AN ANALYSIS OF THE SEQUENCES OF THE VARIABLE REGIONS OF BENCE JONES PROTEINS AND MYELOMA LIGHT CHAINS AND THEIR IMPLICATIONS FOR ANTIBODY COMPLEMENTARITY*

By TAI TE WU, Ph.D., and ELVIN A. KABAT, Ph.D

(*From the Departments of Microbiology, Neurology, and Human Genetics and Development, College of Physicians and Surgeons, Columbia University, and the Neurological Institute, Presbyterian Hospital, New York 10032; the Biomathematics Division, Graduate School of Medical Sciences, Cornell University and the Sloan-Kettering Institute, New York 10021*)

(Received for publication 26 March 1970)

The extraordinary versatility of the antibody-forming mechanism in producing an almost limitless number of specific receptor sites complementary for almost any molecular conformation of matter within a size range (1–3) represented by a hexa- or heptasaccharide as an upper and a mono- or disaccharide as a lower limit, is almost certainly related to the unique structural features of immunoglobulins and differentiates them from all other known proteins. These antibody-combining sites are formed as a consequence of the interaction of two polypeptide chains, a light and a heavy chain (2, 4, 5). The antibodies usually formed to various antigens often represent heterogeneous populations of immunoglobulin molecules of different classes, subclasses, and genetic variants and also show specificities toward different antigenic determinants (1, 2, 6, 7). In some instances, however, relatively homogeneous populations of antibodies with respect to many of these properties have been obtained. Among these have been human antibodies to dextran and levan (8, 9) and rabbit antibodies to the group-specific carbohydrate of streptococcus (10–12), antibodies to the Type III–specific capsular polysaccharide of pneumococcus (13, 14), rabbit antihapten (15), and specimens of antibodies and of Fab′ fragments which crystallized (Nisonoff et al., in references 16, 17), but sequence data on these are not yet available.

The large body of sequence data related to immunoglobulin structure comes from the analysis of urinary Bence Jones proteins and from the monoclonal immunoglobulins found in large amounts in the sera of patients with multiple myeloma and Waldenström macroglobulinemia (16, 18). While a substantial body of evidence was available relating these proteins to immunoglobulins, the recent demonstration that many myeloma globulins have specific ligand-binding properties like those of many antibodies provides increasing confidence that myeloma globulins represent homogeneous populations of antibody molecules (16, 18–27). The ability to produce in BALB/c

* Aided by grants from the National Science Foundation (GB-8341) and the National Cancer Institute (CA-08748), and a general Research Support Grant of the U. S. Public Health Service.

mice myelomas and macroglobulinemias (28) which produce myeloma globulins and Bence Jones proteins like those in the human, provides a source of data from which important evolutionary trends can be inferred.

Thus the extensive sequence data on Bence Jones proteins, which are considered to be light chains of myeloma globulins and Waldenström macroglobulins (29), and on various light and heavy chains, provide information clearly pertinent to the problem of the elucidation of the structure of antibody-combining sites.

The unique finding that distinguishes the immunoglobulins from all other proteins is that the N-terminal half of the light chains and the N-terminal quarter of the heavy chains vary in sequence in samples obtained from individual monoclonal immunoglobulins and that indeed no two such variable regions of any chain and no two myeloma immunoglobulins or Bence Jones proteins have thus far been found to be identical in sequence (30). The constant region, however, is essentially no different from other proteins in that the variation in the amino acids found at any position is ascribable to species and class variations or to genetic variants such as Inv factors. By comparison of sequence data on the variable and constant regions of Bence Jones proteins with amino acid composition of purified human antibodies, it could be shown that most of the compositional variation could only originate in the variable region (see Kabat in reference 18).

From sequence data, a variety of hypotheses have been advanced (7, 31–35) to explain the structural basis of antibody complementarity. All of these are selective theories, i.e. they consider that the information for complementarity is essentially built into the primary sequence of each chain and that a given antigen only triggers the biosynthesis of those antibody molecules having complementary receptor sites. There are two types of selective theories: germ line theories (36) and somatic mutation theories (37–39). At present no hypothesis is generally accepted. Excellent reviews (see above) are available.

The present communication is an extension of earlier efforts from this laboratory (18, p. 87, and 40–43) to locate more precisely those portions of the variable region which are directly responsible for antibody complementarity, that is which make direct contact with the antigenic determinant, and to explain the unique capacity of these proteins to have so many complementary regions.

As in the earlier studies, all human κ, human λ, and mouse κ Bence Jones protein and light chain sequences are aligned for maximum homology (44) and all variable regions are considered as a unit and compared with the constant regions. These earlier studies had called attention to the following:

(*a*) The variable regions had few if any species-specific positions while the constant regions of the human and mouse proteins had 36 species-specific amino acid substitutions per 107 residues (40, 45). A species-specific position is defined as one at which the amino acid residues in the mouse proteins differ from those in the human proteins.

(*b*) When the invariant residues of these two regions were compared, the latest tabulation (45) showed the variable regions to have 10 invariant and almost invariant glycines and no invariant alanines, leucines, valines, histi-

dines, lysines, or serines while the constant regions had 3 each of invariant alanine, leucine, and valine, and 2 invariant histidines, 2 invariant lysines, and 5 invariant serines. It was suggested that the invariant glycines were important in contributing to the flexibility needed by the variable region in accommodating the numerous substitutions (41, 43) at the variable positions. It was also suggested that the invariant glycines near the end of the variable region at positions 99 and 101, plus the almost invariant glycine at position 100, provided a pivot upon which the complementarity-determining regions might move to make better contact with the antigenic determinant (43; 18, p. 87) just as the walls of the lysozyme site have been shown to adjust somewhat to accommodate its hexasaccharide substrate (46). The hydrophobic residues in the constant region were hypothesized to be involved in noncovalent bonding to the heavy chain.

(*c*) From an examination of sequences of the κI, κII, and κIII subgroups (Hood et al. in reference 16) (47, 48) of the human Bence Jones proteins in which many of the proteins in a subgroup had an identical sequence for the first 20–24 residues, it was postulated that there are two kinds of residues in the variable regions, those making direct contact with the antigenic determinant (complementarity determining) and those which are involved only in three-dimensional folding (42). The latter would be expected to have less stringent requirements, and more mutation noise would be permitted than with the complementarity-determining residues. This distinction led to the inspection of the sequences for short stretches showing very high variability and two of these were identified: the most variable beginning at residue 89 and ending at 97, the other running from residue 24 through 34. Each of these two unusually highly variable regions began after an invariant half-cystine and was followed by an invariant phenylalanine (residue 98) and an invariant tryptophane (residue 35) respectively. It is of interest that the two regions are brought close together by the S—S bond I_{23}–II_{88} (45). Milstein (47), Milstein and Pink (7), and Franěk (49) have also called attention to the highly variable positions in these regions and Franěk (49) has noted an additional highly variable region around residues 52–55. It was hypothesized (45) that these first two regions might represent the complementarity-determining regions and that complementarity might be acquired by the insertion of small linear sequences into the light and heavy chains by some episomal or other insertion mechanism. It is striking that the differences in chain length seen in the Bence Jones proteins are confined to these two regions of the chain. The remaining portions of each chain would be essentially under the control of structural genes. The inserted sequences would be drawn from a large but finite set and either inserted under the influence of antigen, if antibody-forming cells are multipotent, or individual sequences might be distributed to immunoglobulin-

forming cells during differentiation if the capacity of individual cells to synthesize antibody is restricted.

This working hypothesis offers several advantages:

(*a*) It is capable of providing the evolutionary stability and accounts for the universality of the antibody-forming mechanism throughout the vertebrates. Germ line theories (34–36) postulate one gene for each of the thousand or more variable regions (30). This would be expected to result in divergence during evolution since the loss by mutation of any one variable region would only minimally affect the capacity to form antibody and survival; thus individuals and populations lacking certain variable regions would arise.

(*b*) It offers a substantial simplification to the problem of producing a very large number of complementary sites. While it is known that in all proteins with specific receptors the site is formed by residues from widely separated portions of the chain, these sites are all formed by single chains. Thus, forming a three-dimensional site must involve residues from various regions. The antibody site being formed by a heavy and a light chain need not necessarily be so restricted.

Since much additional data on the light chains and a number of heavy chain sequences have been accumulated, the present communication represents a further attempt at analyzing the unique features of the variable regions of immunoglobulin chains. Among aspects considered are the role of glycine, invariant residues, and hydrophobicity patterns, and the highly variable portions, with a view to localizing the regions responsible for complementarity and evaluating various theories in terms of evolutionary origin and perpetuation of the antibody-forming mechanism.

Sequence Data Employed—Complete and partial sequence data have been published on 77 Bence Jones proteins and immunoglobulin light chains as well as on a number of heavy chains. Data were available on 24 human κI, 4 human κII, 17 human κIII, 10 human λI, 2 human λII, 6 human λIII, 5 human λIV, 2 human λV, 2 mouse κI, and 5 mouse κII proteins.[1]

The original light chain sequence data may be found in the following references.

HBJ 98: Baglioni, C. 1967. *Biochem. Biophys. Res. Commun.* **26**:82.

Eu: Cunningham, B. A., P. D. Gottlieb, W. H. Konigsberg, and G. M. Edelman. 1968. *Biochemistry.* **7**:1983.

Mil (human κII): Dreyer, W. J., W. R. Gray, and L. Hood. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:353.

Hac, Dob, Pal: Grant, A., and L. Hood. Unpublished work.

Roy, Cum: Hilschman, N., and L. C. Craig. 1965. *Proc. Nat Acad. Sci. U. S. A.* **53**:1403; Hilschmann, N. 1967. *Hoppe-Seyler's Z. Physiol. Chem.* **348**:1077; Hilschmann, N., H. U. Barnikol, M. Hess, B. Langer, H. Ponstingl, M. Steinmetz-Kayne, L. Suter, and S. Watanabe. 1968. *Fed. Eur. Biochem. Soc. Symp., 5th.* In press.

[1] The World Health Organization has recently changed the notation of subgroups so that human κII in this paper will become human κIII and human κIII will become human κII.

HS 78, HS 92, HS 94, HS 68, HS 70, HS 77, HS 86, HS 24: Hood, L., and D. Ein. 1968. *Nature* (*London*). **220:**764.

HBJ 7, HBJ 11, HBJ 2, HBJ 8: Hood, L., W. R. Gray, and W. J. Dreyer. 1966. *J. Mol. Biol.* **22:**179.

MBJ 41, MBJ 70, MBJ 6: Hood, L., W. R. Gray, and W. J. Dreyer. 1966. *Proc. Nat'l Acad. Sci. U. S. A.* **55:**826.

HBJ 10, HBJ 1, HBJ 4, HBJ 6, HBJ 5, HS 4, HBJ 12, HS 6, HBJ 15: Hood, L., W. R. Gray, B. G. Sanders, and W. J. Dreyer. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32:**133.

Ste: Edman, P., and A. G. Cooper. 1968 *Fed. Eur. Biochem Soc. Letters.* **2:**33; Hood, L., and D. W. Talmage. 1969. *In* Developmental Aspects of Antibody Formation and Structure. Prague. In press.

Lay, Mar, Ioc, Wag, How, Koh: Kaplan, A. P. and H. Metzger. 1969. *Biochemistry.* **10:** 3944.

New, III, Mil (human λIV): Langer, B., M. Steinmetz-Kayne, and N. Hilschmann. 1968. *Hoppe-Seyler's Z. Physiol. Chem.* **349:**945.

BJ, Ker: Milstein, C. 1966. *Biochem. J.* **101:**352.

Rad, Fr4: Milstein, C. 1967. *Nature* (*London*) **216:**330.

X: Milstein, C. 1968. *Biochem. J.* **110:**631.

Bel, Man, B6: Milstein, C. 1968. *Fed. Eur. Biochem. Soc. Symp. on γ-globulin*, Prague.

Day, MBJ46, Roy: Atlas of Protein Sequence and Structure, M. O. Dayhoff, Editor. 1969.

Mz: Milstein, C., B. Frangione, and J. R. L. Pink. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32:**31.

Ale, Car, Dee: Milstein, C., C. P. Milstein, and A. Feinstein. 1969. *Nature* (*London*) **221:**151.

Cra, Pap, Lux, Mon, Con, Tra, Nig, Win, Gra, Cas, Smi: Niall, H., and P. Edman. 1967. *Nature* (*London*) **216:**262.

MOPC 149, AdjPC 9, MOPC 157: Perham, R., E. Appella, and M. Potter. 1966. *Science* (*Washington*) **154:**391.

Kern: Ponstingl, H., M. Hess, and N. Hilschmann. 1968. *Hoppe-Seyler's Z. Physiol. Chem.* **349:**867.

Tew: Putnam, F. W. 1969. *Science* (*Washington*). **163:**633.

Ag, Ha, Bo, Sh: Putnam, F. W., K. Titani, M. Wikler, and T. Shinoda. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32:**9; Titani, K., T. Shinoda, and F. W. Putnam. 1969. *J. Biol. Chem.* **244:**3550.

TI: Suter, L., H. U. Barnikol, S. Watanabe, and N. Hilschmann. 1969. *Hoppe-Seyler's Z. Physiol. Chem.* **350:**275.

The accumulation of such large numbers of sequences makes it possible to use statistical criteria in defining the types of residues. Thus in earlier studies, an invariant residue was rigidly defined, e.g., a position at which all samples showed the same amino acid residue sometimes allowing a single exception. The definition of an invariant residue used in this paper is taken as a position at which 88–90% or more of the samples contain the same amino acid. This may allow potential functions to be recognized despite possible errors or uncertainties in sequence, or occasional substitutions compatible with function.

A summary of the sequence data is provided in Table I which lists the amino acids found at any position in any subgroup of human κ-, human λ-, and mouse κ-chains, the number of times each occurs, and the total number of sequences

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.