

Three-Dimensional Structure of the Fab' Fragment of a Human Immunoglobulin at 2.8-Å Resolution

(subunit arrangement/disulfide bonds/hypervariable regions/x-ray crystallography)

R. J. POLJAK, L. M. AMZEL, H. P. AVEY*, B. L. CHEN, R. P. PHIZACKERLEY, AND F. SAUL

Department of Biophysics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Communicated by Max F. Perutz, July 23, 1973

ABSTRACT The structure of the Fab' fragment of a human myeloma immunoglobulin was determined by x-ray crystallographic analysis at 2.8-Å resolution. The Fourier map of the electron density was correlated with the amino acid sequence to obtain a three-dimensional model. Four globular subunits, which correspond to the homology regions of the light and heavy chains, are arranged in a tetrahedral configuration. These subunits closely resemble each other, sharing a basic pattern of polypeptide chain folding. In each subunit, long sequences of tightly packed, hydrogen bonded polypeptide chain run parallel to the major axis of the subunit. No helical conformation can be seen. Different patterns of interchain disulfide linkage and unusual intrachain disulfide bonds that have been observed in other immunoglobulins can be explained with this model. The regions of hypervariable sequences in the light and heavy chains occur at one end of the molecule, in close spatial proximity.

The Fab' fragment of immunoglobulins (molecular weight about 50,000), obtained by enzymatic cleavage of a whole immunoglobulin molecule, consists of a complete light (L) chain and the N-terminal half (Fd') of a heavy (H) chain. This fragment retains the antibody activity and specificity of the parent immunoglobulin molecule. The L and Fd' chains can each be subdivided into a C-terminal half of constant amino acid sequence (C_L and C_{H1} , respectively) and an N-terminal half of variable sequence (V_L and V_H , respectively). The C_L , C_{H1} , V_L , and V_H regions, with about 110 residues each, have homologous amino acid sequences and are referred to as homology regions. Both the V_H and V_L homology regions contain positions of hypervariable amino acid sequence. These hypervariable positions are believed to provide the conformation and specificity of the antigen-binding site and to confer antigenic individuality (idiotype) to immunoglobulins synthesized by a monoclonal cell population.

X-ray crystallographic studies of Fab' New at 6-Å resolution (1, 2) have shown that the molecule consists of two globular domains. One of these domains, V, consists of the V_L and V_H homology regions. The other domain, C, consists of the C_L and C_{H1} homology regions. This paper presents a high-resolution model of the Fab' structure.

Abbreviations: Nomenclature of immunoglobulins, their chains and fragments, is as recommended in *Bull. W.H.O.* **30**, 447 (1964).

* Present address: Tasmanian College of Advanced Education, Box 1415P, Hobart, Australia 7001.

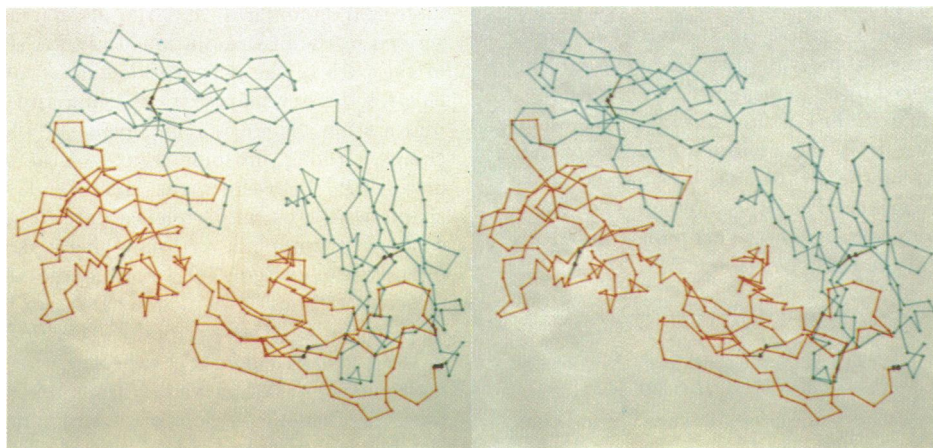
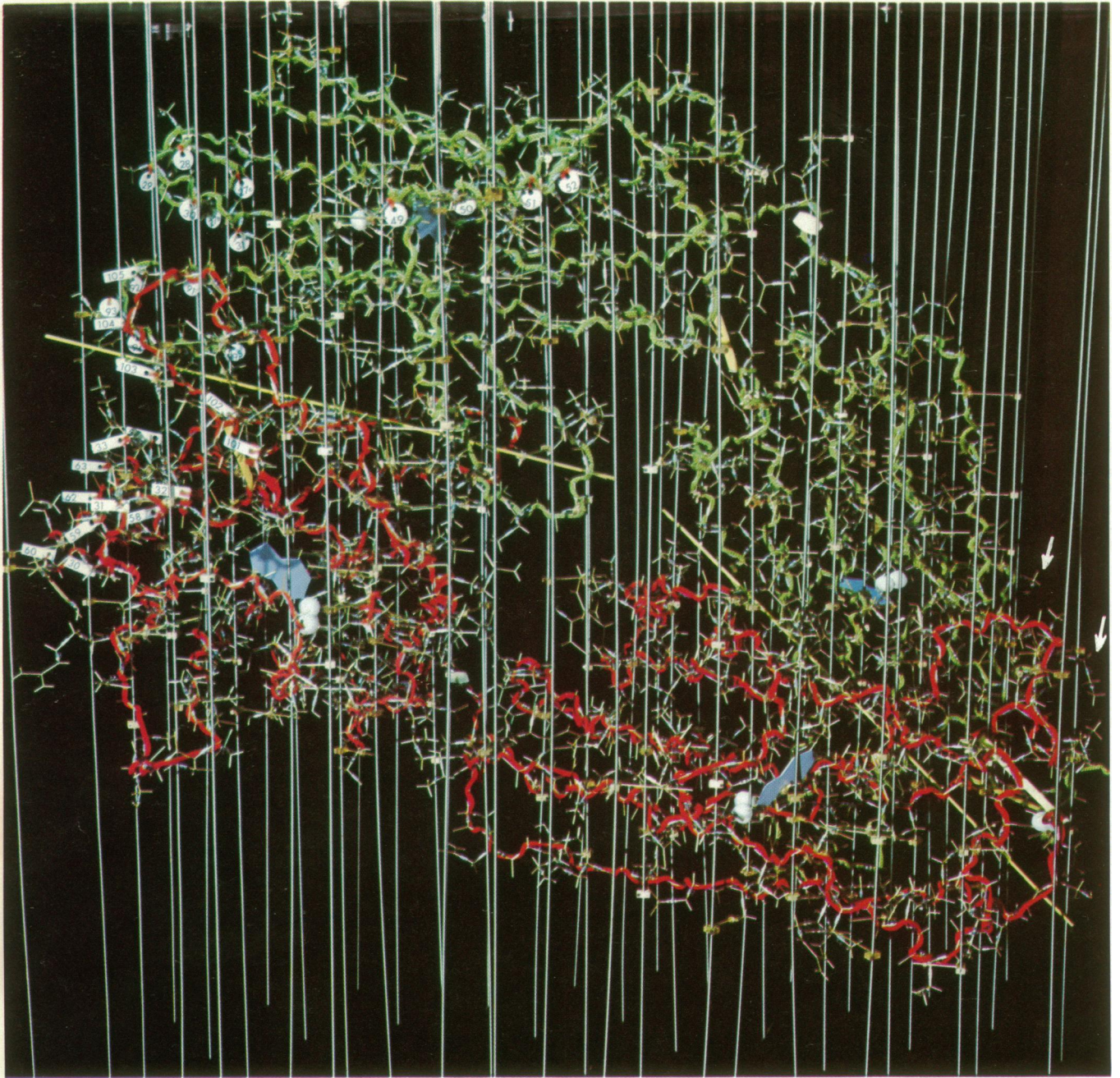
METHODS

IgG1 New and Fab' New were prepared and purified as described (3). The preparation and characterization of heavy-atom derivatives, measurement of diffracted x-ray intensities, and phase-determination procedures have been given (1, 2). The following six heavy-atom compounds were used to obtain isomorphous replacements and phase information to a resolution of 3.5 Å: phenylmercuric acetate, diacetoxymercury dipropylene dioxide, mersalyl 1,4-diacetoxymercuri-2,3-dimethoxybutane, uranyl nitrate, and *p*-acetoxymercurianiline. The first four were used to extend the phase determination to a resolution of 2.8 Å. The average figure of merit obtained was 0.78. The 2.8-Å Fourier map was calculated from about 10,000 reflections which had significant intensity out of the 13,200 that were measured. From the map a model was constructed with an optical comparator (4) and Kendrew skeletal models.

Partially and fully reduced and alkylated H and L chains were separated by gel filtration on Sephadex G-100 columns. Amino acid sequence of the tryptic and chymotryptic peptides of L (λ) chain New was determined by the subtractive Edman and dansyl-Edman (5) procedures. The peptides were aligned by sequence overlaps, homology with other λ -chain sequences (6, 7), and information derived from the Fourier map.

RESULTS AND DISCUSSION

The 2.8-Å Fourier map of Fab' New clearly indicates the molecular boundaries and the arrangement of subunits described in the 6-Å resolution study (2) (Figs. 1 and 2). Much of the path of the polypeptide chains can be followed in the electron-density map without using any amino acid sequence information. There are no breaks in the continuity of the chains, although in some regions of the map it is difficult to distinguish between main chain and closely-packed side chains without using sequence data. The characteristic side chains of tryptophan, tyrosine, phenylalanine, histidine, arginine, lysine, cysteine, and methionine are clearly defined in most cases. Many other side chains can be recognized in the map, including the pyrrolidone carboxylic acid N-termini of the L and H chains. As expected, the inner regions of the molecule consist of packed hydrophobic side chains. In addition, the phenolic oxygens of several tyrosines appear hydrogen bonded to aspartic acid side chains. No α -helical con-



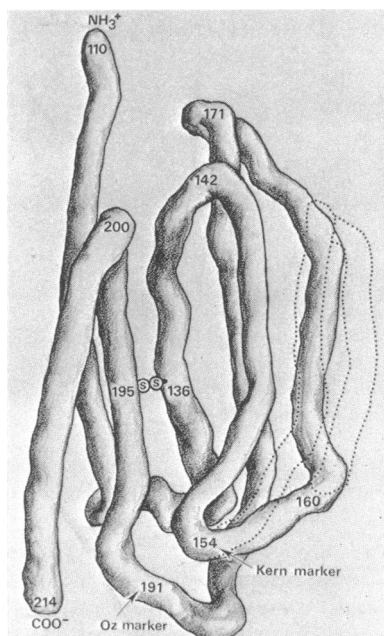


FIG. 4. Diagram of the basic "immunoglobulin fold". *Solid trace* shows the folding of the polypeptide chain in the constant subunits (C_L and C_{H1}). *Numbers* designate L (λ)-chain residues, beginning at " NH_3^+ " which corresponds to residue 110 for the L chain. *Broken lines* indicate the additional loop of polypeptide chain characteristic of the V_L and V_H subunits.

with the serologic nonallelic human λ -chain markers Kern⁻ and Oz⁺, respectively, appear at the outside of the molecule in close spatial proximity, about 8 Å from each other.

The angle between the major axes of the C_L and V_L homology subunits is greater than 90° (100–110°), whereas the corresponding angle between the C_{H1} and V_H subunits is less than 90° (80–85°) (Figs. 1 and 2). The relative arrangement of the four subunits in the frequently observed L-chain dimers may, therefore, be slightly different from the one reported here for Fab'. This angular relationship places the two Fd' chain subunits closer to each other than the two L-chain subunits. Consequently, the "switch" region connecting V_L and C_L appears to be more exposed to the solvent and to attack by proteolytic enzymes than the "switch" region connecting V_H and C_{H1} . An approximate local 2-fold axis of symmetry relates C_L to C_{H1} and another approximate local 2-fold axis relates V_L to V_H (Fig. 1).

The electron density of the polypeptide chain near the interchain disulfide bond appears much lower than in the rest of the molecule. No significant electron density can be found in the map to extend the Fd' chain beyond the interchain disulfide bond towards its C terminus, 14 residues away from this point. This observation explains the fact that no differences can be detected between the x-ray diffraction patterns of Fab New and Fab' New (1) despite the fact that the Fd' chain of Fab' is 10 aminoacid residues longer at the C terminus than the corresponding Fd chain of Fab (1, 4).

In all classes of immunoglobulins the cysteine residue that participates in the normal interchain disulfide link is always at the C terminus of the L chains. In human γ_1 chains the cysteine residue that completes the interchain S-S bond is at about position 214, near the C terminus of Fd'. However, in human γ_2 , γ_3 , γ_4 (11), and μ chains (12), and in rabbit γ

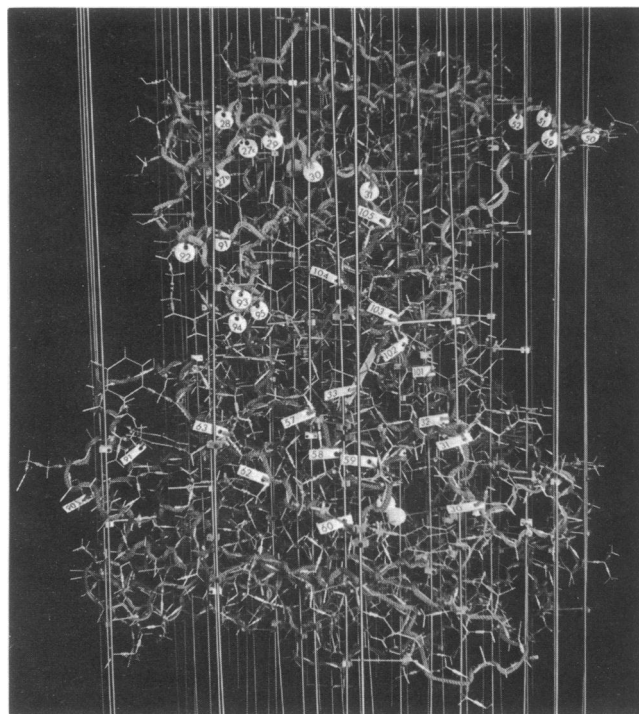


FIG. 5. View of the Fab' model showing the spatial relationship of the hypervariable positions. Numbers in *round and rectangular tags* correspond to L-chain and H-chain hypervariable positions, respectively. The longer axis of the V_H subunit (*bottom*) is horizontal and nearly parallel to the plane of the figure.

(13), mouse γ_2a and γ_2b (11), and guinea pig γ_2a (14) chains, the corresponding cysteinyl residue occurs at about position 131. In the three-dimensional model of Fab' New these three positions (one in the L, two in the H chain) are sufficiently close (about 6 Å) to allow for the alternative 131 (H) – 214 (L) disulfide bond without altering the folding of the polypeptide chain. The additional intra-H chain disulfide link observed in rabbit IgG [positions 131–221, (13)] is also compatible with the close proximity of the homologous residues in the structure of Fab' New. An unusual disulfide bond reported (15) in the V_H region of the human γ_1 chain Daw, linking residues 35–101, can also be explained by the close proximity (about 6 Å) of these residues in the three-dimensional structure. Another unusual disulfide bond observed in the (κ) L chain from the specific antipolysaccharide rabbit antibody 3315 (16) links the V_L to the C_L region of the molecule at positions 80 (V_L) and 171 (C_L) (Strosberg, A. D. & Haber, E., personal communication). A comparison of the sequences of rabbit κ chains and human λ chains suggests that Cys 80 in rabbit κ chain 3315 corresponds to Ala 79 in λ New (Fig. 3) and Cys 171 could correspond to either Asn 172 or Lys 173 in the λ New sequence. In the Fab' New model the side chains of Ala 79 and Asn 172 face each other and the separation of the polypeptide chain at these points is about 5.5 Å, compatible with the spatial requirements of a possible disulfide bond at this point. Thus, it appears that the Fab' model presented here provides an adequate structural framework for the various patterns of interchain and intrachain disulfide links that have been established by chemical analyses (11–16). These observations, together with the similarity of the folding pattern of the homology regions of Fab' (Fig. 4), lead

to the postulate that κ and λ L chains have the same overall three-dimensional structure, and that the V_H and C_H regions of different classes of H chains (α , γ , μ , etc.) also have the same overall pattern of polypeptide-chain folding.

The regions of hypervariable sequences (17) of both L and H chains occur at one end of the molecule and are fully exposed to the solvent (Fig. 5). The overall structural design can be described as one in which the exposed hypervariable positions occur at adjacent bends of tightly packed, linear polypeptide chains. Different immunoglobulin molecules will thus present a unique pattern of aminoacid side chains which are fixed to the rigidly constructed, constant structural frame of the V_L and V_H subunits. The hypervariable positions are related by a vague symmetry, with an approximate 2-fold axis running between the L and H chains, perpendicular to the surface of the molecule. There is no cavity in this region, only a shallow groove ($15 \times 6 \text{ \AA}$, with a depth of 6 \AA) which is shared by both chains. With the exception of this feature, the site could be described as being relatively flat but with protruding side chains that include several tyrosines. Heavy-chain residues 55–65 and 30–33 delineate the "lower" end of this site, and L-chain residues 27a–30 and those around position 50, the "upper" limit. The "sides" are made of L-chain residues around 90–95 to the "left" and H-chain residues 102–107 to the "right." This area is about $25 \times 20 \text{ \AA}$. Evidently, the geometry of the "active" site can be modified by substitution and by deletion (or addition) of aminoacid residues. For example, in human λ chains a subgroup-related "deletion" of three amino acids occurs at positions 27a, 27b, and 27c. In the three-dimensional model of Fab' New these three aminoacid residues appear in a nearly helical loop such that their deletion could be easily accommodated without a major change in the path of the polypeptide chain. However, contacts with antigens should be appreciably modified by this relatively small structural change. Similar considerations apply to positions 96 and 97, which are also considered to be subgroup-related deletions in human λ chains (18), and to deletions (or additions) observed in H-chain sequences. Positions 86–91, which are hypervariable in human subgroup III H-chain sequences (19), are not close to the hypervariable sequences mentioned above and are not in contact with L-chain residues, although they occur at an outside region of the molecule (Fig. 5). The cavity that was found at one end of the molecule in the 6-\AA resolution study (2) is clearly defined in the 2.8-\AA resolution map. It is located at the C-termini of both chains, divided into two parts by the interchain disulfide bond.

Affinity labeling studies of specific antibodies and myeloma proteins (20–25) have tagged residues of the V_L and V_H sequences that coincide with, or are adjacent to, the hypervariable positions. Although many of the positions that were labeled in these studies (20–25) seem (by homology with the sequence of Fab' New) openly accessible to solvent, tyrosines 85 and 86 and cysteine 87 in the L chain and the homologous tyrosines 94 and 95 and cysteine 96 in the H chain are at the core of their respective subunits, and are not accessible from the hypervariable region described above. However, L-chain Tyr 31 (replaced by His in λ New) is openly accessible and Tyr 90 (often replaced by Trp in human λ sequences) is slightly buried inside the "active" region, at the bottom of the 6-\AA deep groove but possibly accessible to affinity labeling reagents.

The location of the isotopic markers (Kern, Oz) on the L chain, the position of the interchain disulfide bond, and the general shape and arrangement of the subunits can be used to generate a plausible model for $F(ab)_2'$ or whole IgG in which some regions of Fab' are inaccessible from the outside of the molecule, and possibly for this reason are poor antigenic determinants. However, the region of the hypervariable sequences occurs on the outside of the molecule and projects several side chains into the solvent. These side chains, e.g., at positions 29, 51, and 92, 93 in the L chain, are at least as prominent as surface determinant groups as the Kern (glycine/serine) and Oz (lysine/arginine) markers in the C_L subunit. On the basis of the structural model presented here, aminoacid sequence changes in the hypervariable region, affecting V_L or V_H or both, could easily be recognized by a molecule of complementary structure (anti-idiotypic antibody) by a "face-to-face" interaction. In addition, the binding of a specific hapten that cannot be deeply buried in the shallow reactive area should in general block recognition of the immunodeterminant groups of the reactive area by an anti-idiotypic antibody (26, 27).

Finally, although the structurally related L and H chains are clearly complementary, it appears that the H chain exposes a larger area at the "active" site or immediately adjacent to it (Figs. 1 and 5). This observation, however, cannot take into account the specific antigen-antibody interactions in each particular immune reaction.

We thank Dr. A. Nisonoff for advice, Dr. R. Sasseti for his generous gift of serum from patient New, Mr. Ronald J. Kendig for his excellent technical help, and Drs. J. J. Cebra and R. L. Humphrey for their critical reading of this manuscript. This research was supported by Grants AI 08202 from the National Institutes of Health, E-638 from the American Cancer Society, and N.I.H. Research Career Development Award AI 70091 to R.J.P.

- Poljak, R. J., Amzel, L. M., Avey, H. P., Becka, L. N., Goldstein, D. J. & Humphrey, R. L. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 421–425.
- Poljak, R. J., Amzel, L. M., Avey, H. P., Becka, L. N. & Nisonoff, A. (1972) *Nature New Biol.* **235**, 137–140.
- Rossi, G., Choi, T. K. & Nisonoff, A. (1969) *Nature* **223**, 837–838.
- Richards, F. M. (1968) *J. Mol. Biol.* **37**, 225–230.
- Gray, W. R. (1967) in *Methods in Enzymology*, eds. Colowick, S. P. & Kaplan, N. O. (Academic Press, New York), Vol. XI, pp. 469–475.
- Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (Nat. Biom. Research Foundation), vol. 5, pp. D229–D274.
- Smith, G. P., Hood, L. & Fitch, W. M. (1971) *Annu. Rev. Biochem.* **40**, 969–1012.
- Ghose, A. C. & Jirgensons, B. (1971) *Biochim. Biophys. Acta* **251**, 14–20.
- Glenner, G. G., Ein, D., Eanes, E. D., Bladen, H. A., Terry, W. & Page, D. L. (1971) *Science* **174**, 712–714.
- Hill, R. L., Delaney, R., Fellows, R. E. & Lebovitz, H. E. (1966) *Proc. Nat. Acad. Sci. USA* **56**, 1762–1769.
- De Preval, C., Pink, J. R. L. & Milstein, C. (1970) *Nature* **228**, 930–932.
- Putnam, F. W., Shimizu, A., Paul, C., Shinoda, T. & Köhler, H. (1971) *Ann. N.Y. Acad. Sci.* **190**, 83–102.
- O'Donnell, I. J., Frangione, B. & Porter, R. R. (1970) *Biochem. J.* **116**, 261–268.
- Birshtein, B. K., Hussain, Q. Z. & Cebra, J. J. (1971) *Biochemistry* **10**, 18–25.
- Press, E. M. & Hogg, N. M. (1970) *Biochem. J.* **117**, 641–660.
- Poulsen, K., Fraser, K. J. & Haber, E. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2495–2499.

17. Wu, T. T. & Kabat, E. A. (1970) *J. Exp. Med.* **132**, 211-250.
18. Garver, F. A. & Hilschmann, N. (1972) *Eur. J. Biochem.* **26**, 10-32.
19. Kehoe, M. J. & Capra, D. J. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2019-2021.
20. Singer, S. J., Martin, N. & Thorpe, N. O. (1971) *Ann. N.Y. Acad. Sci.* **190**, 342-351.
21. Goetzl, E. J. & Metzger, H. (1970) *Biochemistry* **9**, 3862-3871.
22. Franěk, F. (1971) *Eur. J. Biochem.* **19**, 176-183.
23. Ray, A. & Cebra, J. J. (1972) *Biochemistry* **11**, 3647-3657.
24. Givol, D., Strausbauch, P. H., Hurwitz, E., Wilchek, M., Haimovich, J. & Eisen, H. N. (1971) *Biochemistry* **10**, 3461-3466.
25. Fleet, G. W. J., Knowles, J. R. & Porter, R. R. (1972) *Biochem. J.* **128**, 499-508.
26. Brient, B. W. & Nisonoff, A. (1970) *J. Exp. Med.* **132**, 951-962.
27. Carson, D. & Weigert, M. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 235-239.