THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 258, No. 23, Issue of December 10, pp. 14433–14437, 1983
Printed in U.S.A.

Molecular Anatomy of the Antibody Binding Site*

(Received for publication, June 10, 1983)

Jiří Novotný, Robert Bruccoleri, John Newell, David Murphy, Edgar Haber, and Martin Karplus

From the Cellular and Molecular Research Laboratory and Cardiac Computer Center, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114 and the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

The binding region of immunoglobulins, which includes the portion of the molecule having the most variability in its amino acid sequence, is shown to have a surprisingly constant structure that can be characterized in terms of a simple, well-defined model. The binding region is composed of the antigen combining site plus its immediate vicinity and arises by noncovalent association of the light and heavy chain variable domains (VL and VH, respectively). The antigen combining site itself consists of six polypeptide chain segments ("hypervariable loops") which comprise some 80 amino acid residues and are attached to a framework of VL and VH β-sheet bilayers. Having analyzed refined x-ray crystallographic coordinates for three antigen-binding fragments (Fab KOL (Marquart, M., Deisenhofer, J., and Huber, R. (1980) *J. Mol. Biol.* 141, 369–391), MCPC 603 (Segal, D., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M., and Davies, D. R. (1974) *Proc. Natl. Acad. Sci. U. S. A.* 71, 4298–4302), and NEW (Saul, F. A., Amzel, L. M., and Poljak, R. J. (1978) *J. Biol. Chem.* 253, 585–597)) we use the results to introduce a general model for the VL-VH interface forming the binding region. The region consists of two closely packed β-sheets, and its geometry corresponds to a 9-stranded, cylindrical barrel of average radius 0.84 nm with an average angle of −53° between its two constituent β-sheets. The barrel forms the bottom and sides of the antigen combining site. The model demonstrates that the structural variability of the binding region is considerably less than was thought previously. Amino acid residues which are part of the domain-domain interface and appear not to be accessible to solvent or antigen contribute to antibody specificity.

Immunoglobulins have a modular structure in that both the light chains (M_r = 25,000) and the heavy chains (M_r = 50,000) (4, 5) are composed of several domains (6–8), each of which consists of approximately 110 amino acid residues. The NH_2-terminal variable domains (9, 10) of both chains (VL[1] and VH) are formed from two layers of β-sheet (1–3, 11–13) and have been shown to fold independently of the rest of the polypeptide chain (14). Six hypervariable loops of the VL and VH domains (L1, L2, L3, H1, H2, and H3) were predicted by Wu and Kabat (15, 16) to have a role in antigen binding. The loops are relatively short segments of primary structure that occur in homologous positions in all light and heavy chains, but have amino acid sequences and lengths which differ from one antibody molecule to another. The x-ray structure shows that each of the hypervariable loops forms a connection between two antiparallel strands of β-sheet. By noncovalent association of the VL and VH domains, all six loops come into close contact and form a contiguous area on the surface of the VL-VH dimer from which the binding site is constructed (Fig. 1) (2, 11, 17–20).

To obtain a consistent set of refined atomic coordinates from the crystallographic data (1–3, 29), we discarded atoms which form the constant domains of the Fab fragment and energy-minimized the resulting Fv fragments (*i.e.* VL-VH domain dimers, the domains being defined as in Ref. 16) with CHARMM, an empirical energy function program designed for the study of macromolecules (30). The root-mean-square differences between the original and energy-minimized coordinates were 0.031 nm (0.31 Å) for MCPC 603 and 0.021 nm for NEW; such small shifts indicate that the original crystallographic structures were satisfactory and that acceptable values for the potential energy can be achieved by small adjustments of the coordinates.

As a first step in deriving the geometric characteristics of the binding region, the boundaries between the parts of the VL-VH interface forming β-sheet strands and hypervariable loops were determined. The β-sheets were defined by their interstrand backbone (C═O···H—N) hydrogen-bonding pattern. A hydrogen bond list was generated with CHARMM (30) for the four polypeptide chain segments which contribute to the VL-VH interface (Fig. 2), and amino acids with hydrogen bond energy of −4.2 kJ/mol or less were taken to be parts of the β-sheets. By this criterion, portions of the hypervariable loops (as defined in Ref. 16) form part of β-sheets (Fig. 2). Examination of the four β-sheet segments with adjacent hypervariable loops by computer graphics showed that they consist of nine antiparallel β-strands; the light chain β-sheet is 4-stranded, and the heavy chain β-sheet is 5-stranded. The atom-packing density (32) of the sheet-sheet interface is on the order of that found in interiors of other proteins (33, 34). The sheets are strongly curved and wrap counterclockwise around a nearly perfect cylindrical surface (Fig. 3). Although β-barrels are frequently found among protein structures (35), it is unusual to find one formed of sheets contributed by different polypeptide chains. Richardson (35) has noted that "back-to-back β-barrels that share one wall occur in the variable half of immunoglobulin Fab structures."

To provide a quantitative assessment of the geometry of the binding region, a cylindrical surface was least-squares-fitted to the three VL-VH interface barrels (see Fig. 4 for details). Radii of the least-squares-fitted cylinders are very

* The work carried out in the Department of Chemistry was supported in part by a grant from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] The abbreviations used are: VL, light chain variable domain; VH, heavy chain variable domain.

FIG. 1. **Antibody binding region.** A stereo α-carbon drawing of VL-VH domain dimer of mouse Fab fragment MCPC 603 (2). VL domain is at *right*; VH domain is at *left*. The interface β-sheets, represented by *heavy lines* (segments S1, S2, S3, and S4 of Fig. 2), are highly curved and their noncovalent contact forms a nearly perfect cylindrical barrel the upper surface of which constitutes the antigen combining site. Cα atoms of hypervariable residues are represented as *heavy dots*. The nine β-strands which form the VL-VH interface are numbered β1 through β9, in correspondence with Fig. 2.

```
                 25  27a b c        28  30        35          40          45          50          55          60
KOL S1         S C S G T S S N I - - - - G S D S V N W Y Q Q L P G M A P K L L I Y R D A M R P S G V P D R
NEW S1         S C T G S S S N I - - - G A G N H V K W Y Q Q L P G T A P K L L I F - - - - - - - H N N A R
MCPC603 S1     S C R S S Q S L L N S G N Q K N F L A W Y Q Q K P G Q P P K L L I Y G A S T R E S G V P D R
                         |------ L1 ------|     β1                     β2      |--- L2 ---|

                         85       90       95  97a b 98 100
KOL S2                   S D Y Y C A S W N N S D N A Y V F G T G T K V
NEW S2                   A D Y Y C Q S Y D R S - - L R V F G G G T K L
MCPC603 S2               A V Y Y C Q N D H S Y P L - - T F G A G T K L
                             β3      |---- L3 ----|     β4

                 30       35       40        45       50  52a b c 53  55        60        65
KOL S3         F S S Y A M Y W V R Q A P G K G L E W V A I I W D D G S - - D Q H Y A D S V K G R
NEW S3         F D D Y Y S T W V R Q P P G R G L E W I G Y V F Y H G T S D T D T P L R S - - - R
MCPC603 S3     F S D F Y M E W V R Q P P G K R L E W I A A S R N K G N K Y T T E Y S A S V K G R
                 |-H1-|    β5                    β6     |------ H2 ------|   β7

                          90         95     100a b c d e f g h  101      105
KOL S4                   T G V Y F C A R D G G H G F C S S A S C F G P D Y W G Q G T P V
NEW S4                   T A V Y Y C A R D L I A G - - - - - - - - - C I D V W G Q G S L V
MCPC603 S4              T A I Y Y C A R N Y Y G S T W - - - - - - - Y F D V W G A G T T V
                               β8        |--------- H3 ---------|         β9
```

FIG. 2. **Polypeptide-chain segments which constitute the binding region and the VL-VH interface.** Homologous segments from different structures are aligned to maximize amino acid homology. *Boldface letters* indicate hypervariable loops (as defined in Ref. 16). *Boxes* denote β-sheet strands as defined in this paper (see text). Hypervariable loops as redefined in this paper are denoted by *horizontal bars*. Segments S1 and S2 are parts of light chain variable domain; segments S3 and S4 are parts of heavy chain variable domain. The *numbering system* is that of Ref. 16. Amino acids are given in the one-letter code (31).

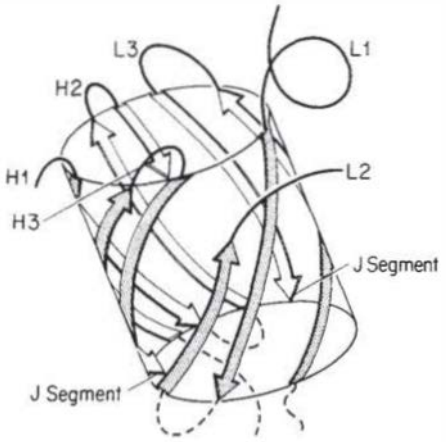

FIG. 3. **Topology of the binding site barrel.** Four hairpin segments, S1 to S4, primary structures of which are given in Fig. 2, organize themselves along a cylindrical surface. Segments S1 and S2 form the light chain β-sheet (on the *right*, *counterclockwise*, $\beta 2$, $\beta 1$, $\beta 3$, and $\beta 4$) as well as the three light chain hypervariable loops L1, L2 and L3. Segments S4 and S3 form the heavy chain β-sheet, consist of five β-strands (on the *left*, *clockwise*, $\beta 9$, $\beta 8$, $\beta 5$, $\beta 6$, and $\beta 7$), and the three heavy chain hypervariable loops H1, H2, and H3. Genetically, the strand $\beta 4$ corresponds to the light chain J gene (21, 22); $\beta 9$ corresponds to the heavy chain J gene (23–26); and the H3 loop is coded for by the D gene segment (27, 28). The shapes of the hypervariable loops are indicated schematically, and only their most general features are given. In all three structures studied, the domain-domain interface is closely packed, leaving no free space between the two β-sheets.

similar (0.82 nm for KOL, 0.86 nm for MCPC 603, and 0.84 nm for NEW), and the root-mean-square deviation of backbone atoms from the fitted surface (0.25 nm for KOL, 0.22 nm for MCPC 603, and 0.21 nm for NEW) is small considering that the fit includes parts of the hypervariable segments. For all three binding-site cylinders, we also computed angles between the backbone atoms of adjacent β-strands; the backbone atoms were represented by straight, least-squares-fitted lines (axes of inertia). The average twist angle between adjacent strands in the same β-sheet is $-20°$, a value corresponding to that found for sheets in other proteins (36–38). If the binding site barrels were perfect cylinders, strands which form sheet-sheet junctions would be expected to subtend the same angle (i.e. $-20°$); the angles are, in fact, significantly larger (the $\beta 4$–$\beta 7$ angle is $-57°$ and the $\beta 2$–$\beta 9$ angle is $33°$).

Interacting β-sheets of a single domain normally pack with the relative orientation of the sheet axes equal to either $-30°$ or $-90°$ (36–38); the sheet-sheet angle of the VL-VH interface by contrast is $-52°$ in KOL, $-54°$ in MCPC 603, and $-52°$ in NEW.

The KOL, MCPC 603, and NEW interface cylinders closely resemble each other (Fig. 4), and their interface β-sheet segments which are adjacent to hypervariable loops have virtually identical primary structures. The same, or very similar, amino acid sequences occur in corresponding segments of all immunoglobulins (16). This suggests that the geometry of the VL-VH interface is an important feature of the antibody binding region and that structurally similar 9-stranded cylinders form the essential element of all the binding regions; the antigen combining site is the concave part of the "upper" surface of the cylinder (Fig. 3).

To ascertain the role which hypervariable residues play in domain-domain association (17, 39, 40), we computed the total contact surface between VL and VH domains. The interdomain contact surfaces were obtained as differences between total solvent-accessible surface (41, 42) of isolated domains and of the domains involved in the dimer. The VL-VH contact surfaces, obtained as sums of the individual VL and VH interfaces, are 17.56 nm² in KOL, 19.39 nm² in MCPC, and 603 and 17.09 nm² in NEW. Thus, the average interface area of a single variable domain is 9 nm² (900 Å²), 36% of which consists of polar residues. The values are on the order of those reported for other protein-protein interfaces (e.g., $\alpha\beta$ oxyhemoglobin dimer) (34). The hypervariable loops contribute about half of the total contact surface (47% in KOL, 53% in MCPC 603, and 42% in NEW).

Single amino acid exchanges in hypervariable positions at the bottom of the binding site and inaccessible to solvent or antigen have a dramatic effect on antibody specificity. Particularly, side chains of residues Leu L96, Glu H35, and Asp H101, known to be indispensable for the specificity of phosphorylcholine-binding myelomas (43–45), are either totally or partially buried in the VL-VH interface. Other hypervariable residues (COOH-terminal parts of L1 and H1, NH_2-terminal parts of J gene-coded segments, i.e. L3 and H3) also form part of the VL-VH contact surface. Thus the domain-domain contact surface, even though it is not accessible in the free antibody structure, is directly involved in antibody specificity. This suggests that, due to the close packing at the interface, changes in volumes of buried side chains may induce changes in the bottom surface of the binding site. Such a result would be consistent with the structural variability found in hemoglobin (46) and the core of immunoglobulin domains (47). Alternatively, structural changes induced by antigen binding (48–58) could expose some of these buried residues.

Visual inspection of the Fv fragments (Figs. 1 and 3) shows that some of the 80 amino acid residues which constitute the hypervariable loops are distant from the binding site surface and may not be directly involved in antigen binding. Others, as indicated in Fig. 2, participate in the well-defined β-sheet secondary structures, so that their conformational freedom is rather limited. Those which are distant from the site include the whole L2 loop (it has been observed in the past that L2 is not directly involved in antigen binding (2, 59, 60)), two NH_2-terminal residues of L1, and five COOH-terminal residues of H2. The NH_2 and COOH termini of L1, L3, H3, and the middle of H2 are involved in backbone-backbone hydrogen bonds and have an extended conformation with backbone torsion angles (Φ and Ψ) in the β-sheet range. Of the 80 amino acids in the hypervariable loops (16), only 47 are intimately involved in the binding site, having at the same time conformations which vary in such a way that they cannot be inferred by analogy to known structures. This greatly reduces the problem of determining the binding site geometry of immunoglobulins with unknown structure, although, due to variations in the lengths of hypervariable loops, the exact number of structurally variable positions is expected to vary somewhat in different molecules.

The model proposed here represents the first part of a project aimed at developing a method for the analysis of the antigen binding region including prediction of its structure and evaluation of the interactions between the antibodies and the antigen.

Acknowledgments—We thank Drs. F. Saul, R. Poljak (Pasteur Institute, Paris), E. Padlan (Johns Hopkins University, Baltimore), D. Davies (National Institutes of Health, Bethesda), M. Marquart, R. Huber (Max-Planck Institut, Munich), and E. Abola (Brookhaven Data Bank) for crystallographic coordinates, Drs. F. M. Richards and

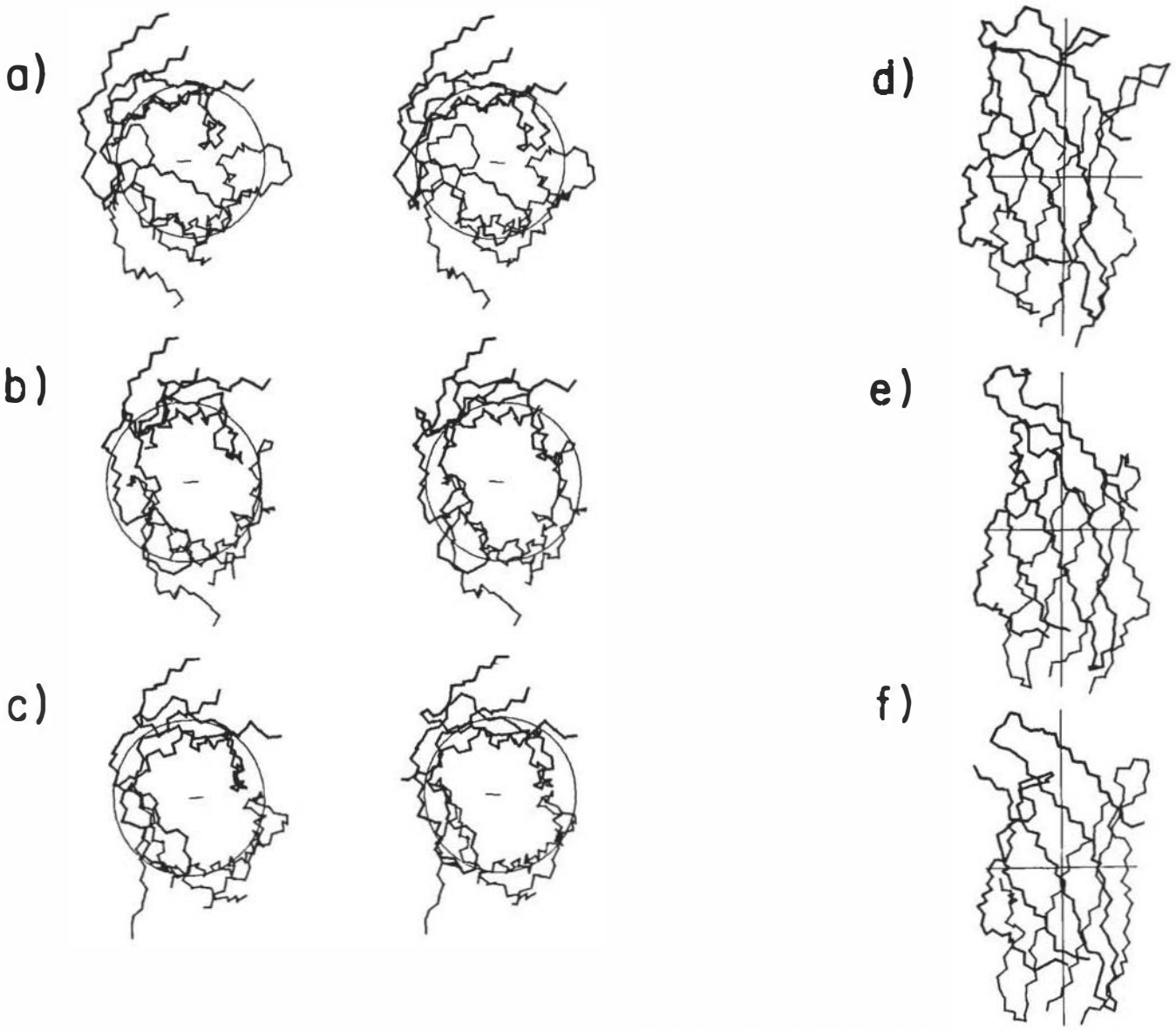

FIG. 4. **Binding site barrels with cylindrical surfaces least squares fitted into them.** Only the backbone atoms N, Cα, and C of the in β-strands including the hypervariable loops L3 and H3 of KOL, MCPC 603, and NEW) were subjected to the surface-fitting algorithm. The cylin vector passing through point **a** with unit direction vector, **λ**, and a radial distance, r, from the axis (7 parameters). The error being minimized is of the i-th particle from the axis ($r_i = \mathbf{p}_i - (\mathbf{p}_i \cdot \boldsymbol{\lambda})\boldsymbol{\lambda}$), where $\mathbf{p}_i = \mathbf{x}_i - \mathbf{a}$, and $\mathbf{x}_i$ is the position of an atom in the polypeptide chain backbone. Stereo to (*a–c*) and perpendicular to (*d–f*) the cylinder axis are given. *Thin lines* represent fitted results (cylinder axis, cylinder radius, and the cylindrica to the light chain β-sheet; and *heavy lines* correspond to the heavy chain β-sheet. *a*) KOL (cylinder radius, $R = 0.82$ nm, root-mean-square distanc cylindrical surface is $d = 0.25$ nm); *b*) MCPC 603 ($R = 0.86$ nm, $d = 0.22$ nm); *c*) NEW ($R = 0.84$ nm, $d = 0.21$ nm); *d*) KOL; *e*) MCPC 603; *f*) N

J. Matthews (Yale University) for computer programs, and R. Ladner and J. Ladner (Harvard University) for computer graphics.

REFERENCES

1. Marquart, M., Deisenhofer, J., and Huber, R. (1980) *J. Mol. Biol.* **141,** 369–391
2. Segal, D., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M., and Davies, D. R. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71,** 4298–4302
3. Saul, F. A., Amzel, L. M., and Poljak, R. J. (1978) *J. Biol. Chem.* **253,** 585–597
4. Porter, R. R. (1973) *Science (Wash. D. C.)* **180,** 713–716
5. Edelman, G. M. (1973) *Science (Wash. D. C.)* **180,** 830–840
6. Singer, S. J., and Doolittle, R. F. (1966) *Science (Wash. D. C.)* **153,** 13–25
7. Hill, R. L., Delaney, R., Fellows, R. E., and Lebovitz, H. E. (1966) *Proc. Natl. Acad. Sci. U. S. A.* **56,** 1762–1769
8. Edelman, G. M. (1970) *Biochemistry* **9,** 3197–3205
9. Hilschmann, N., and Craig, L. C. (1965) *Proc. Natl. Acad. Sci. U. S. A.* **53,** 1403–1409
10. Titani, K., Whitley, E., Avogadro, L., and Putnam, F. M. (1965) *Science (Wash. D. C.)* **149,** 713–716
11. Amzel, L. M., Poljak, R. J., Saul, F., Varga, J. M., and Richards, F. F. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71,** 1427–1430
12. Schiffer, M., Girling, R. L., Ely, K. R., and Edmundson, A. B. (1973) *Biochemistry* **12,** 1620–1631
13. Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M., and Huber, R. (1974) *Eur. J. Biochem.* **45,** 513–524
14. Hochman, J., Inbar, D., and Givol, D. (1973) *Biochemistry* **12,** 1130–1135
15. Wu, T. T., and Kabat, E. A. (1970) *J. Exp. Med.* **132,** 211–250
16. Kabat, E. A., Wu, T. T., and Bilofsky, H. (1983) *Sequences of Proteins of Immunological Interest,* United States Public Health Service, National Institutes of Health, Bethesda
17. Davies, D. R., Padlan, E. A., and Segal, D. (1975) *Annu. Rev. Biochem.* **44,** 639–667
18. Padlan, E. A., Davies, D. R., Rudikoff, S., and Potter, M. (1976) *Immunochemistry* **13,** 945–949
19. Padlan, E. A. (1977) *Quart. Rev. Biophys.* **10,** 35–65
20. Amzel, L. M., and Poljak, R. J. (1979) *Annu. Rev. Biochem.* **48,** 961–997
21. Sakano, H., Hüppi, K., Heinrich, G., and Tonegawa, S. (1979) *Nature (Lond.)* **280,** 288–294
22. Max, E. E., Seidman, J. G., and Leder, P. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 3450–3454
23. Bernard, O., and Gough, N. M. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 3630–3634
24. Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980) *Cell* **19,** 981–992
25. Newell, N., Richards, J. E., Tucker, P. W., and Blattner, F. R. (1980) *Science* **209,** 1128–1132
26. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S. (1980) *Nature (Lond.)* **286,** 676–683
27. Sakano, H., Kurosawa, Y., Weigert, M., and Tonegawa, S. (1981) *Nature (Lond.)* **290,** 562–565
28. Siebenlist, U., Ravetch, J. V., Korsmeyer, S., Waldmann, T., and Leder, P. (1981) *Nature (Lond.)* **294,** 631–635
29. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542
30. Brooks, B., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem.* **4,** 187–217
31. IUPAC-IUB Tentative Rules (1968) *Eur. J. Biochem.* **5,** 151–153
32. Richards, F. M. (1974) *J. Mol. Biol.* **82,** 1–14
33. Chothia, C. (1974) *Nature (Lond.)* **248,** 338–339
34. Chothia, C., and Janin, J. (1975) *Nature (Lond.)* **256,** 705–708
35. Richardson, J. S. (1981) *Adv. Protein Chem.* **34,** 167–339
36. Chothia, C., Levitt, M., and Richardson, D. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 4130–4134
37. Chothia, C., and Janin, J. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4146–4150
38. Cohen, F. E., Sternberg, M. J. E., and Taylor, W. R. (1981) *J. Mol. Biol.* **148,** 253–272
39. Stevens, F. J., Westholm, F. A., Solomon, A., and Schiffer, M. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 1144–1148
40. Horne, C., Klein, M., Polidoulis, I., and Dorrington, K. J. (1982) *J. Immunol.* **129,** 660–664
41. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400
42. Richmond, T. J., and Richards, F. M. (1978) *J. Mol. Biol.* **119,** 537–555
43. Rudikoff, S., Satow, Y., Padlan, E., Davies, D., and Potter, M. (1981) *Mol. Immunol.* **18,** 705–711
44. Cook, W. D., Rudikoff, S., Giusti, A. M., and Scharff, M. D. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 1240–1244
45. Rudikoff, S., Giusti, A. M., Cook, W. D., and Scharff, M. D. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 1979–1983
46. Lesk, M. A., and Chothia, C. (1980) *J. Mol. Biol.* **136,** 225–270
47. Lesk, M. A., and Chothia, C. (1982) *J. Mol. Biol.* **160,** 325–342
48. Roberts, S., and Grabar, P. (1957) *Ann. Inst. Pasteur (Paris)* **92,** 56–61
49. Warner, C., and Shumaker, V. N. (1970) *Biochemistry* **9,** 451–458
50. Holowka, D. A., Strosberg, A. D., Kimball, J. W., Haber, E., and Cathou, R. E. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69,** 3399–3403
51. Liberti, P. A., Stylos, W. A., Maurer, P. H., and Callahan, H. J. (1972) *Biochemistry* **11,** 3321–3326
52. Tumerman, L. A., Nezlin, R. S., and Zagyansky, Y. A. (1972) *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **19,** 290–292
53. Pilz, I., Kratky, O., Licht, A., and Sela, M. (1973) *Biochemistry* **12,** 4998–5005
54. Levison, S. A., Hicks, A. N., Portman, A. J., and Dandliker, W. B. (1975) *Biochemistry* **14,** 3778–3786
55. Schlessinger, J., Steinberg, I. Z., Givol, D., Hochman, J., and Pecht, I. (1975) *Proc. Natl. Acad. Sci. U. S. A.* **72,** 2775–2779
56. Käivärainen, A. I., and Nezlin, R. S. (1976) *Biochem. Biophys. Res. Commun.* **68,** 270–274
57. Lancet, D., and Pecht, I. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73,** 3549–3553
58. Cser, L., Franek, F., Gladkikh, I. A., Nezlin, R. S., Novotny, J., and Ostanevich, Y. M. (1977) *FEBS (Fed. Eur. Biochem. Soc.) Lett.* **80,** 329–331
59. Margolies, M. N., Cannon, L. E., Strosberg, A. D., and Haber, E. (1975) *Proc. Natl. Acad. Sci. U. S. A.* **72,** 2180–2184
60. Novotny, J., and Margolies, M. N. (1983) *Biochemistry* **22,** 1153–1158