the initiation codon could be formed (see Figre 4). This hairpin loop is formed entirely within coding sequences and has a $\Delta G = -7.6$ kcal. No secondary structures were found whichburied the S-D sequence, and none which mutually excluded the hairpin loop described above, and which had a lower ΔG.

To test whether expression of μ protein could be influenced by changes designed to alter potential secondary structure in the ribosome binding region we used synthetic oligonucleotides to mutagenise this region of μ mRNA. The plasmids were constructed by insertion of a pair of oligonucleotides between the blunted Cla I site and the 5' Pst site of μ in pNP2 (see Figure 5). The cloning strategy used to accomplish this, however, was multi-staged due to the presence of four Pst sites in pNP2. The vector pACYC184CM, which lacks Cla I and Pst I sites, was first constructed by cutting pACYC184 at its unique Cla I site, 'filling-in' the cohesive termini using T4 DNA polymerase and religating. Next, PACYC184CM was cut with BamHI, and the BamHI fragment from pNP2, bearing the 5' μ sequence, was inserted into it to form pCMμ (see Figure 4). The vector pCMμ contained unique Cla I and Pst sites which were required for the oligonucleotide cloning (see Figure 5).

pCMμ was cut with Cla I to completion, blunted with nuclease S1, and then cut with Pst I. The cut plasmid was purified by agarose gel electrophoresis and ligated separately with pairs of oligonucleotides, with the latter in a 100-fold molar excess. The pairs of oligonucleotides used were: R131 and R132 (for pNP11); R196 and R197 (for pNP12); and R202 and R203 (for pNP14). These have the sequences given below.

TGCACATATGCAAGTGCAACTGCA (R131), GTTGCACTTGCATATGTGCA (R132), GCTGAACCTGCATATGTGCA (R196), TGCACATATGCAGGTTCAGCTGCA (R197), GTTGCACTTGCATTGATC (R202), GATCAATGCAAGTGCAACTGCA (R203).

The ligation mixture was then used to transform HB101. Recombinant clones were detected by colony hybridisation, on nitrocellulose filters, using one of each pair of oligonucleotides as a probe. Positive clones were sequenced and shown to have the correct nucleotide sequence. The μ sequences containing the cloned oligonucleotides were excised on Eco RV-Bgl II fragments. These were ligated to Eco RV-Bgl II cut pNP2 to reconstruct the full length μ genes (see Figure 5). Three different plasmids were created by these procedures; pNP11, pNP12 and pNP14, having the nucleotide sequences shown in Table 1 above.

In the construct pNP11 the potentially deleterious hairpin loop of pNP2 was abolished by changing the residues in the degenerate position of three codons and by introducing a different A-D to AUD sequence. These changes allow the region between the S-D and AUG to base pair with the 5' end of μ coding sequence to form a hairpin loop of $\Delta G = -7.8$ kcal, approximately equal to that of pNP2, but leaving the AUG and S-D sequences exposed (see Figure 4). The only other mutually exclusive hairpin loops are of $\Delta G = -2.3$, -3.6 and +0.4 kcal.

The construct pNP12 retains the S-D to AUG sequence of pNP11, but has the coding sequence of pNP2, such that the hairpin loop of pNP2 can form (see Figure 4).

The final μ construct was pNP14, which had the S-D to AUG sequence of pNP9, but the 5' coding sequence of pNP11, so that the hairpin loop of pNP12 could not form. The construct pNP14 had no secondary structure burying the S-D or μ AUG.

μ Protein Expression from Constructs with Differing mRNA Secondary Structures

E.coli B strains containing plasmids pNP9, pNP11, pNP12 and pNP14 were grown in induction medium and samples removed for ELISA assay. The concentration of μ was increased 6-7 fold in E.coli B cells containing pNp11 relative to pNP9 (see Table 1, Part b). Plasmid pNP12, which has the S-D to AUG sequence of pNP11, but the coding sequence of pNP9, showed only a two-fold increase in level of μ expression compared to pNP9.

Plasmid pNP14, which has the S-D to AUG sequence of pNP9 but the 5' μ coding sequence of pNP11, and thus differed from pNP9 only in three residues, was found to express μ protein over 90 times that found in pNP9.

A plasmid, pNP8, has also been constructed, from which μ is expressed, also from the trp promoter, as part of a trpE-μ fusion protein containing the amino terminal 53 amino acids of trpE and the carboxyl terminal 503 amino acids of μ. The rationale for making this fusion gene was that μ would be translated from an efficient bacterial ribosome binding site rather than from one of unknown efficiency. Then, by comparing the amounts of trpE-μ produced from pNP8 with those of μ produced from the different constructs, an estimate of relative ribosome binding efficiency may be made. When the trpE-μ gene in pNP8 was induced and the products quantitated by ELISA, it was found that the μ expression in strains containing pNP8 and pNP14 were very similar (see Table 1, part B).

The ribosome binding site (RBS) sequences of pNP11, pNP12 and nNp14 were designed in order to test the hypothesis that sequestering of the initiation codon into secondary structure inhibits translation initiation, and to increase the rate of initiation by releasing the initiation codon. In pNP11, the RBS hairpin loop of pNP2 and pNP9 was abolished by changing the 5' μ coding sequence in the degenerate position of three codons. In addition, the S-D to AUG sequence was changed to allow the base-pairing of this region with the 5' coding sequence. The result of these changes is that the initiation codon is exposed in the loop of a hairpin with the S-D sequence at the base of the stem (see Figure 4). This hairpin loop had a ΔG approximately equivalent to the ΔG of the pNP9 hairpin loop. Strains containing pNP11 produced 6-7 times as much μ as did those with pNP9. To test whether this was due to the change in S-D to ATG sequence characteristics, rather than to the liberation of the initiation codon, pNP12 was constructed as a control. The latter has the S-D to AUG sequence of pNP11, but retains the 5' μ coding sequence of pNP9, so that the RBS hairpin loop of pNP11 cannot form, but the potentially deleterious one of pNP9 can form. The level of μ expression from pNP12 was found to be increased by only two-fold relative to pNP9. This increase is probably due to an altered S-D to AUG distance, a U immediately 5' to the initiation codon , and the S-D to AUG sequence being G-poor each of which has been shown to be advantageous. It is therefore likely that

the increased expression from pNP11 is due to a preferable mRNA secondary structure. The RBS hairpin loop of pNP9 probably acts to inhibit translation while creation of another hairpin loop which exposes the initiation codon is responsible for the increased expression.

The plasmid pNP14 was constructed, such that it had the S-D to AUG sequence of pNP9, but the coding sequence of pNP11. The expression plasmid pNP14, therefore, differed from pNP9 in only three residues, those in the degenerate position of three codons, yet strains containing pNP14 expressed over ninety times more μ than those containing pNP9. This increase in expression cannot be explained on the basis of the three residue changes optimizing codon usage, for they introduce less favoured codons, as judged by the frequency of their occurrence in strongly expressed E.coli genes.

The increased expression is best explained by the abolition of the RBS hairpin loop in pNP9. The magnitude of the increases in expression observed, and the few residue changes introduced, makes it difficult for these results to be interpreted in terms of changes in mRNA stability.

The expression of pNP8, which encodes the trpE-μ fusion protein from a native trpE RBS, was also examined, and found to express at approximately the same level as pNP14. This may indicate that the RBS of pNP14 is equivalent to that of trpE in terms of its efficiency in directing translation initiation. The native trpE RBS of pNP8 and that of pNP14 were found to have no secondary structures hindering the use of their RBS sequences.

Expression of μ Protein in E. Coli

E. Coli B cells containing the μ expression plasmid pNP11 were grown under inducing conditions and soluble and insoluble extracts prepared, and analysed by SDS-PAGE. A novel band was seen after staining the gel with Coomassie blue in the lane containing proteins from the insoluble fraction (see Figure 6, lane 2). This band was not seen in the negative control lane which contained proteins from the same fraction from cells harbouring pCT70 (see Figure 6, lane 3). The novel band was found to migrate to a position corresponding to a protein of a molecular weight within less than 5% of the actual molecular weight of non-glycosylated μ of 62.5Kd. A duplicate set of lanes were transferred to nitrocellulose, and Western blotted. Alignment of the stained gel and the blot autoradiogram confirms that this novel band is antigenically related to IgM (see Figure 6, lanes 4 and 8). No band was found in extracts from cells containing pCT70 (see Figure 6, lanes 5 and 7). Only a low amount of μ was found in the soluble fraction (see Figure 6, lane 6).

A greatly increased level of expression of μ was found in E. Coli B compared to HB101. Pulse chase analysis demonstrated that in E. Coli B, a similar level of μ protein was detected after a 60 minute chase (figure 7, lane 3) as was seen after the initial labelling period (Figure 7, lane 1). In HB101, however, very little μ protein could be seen after a 10 minute chase (Figure 7, lane 7), and none after 30 minutes (Figure 7, lane 8), compared to the amount detected after the initial labelling period (Figure 7, lane 5).

For pulse chase analysis, inductions were set up as described above, except that the medium used consisted of: proline (0.3g/L), leucine (0.1g/L), Difco methionine assay medium (5g/L), glucose (60mg/L), thiamine (10mg/L),

$CaCl_2$ (22mg/L), $MgSO_4$ (0.25g/L)m and carbenicillin (0.1g/L). During exponential growth cells were pulse labelled with 30 µCi/ml L-[^{35}S] methionine for 2 minutes, after which unlabelled methione (100 µg/ml) was added and the incubation continued for the times indicated.

Induced E. Coli B cells harbouring pNP14 when examined by phase contrast microscopy were found to contain inclusion bodies.

Expression of λ Light Chain in E.Coli

A fresh 1 ml overnight culture of E.Coli HB101 or RV308 containing the plasmid under study was grown in L Broth (Maniatis et al,, Molecular Cloning, Cold Spring Harbor Laboratory, 1982), supplemented with carbenicillin to 100 µg/ml.

A 1:100 dilution of this culture was made into M9 medium (Maniatis, 1982, op.cit.) supplemented with glucose, vitamin B_1, carbenicillin, leucine and proline and the

cultures shaken at 37° for 5-6 hours. At this time ^{35}S-methionine was added for 5-20 minutes. The cells were then harvested by centrifugation, lysed by boiling in 1% SDS for 2 minutes and diluted by addition of a buffer consisting of 2% Triton X-100, 50 mM Tris pH 8, 0.15 M NaCl and 0.1 mM EDTA. Immunoprecipitations were carried out by addition of antisera to aliquots of labelled E.Coli extracts and, after incubation at 4° overnight, immune complexes were isolated by binding to Staphylococcus aureus fixed cells. The complexes were dissociated by boiling in 60 mM Tris pH 6.8 buffer containing 10% glycerol, 3% SDS and 5% mercaptoethanol and the liberated proteins analysed on 10% or 12.5% acrylamide/SDS gels (Laemmli, UK. Nature, 227, 680-685, 1970). Gels were stained with Coomassie Brilliant blue to visualise the protein bands while labelled proteins were detected on Fuji-RX film by fluorography using 1 M sodium salicylate.

HB101 cells containing pNP3 and other plasmids were grown under inducing conditions to an OD_{600} = 0.6 and the proteins present examined by specific immunoprecipitation substantially as described above. The results obtained are given in Figure 8, in which lanes 1 and 2 respectively were extracts from pCT54-19 and pNP3, immunoprecipitated with normal rabbit serum; lanes 3, 4 and 5 were extracts immunoprecipitated with rabbit anti-lambda serum and represent pNP4 (lane 3), pNP3 (lane 4) and pCT54-19 (lane 5). The position of unlabelled lambda protein from MOPC104E is indicated on the left hand side of Figure 8. All of the plasmids produced a protein which reacted with rabbit anti-mouse λ light chain serum and comigrated with authentic λ_1 light chain from the mouse myeloma MOPC104E. pNP3 however produced the most of this protein (compare lane 4 with lanes 3 and 5 in Figure 8). No such band was detected with control immunoprecipitations using normal rabbit serum (Figure 8, lanes 1 and 2). Studies using brief pulses with ^{35}S-methionine followed by chasing with

excess cold methionine indicated that the recombinant λ light chain had a 1/2 life of about 20 minutes.

In vitro transcription/translation (Pratt et al, Nucleic Acids Research, 9, 4459-447, 1981) confirmed that pNP3 coded for a protein which comigrated with authentic λ light chain. This 25 Kd product was synthesised in vitro at a rate comparable to that of β-lactamase indicating that it is synthesised in vivo at a level of 0.5% of total E.Coli proteins. This figure is in good agreement with the percentage of recombinant λ product synthesised in vivo (0.4-2%) as determined from the following equation:

$$\% \text{ specific CPM} = \frac{100 \times (\text{CPM with anti-}\lambda) - (\text{CPM with normal serum})}{\text{total CPM}}$$

The procedure used to disrupt cells was as follows (Emtage J S et al, Proc. Natl. Acad. Sci., 80, 3671-3715, 1983). E.Coli HB101/pNP3 grown under inducing conditions were harvested and resuspended in 0.05 M Tris pH 8, 0.233 M NaCl, 5% glycerol containing 130 g/ml of lysozyme and incubated at 4°C or room temperature for 20 minutes. Sodium deoxycholate was then added to a final concentration of 0.05% and 10 µg of DNAase 1 (from bovine pancreas) was added per g wet wt of E.Coli. The solution was incubated at 15°C for 30 minutes by which time the viscosity of the solution had decreased markedly. The extract was centrifuged (at 10,000 x g for 15 minutes for small volumes (1 ml) or 1 hour for larger volumes) to produce a soluble and an insoluble fraction.

For immunoprecipitations, the soluble fraction was diluted in the Triton-containing buffer described above and the insoluble fraction solubilised by boiling in 1% SDS followed by dilution in Triton-containing buffer. HB101 cells containing pNP3 or pNP4 were grown under inducing

conditions, pulse labelled with [^{35}S] methionine, separated into soluble and insoluble fractions and analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis. The results obtained are shown in Figure 9 in which lanes 1 and 3 correspond to the soluble fraction from HB101-pNP3 immunoprecipitated with normal rabbit serum and rabbit anti-lambda serum respectively; lanes 2 and 4 correspond to the insoluble fraction from HB101-pNP3 immunoprecipitated with normal rabbit serum and rabbit anti-lambda serum respectively, lanes 5 and 7 correspond to the soluble fraction from HB101-pNP4 immunoprecipitated with normal rabbit serum and rabbit anti-lambda serum respectively; lane 9 corresponds to the soluble fraction from HB101-pNP3, and lane 10 corresponds to the insoluble fraction from HB101-pNP3. The position of unlabelled lambda protein from MOPC104E is indicated on the left hand side of Figure 9. When such a procedure was carried out using pNP3, the recombinant λ light chain protein was present in the insoluble fraction (Figure 9, lane 4) rather than the soluble fraction (Figure 9, lane 3). No 25 Kd band comigrating with authentic λ_1 light chain was present when either the soluble or insoluble fraction were immunoprecipitated using normal rabbit serum (Figure 9, lanes 1 and 2). By inference it is likely that all Ig light chains will fractionate as in this specific example.

In the absence of specific immunoprecipitation a novel protein band was not visible from extracts of HB101 containing pNP3 nor was the λ protein found to accumulate. However, there was a dramatic difference when pNP3 was induced in the K12 strain E103s. In this strain λ protein was found to accumulate during induction until the cells reaches stationary phase (figure 10, lanes 3-5) to a level of about 150 times that found in HB101 as determined by an ELISA (enzyme linked immunosorbent assay). These cells were found to contain inclusion bodies which appeared refractile under light microscopy, a phenomenon characteristic of high level expression of foreign proteins. An estimate of the percentage of total E. Coli protein represented by recombinant λ protein was obtained by separating the proteins by gel electrophoresis, staining them, with Coomassie blue and scanning the stained gel with a Joyce-Loebl chromoscan 3. This method showed that λ was the major protein present (Figure 10, lane 5) and represented 13% of total E. Coli protein. The λ protein had a half-life of 20 minute in HB101 but accumulated to very high levels in E103s, suggesting that Lambda protein was much more stable in the latter strain. After cell lysis and centrifugation of HB101 or E103s containing pNP3, λ light chain was detected in the insoluble (Figure 10, lanes 7 and 10) but not in the soluble fractions ((Figure 10, lanes 6), as determined by Coomassie blue staining. The identity of the major Coomassie blue stained band as λ protein was confirmed by Western blot analysis (Figure 10, lanes 8-10). The presence of such immunoreactive bands was specific to pNP3 containing cells. When extracts from cells containing pCT70, a prochymosin expressing plasmid, were subjected to the same analysis no bands were detected. This more sensitive technique showed that a small amount of the λ protein was in the soluble fraction (Figure 10, lane 9). The presence of a number of distinct immuno-

reactive proteins all smaller than full-length protein were also detected. These may result from proteolytic degradation of of λ protein, from premature termination of transcription or from internal initiation of translation.

Expression of μ and Lambda Polypeptides in the same Bacterial Cell

Each of the Ig μ and λ genes in expression plasmids were transformed into the same E.Coli cell to direct the synthesis of both Ig μ and λ polypeptides. In order to overcome plasmid incompatibility and provide a second antibiotic resistance marker, the trp promoter and λ sequences were excised from pNP3 on a Hind III - Bam HI fragment and inserted into the Hind III - Bam HI fragment of pACYC 184. The resultant plasmid pACYC λ caused the E.Coli to grow very poorly. This weak growth was thought to result from read-through of RNA polymerase into the origin of replication. However, the inhibition of growth was virtually eliminated by the cloning in of the bacteriophage T7 transcriptional terminator at the Hind III site of plasmid pACYC λ .

This terminator functions in both orientations. The resultant plasmid pAC λ T7-1 has a chloramphenicol resistance gene and an origin compatible with the pBR 322-derived origin on pNP14, the Ig μ expressing plasmid. Transformation of both plasmids in the same E.Coli B was achieved in two steps. Firstly pNP14 was introduced followed by pAC λ T7 - I in two sequential transformations to give ampicillin and chloramphenicol resistant clones.

E.Coli B cells derived form this double-transfomant clone showed the presence of inclusion bodies and two novel polypeptide bands on stained gels of the insoluble fraction after lysis. These two bands correspond both with immunological activity by Western blotting for Ig μ and Ig λ and in their expected molecular weights. It has thus been shown that the double-transformant clone expresses both the heterologous genes. This had not hitherto been shown.

The presence of λ light chain in the insoluble fraction was a useful purification step since it both concentrated the protein and separated it from the bulk of E. Coli soluble proteins.

For further purification of the λ light chain, the cell debris were dissolved in 10mM Tris-HCl pH8.0, 25% formamide, 7M urea, 1mM EDTA and 2mM dithiothreitol. This material was loaded onto a DEAE Sephacel column (Pharmacia) (1x 25cm at a flow rate of 5ml/hr) which had been equilibrated in 9M urea, 10mM Tris-HCl pH8.0, 1mM EDTA and 2mM DTT. The DEAE Sephacel column was developed using a 0-150mM NaCl gradient in loading buffer. The eluted peak λ light chain immunoreactivity, corresponding to the major peak of protein, was diluted to a final concentration of 2.25M urea, 10mM Tris-Hcl pH8.0, 1mM EDTA, 2mM DTT and loaded onto an octyl-Sepharose column (Pharmacia) (2.5 x 10cm). Material was eluted by use of a urea gradient of 2.25-9M urea. The peak material was pooled, dialysed into ammonium bicarbonate and lyophilised. Following this step, only a single band of Coomassie blue stainable material corresponding to recombinant λ protein was visualised by SDS-PAGE.

The μ heavy chain was purified from 9M urea solubilised pellets by anion exchange chromatogrpahy and chromato-focussing (Pharmacia).

It was of great interest to determine whether the concomitant expression of μ and λ would lead to the formation of functional IgM. In order to determine this, extracts were made from E. Coli containing both Ig μ and λ polypeptides and these tested for antigen binding. We used a two-site sandwich ELISA which detects μ chain binding to haptenalated bovine serum albumin (NIP-caproate-BSA).

In the NIP binding assay, bovine serum albumin (BSA) is reacted with an equimolar amount of NIP-Cap-N-hydroxy-succinamide at pH 7.5 in 10mM phosphate buffer. The resulting NIP-Cap-BSA is separated from free NIP-Cap on a G-50 Sephadex column.

Microtitre plates (96 well Nunc Immuno Plate 1) are coated with 100 µl of a solution of 10 µg/ml NIP-BSA in sodium carbonate/sodium bicarbonate 0.1M, pH 9.6 buffer (coating buffer) overnight at 4°C.

The coated plates are then blocked for non-specific binding by addition of 100 µl of 0.5% casein in coating buffer (blocking buffer) and incubating at 37°C for 1 hour.

The coated and blocked plates are then washed three times in 150 mM NaCl, 0.05% NP40, 20mM Tris-HCl, pH 7.8 buffer (washing buffer).

The washed plates are shaken free of excess washing buffer and samples are added in their buffer or typically in 0.5% casein in washing buffer (sample buffer) to a 100 µl final volume.

In order to demonstrate the nature of the binding to NIP-BSA, samples for testing were supplemented with either NIP or NP is free solution at various concentrations (from 60 to 0.3 µM for NIP or from 600 to 3 µM for NP).

The plates with samples are then incubated at 37°C for one hour and thereafter washed three times with washing buffer.

The washed plates are then innoculated with 100 µl of anti µ-peroxidase conjugate (1:1000 dilution; TAGO Inc) in sample buffer, and incubated for one hour at 37°C. The plates are then washed three times in washing buffer.

The washed plates are innoculated with 100 µl of 0.1M Na acetate-citrate, pH 6.0, 0.1 mg/ml tetramethylbenzidine, 13mM H_2O_2 (peroxidase substrate) and incubated at room temperature for one hour. The reaction is terminated by the addition of 25µl of 2.5M H_2SO_4.

The microtitre plates are then read in a plate reader (Dynatech) at 450 nm with a reference of 630 nm. The $^{A}450$ was related to the level of a standard protein, B1-8 anti-NIP IgM.

This assay demonstrates sensitivty to 60pg of B1-8 IgM. The extracts were prepared as soluble and insoluble material. The insoluble material was solubilized in the same buffer used in lysis but containing 8M urea followed by its dilution for assay.

In order to obtain activity for the Ig µ and λ, extracts were made of the insoluble fraction and these dialysed into buffer conditions in which disulphide interchange will occur at a higher frequency.

Production of functional antibodies from E. Coli expressing both heavy and light chains was achieved by lysing the cells and clarifying the supernatant by centrifugation. The insoluble material was washed, followed by sonication (3 times for 3 minutes), and finally dissolved in 9M urea, 50mM glycine-NaOH pH10.8,

1mM EDTA, and 20mM 2-mercaptoethanol. This extract was dialysed for 40 hours against 3 changes of 20 vols. of 100mM KCl, 50mM glycine-NaOH pH10.8, 5% glycerol, 0.05mM EDTA, 0.5mM reduced glutathione and 0.1mM oxidised glutathione. The dialysate was cleared by centrifugation at 30,000g for 15 minutes and loaded directly onto DEAE Sephacel, followed by development with a 0-0.5M KCl linear gradient in 10mM Tris-HCl, 0.5mM EDTA, pH8.0.

The purified Igμ and λ were treated as above, except that no anion exchange chromatography was carried out. The preparation was finally dialysed into phosphate buffered saline, 5% glycerol, 0.01% sodium azide and 0.5mM EDTA pH7.4.

The results from assays of material processed in this way indicated that some activity was obtained. The level of activity obtained in this way was too low to do any detailed studies on, so the resultant dialysate was purified by anion exchange chromatography (Figure 11). This process resulted in the isolation of significant NIP-cap-BSA binding activity over that of background binding to BSA (Figure 11). The assay of the fractions for the level of Igμ, expressed as Bl-8 IgM equivalents demonstrated two peaks of activity. This was not found to correlate with full length Igμ by Western blotting. The first peak observed may represent a fragment of Igμ. The separation of NIP-CAP-BSA binding activity from the majority of full length Igμ and protein indicates that the hapten binding activity is contained within a particular molecular species formed at low efficiency.

The processing of insoluble material obtained from Igμ expression in E. Coli produced a similar IgM protein

profile but without NIP-cap-BSA binding activity. This demonstrates that the activity recovered was a property of the combined immunoglobulin expression, not of some E. Coli factor, or of the Ig μ heavy chain alone.

Further studies of the characteristics of the hapten (NIP-cap-BSA) binding were carried out. As samples were diluted they showed less binding to hapten in a very similar way to the original antibody (Figure 12). Free hapten was found to inhibit most of the binding activity in both undiluted and diluted samples. Using B1-8 antibody as a standard for both IgM and hapten binding, the specific activity of the assembled antibody was calculated to be 1.4×10^4 gm/gm of IgM equivalents. This value demonstrates the inefficient recovery of activity, but possibly represents an underestimate of the specific activity due to an overestimate of full-length Ig μ in these fractions, as described above.

Heteroclitic Nature of Recombined Antibody

Detailed specificity of binding to NIP-cap-BSA was investigated by comparing the assembled antibodies with B1-8 IgM in the presence of free NIP-cap and NP-cap (Figure 13). Both B1-8 IgM and the assembled antibodies showed that higher NP-cap than NIP-cap concentrations were required to inhibit NIP-cap-BSA binding.

The heteroclitic nature is demonstrated by the molar ratio of NIP to NP at 50% inhibition. The concentrations of NIP and NP at 50% inhibition (I50) were found to be similar for both B1-8 and the assembled antibodies as shown in Table 2. Also the NP I50/NIP I50 ratios were similar (Table 2).

Table 2. Hapten concentration at 50% inhibition (I50) of binding of antibodies to NIP-cap-BSA solid phase.

SD = Standard Deviation.

	NIP I50 uM	NP I50 uM	NP I50 / NIP I50
B1-8 IgM	0.13 (SD, 0.05)	3.7 (SD,2.9)	29
Fraction 26	0.34 (SD, 0.09)	1.9 (SD,0.4)	6
Fraction 27 and 28	0.11 (SD,0.02)	1.1 (SD,0.3)	10
Purified μ and λ	0.4	0.84	22

It has thus been shown that not only is it possible to express μ and λ proteins having intact variable domains, but also that it is possible to express both μ and λ proteins on compatible plasmids in the same cell. This latter has not hitherto been disclosed or even suggested. Moreover, it proved possible to derive functional Ig molecules from E. Coli cells which expressed both μ and λ proteins.

It has thus been shown that the process of the present invention produces a functional Ig molecule by recombinant DNA techniques.

Construction of Expression Plasmids for Yeast

The plasmids used for expression in yeast (Saccharomyces cerevisiae) are based on pBR322 and the yeast 2 micron plasmid. These plasmids are the subject of copending European Patent Application EP 0 073 635, the disclosure of which is incorporated hereby reference. The yeast phosphoglycerate kinase gene (PGK) provides the 5' sequence necessary for initiation of transcription and insertion of genes can be made at the unique BglII site in the plasmid pMA 3013 described in the above mentioned European patent application. This plasmid has now been renamed pMA91 as is so referred to hereinafter.

a) met-mu (μ)

BglII and BclI produce compatible 5'-GATC ends. The met-μ gene from pNP2 was excised on a partial BclI fragment and ligated into the unique BglII site of pMA91.

b) Pre-mu (μ)

The plasmid pCT54 u containing the full length pre-μ cDNA was digested with Hind III. This cuts at the 3' side of the μ cDNA. This Hind III site was changed to a BclI site by incubation with T_4 DNA polymerase in the presence

of all 4 nucleotides. The linker R107 of sequence TTTTGATCAAAA, which contains an internal BclI was ligated to the DNA obtained above. After ligation one aliquot of the resultant DNA was digested with BclI and AccI and the μ AccI-BclI fragment was isolated from the resultant mixture by gel electrophoresis. A separate aliquot of the resultant DNA was digested with MboII and ligated with chemically synthesised oligonucleotide linkers R121 and R112 which have cohesive BclI and MboII ends.

R121 5' GATCAATGGGATGGAGCTGT 3'

R112 5' TTACCCTACCTCGAC 3'

The ligation reaction was terminated and the resultant DNA digested with AccI to give a μ MboII-AccI fragment which was isolated from the resultant mixture by gel electrophoresis on a 5% polyacrylamide gel.

The two μ fragments obtained above were ligated together and then digested with BclI to eliminate unwanted ligation products after which the resultant pre-μ DNA was ligated to BglII cut plasmid pMA91.

The resultant ligation mix was used to transform E.Coli strain HB101 to ampicillin resistance. A colony of the transformed bacteria was isolated and was found to contain a plasmid exhibiting the predicted enzyme digestion pattern. The 5' end of the inserted DNA from this plasmid was sequenced and shown to have the anticipated sequence for a pre-μ cDNA.

c) met-lambda (λ)

The met-λ DNA was cloned into the BglII site of plasmid pMA91. Plasmid pCT54 Clone 1 was cut with Hind III to the 3' side of the met-λ cDNA and this cleavage site was altered to

a BclI site in a similar manner as previously described for pNP2. Following this the plasmid DNA was cut 5' of the λ gene at the BclI site and the resultant λ DNA was cloned into the BglII site of pMA91.

d) Pre-lambda (λ)

Pre-λ cDNA was reconstructed into pMA 91 as follows. The plasmid pCT54 was digested with BclI and Hind III and the resultant vector DNA was isolated by gel electrophoresis and ligated together with two synthetic oligonucleotides, R162 and R163, and the two fragments from plasmid pAT λ1-15.

R162 5' GATCAATGGCCTGGATT 3'

R163 5' GTGAAATCCAGGCCATT 3'

These fragments were produced by digesting pATλ1-15 with FokI and Hind III and isolating the 5' 300 base pair FokI fragment and the 3' 600 base pair FokI-Hind III fragment by gel electrophoresis. The resultant ligation mixture was used to transform E.Coli strain HB101 to ampicillin resistance and a colony of bacteria containing a plasmid having the predicted correct restriction enzyme digestion pattern was isolated. The 5' end of the inserted DNA was sequenced and found to have the anticipated sequence of a pre-λ cDNA. This plasmid was digested with Hind III and the resultant Hind III cleaved ends converted to a BclI site by blunting with T_4 polymerase and ligating with the linker R107. The resultant plasmid was then digested with BclI and the pre-λ cDNA isolated on a BclI fragment by gel electrophoresis. This BclI fragment was then ligated to BglI cut pMA91 to give a pMA 91 plasmid containing the full length pre-λ cDNA.

Expression of Immunogloublin Genes in Yeast

The pMA91 derivative plasmids containing the pre-λ and pre-μ genes as prepared above were used to transform

Saccharomyces cerevisiae yeast host organisms and the pre-λ and pre-μ genes were expressed by the transformed cells.

Saccharomyces cerevisiae strain MD46 when transformed with pMA91 pre-λ or pMA91 pre-μ gave rise to colonies which expressed immunoreactive proteins as revealed on Western blots. Yeast cells containing plasmid pMA91 pre-λ produced a protein which reacted with anti-λ antiserum and co-migrated on polyacrylamide gels with bacterially synthesised mature λ . Similarly yeast cells containing pMA 91 pre-μ produced a protein which reacted with anti-μ antiserum and co-migrated on polyacrylamide gels with bacterially synthesised mature μ protein. These observations indicate that both pre-λ and pre-μ are being processed to the corresponding mature proteins within the yeast host cell environment.

Additionally the μ protein product (but not the λ protein) was shown to be glycosylated. Cells were incubated in the presence or absence of tunicamycin at a concentration of 15 μg/ml. This compound specifically arrests the N-linked glycosylation of proteins. Cell extracts derived from cells incubated in the absence of tunicamycin showed higher molecular weight bands (as revealed by Western blotting) whilst extracts from cells incubated in the presence of tunicamycin showed no such higher molecular weight bands.

When cells were lysed with glass beads and the soluble and insoluble fractions examined it was found that the μ and λ proteins were exclusively in the insoluble fractions, as determined by Western blotting.

Secretion

After incubation transformed yeast cells were spun down and 1 ml volumes of supernatant were removed and passed through BSA-coated 0.2 μM Millipore filters to remove any remaining cellular material. ELISA assays for μ and λ protein were then carried out on the filtered supernatants. Only supernatants from pMA91 pre-λ harbouring cells showed detectable levels of immunoglobulin protein. Increased amounts of λ protein was detected in the supernatant when cells were grown in minimal medium to OD_{660} = 0.2 and then spun down and resuspended in YPAB, and harvested at OD_{660} = 1.0-1.5. Thus the λ protein but not the μ protein is secreted from yeast cells.

Intracellular Location

After incubation cells containing pMA91 pre-μ were converted to spheroplasts and fixed in 5% acetic acid/95% ethanl (v/v) and incubated with fluorescein conjugated goat anti-mouse μ. The fixed cells were examined by fluorescent microscopy and the μ protein was found to be localised in the periphery of the spheroplasts and especially in vacuoles.

Expression of both μ and λ Proteins in the Same Cells

In order to express both genes in the same host cells it was necessary to provide compatible plasmids for use in transformation. The pre-λ gene was excised from pMA91 pre-λ on a Hind III fragment and inserted into the Hind III site of plasmid pLG89 (Griti L, and Davies J, Gene, 25 179-188, 1983). This plasmid contains a ura3 marker which can be used as a positive selection for transformed host cells. A convenient $ura3^-$ host organism is S. cerevisiae strain X4003-5B. Both plasmids, pMA91 pre-μ

and pLG89 pre-λ , were transformed into this strain, and colonies were grown which contained together both the ura3 and leu2 markers. After incubation of the transformed cells both λ and μ proteins together were detected in the same cultures of X4003-5B using ELISA techniques. Levels of expression were comparable to those obtained for the individual genes in MD46 cells and for the λ gene alone in X4003-5B.

In addition, to check for assembly of the λ and μ proteins in vivo the following procedure was followd.

After the growth to an $OD_{660} \cong 1$, the transformed X4003-5B cells were spun down and resuspended in buffer, either 5 mM borate buffer at pH 8.0 or phosphate buffer at pH 5.8 either with or without detergent (e.g. 0.5% Triton X). The suspended cells were then lysed by vortexing with glass beads in buffer as above and the insoluble material was spun down. After centrifugation aliquots of the supernatant were assayed in a NIP binding assay (as described previously for assembled E.Coli μ and λ proteins). Specific antigen binding activity was detected in the supernatant and this activity was shown to be specifically competed out by free NIP (the specific antigen). These results indicated that μ and λ protein were expressed within the X4003-5B cells and assembled into functional immunogloublin molecules in vivo.

In the above experiments the culture medium used for incubating the transformed cells was yeast minimal medium (containing, per litre, 6.7 g DIFCO yeast nitrogen base without amino acids, 10 g glucose, and 200 mg each of histidine, tryptophan, methionine and adenine). When the λ protein alone was being expressed the medium contained in additon 200 mg/l of leucine, and when the μ protein alone was being expressed the medium contained in addition 200 mg/l of uracil.

CLAIMS:

1. A process for producing a heterologous multi chain polypeptide or protein in a single host cell, which comprises transforming the host cell with DNA sequences coding for each of the polypeptide chains and expressing said polypeptide chains in said transformed host cells.

2. A process according to claim 1, wherein each DNA sequence is inserted individually into a vector and the individual vectors are used to transform the single host cell.

3. A process according to claim 1, wherein at least two of the DNA sequences are inserted into separate parts of a vector which is used to transform the single host cell.

4. A process according to claim 2 or claim 3 wherein the or each vector is a plasmid.

5. A process according to claim 4, wherein the plasmid is col E1, pcR1, pBR322, RP4 or phage DNA.

6. A process according to any one of claims 1 to 5 wherein the host cell is a bacterium or a yeast.

7. A process according to claim 6, wherein the host cell is E.Coli, B. subtilis or S.cerevisiae.

8. A process according to claim 7, wherein the host cell is E.Coli strain HB101, X1776, X2887, PS410, MRC 1 RV308, 2 or E1035.

9. A process according to any one of claims 1 to 8, wherein the polypeptide chains are expressed and secreted by the host cell.

10. A process according to any one of claims 1 to 8, wherein the polypeptide chains expressed by the host cell are solubilised under conditions which facilitate assembly of the active multi chain polypeptide or protein.

11. A process according to any one of claims 1 to 10, for producing an Ig molecule or an immunologically functional Ig fragment, wherein the host cell is transformed with DNA sequences coding for at least the variable domains of the Ig heavy and light chains.

12. A process according to claim 11, wherein the DNA sequences code for the complete Ig heavy and light chains.

13. A process according to claim 11 or claim 12, for producing an Ig molecule or fragment having at least one constant domain, wherein the constant domain is derived from the same source as the variable domain to which it is attached.

14. A process according to claim 11 or claim 12, for producing an Ig molecule or fragment having at least one constant domain, wherein the constant domain is derived from a different source for that from which the variable domain to which it is attached is derived.

15. A process according to any one of claims 11 to 14, wherein the DNA coding sequences for the Ig molecule

or fragment are derived from one or more monoclonal antibody producing hybridomas.

16. A process for increasing the level of protein expression in a transformed host cell comprising altering the vector DNA sequence in the region of the initiator ATG codon to one which substantially eliminates RNA secondary structure.

17. A process according to claim 16 wherein the DNA sequence is altered to minimise the absolute value of the ΔG value of the sequence.

18. A heterologous multi chain polypeptide or protein produced by recombinant DNA technology from a single host cell.

19. An Ig molecule or an immunologically functional Ig fragment produced by recombinant DNA technology from a single host cell.

20. An Ig molecule or fragment according to claim 19, including at least one constant domain derived from the same source as the variable domain to which it is attached.

21. An Ig molecule or fragment according to claim 19, including at least one constant domain derived from a source different from the source from which the variable domain to which it is attached was derived.

22. An Ig fragment according to claim 20 or claim 21, comprising only one constant domain attached to each variable domain.

23. An Ig molecule according to claim 20 or claim 21, comprising sufficient constant domains attached to each variable domain to produce a complete Ig molecule.

24. An Ig heavy or light chain or a fragment thereof having a complete variable domain, as a product of recombinant DNA technology.

25. A vector for use in a process according to any one of claims 1 to 17.

26. A vector for use in transforming a host cell to enable it to produce a molecule according to any one of claims 18 to 24.

27. A vector according to claim 25 or claim 26, which is a plasmid.

28. A host cell transformed with a vector according to any one of claims 25 to 27.

29. A therapeutic composition comprising a molecule according to any one of claims 18 to 24 and a pharmaceutically acceptable carrier.

30. A process for preparing a composition according to claim 29, comprising mixing a molecule according to any one of claims 18 to 24 with a pharmaceutically acceptable carrier.

31. An affinity chromatography medium comprising a molecule according to any one of claims 18 to 24 immobilised on a chromatographic medium.

32. An immunological assay method in which the or each antigen binding Ig molecule or fragment is a molecule or fragment according to any one of claims 19 to 24.

33. An Ig molecule or fragment according to any one of claims 18 to 24 having attached thereto a cytotoxic, enzymic or structural polypeptide moiety synthesised in the cell with the Ig molecule or fragment

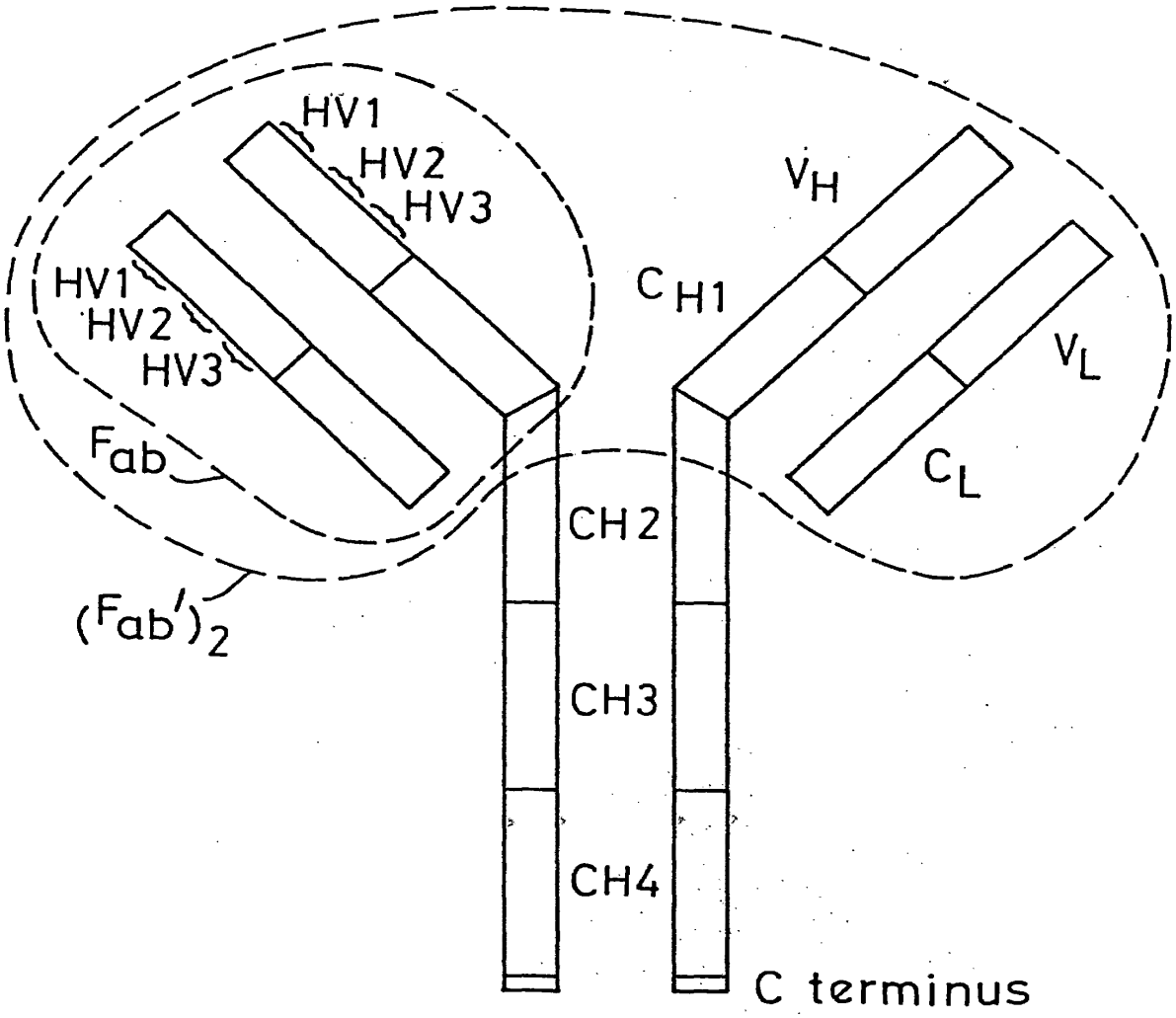
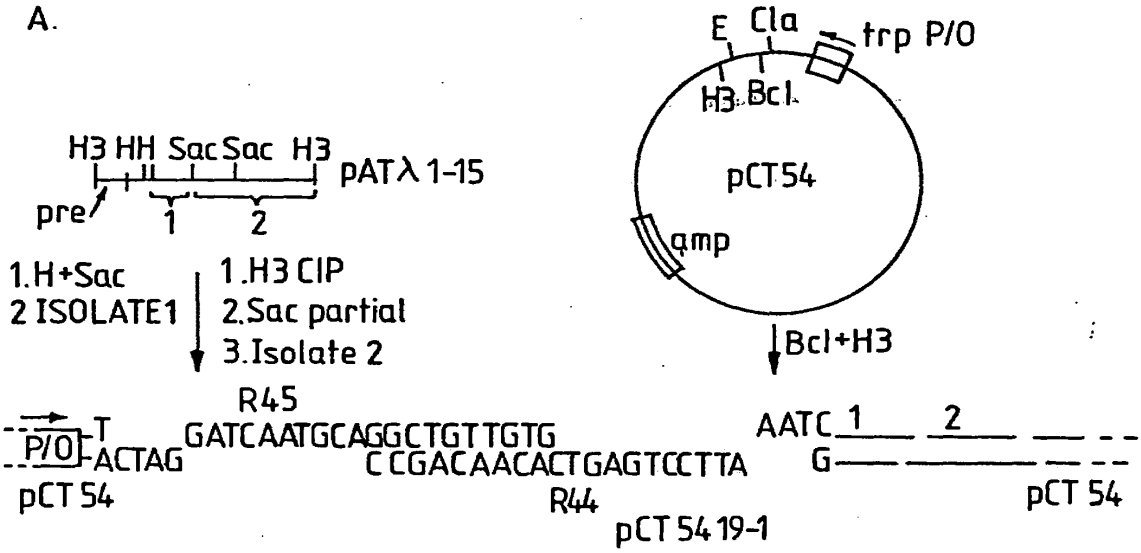

HV1
HV2
HV3
HV1
HV2
HV3
Fab
(Fab')2
VH
CH1
VL
CL
CH2
CH3
CH4
C terminus

FIG. 1



B.

pCT 54 19-1	AAGGGTATCGATTGATCAATG	
pNP3	AAGGGTA	TTGATCAATG
pNP4	AAGGGT	TTGATCAATG

Legend. E = E corRI

H = HinfI

H3= Hind III

FIG. 2

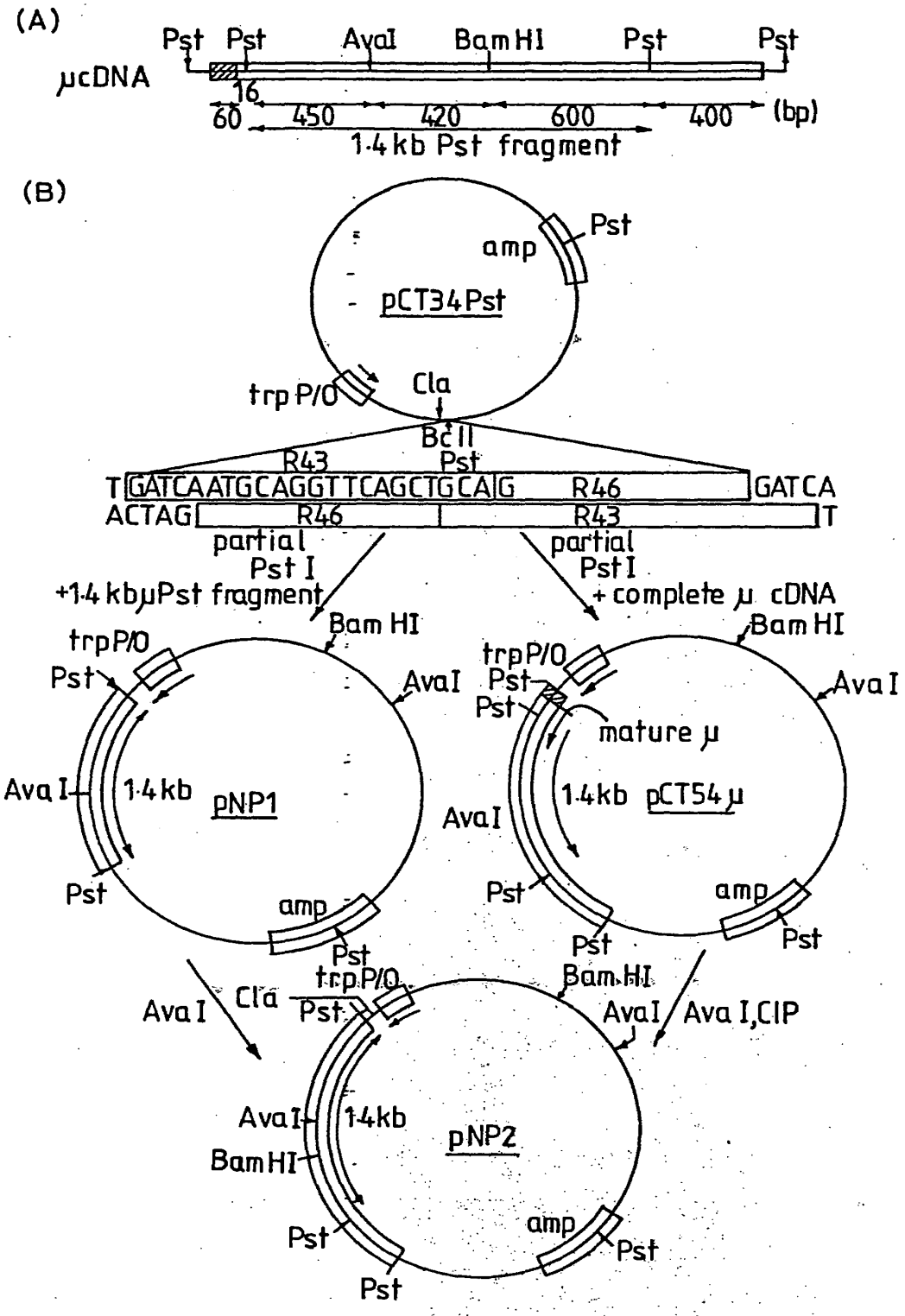
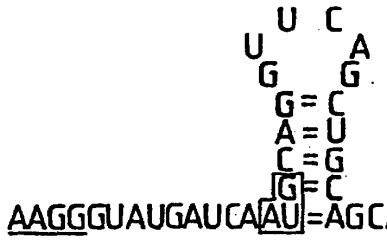

(A)
μcDNA
Pst Pst AvaI BamHI Pst Pst
16
60 450 420 600 400 (bp)
1.4kb Pst fragment
(B)
amp
Pst
pCT34Pst
trp P/O
Cla
BcII
Pst
R43
T GATCAATGCAGGTTCAGCTGCA G R46 GATCA
ACTAG R46 R43 T
partial
Pst I
+1.4kbμPst fragment
partial
Pst I
+ complete μ cDNA
Bam HI
trpP/O
Pst
AvaI
AvaI
1.4kb
pNP1
Pst
amp
Pst
Bam HI
trpP/O
Pst
Pst
AvaI
mature μ
1.4kb
pCT54μ
AvaI
Pst
amp
Pst
Pst
AvaI
Cla
trpP/O
Pst
Bam HI
AvaI
Ava I,CIP
AvaI
1.4kb
BamHI
pNP2
Pst
amp
Pst
Pst

FIG. 3

Possible 2° structures of μ mRNAs

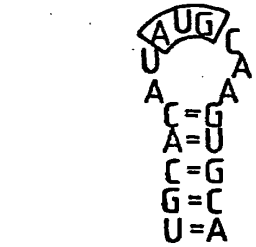
5′→3′

pNP9.



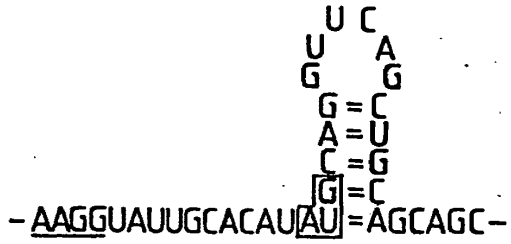
ΔG= 7·6 K.cal.
(1·0 ru)

pNP11.



ΔG=7·8 K cal.
(6·9ru)

pNP12



ΔG= 7·6 K cal.
(1·8 ru)

pNP14.

(101·2 ru)

-AAGGGUAUGAUCAAUGCAAGUGCAACUGCAG

FIG. 4

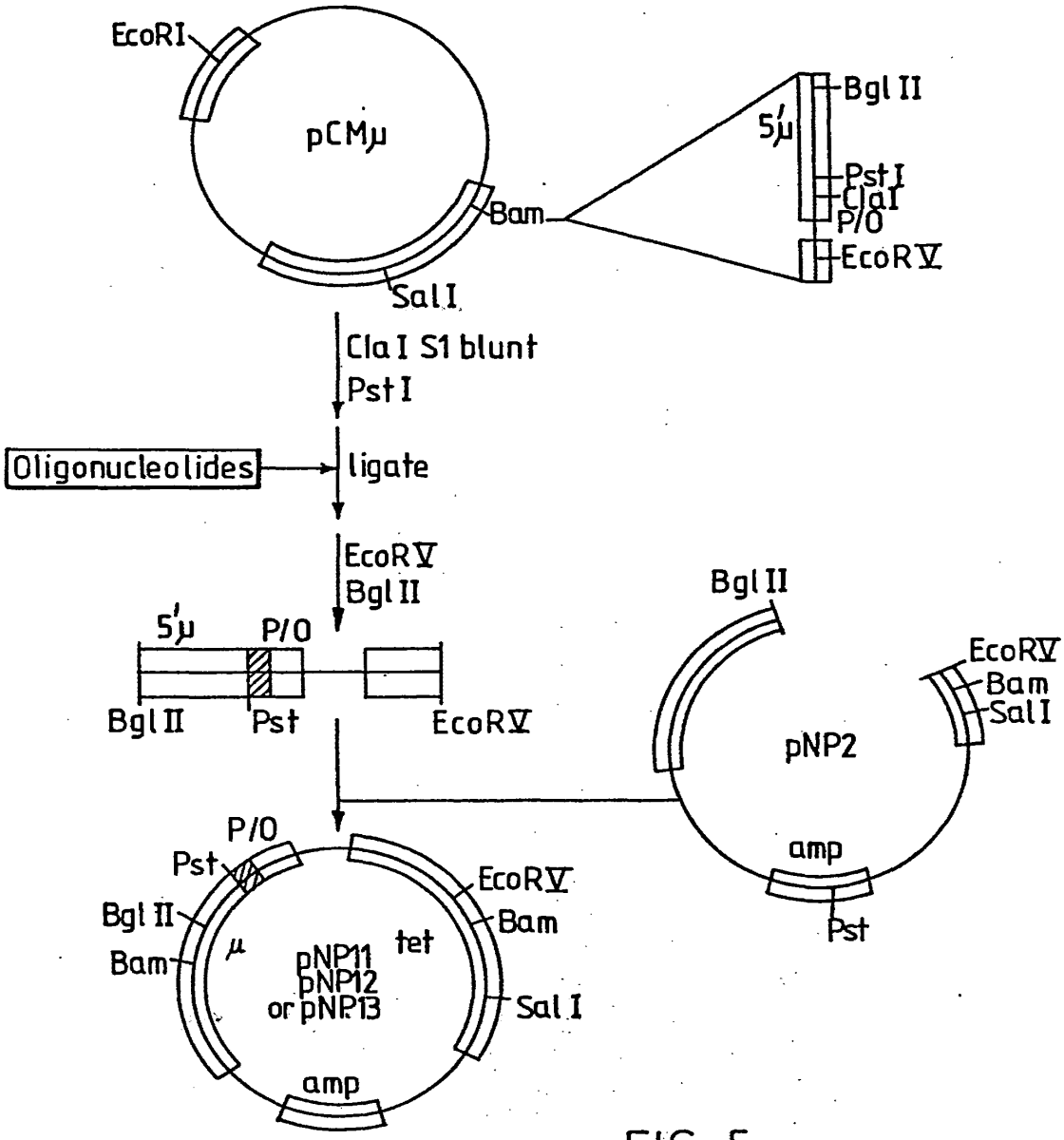

EcoRI
pCMμ
Bam
Sal I
5'μ
Bgl II
Pst I
Cla I
P/O
EcoR V
Cla I S1 blunt
Pst I
Oligonucleolides
ligate
EcoR V
Bgl II
5'μ
P/O
Bgl II
Pst
EcoR V
Bgl II
EcoR V
Bam
Sal I
pNP2
amp
Pst
P/O
Pst
Bgl II
Bam
μ
pNP11
pNP12
or pNP13
tet
EcoR V
Bam
Sal I
amp

FIG. 5

FIG. 6

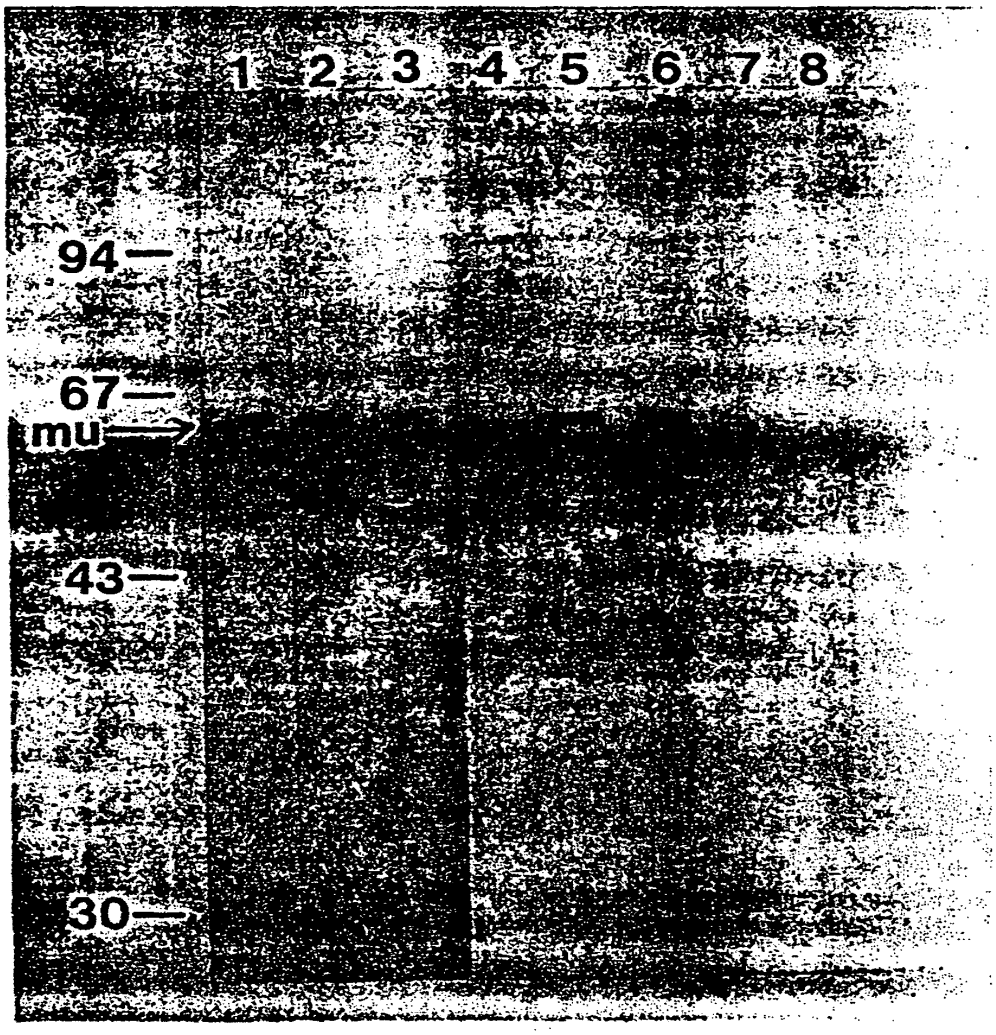


FIG. 7

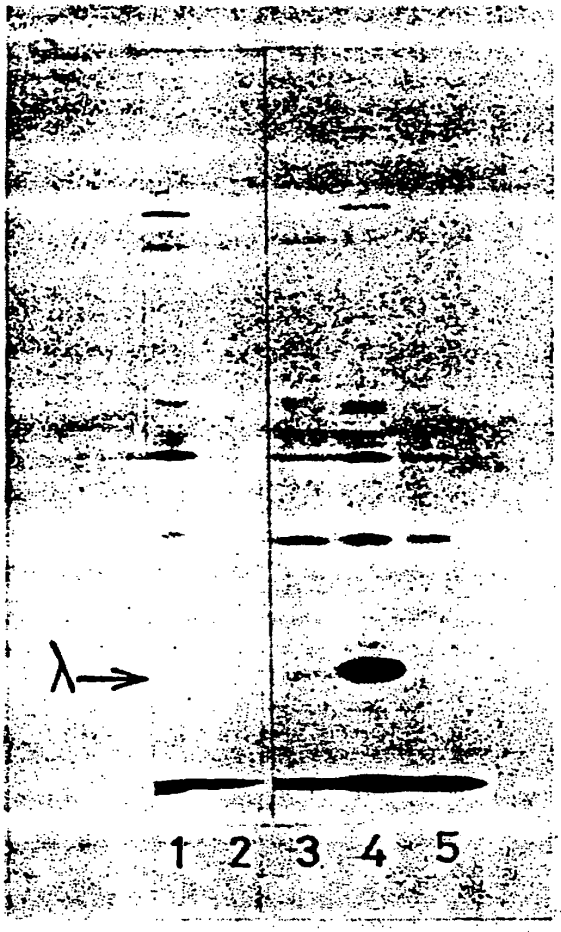


FIG. 8

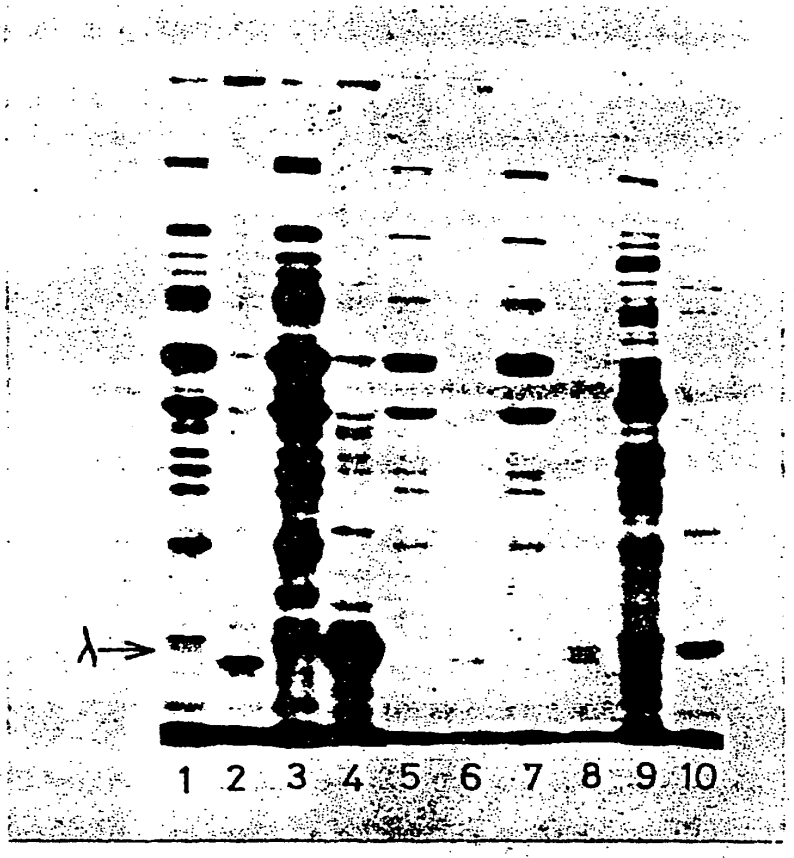


FIG. 9

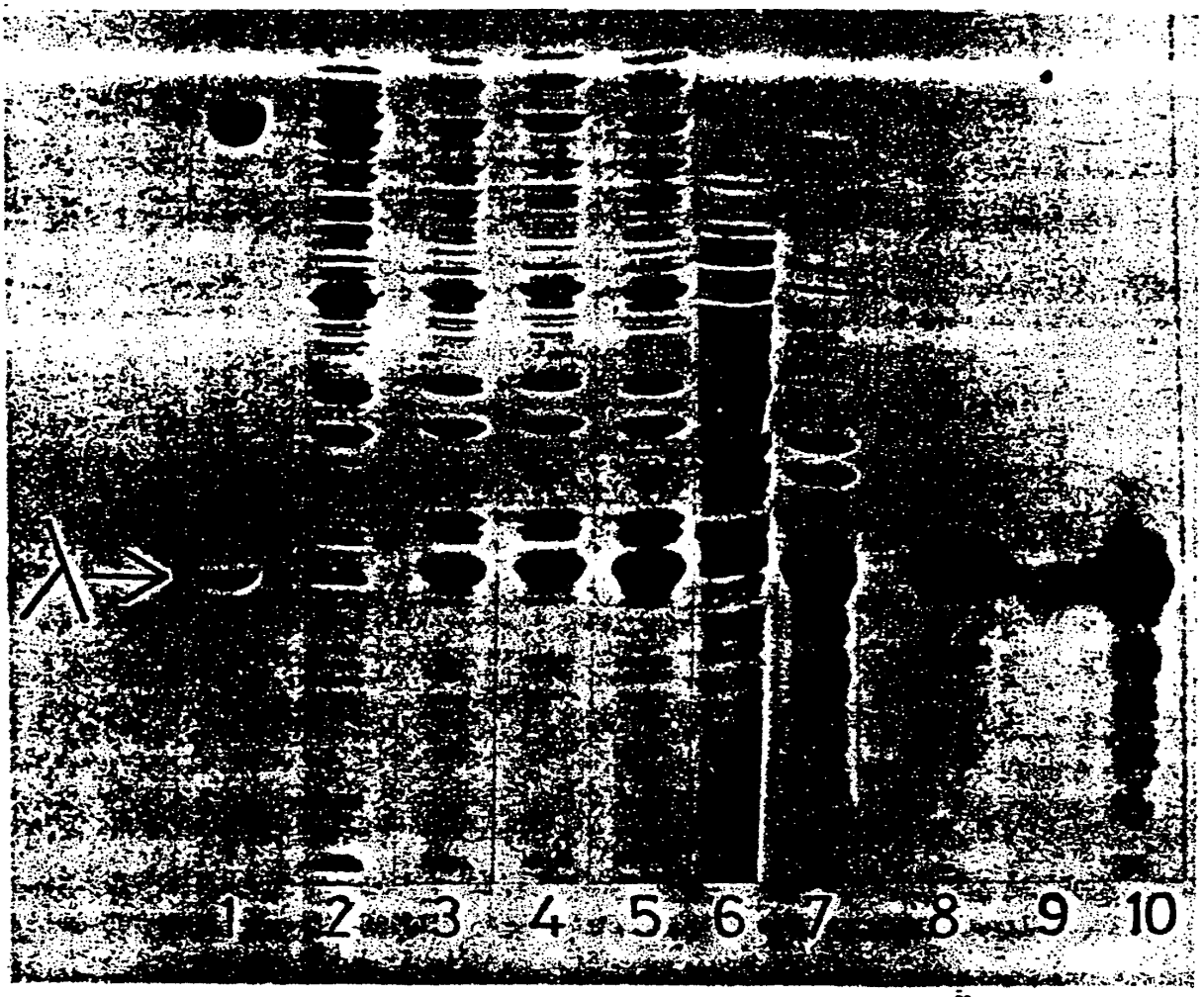


FIG. 10

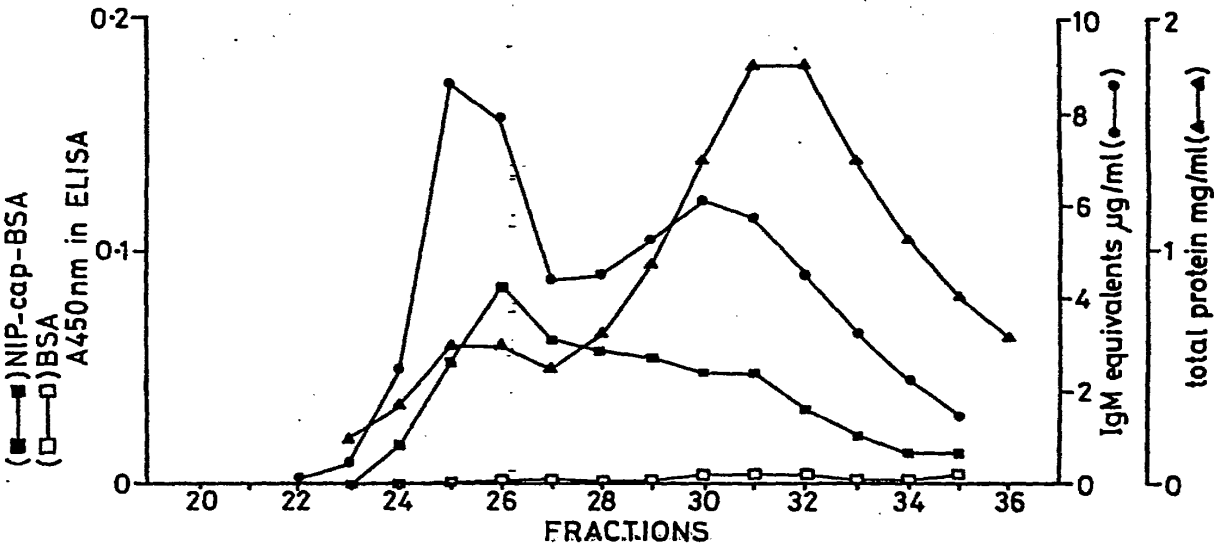

0·2
0·1
0
(■—■)NIP-cap-BSA
(□—□)BSA
A450nm in ELISA
20
22
24
26
28
30
32
34
36
FRACTIONS
10
8
6
4
2
0
IgM equivalents µg/ml(●—●)
2
1
0
total protein mg/ml(▲—▲)

FIG. 11

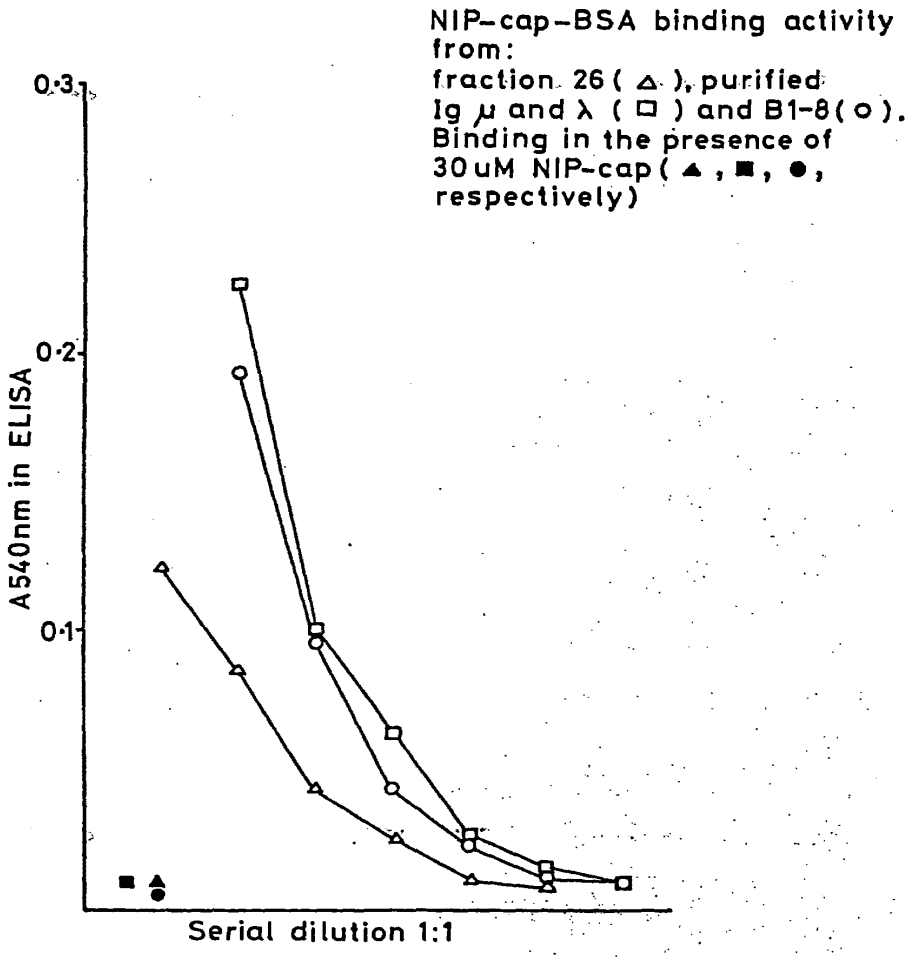
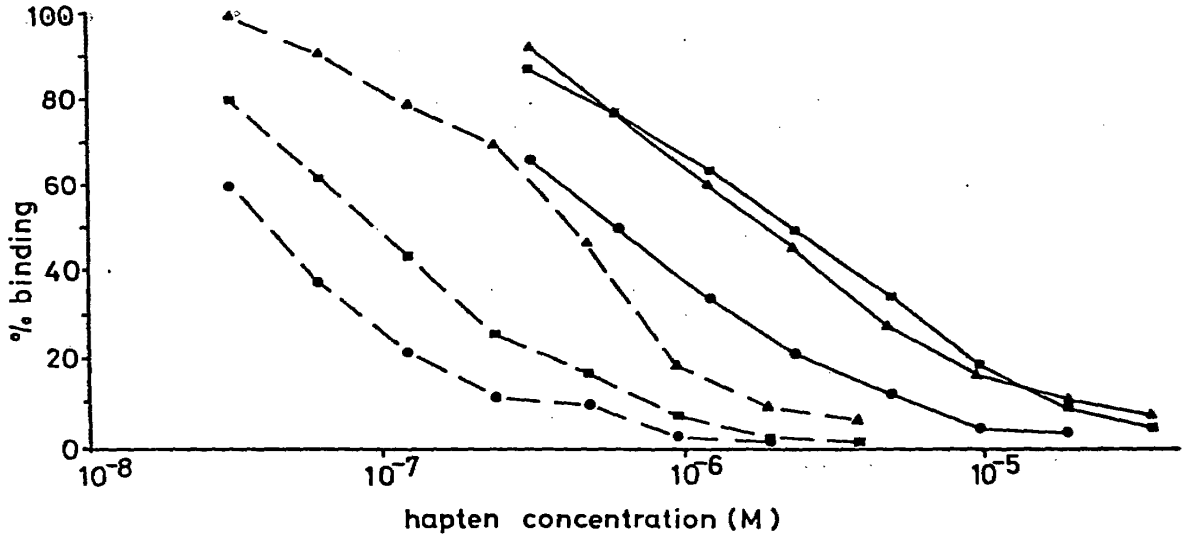

NIP-cap-BSA binding activity from:
fraction 26 (△), purified Ig μ and λ (□) and B1-8(○).
Binding in the presence of 30uM NIP-cap(▲ , ■ , ● , respectively)
0·3
0·2
0·1
A540nm in ELISA
Serial dilution 1:1

FIG. 12



Binding of antibodies to NIP-cap BSA B1-8 IgM (■), fraction 26 (▲), purified Ig μ or λ (●), in the presence of free NIP-cap (----) or NIP-cap (——).

FIG. 13

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number 0 125 023 A1

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 84302368.0

(22) Date of filing: 06.04.84

(51) Int. Cl.³: C 12 N 15/00
C 12 P 21/00, A 61 K 39/395

(30) Priority: 08.04.83 US 483457

(43) Date of publication of application
14.11.84 Bulletin 84/46

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: GENENTECH, INC.
460 Point San Bruno Boulevard
South San Francisco California 94080(US)

(71) Applicant: CITY OF HOPE
1450 East Duarte Road
Duarte California 91010(US)

(72) Inventor: Cabilly, Shmuel
325 South Second Avenue
Arcadia California 91006(US)

(72) Inventor: Holmes, William Evans
29 Eastlake
Pacifica California 94044(US)

(72) Inventor: Wetzel, Ronald Burnell
455 Urbano Drive
San Francisco California 94127(US)

(72) Inventor: Heyneker, Herbert Louis
2621 Easton Drive
Burlingame California 94010(US)

(72) Inventor: Riggs, Arthur Dale
4852 St. Andres Avenue
La Verne California 91750(US)

(74) Representative: Armitage, Ian Michael et al,
MEWBURN ELLIS & CO. 2/3 Cursitor Street
London EC4A 1BQ(GB)

(54) Recombinant immunoglobulin preparations, methods for their preparation, DNA sequences, expression vectors and recombinant host cells therefor.

(57) Recombinant DNA techniques are used to produce both immunoglobulins which are analogous to those normally found in vertebrate systems and to take advantage of these gene modification techniques to construct chimeric or other modified forms.

EP 0 125 023 A1

Croydon Printing Company Ltd

Docket 100/150

RECOMBINANT IMMUNOGLOBULIN PREPARATIONS, METHODS FOR THEIR PREPARATION, DNA SEQUENCES, EXPRESSION VECTORS AND RECOMBINANT HOST CELLS THEREFOR

Background of the Invention

This invention relates to the field of immunoglobulin production and to modification of naturally occuring immunoglobulin amino acid sequences. Specifically, the invention relates to using recombinant techniques to produce both immunoglobulins which are analogous to those normally found in vertebrate systems and to take advantage of these gene modification techniques to construct chimeric or other modified forms.

A. Immunoglobulins and Antibodies

Antibodies are specific immunoglobulin polypeptides produced by the vertebrate immune system in response to challenge by foreign proteins, glycoproteins, cells, or other antigenic foreign substances. The sequence of events which permits the organism to overcome invasion by foreign cells or to rid the system of foreign substances is at least partially understood. An important part of this process is the manufacture of antibodies which bind specifically to a particular foreign substance. The binding specificity of such polypeptides to a particular antigen is highly refined, and the multitude of specificities capable of being generated by the individual vertebrate is remarkable in its complexity and variability. Thousands of antigens are capable of

eliciting responses, each almost exclusively directed to the particular antigen which elicited it.

Immunoglobulins include both antibodies, as above described, and analogous protein substances which lack antigen specificity. The latter are produced at low levels by the lymph system and in increased levels by myelomas.

A.1 Source and Utility

Two major souces of vertebrate antibodies are presently utilized--generation in situ by the mammalian B lymphocytes and in cell culture by B-cell hybrids. Antibodies are made in situ as a result of the differentiation of immature B lymphocytes into plasma cells, which occurs in response to stimulation by specific antigens. In the undifferentiated B cell, the portions of DNA coding for the various regions on the immunoglobulin chains are separated in the genomic DNA. The sequences are reassembled sequentially prior to transcription. A review of this process has been given by Gough, Trends in Biochem Sci, 6: 203 (1981). The resulting rearranged genome is capable of expression in the mature B lymphocyte to produce the desired antibody. Even when only a single antigen is introduced into the sphere of the immune system for a particular mammal, however, a uniform population of antibodies does not result. The in situ immune response to any particular antigen is defined by the mosaic of responses to the various determinants which are present on the antigen. Each subset of homologous antibody is contributed by a single population of B cells--hence in situ generation of antibodies is "polyclonal".

This limited but inherent heterogeneity has been overcome in numerous particular cases by use of hybridoma technology to create "monoclonal" antibodies (Kohler, et al., Eur. J. Immunol., 6: 511 (1976)). In this process, splenocytes or lymphocytes from a mammal which has been injected with antigen are fused with a tumor cell

line, thus producing hybrid cells or "hybridomas" which are both immortal and capable of producing the genetically coded antibody of the B cell. The hybrids thus formed are segregated into single genetic strains by selection, dilution, and regrowth, and each strain thus represents a single genetic line. They therefore produce immunoreactive antibodies against a desired antigen which are assured to be homogenous, and which antibodies, referencing their pure genetic parentage, are called "monoclonal". Hybridoma technology has to this time been focused largely on the fusion of murine lines, but human-human hybridomas (Olsson, L. et al., Proc. Natl. Acad. Sci. (USA), 77: 5429 (1980)); human-murine hybridomas (Schlom, J., et al. (ibid) 77: 6841 (1980)) and several other xenogenic hybrid combinations have been prepared as well. Alternatively, primary, antibody producing, B cells have been immortalized in vitro by transformation with viral DNA.

Polyclonal, or, much more preferably, monoclonal, antibodies have a variety of useful properties similar to those of the present invention. For example, they can be used as specific immunoprecipitating reagents to detect the presence of the antigen which elicited the initial processing of the B cell genome by coupling this antigen-antibody reaction with suitable detection techniques such as labeling with radioisotopes or with enzymes capable of assay (RIA, EMIT, and ELISA). Antibodies are thus the foundation of immuno diagnostic tests for many antigenic substances. In another important use, antibodies can be directly injected into subjects suffering from an attack by a substance or organism containing the antigen in question to combat this attack. This process is currently in its experimental stages, but its potential is clearly seen. Third, whole body diagnosis and treatment is made possible because injected antibodies are directed to specific target disease tissues, and thus can be used either to determine the presence of the disease by carrying with them a

suitable label, or to attack the diseased tissue by carrying a suitable drug.

Monoclonal antibodies produced by hybridomas, while theoretically effective as suggested above and clearly preferable to polyclonal antibodies because of their specificity, suffer from certain disadvantages. First, they tend to be contaminated with other proteins and cellular materials of hybridoma, (and, therefore, mammalian) origin. These cells contain additional materials, notably nucleic acid fragments, but protein fragments as well, which are capable of enhancing, causing, or mediating carcinogic responses. Second, hybridoma lines producing monoclonal antibodies tend to be unstable and may alter the structure of antibody produced or stop producing antibody altogether (Kohler, G., et al., Proc. Natl. Acad. Sci (USA) 77: 2197 (1980); Morrison, S.L., J. Immunol. 123: 793 (1979)). The cell line genome appears to alter itself in response to stimuli whose nature is not currently known, and this alteration may result in production of incorrect sequences. Third, both hybridoma and B cells inevitably produce certain antibodies in glycosylated form (Melchers, F., Biochemistry, 10: 653 (1971)) which, under some circumstances, may be undesirable. Fourth, production of both monoclonal and polyclonal antibodies is relatively expensive. Fifth, and perhaps most important, production by current techniques (either by hybridoma or by B cell response) does not permit manipulation of the genome so as to produce antibodies with more effective design components than those normally elicited in response to antigens from the mature B cell in situ. The antibodies of the present invention do not suffer from the foregoing drawbacks, and, furthermore, offer the opportunity to provide molecules of superior design.

Even those immunoglobulins which lack the specificity of antibodies are useful, although over a smaller spectrum of potential uses than the antibodies themselves. In presently understood

applications, such immunoglobulins are helpful in protein replacement therapy for globulin related anemia. In this context, an inability to bind to antigen is in fact helpful, as the therapeutic value of these proteins would be impaired by such functionality. At present, such non-specific antibodies are derivable in quantity only from myeloma cell cultures suitably induced. The present invention offers an alternative, more economical source. It also offers the opportunity of cancelling out specificity by manipulating the four chains of the tetramer separately.

A.2 General Structure Characteristics

The basic immunoglobin structural unit in vertebrate systems is now well understood (Edelman, G.M., Ann. N.Y. Acad. Sci., 190: 5 (1971)). The units are composed of two identical light polypeptide chains of molecular weight approximately 23,000 daltons, and two identical heavy chains of molecular weight 53,000 - 70,000. The four chains are joined by disulfide bonds in a "Y" configuration wherein the light chains bracket the heavy chains starting at the mouth of the Y and continuing through the divergent region as shown in figure 1. The "branch" portion, as there indicated, is designated the Fab region. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, with some subclasses among them, and the nature of this chain, as it has a long constant region, determines the "class" of the antibody as IgG, IgM, IgA, IgD, or IgE. Light chains are classified as either kappa or lambda. Each heavy chain class can be prepared with either kappa or lambda light chain. The light and heavy chains are covalently bonded to each other, and the "tail" portions of the two heavy chains are bonded to each other by covalent disulfide linkages when the immunoglobulins are generated either by hybridomas or by B cells. However, if non-covalent association of the chains can be effected in the correct geometry, the aggregate will still be capable of reaction

with antigen, or of utility as a protein supplement as a non-specific immunoglobulin.

The amino acid sequence runs from the N-terminal end at the top of the Y to the C-terminal end at the bottom of each chain. At the N-terminal end is a variable region which is specific for the antigen which elicited it, and is approximately 100 amino acids in length, there being slight variations between light and heavy chain and from antibody to antibody. The variable region is linked in each chain to a constant region which extends the remaining length of the chain. Linkage is seen, at the genomic level, as occuring through a linking sequence known currently as the "J" region in the light chain gene, which encodes about 12 amino acids, and as a combination of "D" region and "J" region in the heavy chain gene, which together encode approximately 25 amino acids.

The remaining portions of the chain are referred to as constant regions and within a particular class do not to vary with the specificity of the antibody (i.e., the antigen eliciting it).

As stated above, there are five known major classes of constant regions which determine the class of the immunoglobulin molecule (IgG, IgM, IgA, IgD, and IgE corresponding to γ, μ, α, δ, and ϵ heavy chain constant regions). The constant region or class determines subsequent effector function of the antibody, including activation of complement (Kabat, E.A., Structural Concepts in Immunology and Immunochemistry, 2nd Ed., p. 413-436, Holt, Rinehart, Winston (1976)), and other cellular responses (Andrews, D.W., et al., Clinical Immunobiology pp 1-18, W.B. Sanders (1980); Kohl, S., et al., Immunology, 48: 187 (1983)); while the variable region determines the antigen with which it will react.

B. Recombinant DNA Technology

Recombinant DNA technology has reached sufficient sophistication

that it includes a repertoire of techniques for cloning and expression of gene sequences. Various DNA sequences can be recombined with some facility, creating new DNA entities capable of producing heterologous protein product in transformed microbes and cell cultures. The general means and methods for the in vitro ligation of various blunt ended or "sticky" ended fragments of DNA, for producing expression vectors, and for transforming organisms are now in hand.

DNA recombination of the essential elements (i.e., an origin of replication, one or more phenotypic selection characteristics, expression control sequence, heterologous gene insert and remainder vector) generally is performed outside the host cell. The resulting recombinant replicable expression vector, or plasmid, is introduced into cells by transformation and large quantities of the recombinant vehicle is obtained by growing the transformant. Where the gene is properly inserted with reference to portions which govern the transcription and translation of the encoded DNA message, the resulting expression vector is useful to produce the polypeptide sequence for which the inserted gene codes, a process referred to as "expression." The resulting product may be obtained by lysis, if necessary, of the host cell and recovery of the product by appropriate purifications from other proteins.

In practice, the use of recombinant DNA technology can express entirely heterologous polypeptides--so-called direct expression--or alternatively may express a heterologous polypeptide fused to a portion of the amino acid sequence of a homologous polypeptide. In the latter cases, the intended bioactive product is sometimes rendered bioinactive within the fused, homologous/heterologous polypeptide until it is cleaved in an extracellular environment.

The art of maintaining cell or tissue cultures as well as microbial systems for studying genetics and cell physiology is well

established. Means and methods are available for maintaining permanent cell lines, prepared by successive serial transfers from isolated cells. For use in research, such cell lines are maintained on a solid support in liquid medium, or by growth in suspension containing support nutriments. Scale-up for large preparations seems to pose only mechanical problems.

Summary of the Invention

The invention relates to antibodies and to non-specific immunoglobulins (NSIs) formed by recombinant techniques using suitable host cell cultures. These antibodies and NSIs can be readily prepared in pure "monoclonal" form. They can be manipulated at the genomic level to produce chimeras of variants which draw their homology from species which differ from each other. They can also be manipulated at the protein level, since all four chains do not need to be produced by the same cell. Thus, there are a number of "types" of immunoglobulins encompassed by the invention.

First, immunoglobulins, particularly antibodies, are produced using recombinant techniques which mimic the amino acid sequence of naturally occuring antibodies produced by either mammalian B cells in situ, or by B cells fused with suitable immortalizing tumor lines, i.e., hybridomas. Second, the methods of this invention produce, and the invention is directed to, immunoglobulins which comprise polypeptides not hitherto found associated with each other in nature. Such reassembly is particularly useful in producing "hybrid" antibodies capable of binding more than one antigen; and in producing "composite" immunoglobuins wherein heavy and light chains of different origins essentially damp out specificity. Third, by genetic manipulation, "chimeric" antibodies can be formed wherein, for example, the variable regions correspond to the amino acid sequence from one mammalian model system, whereas the constant region mimics the amino acid sequence of another. Again, the derivation of these two mimicked sequences may be from different

species. Fourth, also by genetic manipulation, "altered" antibodies with improved specificity and other characteristics can be formed.

Two other types of immunoglobulin-like moieties may be produced: "univalent" antibodies, which are useful as homing carriers to target tissues, and "Fab proteins" which include only the "Fab" region of an immunoglobulin molecule i.e, the branches of the "Y". These univalent antibodies and Fab fragments may also be "mammalian" i.e., mimic mammalian amino acid sequences; novel assemblies of mammalian chains, or chimeric, where for example, the constant and variable sequence patterns may be of different origin. Finally, either the light chain or heavy chain alone, or portions thereof, produced by recombinant techniques are included in the invention and may be mammalian or chimeric.

In other aspects, the invention is directed to DNA which encodes the aforementioned NSIs, antibodies, and portions thereof, as well as expression vectors or plasmids capable of effecting the production of such immunoglobulins in suitable host cells. It includes the host cells and cell cultures which result from transformation with these vectors. Finally, the invention is directed to methods of producing these NSIs and antibodies, and the DNA sequences, plasmids, and transformed cells intermediate to them.

Brief Description of the Drawings

Figure 1 is a representation of the general structure of immunoglobulins.

Figure 2 shows the detailed sequence of the cDNA insert of pK17G4 which encodes kappa anti CEA chain.

Figure 3 shows the coding sequence of the fragment shown in Figure 2, along with the corresponding amino acid sequence.

Figure 4 shows the combined detailed sequence of the cDNA inserts of $p_{\gamma}298$ and $p_{\gamma}11$ which encode gamma anti CEA chain.

Figure 5 shows the corresponding amino acid sequence encoded by the fragment in Figure 4.

Figures 6 and 7 outline the construction of expression vectors for kappa and gamma anti-CEA chains respectively.

Figures 8A, 8B, and 8C show the results of sizing gels run on extracts of E. coli expressing the genes for gamma chain, kappa chain, and both kappa and gamma chains respectively.

Figure 9 shows the results of western blots of extracts of cells transformed as those in Figures 8.

Figure 10 shows a standard curve for ELISA assay of anti CEA activity.

Figures 11 and 12 show the construction of a plasmid for expression of the gene encoding a chimeric heavy chain.

Figure 13 shows the construction of a plasmid for expression of the gene encoding the Fab region of heavy chain.

Detailed Description

A. Definitions

As used herein, "antibodies" refers to tetramers or aggregates thereof which have specific immunoreactive activity, comprising light and heavy chains usually aggregated in the "Y" configuration of Figure 1, with or without covalent linkage between them; "immunoglobulins" refers to such assemblies whether or not specific immunoreactive activity is a property. "Non-specific

immunoglobulin" ("NSI") means those immunoglobulins which do not possess specificity--i.e., those which are not antibodies.

"Mammalian antibodies" refers to antibodies wherein the amino acid sequences of the chains are homologous with those sequences found in antibodies produced by mammalian systems, either in situ, or in hybridomas. These antibodies mimic antibodies which are otherwise capable of being generated, although in impure form, in these traditional systems.

"Hybrid antibodies" refers to antibodies wherein chains are separately homologous with referenced mammalian antibody chains and represent novel assemblies of them, so that two different antigens are precipitable by the tetramer. In hybrid antibodies, one pair of heavy and light chain is homologous to antibodies raised against one antigen, while the other pair of heavy and light chain is homologous to those raised against another antigen. This results in the property of "divalence" i.e., ability to bind two antigens simultaneously. Such hybrids may, of course, also be formed using chimeric chains, as set forth below.

"Composite" immunoglobulins means those wherein the heavy and light chains mimic those of different species origins or specificities, and the resultant is thus likely to be a non-specific immunoglobulin (NSI), i.e.--lacking in antibody character.

"Chimeric antibodies" refers to those antibodies wherein one portion of each of the amino acid sequences of heavy and light chains is homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular class, while the remaining segment of the chains is homologous to corresponding sequences in another. Typically, in these chimeric antibodies, the variable region of both light and heavy chains mimics the variable regions of antibodies derived from one species

of mammals, while the constant portions are homologous to the sequences in antibodies derived from another. One clear advantage to such chimeric forms is that, for example, the variable regions can conveniently be derived from presently known sources using readily available hybridomas or B cells from non human host organisms in combination with constant regions derived from, for example, human cell preparations. While the variable region has the advantage of ease of preparation, and the specificity is not affected by its source, the constant region being human, is less likely to elicit an immune response from a human subject when the antibodies are injected than would the constant region from a non-human source.

However, the definition is not limited to this particular example. It includes any antibody in which either or both of the heavy or light chains are composed of combinations of sequences mimicking the sequences in antibodies of different sources, whether these sources be differing classes, differing antigen responses, or differing species of origin and whether or not the fusion point is at the variable/constant boundary. Thus, it is possible to produce antibodies in which neither the constant nor the variable region mimic known antibody sequences. It then becomes possible, for example, to construct antibodies whose variable region has a higher specific affinity for a particular antigen, or whose constant region can elicit enhanced complement fixation or to make other improvements in properties possessed by a particular constant region.

"Altered antibodies" means antibodies wherein the amino acid sequence has been varied from that of a mammalian or other vertebrate antibody. Because of the relevance of recombinant DNA techniques to this invention, one need not be confined to the sequences of amino acids found in natural antibodies; antibodies can be redesigned to obtain desired characteristics. The possible variations are many and range from the changing of just one or a few

amino acids to the complete redesign of, for example, the constant region. Changes in the constant region will, in general, be made in order to improve the cellular process characteristics, such as complement fixation, interaction with membranes, and other effector functions. Changes in the variable region will be made in order to improve the antigen binding characteristics. The antibody can also be engineered so as to aid the specific delivery of a toxic agent according to the "magic bullet" concept. Alterations, can be made by standard recombinant techniques and also by oligonucleotide-directed mutagenesis techniques (Dalbadie-McFarland, et al Proc. Natl. Acad. Sci. (USA), 79:6409 (1982)).

"Univalent antibodies" refers to aggregations which comprise a heavy chain/light chain dimer bound to the Fc (or stem) region of a second heavy chain. Such antibodies are specific for antigen, but have the additional desirable property of targeting tissues with specific antigenic surfaces, without causing its antigenic effectiveness to be impaired--i.e., there is no antigenic modulation. This phenomenon and the property of univalent antibodies in this regard is set forth in Glennie, M.J., et al., Nature, 295: 712 (1982). Univalent antibodies have heretofore been formed by proteolysis.

"Fab" region refers to those portions of the chains which are roughly equivalent, or analogous, to the sequences which comprise the Y branch portions of the heavy chain and to the light chain in its entirety, and which collectively (in aggregates) have been shown to exhibit antibody activity. "Fab protein", which protein is one of the aspects of the invention, includes aggregates of one heavy and one light chain (commonly known as Fab'), as well as tetramers which correspond to the two branch segments of the antibody Y, (commonly known as $F(ab)_2$), whether any of the above are covalently or non-covalently aggregated, so long as the aggregation is capable of selectively reacting with a particular antigen or

antigen family. Fab antibodies have, as have univalent ones, been formed heretofore by proteolysis, and share the property of not eliciting antigen modulation on target tissues. However, as they lack the "effector" Fc portion they cannot effect, for example, lysis of the target cell by macrophages.

"Fab protein" has similar subsets according to the definition of the present invention as does the general term "antibodies" or "immunoglobulins". Thus, "mammalian" Fab protein, "hybrid" Fab protein "chimeric" Fab and "altered" Fab protein are defined analogously to the corresponding definitions set forth in the previous paragraphs for the various types of antibodies.

Individual heavy or light chains may of course be "mammalian", "chimeric" or "altered" in accordance with the above. As will become apparent from the detailed description of the invention, it is possible, using the techniques disclosed to prepare other combinations of the four-peptide chain aggregates, besides those specifically defined, such as hybrid antibodies containing chimeric light and mammalian heavy chains, hybrid Fab proteins containing chimeric Fab proteins of heavy chains associated with mammalian light chains, and so forth.

"Expression vector" includes vectors which are capable of expressing DNA sequences contained therein, i.e., the coding sequences are operably linked to other sequences capable of effecting their expression. It is implied, although not always explicitly stated, that these expression vectors must be replicable in the host organisms either as episomes or as an integral part of the chromosomal DNA. Clearly a lack of replicability would render them effectively inoperable. A useful, but not a necessary, element of an effective expression vector is a marker encoding sequence -- i.e. a sequence encoding a protein which results in a phenotypic property (e.g. tetracycline resistance) of the cells

containing the protein which permits those cells to be readily identified. In sum, "expression vector" is given a functional definition, and any DNA sequence which is capable of effecting expression of a specified contained DNA code is included in this term, as it is applied to the specified sequence. As at present, such vectors are frequently in the form of plasmids, thus "plasmid" and "expression vector" are often used interchangeably. However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which may, from time to time become known in the art.

"Recombinant host cells" refers to cells which have been transformed with vectors constructed using recombinant DNA techniques. As defined herein, the antibody or modification thereof produced by a recombinant host cell is by virtue of this transformation, rather than in such lesser amounts, or more commonly, in such less than detectable amounts, as would be produced by the untransformed host.

In descriptions of processes for isolation of antibodies from recombinant hosts, the terms "cell" and "cell culture" are used interchangeably to denote the source of antibody unless it is clearly specified otherwise. In other words, recovery of antibody from the "cells" may mean either from spun down whole cells, or from the cell culture containing both the medium and the suspended cells.

B. Host Cell Cultures and Vectors

The vectors and methods disclosed herein are suitable for use in host cells over a wide range of prokaryotic and eukaryotic organisms. In general, of course, prokaryotes are preferred for cloning of DNA sequences in constructing the vectors useful in the invention. For example, E. coli K12 strain 294 (ATCC No. 31446) is particularly useful. Other microbial strains which may be used include E. coli strains such as E. coli B, and E. coli X1776 (ATTC No. 31537).

These examples are, of course, intended to be illustrative rather than limiting.

Prokaryotes may also be used for expression. The aforementioned strains, as well as E. coli W3110 (F^-, λ^-, prototrophic, ATTC No. 27325), bacilli such as Bacillus subtilus, and other enterobacteriaceae such as Salmonella typhimurium or Serratia marcesans, and various Pseudomonas species may be used.

In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell are used in connection with these hosts. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, E. coli is typically transformed using pBR322, a plasmid derived from an E. coli species (Bolivar, et al., Gene 2: 95 (1977)). pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells. The pBR322 plasmid, or other microbial plasmid must also contain, or be modified to contain, promoters which can be used by the microbial organism for expression of its own proteins. Those promoters most commonly used in recombinant DNA construction include the β-lactamase (penicillinase) and lactose promoter systems (Chang et al, Nature, 275: 615 (1978); Itakura, et al, Science, 198: 1056 (1977); (Goeddel, et al Nature 281: 544 (1979)) and a tryptophan (trp) promoter system (Goeddel, et al, Nucleic Acids Res., 8: 4057 (1980); EPO Appl Publ No. 0036776). While these are the most commonly used, other microbial promoters have been discovered and utilized, and details concerning their nucleotide sequences have been published, enabling a skilled worker to ligate them functionally with plasmid vectors (Siebenlist, et al, Cell 20: 269 (1980)).

In addition to prokaryates, eukaryotic microbes, such as yeast cultures may also be used. Saccharomyces cerevisiae, or common

baker's yeast is the most commonly used among eukaryotic microorganisms, although a number of other strains are commonly available. For expression in Saccharomyces, the plasmid YRp7, for example, (Stinchcomb, et al, Nature, 282: 39 (1979); Kingsman et al, Gene, 7: 141 (1979); Tschemper, et al, Gene, 10: 157 (1980)) is commonly used. This plasmid already contains the trp1 gene which provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example ATCC No. 44076 or PEP4-1 (Jones, Genetics, 85: 12 (1977)). The presence of the trp1 lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman, et al., J. Biol. Chem., 255: 2073 (1980)) or other glycolytic enzymes (Hess, et al, J. Adv. Enzyme Reg., 7: 149 (1968); Holland, et al, Biochemistry, 17: 4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In constructing suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination. Other promoters, which have the additional advantage of transcription controlled by growth conditions are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization (Holland, ibid.). Any plasmid vector containing yeast-compatible promoter, origin of replication and termination sequences is suitable.

In addition to microorganisms, cultures of cells derived from multicellular organisms may also be used as hosts. In principle, any such cell culture is workable, whether from vertebrate or invertebrate culture. However interest has been greatest in vertebrate cells, and propogation of vertebrate cells in culture (tissue culture) has become a routine procedure in recent years (Tissue Culture, Academic Press, Kruse and Patterson, editors (1973)). Examples of such useful host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, and W138, BHK, COS-7 and MDCK cell lines. Expression vectors for such cells ordinarily include (if necessary) an origin of replication, a promoter located in front of the gene to be expressed, along with any necessary ribosome binding sites, RNA splice sites, polyadenylation site, and transcriptional terminator sequences.

For use in mammalian cells, the control functions on the expression vectors are often provided by viral material. For example, commonly used promoters are derived from polyoma, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The early and late promoters of SV40 virus are particularly useful because both are obtained easily from the virus as a fragment which also contains the SV40 viral origin of replication (Fiers, et al, Nature, 273: 113 (1978)) incorporated herein by reference. Smaller or larger SV40 fragments may also be used, provided there is included the approximately 250 bp sequence extending from the Hind III site toward the Bgl I site located in the viral origin of replication. Further, it is also possible, and often desirable, to utilize promoter or control sequences normally associated with the desired gene sequence, provided such control sequences are compatible with the host cell systems.

An origin of replication may be provided either by construction of the vector to include an exogenous origin, such as may be derived from SV40 or other viral (e.g. Polyoma, Adeno, VSV, BPV, etc.)

source, or may be provided by the host cell chromosomal replication mechanism. If the vector is integrated into the host cell chromosome, the latter is often sufficient.

It will be understood that this invention, although described herein in terms of a preferred embodiment, should not be construed as limited to those host cells, vectors and expression systems exemplified.

C. Methods Employed

C.1 Transformation:

If cells without formidable cell wall barriers are used as host cells, transfection is carried out by the calcium phosphate precipitation method as described by Graham and Van der Eb, Virology, 52: 546 (1978). However, other methods for introducing DNA into cells such as by nuclear injection or by protoplast fusion may also be used.

If prokaryotic cells or cells which contain substantial cell wall constructions are used, the preferred method of transfection is calcium treatment using calcium chloride as described by Cohen, F.N. et al Proc. Natl. Acad. Sci. (USA), 69: 2110 (1972).

C.2 Vector Construction

Construction of suitable vectors containing the desired coding and control sequences employ standard ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to form the plasmids required. The methods employed are not dependent on the DNA source, or intended host.

Cleavage is performed by treating with restriction enzyme (or enyzmes) in suitable buffer. In general, about 1 µg plasmid or DNA fragments is used with about 1 unit of enzyme in about 20 µl of buffer solution. (Appropriate buffers and substrate amounts for

particular restriction enzymes are specified by the manufacturer.) Incubation times of about 1 hour at 37°C are workable. After incubations, protein is removed by extraction with phenol and chloroform, and the nucleic acid is recovered from the aqueous fraction by precipitation with ethanol.

If blunt ends are required, the preparation is treated for 15 minutes at 15° with 10 units of E. coli DNA Polymerase I (Klenow), phenol-chloroform extracted, and ethanol precipitated.

Size separation of the cleaved fragments is performed using 6 percent polyacrylamide gel described by Goeddel, D., et al, Nucleic Acids Res., 8: 4057 (1980) incorporated herein by reference.

For ligation, approximately equimolar amounts of the desired components, suitably end tailored to provide correct matching are treated with about 10 units T4 DNA ligase per 0.5 µg DNA. (When cleaved vectors are used as components, it may be useful to prevent religation of the cleaved vector by pretreatment with bacterial alkaline phosphatase.)

In the examples described below correct ligations for plasmid construction are confirmed by transforming E. coli K12 strain 294 (ATCC 31446) with the ligation mixture. Successful transformants were selected by ampicillin or tetracycline resistance depending on the mode of plasmid construction. Plasmids from the transformants were then prepared, analyzed by restriction and/or sequenced by the method of Messing, et al, Nucleic Acids Res., 9:309 (1981) or by the method of Maxam, et al, Methods in Enzymology, 65:499 (1980).

D. Outline of Procedures

D.1 Mammalian Antibodies

The first type of antibody which forms a part of this invention, and is prepared by the methods thereof, is "mammalian antibody"-one

wherein the heavy and light chains mimic the amino acid sequences of an antibody otherwise produced by a mature mammalian B lymphocyte either *in situ* or when fused with an immortalized cell as part of a hybridoma culture. In outline, these antibodies are produced as follows:

Messenger RNA coding for heavy or light chain is isolated from a suitable source, either mature B cells or a hybridoma culture, employing standard techniques of RNA isolation, and the use of oligo-dT cellulose chromatography to segregate the poly-A mRNA.. The poly-A mRNA may, further, be fractionated to obtain sequences of sufficient size to code for the amino acid sequences in the light or heavy chain of the desired antibody as the case may be.

A cDNA library is then prepared from the mixture of mRNA using a suitable primer, preferably a nucleic acid sequence which is characteristic of the desired cDNA. Such a primer may be hypothesized and synthesized based on the amino acid sequence of the antibody if the sequence is known. In the alternative cDNA from unfractionated poly-A mRNA from a cell line producing the desired antibody or poly-dT may also be used. The resulting cDNA is optionally size fractionated on polyacrylamide gel and then extended with, for example, dC residues for annealing with pBR322 or other suitable cloning vector which has been cleaved by a suitable restriction enzyme, such as Pst I, and extended with dG residues. Alternative means of forming cloning vectors containing the cDNA using other tails and other cloning vector remainder may, of course, also be used but the foregoing is a standard and preferable choice. A suitable host cell strain, typically *E. coli*, is transformed with the annealed cloning vectors, and the successful transformants identified by means of, for example, tetracycline resistance or other phenotypic characteristic residing on the cloning vector plasmid.

Successful transformants are picked and transferred to microtiter dishes or other support for further growth and preservation. Nitrocellulose filter imprints of these growing cultures are then probed with suitable nucleotide sequences containing bases known to be complementary to desired sequences in the cDNA. Several types of probe may be used, preferably synthetic single stranded DNA sequences labeled by kinasing with ATP^{32}. The cells fixed to the nitrocellulose filter are lysed, the DNA denatured, and then fixed before reaction with kinased probe. Clones which successfully hybridize are detected by contact with a photoplate, then plasmids from the growing colonies isolated and sequenced by means known in the art to verify that the desired portions of the gene are present.

The desired gene fragments are excised and tailored to assure appropriate reading frame with the control segments when inserted into suitable expression vectors. Typically, nucleotides are added to the 5' end to include a start signal and a suitably positioned restriction endonuclease site.

The tailored gene sequence is then positioned in a vector which contains a promoter in reading frame with the gene and compatible with the proposed host cell. A number of plasmids such as those described in U.S. Pat. Appln. Ser. Nos. 307473; 291892; and 305657 (EPO Publ. Nos. 0036776; 0048970 and 0051873) have been described which already contain the appropriate promoters, control sequences, ribosome binding sites, and transcription termination sites, as well as convenient markers.

In the present invention, the gene coding for the light chain and that coding for the heavy chain are recovered separately by the procedures outlined above. Thus they may be inserted into separate expression plasmids, or together in the same plasmid, so long as each is under suitable promoter and translation control.

The expression vectors constructed above are then used to transform suitable cells. The light and heavy chains may be transformed into separate cell cultures, either of the same or of differing species; separate plasmids for light and heavy chain may be used to co-transform a single cell culture, or, finally, a single expression plasmid containing both genes and capable of expressing the genes for both light and heavy chain may be transformed into a single cell culture.

Regardless of which of the three foregoing options is chosen, the cells are grown under conditions appropriate to the production of the desired protein. Such conditions are primarily mandated by the type of promoter and control systems used in the expression vector, rather than by the nature of the desired protein. The protein thus produced is then recovered from the cell culture by methods known in the art, but choice of which is necessarily dependent on the form in which the protein is expressed. For example, it is common for mature heterologous proteins expressed in E. coli to be deposited within the cells as insoluble particles which require cell lysis and solubilization in denaturant to permit recovery. On the other hand, proteins under proper synthesis circumstances, in yeast and bacterial strains, can be secreted into the medium (yeast and gram positive bacteria) or into the periplasmic space (gram negative bacteria) allowing recovery by less drastic procedures. Tissue culture cells as hosts also appear, in general, to permit reasonably facile recovery of heterologous proteins.

When heavy and light chain are coexpressed in the same host, the isolation procedure is designed so as to recover reconstituted antibody. This can be accomplished in vitro as described below, or might be possible in vivo in a microorganism which secretes the IgG chains out of the reducing environment of the cytoplasm. A more detailed description is given in D.2, below.

D.2 Chain Recombination Techniques

The ability of the method of the invention to produce heavy and light chains or portions thereof, in isolation from each other offers the opportunity to obtain unique and unprecedented assemblies of immunoglobulins, Fab regions, and univalent antibodies. Such preparations require the use of techniques to reassemble isolated chains. Such means are known in the art, and it is, thus, appropriate to review them here.

While single chain disulfide bond containing proteins have been reduced and reoxidized to regenerate in high yield native structure and activity (Freedman, R.B., et al. In Enzymology of Post Translational Modification of Proteins, I: 157-212 (1980) Academic Press, NY.), proteins which consist of discontinuous polypeptide chains held together by disulfide bonds are more difficult to reconstruct in vitro after reductive cleavage. Insulin, a cameo case, has received much experimental attention over the years, and can now be reconstructed so efficiently that an industrial process has been built around it (Chance, R.E., et al., In Peptides: Proceedings of the Seventh Annual American Peptide Symposium (Rich, D.H. and Gross, E., eds.) 721-728, Pierce Chemical Co., Rockford, IL. (1981)).

Immunoglobulin has proved a more difficult problem than insulin. The tetramer is stabilized intra and intermolecularly by 15 or more disulfide bonds. It has been possible to recombine heavy and light chains, disrupted by cleavage of only the interchain disulfides, to regain antibody activity even without restoration of the inter-chain disulfides (Edelman, G.M., et al., Proc. Natl. Acad. Sci. (USA) 50: 753 (1963)). In addition, active fragments of IgG formed by proteolysis (Fab fragments of ~50,000 MW) can be split into their fully reduced heavy chain and light chain components and fairly efficiently reconstructed to give active antibody (Haber, E., Proc. Natl. Acad. Sci. (USA) 52: 1099 (1964); Whitney, P.L.,

et al., Proc. Natl. Acad. Sci. (USA) 53: 524 (1965)). Attempts to reconstitute active antibody from fully reduced native IgG have been largely unsuccessful, presumably due to insolubility of the reduced chains and of side products or intermediates in the refolding pathway (see discussion in Freedman, M.H., et al., J. Biol. Chem. 241: 5225 (1966)). If, however, the immunoglobulin is randomly modified by polyalanylation of its lysines before complete reduction, the separated chains have the ability to recover antigen-combining activity upon reoxidation (ibid).

A particularly suitable method for immunoglobulin reconstitution is derivable from the now classical insulin recombination studies, wherein starting material was prepared by oxidative sulfitolysis, thus generating thiol-labile S-sulfonate groups at all cysteines in the protein, non-reductively breaking disulfides (Chance et al. (supra)). Oxidative sulfitolysis is a mild disulfide cleavage reaction (Means, G.E., et al., Chemical Modification of Proteins, Holden-Day, San Francisco (1971)) which is sometimes more gentle than reduction, and which generates derivatives which are stable until exposed to mild reducing agent at which time disulfide reformation can occur via thiol-disulfide interchange. In the present invention the heavy and light chain S-sulfonates generated by oxidative sulfitolysis were reconstituted utilizing both air oxidation and thiol-disulfide interchange to drive disulfide bond formation. The general procedure is set forth in detail in U.S. Serial No. 452,187, filed Dec. 22, 1982 (EPO Appln. No. 83.307840.5), incorporated herein by reference.

D.3 Variants Permitted by Recombinant Technology

Using the techniques described in paragraphs D.1 and D.2, additional operations which were utilized to gain efficient production of mammalian antibody can be varied in quite straightforward and simple ways to produce a great variety of

modifications of this basic antibody form. These variations are inherent in the use of recombinant technology, which permits modification at a genetic level of amino acid sequences in normally encountered mammalian immunoglobulin chains, and the great power of this approach lies in its ability to achieve these variations, as well as in its potential for economic and specific production of desired scarce, and often contaminated, molecules. The variations also inhere in the ability to isolate production of individual chains, and thus create novel assemblies.

Briefly, since genetic manipulations permit reconstruction of genomic material in the process of construction of expression vectors, such reconstruction can be manipulated to produce new coding sequences for the components of "natural" antibodies or immunoglobulins. As discussed in further detail below, the coding sequence for a mammalian heavy chain may not be derived entirely from a single source or single species, but portions of a sequence can be recovered by the techniques described in D.1 from differing pools of mRNA, such as murine-murine hybridomas, human-murine hybridomas, or B cells differentiated in response to a series of antigen challenges. The desired portions of the sequences in each case can be recovered using the probe and analysis techniques described in D.1, and recombined in an expression vector using the same ligation procedures as would be employed for portions of the same model sequence. Such chimeric chains can be constructed of any desired length; hence, for example, a complete heavy chain can be constructed, or only sequence for the Fab region thereof.

The additional area of flexibility which arises from the use of recombinant techniques results from the power to produce heavy and light chains or fragments thereof in separate cultures or of unique combinations of heavy and light chain in the same culture, and to prevent reconstitution of the antibody or immunoglobulin aggregation until the suitable components are assembled. Thus, while normal

antibody production results automatically in the formation of "mammalian antibodies" because the light and heavy chain portions are constructed in response to a particular determinant in the same cell, the methods of the present invention present the opportunity to assemble entirely new mixtures. Somewhat limited quantities of "hybrid" antibodies have been produced by "quadromas" i.e., fusions of two hybridoma cell cultures which permit random assemblies of the heavy and light chains so produced.

The present invention permits a more controlled assembly of desired chains, either by mixing the desired chains in vitro, or by transforming the same culture with the coding sequences for the desired chains.

D.4 Composite Immunoglobulins

The foregoing procedure, which describes in detail the recombinant production of mammalian antibodies is employed with some modifications to construct the remaining types of antibodies or NSIs encompassed by the present invention. To prepare the particular embodiment of composite non-specific immunoglobulin wherein the homology of the chains corresponds to the sequences of immunoglobulins of different specificities, it is of course, only necessary to prepare the heavy and light chains in separate cultures and reassemble them as desired.

For example, in order to make an anti-CEA light chain/anti-hepatitis heavy chain composite antibody, a suitable source for the mRNA used as a template for the light chain clone would comprise, for instance, the anti CEA producing cell line of paragraph E.1. The mRNA corresponding to heavy chain would be derived from B cells raised in response to hepatitis infection or from hybridoma in which the B cell was of this origin. It is clear that such composites can be assembled using the methods of the invention almost at will, and are limited only by available sources of mRNA suitable for use as

templates for the respective chains. All other features of the process are similar to those described above.

D.5 Hybrid Antibodies

Hybrid antibodies are particularly useful as they are capable of simultaneous reaction with more than one antigen. Pairs of heavy and light chains corresponding to chains of antibodies for different antigens, such as those set forth in paragraph D.4 are prepared in four separate cultures, thus preventing premature assembly of the tetramer. Subsequent mixing of the four separately prepared peptides then permits assembly into the desired tetramers. While random aggregation may lead to the formation of considerable undesired product, that portion of the product in which homologous light and heavy chains are bound to each other and mismatched to another pair gives the desired hybrid antibody.

D.6 Chimeric Antibodies

For construction of chimeric antibodies (wherein, for example, the variable sequences are separately derived from the constant sequences) the procedures of paragraph D.1 and D.2 are again applicable with appropriate additions and modifications. A preferred procedure is to recover desired portions of the genes encoding for parts of the heavy and light chains from suitable, differing, sources and then to religate these fragments using restriction endonucleases to reconstruct the gene coding for each chain.

For example, in a particularly preferred chimeric construction, portions of the heavy chain gene and of the light chain gene which encode the variable sequences of antibodies produced by a murine hybridoma culture are recovered and cloned from this culture and gene fragments encoding the constant regions of the heavy and light chains for human antibodies recovered and cloned from, for example, human myeloma cells. Suitable restriction enzymes may then be used

to ligate the variable portions of the mouse gene to the constant regions of the human gene for each of the two chains. The chimeric chains are produced as set forth in D.1, aggregated as set forth in D.2 and used in the same manner as the non-chimeric forms. Of course, any splice point in the chains can be chosen.

D.7 Altered Antibodies

Altered antibodies present, in essence, an extension of chimeric ones. Again, the techniques of D.1 and D.2 are applicable; however, rather than splicing portions of the chain(s), suitable amino acid alterations, deletions or additions are made using available techniques such as mutagenesis (supra). For example, genes which encode antibodies having diminished complement fixation properties, or which have enhanced metal binding capacities are prepared using such techniques. The latter type may, for example, take advantage of the known gene sequence encoding metalothionein II (Karin, M., et al., Nature, 299: 797 (1982)). The chelating properties of this molecular fragment are useful in carrying heavy metals to tumor sites as an aid in tumor imaging (Scheinberg, D.A., et al., Science, 215: 19 (1982).

D.8 Univalent Antibodies

In another preferred embodiment, antibodies are formed which comprise one heavy and light chain pair coupled with the Fc region of a third (heavy) chain. These antibodies have a particularly useful property. They can, like ordinary antibodies, be used to target antigenic surfaces of tissues, such as tumors, but, unlike ordinary antibodies, they do not cause the antigenic surfaces of the target tissue to retreat and become non-receptive. Ordinary antibody use results in aggregation and subsequent inactivation, for several hours, of such surface antigens.

The method of construction of univalent antibodies is a straightforward application of the invention. The gene for heavy

chain of the desired Fc region is cleaved by restriction enzymes, and only that portion coding for the desired Fc region expressed. This portion is then bound using the technique of D.2 to separately produced heavy chain the desired pairs separated from heavy/heavy and Fc/Fc combinations, and separately produced light chain added. Pre-binding of the two heavy chain portions thus diminishes the probability of formation of ordinary antibody.

D.9 Fab Protein

Similarly, it is not necessary to include the entire gene for the heavy chain portion. All of the aforementioned variations can be superimposed on a procedure for Fab protein production and the overall procedure differs only in that that portion of the heavy chain coding for the amino terminal 220 amino acids is employed in the appropriate expression vector.

E. Specific Examples of Preferred Embodiments

The invention has been described above in general terms and there follow several specific examples of embodiments which set forth details of experimental procedure in producing the desired antibodies. Example E.1 sets forth the general procedure for preparing anti CEA antibody components, i.e. for a "mammalian antibody". Example E.3 sets forth the procedure for reconstitution and thus is applicable to preparation of mammalian, composite, hybrid and chimeric immunoglobulins, and Fab proteins and univalent antibodies. Example E.4 sets forth the procedure for tailoring the heavy or light chain so that the variable and constant regions may be derived from different sources. Example E.5 sets forth the method of obtaining a shortened heavy chain genome which permits the production of the Fab regions and, in an analogous manner, Fc region.

The examples set forth below are included for illustrative purposes and do not limit the scope of the invention.

E.1 Construction of Expression Vectors for Murine anti-CEA Antibody Chains and Peptide Synthesis

Carcinoembryonic antigen (CEA) is associated with the surface of certain tumor cells of human origin (Gold, P., et al., J. Exp. Med., 122: 467 (1965)). Antibodies which bind to CEA (anti-CEA antibodies) are useful in early detection of these tumors (Van Nagell, T.R., et al., Cancer Res. 40: 502 (1980)), and have the potential for use in treatment of those human tumors which appear to support CEA at their surfaces. A mouse hybridoma cell line which secretes anti-CEA antibodies of the $Ig\gamma_1$ class, CEA.66-E3, has been prepared as described by Wagener, C. et al., J. Immunol. 130, 2308 (1983) which is incorporated herein by reference, and was used as mRNA source. The production of anti CEA antibodies by this cell line was determined. The N-terminal sequences of the antibodies produced by these cells was compared with those of monoclonal anti CEA as follows. Purified IgG was treated with PCAse (Podell, D.N., et al., Biochem. Biophys. Res. Commun. 81: 176 (1978)), and then dissociated in 6M guanidine hydrochloride, 10 mM 2-mercaptoethanol (1.0 mg of immunoglobulin, 5 min, 100°C water bath). The dissociated chains were separated on a Waters Associates alkyl phenyl column using a linear gradient from 100 percent A (0.1 percent TFA-water) to 90 percent B (TFA/H_2O/MeCN 0.1/9.9/90) at a flow rate of 0.8 ml/min. Three major peaks were eluted and analyzed on SDS gels by silver staining. The first two peaks were pure light chain (MW 25,000 daltons), the third peak showed a (7:3) mixture of heavy and light chain. 1.2 nmoles of light chain were sequenced by the method of Shively, J.E., Methods in Enzymology, 79: 31 (1981), with an NH_2-terminal yield of 0.4 nmoles. A mixture of heavy and light chains (3 nmoles) was also sequenced, and sequence of light chain was deducted from the double sequence to yield the sequence of the heavy chain.

In the description which follows, isolation and expression of the genes for the heavy and light chains for anti CEA antibody produced by CEA.66-E3 are described. As the constant regions of these chains

belong to the gamma and kappa families, respectively, "light chain" and "kappa chain", and "heavy chain" and "gamma chain", respectively, are used interchangeably below.

E.1.1 Isolation of Messenger RNA for Anti CEA Light and Heavy (Kappa and Gamma) Chains

Total RNA from CEA.66-E3 cells was extracted essentially as reported by Lynch et al, Virology, 98: 251 (1979). Cells were pelleted by centrifugation and approximately 1 g portions of pellet resuspended in 10 ml of 10 mM NaCl, 10 mM Tris HCl (pH 7.4), 1.5 mM $MgCl_2$. The resuspended cells were lysed by addition of non-ionic detergent NP-40 to a final concentration of 1 percent, and nuclei removed by centrifugation. After addition of SDS (pH 7.4) to 1 percent final concentration, the supernatant was extracted twice with 3 ml portions of phenol (redistilled)/chloroform: isoamyl alcohol 25:1 at 4°C. The aqueous phase was made 0.2 M in NaCl and total RNA was precipitated by addition of two volumes of 100 percent ethanol and overnight storage at -20°C. After centrifugation, polyA mRNA was purified from total RNA by oligo-dT cellulose chromatography as described by Aviv and Leder, Proc. Nat'l. Acad. Sci. (USA), 69: 1408 (1972). 142 µg of polyA mRNA was obtained from 1 g cells.

E.1.2 Preparation of E. coli Colony Library Containing Plasmids with Heavy and Light DNA Sequence Inserts

5 µg of the unfractionated polyA mRNA prepared in paragraph E.1.1 was used as template for oligo-dT primed preparation of double-stranded (ds) cDNA by standard procedures as described by Goeddel et al., Nature 281: 544 (1979) and Wickens et al., J. Biol. Chem. 253: 2483 (1978) incorporated herein by reference. The cDNA was size fractionated by 6 percent polyacrylamide gel electrophoresis and 124 ng of ds cDNA greater than 600 base pairs in length was recovered by electroelution. A 20 ng portion of ds cDNA was extended with deoxy C residues using terminal deoxynucleotidyl

transferase as described in Chang et al., Nature 275: 617 (1978) incorporated herein by reference, and annealed with 200 ng of the plasmid pBR322 (Bolivar et al., Gene 2: 95 (1977)) which had been cleaved with Pst I and tailed with deoxy G. Each annealed mixture was then transformed into E. coli K12 strain 294 (ATCC No. 31446). Approximately 8500 ampicillin sensitive, tetracycline resistant transformants were obtained.

E.1.3 Preparation of Synthetic Probes

The 14mer, 5' GGTGGGAAGATGGA 3' complementary to the coding sequence of constant region for mouse MOPC21 kappa chain which begins 25 basepairs 3' of the variable region DNA sequence was used as kappa chain probe. A 15 mer, 5' GACCAGGCATCCCAG 3', complementary to a coding sequence located 72 basepairs 3' of the variable region DNA sequence for mouse MOPC21 gamma chain was used to probe gamma chain gene.

Both probes were synthesized by the phosphotriester method described in German Offenlegungschrift 2644432, incorporated herein by reference, and made radioactive by kinasing as follows: 250 ng of deoxyoligonucleotide were combined in 25 µl of 60 mM Tris HCl (pH 8), 10 mM $MgCl_2$, 15 mM beta-mercaptoethanol, and 100 µCi (γ-^{32}P) ATP (Amersham, 5000 Ci/mMole). 5 units of T4 polynucleotide kinase were added and the reaction was allowed to proceed at 37°C for 30 minutes and terminated by addition of EDTA to 20 mM.

E.1.4 Screening of Colony Library for Kappa or Gamma Chain Sequences

~2000 colonies prepared as described in paragraph E.1.2 were individually inoculated into wells of microtitre dishes containing LB (Miller, Experiments in Molecular Genetics, p. 431-3, Cold Spring Harbor Lab., Cold Spring Harbor, New York (1972)) + 5 µg/ml tetracycline and stored at -20°C after addition of DMSO to 7

percent. Individual colonies from this library were transferred to duplicate sets of Schleicher and Schuell BA85/20 nitrocellulose filters and grown on agar plates containing LB + 5 µg/ml tetracycline. After ~10 hours growth at 37°C the colony filters were transferred to agar plates containing LB + 5 µg/ml tetracycline and 12.5 µg/ml chloramphenicol and reincubated overnight at 37°C. The DNA from each colony was then denatured and fixed to the filter by a modification of the Grunstein-Hogness procedure as described in Grunstein et al., Proc. Natl. Acad. Sci. (USA) 72: 3961 (1975), incorporated herein by reference. Each filter was floated for 3 minutes on 0.5 N NaOH, 1.5 M NaCl to lyse the colonies and denature the DNA then neutralized by floating for 15 minutes on 3 M NaCl, 0.5 M Tris HCl (pH 7.5). The filters were then floated for an additional 15 minutes on 2XSSC, and subsequently baked for 2 hours in an 80°C vacuum oven. The filters were prehybridized for ~2 hours at room temperature in 0.9 M NaCl, 1X Denhardts, 100 mM Tris HCl (pH 7.5), 5 mM Na-EDTA, 1 mM ATP, 1 M sodium phosphate (dibasic), 1 mM sodium pyrophosphate, 0.5 percent NP-40, and 200 µg/ml E. coli t-RNA, and hybridized in the same solution overnight, essentially as described by Wallace et al. Nucleic Acids Research 9: 879 (1981) using ~40x10^6 cpm of either the kinased kappa or gamma probe described above.

After extensive washing at 37°C in 6X SSC, 0.1 percent SDS, the filters were exposed to Kodak XR-5 X-ray film with DuPont Lightning-Plus intensifying screens for 16-24 hours at -80°C. Approximately 20 colonies which hybridized with kappa chain probe and 20 which hybridized with gamma chain probe were characterized.

E.1.5 Characterization of Colonies which Hybridize to Kappa DNA Sequence Probe

Plasmid DNAs isolated from several different transformants which hybridized to kappa chain probe were cleaved with Pst I and fractionated by polyacrylamide gel electrophoresis (PAGE). This

analysis demonstrated that a number of plasmid DNAs contained cDNA inserts large enough to encode full length kappa chain. The complete nucleotide sequence of the cDNA insert of one of these plasmids was determined by the dideoxynucleotide chain termination method as described by Smith, Methods Enzymol. 65, 560 (1980) incorporated herein by reference after subcloning restriction endonuclease cleavage fragments into M13 vectors (Messing et al., Nucleic Acids Research 9: 309 (1981). Figure 2 shows the nucleotide sequence of the cDNA insert of pK17G4 and Figure 3 shows the gene sequence with the corresponding amino acid sequence. Thus, the entire coding region of mouse anti-CEA kappa chain was isolated on this one large DNA fragment. The amino acid sequence of kappa chain, deduced from the nucleotide sequence of the pK17G4 cDNA insert, corresponds perfectly with the first 23 N-terminal amino acids of mature mouse anti-CEA kappa chain as determined by amino acid sequence analysis of purified mouse anti-CEA kappa chain. The coding region of pK17G4 contains 27 basepairs or 9 amino acids of the presequence and 642 basepairs or 214 amino acids of the mature protein. The mature unglycosylated protein (MW 24,553) has a variable region of 119 amino acids, including the J1 joining region of 12 amino acids, and a constant region of 107 amino acids. After the stop codon behind amino acid 215 begins 212 basepairs of 3' untranslated sequence up to the polyA addition. The kappa chain probe used to identify pK17G4 hybridizes to nucleotides 374-388 (figure 2).

E.1.6 Characterization of Colonies which Hybridize to Gamma 1 DNA Probe

Plasmid DNA isolated from several transformants positive for hybridization with the heavy chain gamma 1 probe was subjected to Pst I restriction endonuclease analysis as described in E.1.5. Plasmid DNAs demonstrating the largest cDNA insert fragments were selected for further study. Nucleotide sequence coding for mouse heavy (gamma-1) chain, shows an NcoI restriction endonuclease

cleavage site near the junction between variable and constant region. Selected plasmid DNAs were digested with both PstI and NcoI and sized on polyacrylamide. This analysis allowed identification of a number of plasmid DNAs that contain NcoI restriction endonuclease sites, although none that demonstrate cDNA insert fragments large enough to encode the entire coding region of mouse anti-CEA heavy chain.

In one plasmid isolated, p γ298 the cDNA insert of about 1300 bp contains sequence information for the 5' untranslated region, the signal sequence and the N-terminal portion of heavy chain. Because pγ298 did not encode the C-terminal sequence for mouse anti-CEA gamma 1 chain, plasmid DNA was isolated from other colonies and screened with PstI and NcoI. The C-terminal region of the cDNA insert of pγ11 was sequenced and shown to contain the stop codon, 3' untranslated sequence and that portion of the coding sequence missing from p γ298.

Figure 4 presents the entire nucleotide sequence of mouse anti-CEA heavy chain (as determined by the dideoxynucleotide chain termination method of Smith, Methods Enzymol., 65: 560 (1980)) and Figure 5 includes the translated sequence.

The amino acid sequence of gamma 1 (heavy chain) deduced from the nucleotide sequence of the pγ298 cDNA insert corresponds perfectly to the first 23 N-terminal amino acids of mature mouse anti-CEA gamma 1 chain as determined by amino acid sequence analysis of purified mouse anti-CEA gamma-1 chain. The coding region consists of 57 basepairs or 19 amino acids of presequences and 1346 basepairs or 447 amino acids of mature protein. The mature unglycosolated protein (MW 52,258) has a variable region of 135 amino acids, including a D region of 12 amino acids, and a J4 joining region of 13 amino acids. The constant region is 324 amino acids. After the stop codon behind amino acid 447 begins 96 bp of

3' untranslated sequences up to the polyA addition. The probe used to identify $p_{\gamma}298$ and $p_{\gamma}11$ hybridized to nucleotides 528-542 (Figure 4).

E.1.7 Construction of a Plasmid For Direct Expression of Mouse Mature Anti-CEA Kappa Chain Gene, pKCEAtrp207-1*

Figure 6 illustrates the construction of pKCEAtrp207-1*

First, an intermediate plasmid pHGH207-1*, having a single trp promoter, was prepared as follows:

The plasmid pHGH 207 (described in U.S. Pat. Appl. Serial No. 307,473, filed Oct. 1, 1981 (EPO Publn. No. 0036776)) has a double lac promoter followed by the trp promoter, flanked by EcoR I sites and was used to prepare pHGH207-1. pHGH207 was digested with BamH 1, followed by partial digestion with EcoR I. The largest fragment, which contains the entire trp promoter, was isolated and ligated to the largest EcoR I- BamH I fragment from pBR322, and the ligation mixture used to transform E. coli 294. Tet^R Amp^R colonies were isolated, and most of them contained pHGH207-1. pHGH207-1* which lacks the EcoR1 site between the amp^R gene and the trp promoter, was obtained by partial digestion of pHGH207-1 with EcoR I, filling in the ends with Klenow and dNTPs, and religation.

5 µg of pHGH207-1* was digested with EcoRI, and the ends extended to blunt ends using 12 units of DNA Polymerase I in a 50 µl reaction containing 60 mM NaCl, 7 mM $MgCl_2$, 7 mM Tris HCl (pH 7.4) and 1 mM in each dNTP at 37°C for 1 hour, followed by extraction with phenol/$CHCl_3$ and precipitation with ethanol. The precipitated DNA was digested with BamH I, and the large vector fragment (fragment 1) purified using 5 percent polyacrylamide gel electrophoresis, electroelution, phenol/$CHCl_3$ extraction and ethanol precipitation.

The DNA was resuspended in 50 µl of 10 mM Tris pH 8, 1 mM EDTA and treated with 500 units Bacterial Alkaline Phosphatase (BAP) for 30' at 65° followed by phenol/$CHCl_3$ extraction and ethanol precipitation.

A DNA fragment containing part of the light chain sequence was prepared as follows: 7 µg of pK17G4 DNA was digested with Pst I and the kappa chain containing cDNA insert was isolated by 6 percent gel electrophoresis, and electroelution. After phenol/$CHCl_3$ extraction, ethanol precipitation and resuspension in water, this fragment was digested with Ava II. The 333 bp Pst I-Ava II DNA fragment was isolated and purified from a 6 percent polyacrylamide gel.

A 15 nucleotide DNA primer was synthesized by the phosphotriester method G. O. 2,644,432 (supra) and has the following sequence:

Met Asp Ile Val Met
5' ATG GAC ATT GTT ATG 3'

The 5' methionine serves as the initiation codon. 500 ng of this primer was phosphorylated at the 5' end with 10 units T4 DNA kinase in 20 µl reaction containing 0.5 mM ATP. ~200 ng of the Pst I-Ava II DNA fragment was mixed with the 20 µl of the phosphorylated primer, heated to 95°C for 3 minutes and quick frozen in a dry-ice ethanol bath. The denatured DNA solution was made 60mM NaCl, 7mM $MgCl_2$, 7 mM Tris HCl (pH 7.4), 12 mM in each dNTP and 12 units DNA Polymerase I-Large Fragment was added. After 2 hours incubation at 37°C this primer repair reaction was phenol/$CHCl_3$ extracted, ethanol precipitated, and digested to completion with Sau 3A. The reaction mixture was then electrophoresed on a 6 percent polyacrylamide gel and ~50 ng of the 182 basepair amino-terminal blunt-end to Sau 3A fragment (fragment 2) was obtained after electroelution.

100 ng of fragment 1 (supra) and 50 ng of fragment 2 were combined in 20 µl of 20 mM Tris HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 2.5 mM ATP and 1 unit of T4 DNA ligase. After overnight ligation at 14°C the reaction was transformed into E. coli K12 strain 294. Restriction endonuclease digestion of plasmid DNA from a number of ampicillin resistant transformants indicated the proper construction and DNA sequence analysis proved the desired nucleotide sequence through the initiation codon of this new plasmid, pKCEAIntl (Figure 6).

The remainder of the coding sequence of the kappa light chain gene was prepared as follows:

The Pst I cDNA insert fragment from 7 µg of K17G4 DNA was partially digested with Ava II and the Ava II cohesive ends were extended to blunt ends in a DNA Polymerase I large fragment reaction. Following 6 percent polyacrylamide gel electrophoresis the 686 basepair Pst I to blunt ended Ava II DNA fragment was isolated, purified and subjected to Hpa II restriction endonuclease digestion. The 497 basepair Hpa II to blunt ended Ava II DNA fragment (fragment 3) was isolated and purified after gel electrophoresis.

10 µg of pKCEAIntl DNA was digested with Ava I, extended with DNA polymerase I large fragment, and digested with Xba I. Both the large blunt ended Ava I to Xba I vector fragment and the small blunt ended Ava I to Xba I fragment were isolated and purified from a 6 percent polyacrylamide gel after electrophoresis. The large vector fragment (fragment 4) was treated with Bacterial Alkaline Phosphatase (BAP), and the small fragment was digested with Hpa II, electrophoresed on a 6 percent polyacrylamide and the 169 basepair Xba I-Hpa II DNA fragment (fragment 5) was purified. ~75 ng of fragment 4, ~50 ng of fragment 3 and ~50 ng of fragment 5 were combined in a T4 DNA ligase reaction and incubated overnight at 14°,

and the reaction mixture transformed into E. coli K12 strain 294. Plasmid DNA from six ampicillin resistant transformants were analyzed by restriction endonuclease digestion. One plasmid DNA demonstrated the proper construction and was designated pKCEAInt2.

Final construction was effected by ligating the K-CEA fragment, including the trp promoter from pKCEAInt2 into pBR322(XAP). (pBR322(XAP) is prepared as described in U.S. Application 452,227, filed December 22, 1982; from pBR322 by deletion of the AvaI-PvuII fragment followed by ligation.)

The K-CEA fragment was prepared by treating pKCEAInt2 with Ava I, blunt ending with DNA polymerase I (Klenow fragment) in the presence of DNTPs, digestion with Pst I and isolation of the desired fragment by gel electrophoresis and electroelution.

The large vector fragment from pBR322(XAP) was prepared by successive treatment with EcoR I, blunt ending with polymerase, and redigestion with Pst I, followed by isolation of the large vector fragment by electrophoresis and electroelution.

The K-CEA and large vector fragments as prepared in the preceding paragraphs were ligated with T4 DNA ligase, and the ligation mixture transformed into E. coli as above. Plasmid DNA from several ampicillin resistant transformants were selected for analysis, and one plasmid DNA demonstrated the proper construction, and was designated pKCEAtrp207-I*.

E.1.8 Construction of a Plasmid Vector for Direct Expression of Mouse Mature Anti-CEA Heavy (Gamma 1) Chain Gene, p_{γ}CEAtrp207-1*

Figure 7 illustrates the construction of p_{γ}CEAtrp207-1*. This plasmid was constructed in two parts beginning with construction of the C-terminal region of the gamma 1 gene.

5 µg of plasmid pHGH207-1* was digested with Ava I, extended to blunt ends with DNA polymerase I large fragment (Klenow fragment), extracted with phenol/$CHCl_3$, and ethanol precipitated. The DNA was digested with BamH I treated with BAP and the large fragment (fragment A) was purified by 6 percent polyacrylamide gel electrophoresis and electroelution.

~5 µg of pγ11 was digested with Pst I and the gamma chain cDNA insert fragment containing the C-terminal portion of the gene was purified, digested with Ava II followed by extension of the Ava II cohesive ends with Klenow, followed by Taq I digestion. The 375 basepair blunt ended Ava II to Taq I fragment (fragment B) was isolated and purified by gel electrophoresis and electroelution.

9 µg of pγ298 was digested with Taq I and BamH I for isolation of the 496 basepair fragment (fragment C).

Approximately equimolar amounts of fragments A, B, and C were ligated overnight at 14° in 20µl reaction mixture, then transformed into E. coli strain 294. The plasmid DNA from six ampicillin resistant transformants was committed to restriction endonuclease analysis and one plasmid DNA, named pγCEAInt, demonstrated the correct construction of the C-terminal portion of gamma 1 (Figure 5).

To obtain the N-terminal sequences, 30 µg of pγ298 was digested with Pst I and the 628 basepair DNA fragment encoding the N-terminal region of mouse anti-CEA gamma chain was isolated and purified. This fragment was further digested with Alu I and Rsa I for isolation of the 280 basepair fragment. A 15 nucleotide DNA primer

met glu val met leu
5' ATG GAA GTG ATG CTG 3'

was synthesized by the phosphotriester method (supra).

The 5' methionine serves as the initiation codon. 500 ng of this synthetic oligomer primer was phosphorylated at the 5' end in a

reaction with 10 units T4 DNA kinase containing 0.5 mM ATP in 20µl reaction mixture. ~500 ng of the 280 basepair Alu I-Rsa I DNA fragment was mixed with the phosphorylated primer. The mixture was heat denatured for 3 minutes at 95° and quenched in dry-ice ethanol. The denatured DNA solution was made 60mM NaCl, 7mM $MgCl_2$, 7 mM Tris HCl (pH 7.4), 12 mM in each dNTP and 12 units DNA Polymerase I-Large Fragment was added. After 2 hours incubation at 37°C, this primer repair reaction was phenol/$CHCl_3$ extracted, ethanol precipitated, and digested to completion with HpaII. ~50 ng of the expected 125 basepair blunt-end to Hpa II DNA fragment (fragment D) was purified from the gel.

A second aliquot of pγ298 DNA was digested with Pst I, the 628 basepair DNA fragment purified by polyacrylamide gel electrophoresis, and further digested with BamH I and Hpa II. The resulting 380 basepair fragment (fragment E) was purified by gel electrophoresis.

~5 µg of pγCEAIntI was digested with EcoR I, the cohesive ends were made flush with DNA polymerase I (Klenow), further digested with BamH I, treated with BAP and electrophoresed on a 6 percent polyacrylamide gel. The large vector fragment (fragment F) was isolated and purified.

In a three fragment ligation, 50 ng fragment D, 100 ng fragment E, and 100 ng fragment F were ligated overnight at 4° in a 20 µl reaction mixture and used to transform E. coli K12 strain 294. The plasmid DNAs from 12 ampicillin resistant transformants were analyzed for the correct construction and the nucleotide sequence surrounding the initiation codon was verified to be correct for the plasmid named pγCEAInt2.

The expression plasmid, pγCEAtrp207-I* used for expression of the heavy chain gene is prepared by a 3-way ligation using the large.

vector fragment from pBR322(XAP) (supra) and two fragments prepared from pγCEAInt2.

pBR322(XAP) was treated as above by digestion with EcoR1, blunt ending with DNA polymerase (Klenow) in the presence of dNTPs, followed by digestion with Pst I, and isolation of the large vector fragment by gel electrophoresis. A 1543 base pair fragment from pγCEAInt2 containing trp promoter linked with the N-terminal coding region of the heavy chain gene was isolated by treating pγCEAInt2 with Pst I followed by BamH I, and isolation of the desired fragment using PAGE. The 869 base pair fragment containing the C-terminal coding portion of the gene was prepared by partial digestion of pγCEAInt2 with Ava I, blunt ending with Klenow, and subsequent digestion with BamH I, followed by purification of the desired fragment by gel electrophoresis.

The aforementioned three fragments were then ligated under standard conditions using T4 DNA ligase, and a ligation mixture used to transform E. coli strain 294. Plasmid DNAs from several tetracycline resistant transformants were analyzed; one plasmid DNA demonstrated the proper construction and was designated pγCEAtrp207-1*.

E.1.9 Production of Immunoglobulin Chains by E. coli

E. coli strain W3110 (ATTC No. 27325) was transformed with pγCEAtrp207-1* or pKCEAtrp207-1* using standard techniques.

To obtain double transformants, E. coli strain W3110 cells were transformed with a modified pKCEAtrp207-1*, pKCEAtrp207-1*Δ, which had been modified by cleaving a Pst I-Pvu I fragment from the amp^R gene and religating. Cells transformed with pKCEAtrp207-1*Δ are thus sensitive to ampicillin but still resistant to tetracycline. Successful transformants were retransformed using pγCEAInt2 which

confers resistance to ampicillin but not tetracycline. Cells containing both pKCEAtrp207-1*Δ and pγCEAInt2 thus identified by growth in a medium containing both ampicillin and tetracycline.

To confirm the production of heavy and/or light chains in the transformed cells, the cell samples were inoculated into M9 tryptophan free medium containing 10μg/ml tetracycline, and induced with indoleacrylic acid (IAA) when the OD 550 reads 0.5. The induced cells were grown at 37°C during various time periods and then spun down, and suspended in TE buffer containing 2 percent SDS and 0.1 M β-mercaptoethanol and boiled for 5 minutes. A 10 x volume of acetone was added and the cells kept at 22°C for 10 minutes, then centrifuged at 12,000 rpm. The precipitate was suspended in O'Farrell SDS sample buffer (O'Farrell, P.H., J. Biol. Chem., 250: 4007 (1975)); boiled 3 minutes, recentrifuged, and fractionated using SDS PAGE (10 percent), and stained with silver stain (Goldman, D. et al., Science 211: 1437 (1981)); or subjected to Western blot using rabbit anti-mouse IgG (Burnett, W. N., et al., Anal. Biochem. 112: 195 (1981)), for identification light chain and heavy chain.

Cells transformed with pγCEAtrp207-1* showed bands upon SDS PAGE corresponding to heavy chain molecular weight as developed by silver stain. Cells transformed with pKCEAtrp207-1* showed the proper molecular weight band for light chain as identified by Western blot; double transformed cells showed bands for both heavy and light chain molecular weight proteins when developed using rabbit anti-mouse IgG by Western blot. These results are shown in Figures 8A, 8B, and 8C.

Figure 8A shows results developed by silver stain from cells transformed with pγCEAtrp207-1*. Lane 1 is monoclonal anti-CEA heavy chain (standard) from CEA.66-E3. Lanes 2b-5b are timed samples 2 hrs, 4 hrs, 6 hrs, and 24 hrs after IAA addition. Lanes

2a-5a are corresponding untransformed controls; Lanes 2c-5c are corresponding uninduced transformants.

Figure 8B shows results developed by Western blot from cells transformed with pKCEAtrp207-1*. Lanes 1b-6b are extracts from induced cells immediately, 1hr, 3.5 hrs, 5 hrs, 8 hrs, and 24 hrs after IAA addition, and 1a-6a corresponding uninduced controls. Lane 7 is an extract from a pγCEAtrp207-1* control, lanes 8, 9, and 10 are varying amounts of anti CEA-kappa chain from CEA.66-E3 cells.

Figure 8C shows results developed by Western blot from four colonies of double transformed cells 24 hours after IAA addition (lanes 4-7). Lanes 1-3 are varying amounts of monoclonal gamma chain controls, lanes 8 and 9 are untransformed and pγCEAtrp207-1* transformed cell extracts, respectively.

In another quantitative assay, frozen, transformed E. coli cells grown according to E.1.10 (below) were lysed by heating in sodium dodecyl sulfate (SDS)/β-mercaptoethanol cell lysis buffer at 100°. Aliquots were loaded on an SDS polyacrylamide gel next to lanes loaded with various amounts of hybridoma anti-CEA. The gel was developed by the Western blot, Burnett (supra), using ^{125}I-labeled sheep anti-mouse IgG antibody from New England Nuclear. The results are shown in Figure 9. The figure shows that the E. coli products co-migrate with the authentic hybridoma chains, indicating no detectable proteolytic degradation in E. coli. Heavy chain from mammalian cells is expected to be slightly heavier than E. coli material due to glycosylation in the former. Using the hybridoma lanes as a standard, the following estimates of heavy and light chain production were made:

	(Per gram of cells)
E. coli (W3110/pγCEAtrp207-1*)	5 mg γ
E. coli (W3110/pKCEAtrp207-1*)	1.5 mg K
E. coli (W3110/pKCEAtrp207-1*Δ, pγCEAInt2)	0.5 mg K, 1.0 mg γ

E.1.10 Reconstitution of Antibody from Recombinant K and Gamma Chains

In order to obtain heavy and light chain preparations for reconstitution, transformed cells were grown in larger batches, harvested and frozen. Conditions of growth of the variously transformed cells were as follows:

E. coli (W3110/pγCEAtrp207-1*) were inoculated into 500 ml LB medium containing 5µg/ml tetracycline and grown on a rotary shaker for 8 hours. The culture was then transferred to 10 liters of fermentation medium containing yeast nutrients, salts, glucose, and 2µg/ml tetracycline. Additional glucose was added during growth and at OD 550 = 20, indoleacrylic (IAA), a trp derepressor, was added to a concentration of 50 µg/ml. The cells were fed additional glucose to a final OD 550 = 40, achieved approximately 6 hours from the IAA addition.

E. coli (W3110) cells transformed with pKCEA trp 207-1* and double transformed (with pKCEAtrp207-1*Δ and pγCEAInt2) were grown in a manner analogous to that described above except that the OD 550 six hours after IAA addition at harvest was 25-30.

The cells were then harvested by centrifugation, and frozen.

E.2 Assay Method for Reconstituted Antibody

Anti-CEA activity was determined by ELISA as a criterion for successful reconstitution. Wells of microtiter plates (Dynatech Immulon) were saturated with CEA by incubating 100 µl of 2-5 µg CEA/ml solution in 0.1M carbonate buffer, pH 9.3 for 12 hours at room temperature. The wells were then washed 4 times with phosphate buffered saline (PBS), and then saturated with BSA by incubating 200 µl of 0.5 percent BSA in PBS for 2 hours at 37°C, followed by

washing 4 times with PBS. Fifty microliters of each sample was applied to each well. A standard curve (shown in Figure 10), was run, which consisted of 50 µl samples of 10 µg, 5 µg, 1 µg, 500 ng, 100 ng, 50 ng, 10 ng, 5 ng and 1 ng anti-CEA/ml in 0.5 percent BSA in PBS, plus 50 µl of 0.5 percent BSA in PBS alone as a blank. All of the samples were incubated in the plate for 90 minutes at 37°C.

The plates were then washed 4 times with PBS, and sheep anti-mouse IgG-alkaline phosphate (TAGO, Inc.) was applied to each well by adding 100 µl of an enzyme concentration of 24 units/ml in 0.5 percent BSA in PBS. The solution was incubated at 37°C for 90 minutes. The plates were washed 4 times with PBS before adding the substrate, 100 µl of a 0.4 mg/ml solution of p-nitrophenylphosphate (Sigma) in ethanolamine buffered saline, pH 9.5. The substrate was incubated 90 minutes at 37°C for color development.

The A_{450} of each well was read by the Microelisa Auto Reader (Dynatech) set to a threshold of 1.5, calibration of 1.0 and the 0.5 percent BSA in PBS (Blank) well set to 0.000. The A_{450} data was tabulated in RS-1 on the VAX system, and the standard curve data fitted to a four-parameter logistic model. The unknown samples' concentrations were calculated based on the A_{450} data.

E.3 Reconstitution of Recombinant Antibody and Assay

Frozen cells prepared as described in paragraph E.1.10 were thawed in cold lysis buffer [10mM Tris HCl, pH 7.5, 1mM EDTA, 0.1M NaCl, 1mM phenylmethylsulfonyl fluoride (PMSF)] and lysed by sonication. The lysate was partially clarified by centrifugation for 20 mins at 3,000 rpm. The supernatant was protected from proteolytic enzymes by an additional 1mM PMSF, and used immediately or stored frozen at -80°C; frozen lysates were never thawed more than once.

The S-sulfonate of E. coli produced anti-CEA heavy chain (γ) was prepared as follows: Recombinant E. coli cells transformed with