Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 7154–7158, August 1992
Immunology

Structure of a human monoclonal antibody Fab fragment against gp41 of human immunodeficiency virus type 1

(antibody structure/AIDS)

XIAO MIN HE*, FLORIAN RÜKER†, ELENA CASALE*, AND DANIEL C. CARTER*‡

*The Space Science Laboratory, The Microgravity Science and Applications Division, ES76 Biophysics, Marshall Space Flight Center, Huntsville, AL 35812; and †The Institute of Applied Microbiology, University of Agriculture and Forestry, Nussdorfer Lände 11, A-1190 Vienna, Austria

Communicated by Frank W. Putnam, April 6, 1992 (received for review November 20, 1991)

ABSTRACT **The three-dimensional structure of a human monoclonal antibody (Fab), which binds specifically to a major epitope of the transmembrane protein gp41 of the human immunodeficiency virus type 1, has been determined by crystallographic methods to a resolution of 2.7 Å. It has been previously determined that this antibody recognizes the epitope SGKLICTTAVPWNAS, belongs to the subclass IgG1 (κ), and exhibits antibody-dependent cellular cytotoxicity. The quaternary structure of the Fab is in an extended conformation with an elbow bend angle between the constant and variable domains of 175°. Structurally, four of the hypervariable loops can be classified according to previously recognized canonical structures. The third hypervariable loops of the heavy (H3) and light chain (L3) are structurally distinct. Hypervariable loop H3, residues 102H–109H, is unusually extended from the surface. The complementarity-determining region forms a hydrophobic binding pocket that is created primarily from hypervariable loops L3, H3, and H2.**

To date there have been several antibody structures complexed with antigens (1–7) and numerous Fab structures determined by crystallographic methods. Details of the intermolecular contacts in these complexes vary considerably and may include a number of salt bridges, hydrogen bonds, and hydrophobic interactions (8). Despite the small number of completed structures compared with the virtually unlimited number of possible paratopes, evidence for discrete conformational patterns among five of the six antibody hypervariable regions (L1, L2, L3, H1, H2) is emerging (9, 10). Recognition of these canonical structures, and the development of new accurate *ab-initio* (11) and database modeling algorithms (12–14), has aided the development of predictive methods that allow antibody combining sites to be more accurately modeled from sequence data alone (15).

Currently, the growing body of structural data on immunoglobulins and their complexes with antigens, together with the implementation of site-directed mutagenesis and the creation of smaller Fv proteins by recombinant methods, is providing a more complete understanding of antibody structure and antibody–antigen interaction (16–20). These advances in knowledge also promise avenues for new therapeutic approaches to various immunologically recognizable illnesses, including the acquired immunodeficiency syndrome (AIDS) (21).

This structural work is part of a series that aims to elucidate the detailed nature of human monoclonal antibodies expressed against the AIDS virus together with their respective antigen complexes. Atomic coordinates produced from these studies should (*i*) provide important information for future recombinant experiments with genetically engineered antibodies and smaller Fv proteins, (*ii*) guide chemists in the selection of superior antigenic peptides for human immunodeficiency virus (HIV) detection, (*iii*) yield further insight into the details of antibody–antigen recognition, and (*iv*) aid in future crystal structure solutions of human IgG antibodies of interest. Here we report the atomic structure of an Fab (3D6)[§] from a human monoclonal antibody that binds specifically to the coat protein gp41 of the HIV type 1 (HIV-1). Fab 3D6 was produced from a xeno hybridoma line 3D6 (22) in a research effort involving the large-scale production of monoclonal antibodies to HIV-1 (23). It has been previously determined that 3D6 recognizes a major epitope of the viral transmembrane protein gp41 (a subcomponent of gp160) corresponding to the amino acid sequence SGKLICTTAVPWNAS (24). 3D6 readily forms complexes with peptides containing all or part of the epitope GCSGKLICTTAVPWNAS and a larger 80-amino acid fragment of gp41 also known as P121 (25). A portion of the epitope, GCSGKLIC, is currently thought to represent a solvent-exposed disulfide loop of gp41. The nucleotide sequence of 3D6 has been previously determined (26) and the amino acid sequence of the variable (V) domains and complementarity-determining region (CDR) is shown in Fig. 1. 3D6 belongs to the subclass IgG1 (κ) and exhibits antibody-dependent cellular cytotoxicity.

MATERIALS AND METHODS

Crystallization. The preparation, purification, and crystallization of Fab 3D6 have been reported (27). Briefly, the crystals grow as clusters of plates in the space group $P2_12_12_1$ (a = 66.5 Å, b = 74.3 Å, c = 105.3 Å) from solutions of PEG 3350 at neutral pH. Diffraction data were collected on a Siemens/Nicolet multiwire area detector operating on a Rigaku RU200 rotating anode x-ray generator (λ = 1.5418 Å, 40 kV, 70 mA) and reduced with the program package XENGEN (28). Native data were collected and merged from the best two out of three crystals. A total of 77,314 reflections was collected and merged to produce 13,529 unique observed data to 2.7 Å out of 15,019 that are theoretically possible. The merging unweighted R factor on F_{hkl} for these data is 7.33%. There is one Fab molecule in the asymmetric unit.

Structure Determination and Refinement. The structural solution of Fab 3D6 was obtained by the molecular replacement method (29) using the software package MERLOT (30). Several known Fab structures were tested as model structures. In searching for rotation function solutions using the fast rotation function calculation (31), heterodimers $V_L + V_H$ and $C_L + C_H1$ (L = light chain, H = heavy chain, C = constant) were first used as model structures, followed by single domains V_L, C_L, V_H, and C_H1. All solutions from these calcu-

Abbreviations: CDR, complementarity-determining region; C, constant; V, variable; H, heavy chain; L, light chain; SA, simulated annealing; CC, correlation coefficient.

‡To whom reprint requests should be addressed.

§The atomic coordinates have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1DFB).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

```
                    Light Chain
          5         10        15        20        25        30
1  D I Q M T Q S P S T L S A S V G D R V T I T C R A S Q S I S
                                                   L1
31 R W L A W Y Q Q K P G K V P K L L I Y K A S S L E S G V P S
                                        L2
61 R F S G S G S G T E F T L T I S S L Q P D D F A T Y Y C Q Q
91 Y N S Y S F G P G T K V D I K R T V A A P S V F I F P P S D
   L3

                    Heavy Chain
1  E V Q L V E S G G G L V Q P G R S L R L S C A A S G F T F N
31 D Y A M H W V R Q A P G K G L E W V S G I S W D S S S I G Y
   H1                                   H2
61 A D S V K G R F T I S R D N A K N S L Y L Q M N S L R A E D
91 M A L Y Y C V K G R D Y Y D S G G Y F T V A F D I W G Q G T
                        H3
```

FIG. 1. Amino acid sequence of 3D6 illustrating the sequences of the V regions. The six CDRs (L1, L2, H1, H2, H3) are designated after Chothia and Lesk (10) and the numbering scheme for the amino acid sequence (26) is sequential.

lations were examined for contrast and consistency. The orientations of individual domains were then fine-tuned using a localized search by further utilization of the rotation function (32). Patterson translation function (33) calculations for individual domains gave consistent solutions, although some had lower contrast. This translation function solution was further confirmed by a translational correlation coefficient (CC)¶ searching method with very good contrast. Among these trials the structure of Fab BV0401 (34) provided the best solutions for orientation and translation and was finally chosen. At this time the V_H domain was replaced by that of KOL [Protein Data Bank (PDB) code: 2FB4] (35) and C_H1 domain was replaced by that of NEW (PDB code: 3FAB) (36) by superposition of these domains on a graphics workstation using FRODO (37). The model thus formed contains V_L (BV0401), C_L (BV0401), V_H (KOL), and C_H1 (NEW). Rigid-body least squares refinement with individual domains as rigid groups gave R factor 0.39 and CC = 0.57 for 4391 reflections with Fobs. $> 3\sigma$ in the resolution range of 20 Å to 4 Å. Where the amino acids in the model differ from the amino acid sequence of 3D6, replacements were made with alanines or glycines. This model underwent a cycle of simulated annealing (SA) process for data within a resolution range of 10.0 Å to 3.2 Å using the program XPLOR (38). The R factor at this time for 8452 reflections was 0.20. Correct side-chain atoms were fitted into the molecular model utilizing a series of difference Fourier maps. Another cycle of SA with the data extended to a resolution range of 10.0 Å to 2.7 Å gave an R = 0.216 for 13,190 reflections. The crystal structure was carefully examined revealing no unusual or disallowed crystal packing interactions. At this stage two different approaches were adopted to eliminate possible model bias. In the first, 40 consecutive amino acids at a time were deleted from the structure followed by several cycles of constrained-restrained refinement of the remaining model against the reflection data (39). The phase sets thus obtained were used to calculate the difference Fourier synthesis (usually called omit maps or "unbiased" difference maps). These omit maps indicated that the entire molecular structure was well determined with one exception at the third hypervariable region of the heavy chain, H3. In this region, the electron density was spurious and unrecognizable from amino acid Tyr-102H to Phe-109H and no alternative interpretation was revealed by the electron density. This suggested that this loop is flexible or disordered in the crystal lattice and/or that the model bias in this region had not been completely eliminated. It was then decided to use a second approach in calculating the omit maps. An additional six omit maps were calculated, each with one of the six hypervariable regions deleted from the model. These individual structural models underwent one cycle of SA process. In omit maps calculated from these models, five of six hypervariable loops gave well-defined electron density in accordance with the

¶CC = $[\langle FoFc\rangle - \langle Fo\rangle\langle Fc\rangle]/\{[\langle Fo^2\rangle - \langle Fo\rangle^2]^{1/2}[\langle Fc^2\rangle - \langle Fc\rangle^2]^{1/2}\}$.

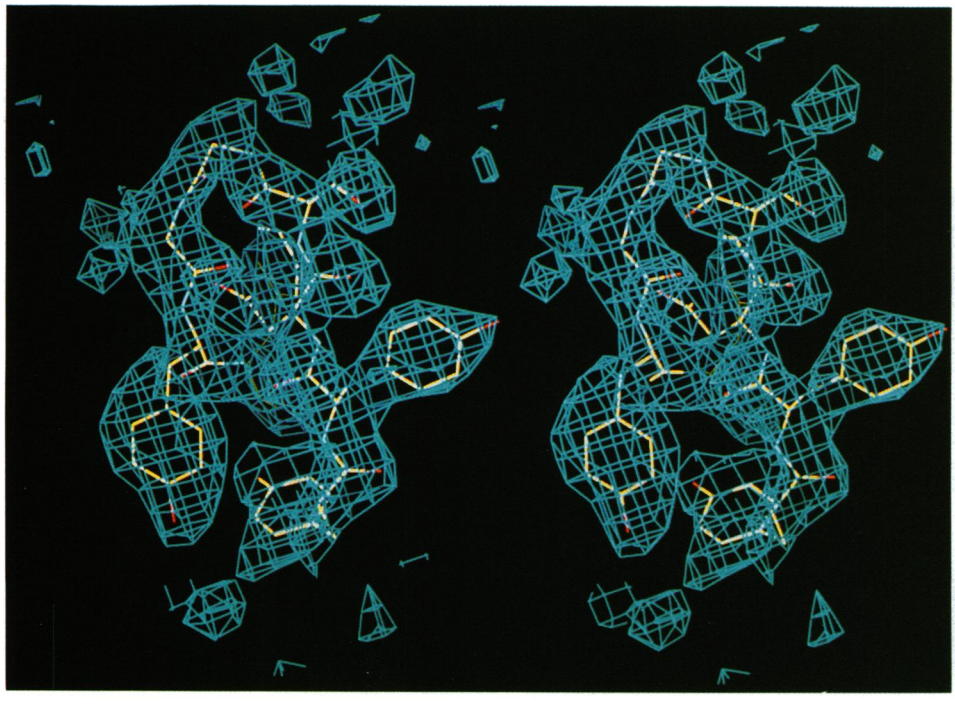

FIG. 2. Stereo view of the omit map illustrating the electron density of the polypeptide region of the third hypervariable loop of the heavy chain, H3, contoured at 2σ (where σ is the square root of the map variance).

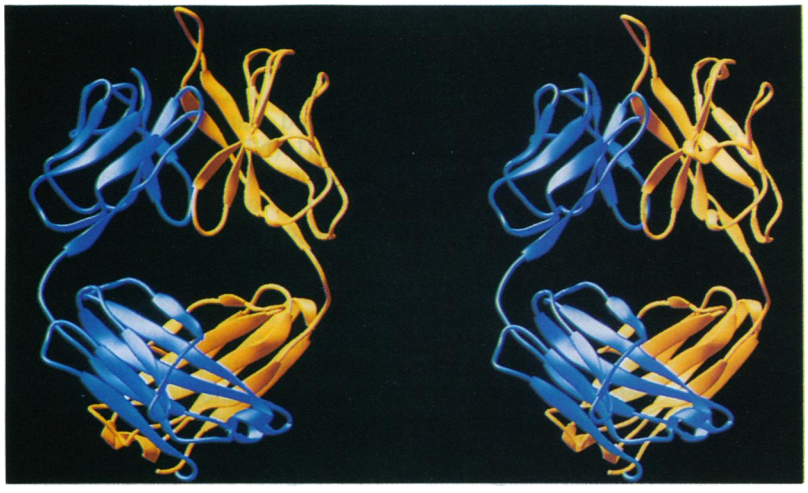

FIG. 3. Stereo view of the ribbon tracing of Fab 3D6 as determined at 2.7-Å resolution illustrating the unusually extended nature of the third hypervariable loop of H3. L and H are shown in blue and yellow, respectively. This figure was produced with the aid of the computer program RIBBONS (15).

folding of the amino acids in the previous model of the missing loop. H3, however, now gave reasonable, though weak, electron density for Tyr-102H through Phe-109H, which was consistent with the amino acid sequence but different from the previous model. Subsequent model refinement showed a significant reduction in the R factor and good electron density produced from unbiased (2FO-FC) and omit map electron density calculations (Fig. 2).

The final refinement with tightened geometrical restraints gave R = 0.191 for data from 6.0 Å to 2.7 Å (12,079 reflections) with fixed temperature factors and R = 0.177 with temperature factors refined. Refinement of the individual temperature factors was done with data ranging from 6 Å to 2.7 Å using program XPLOR and produced B values in the range of 2 $Å^2$ to 28 $Å^2$. B values over 20 $Å^2$ were located primarily in the longer, solvent-exposed side chains. The majority of the main chain atoms possess low B values with the exception of the four termini and residues 102H to 107H (H3) and 140H to 146H; in these cases the main chain atoms have B values around 20 $Å^2$. Strictly speaking, with the limited resolution range of 2.7-Å data, the B values may not be entirely accurate. However, they do give us some indication of the mobility of different regions of the molecule, which, as previously described, are reasonable. No water molecules were included in the refinement. Deviations from ideality in bond lengths and angles are 0.02 Å and 4°, respectively.

RESULTS AND DISCUSSION

The overall molecular structure of 3D6 assumes the standard immunoglobulin fold on which six hypervariable loops are supported, three from the V_L region (L1, L2, L3), and three from the V_H region (H1, H2, H3) (Fig. 3). Although the topology of the structure is unambiguous, the orientations of some amide planes in the exposed loops could not be definitively placed at this resolution. In addition, there is a small break in the electron density at Ser-143H. The remainder of the structure is very well determined. The elbow bend angle of 175° is within a normal range for observed Fab structures, and, consistent with this observation, Fab 3D6 is in an extended conformation with the V and C heterodimers well separated (Fig. 3). Structurally, four of the hypervariable loops can be classified after Chothia *et al.* (9, 10) according to the previously recognized canonical structures (L1, type 2; L2; H1, type 1; and H2, type 3). L3 is structurally distinct and may represent a new canonical structure (Fig. 4). For example, all of the previously determined structures described by Chothia and Lesk (10) contain prolines at position 94 or 95. The absence of the characteristic proline in L3, as in 3D6, is rarely observed in the amino acid sequences of κ chains, adding further support for the observed distinction. H3, an unusually long 17-amino acid loop, is identical in length with H3 of KOL (36); however, there is no sequence or structural homology between them. A further conformational analysis of these hypervariable loops as it pertains to *ab-initio* model building will be described elsewhere.

The most interesting and striking feature of the structure is the surface of the antibody combining site (Fig. 5). The loops L3, H3, and H2, formed by residues 93L–102L, 101H–110H, and 52H–57H, respectively, protrude from the surface, forming the major portions of the outer walls of a centrally located,

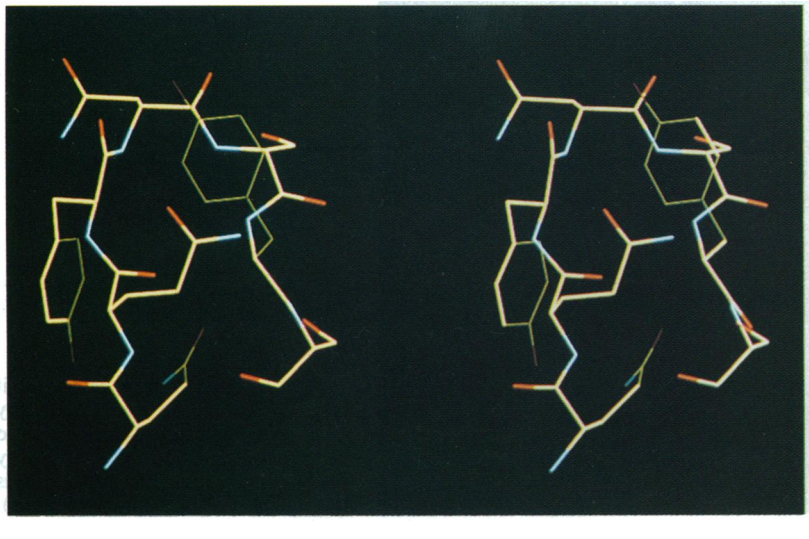

FIG. 4. Stereo view of hypervariable loop L3 of 3D6 oriented after Chothia *et al.* (9) illustrating the distinction from previous canonical L3 structures. Amino acids from left to right are Gln-89L, Gln-90L, Tyr-91L, Asn-92L, Ser-93L, Tyr-94L, and Ser-95L.

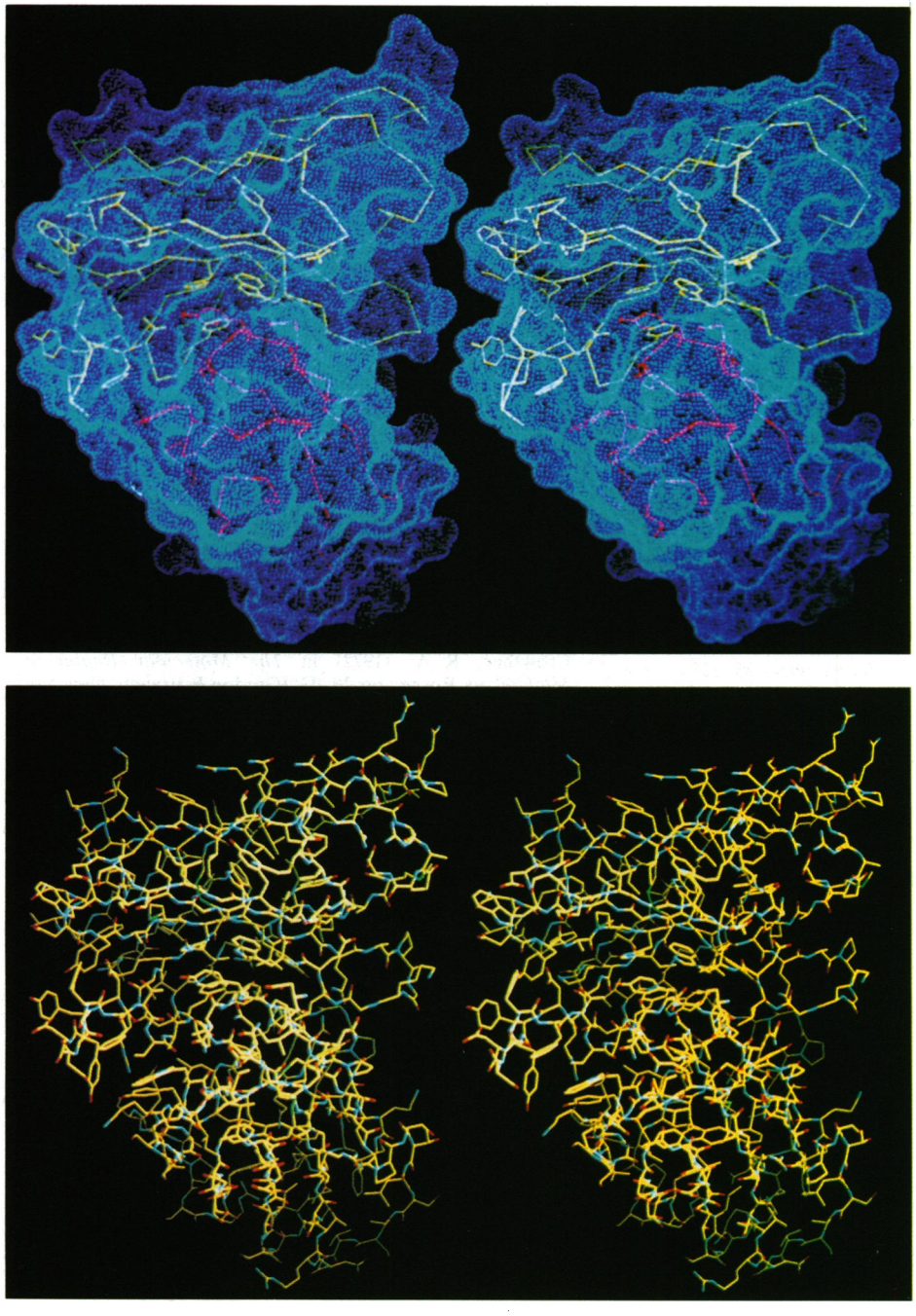

FIG. 5. (*Upper*) Stereo view of the Van der Waals surface of the hypervariable region superimposed on the Cα-carbon backbone of the polypeptide illustrating details of the hydrophobic pocket (see text). The H and L hypervariable regions are shown in yellow and red, respectively. (*Lower*) Stereo view in the same orientation as in *Upper* showing all of the atoms (yellow, carbon; red, oxygen; and blue, nitrogen).

triangular-shaped pocket. The unusually extended nature of H3 is stabilized by β-sheet formation and hydrophobic interactions between Phe-109H and Tyr-103H of H3 and Trp-32L of L1. The base of the pocket is lined with aromatic residues: Trp-47H, His-35H, and Tyr-94L. The back side of the pocket is completed with a section of extended polypeptide chain from H1. Thus, with the exception of L2, all of the hypervariable loops make structural contributions to the presumed binding pocket. The amino acids that contribute side chains to the surface of the pocket and surrounding area are Asp-1L, Trp-32L, Tyr-91L, Ser-93L, Tyr-94L, and Ser-96L for L and Asn-30H, Asp-31H, Tyr-32H, Ala-33H, His-35H, Val-37H, Trp-47H, Ser-52H, Trp-53H, Asp-54H, Ser-56H, Ser-57H, Tyr-60H, Asp-62H, Tyr-95H, Asp-101H, Tyr-102H, Ser-105H, Tyr-108H, Phe-109H, Thr-110H, and Phe-113H for H. It is interesting to note that the epitope recognized by 3D6, SGKLCTTAVPWNAS, contains only one basic residue, Lys, and one Asn. The remaining amino acids are primarily hydrophobic in nature. Two charged residues are buried in the triangular pocket of the CDR of 3D6, Asp-101H from H3 and His-35H from H1, while the remaining residues of the immediate pocket are primarily hydrophobic. Although the exact chemical nature of the antigen and antibody complex can only be determined experimentally, the potential for a complementary interaction is striking and provides an important framework for future experiments—e.g., site-directed mutagenesis. To elucidate the atomic details of this antibody complexed with its antigen, research must continue in an effort to form crystalline complexes with large aqueous soluble fragments of gp41 with this antibody.

J. Herron is thanked for many helpful discussions, for comments on the manuscript, and for supplying the coordinates of BV0401. This research was supported in part by the Office of Space Science and Applications of the National Aeronautics and Space Administration (NASA) (RTOP 674-23-08-17) and the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (Project P7556-BIO). X.-m.H.

and E.C. were supported by NASA under contract with Universities Space Research Association (USRA).

1. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. (1986) *Science* **233**, 747–753.
2. Colman, P. M., Laver, W. G., Varghese, J. N., Baker, A. T., Tulloch, P. A., Air, G. M. & Webster, R. G. (1987) *Nature (London)* **326**, 358–363.
3. Colman, P. M., Tulip, W. R., Varghese, J. N., Tulloch, P. A., Baker, A. T., Laver, W. G., Air, G. M. & Webster, R. G. (1989) *Philos. Trans. R. Soc. London Ser. B* **323**, 511–518.
4. Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. & Davies, D. R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8075–8079.
5. Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J. & Davies, D. R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5938–5942.
6. Stanfield, R. L., Fieser, T. M., Lerner, R. A. & Wilson, I. A. (1990) *Science* **248**, 712–719.
7. Bentley, G. A., Boulot, G., Riottot, M. M. & Poljak, R. J. (1990) *Nature (London)* **348**, 254–257.
8. Davies, D. R., Sheriff, S. & Padlan, E. A. (1988) *J. Biol. Chem.* **263**, 10541–10544.
9. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989) *Nature (London)* **342**, 877–883.
10. Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196**, 901–917.
11. Bruccoleri, R. E. & Karplus, M. (1987) *Biopolymers* **26**, 137–168.
12. Martin, A. C. R., Cheetham, J. C. & Rees, A. R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9268–9272.
13. Martin, A. C. R., Cheetham, J. C. & Rees, A. R. (1991) *Methods Enzymol.* **203**, 121–152.
14. Roberts, S., Cheetham, J. C. & Rees, A. R. (1987) *Nature (London)* **328**, 731–734.
15. Carson, M. (1987) *J. Mol. Graphics* **5**, 103–106.
16. Glockshuber, R., Steipe, B., Huber, R. & Plückthun, A. (1990) *J. Mol. Biol.* **213**, 613–615.
17. Skerra, A., Glockshuber, R. & Plückthun, A. (1990) *FEBS Lett.* **271**, 203–206.
18. Skerra, A. & Plückthun, A. (1988) *Science* **240**, 1038–1041.
19. Bird, R. E., Hardman, K. D., Jacobson, G. W., Johnson, S., Kaufman, B. M., Lee, S.-M., Lee, T., Pope, S. H., Riordan, G. S. & Whitlow, M. (1988) *Science* **242**, 423–426.
20. Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M.-S., Novotny, J., Margolies, M. N., Ridge, R. S., Bruccoleri, R. E., Haber, E., Crea, R. & Oppermann, H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5879–5883.
21. Waldmann, T. A. (1991) *Science* **252**, 1657–1662.
22. Grunow, R., Jahn, S., Porstmann, T., Kiessig, S. S., Steinkellner, H., Steindl, F., Mattanovich, D., Gürtler, L., Deinhardt, F., Katinger, H. & Von Baehr, R. J. (1988) *Immunol. Methods* **106**, 257–265.
23. Jungbauer, A., Tauer, C., Wenisch, E., Steindl, F., Purtscher, M., Reiter, M., Unterluggauer, F., Buchacher, A., Uhl, K. & Katinger, H. (1989) *J. Biochem. Biophys. Methods* **19**, 223–240.
24. Döpel, S.-H., Porstmann, T., Henklein, P. & Von Baehr, R. (1989) *J. Virol. Met.* **25**, 167–177.
25. Chang, T. W., Kato, I., McKinney, S., Chanda, P., Barone, A. D., Wong-Staal, F., Gallo, R. C. & Chang, N. T. (1985) *BioTechnology* **3**, 905–909.
26. Felgenhauer, M., Kohl, J. & Rüker, F. (1990) *Nucleic Acids Res.* **18**, 4927.
27. Casale, E., Wenisch, E., He, X.-M., Righetti, P. G., Snyder, R. S., Jungbauer, A., Tauer, C., Rüker, F. & Carter, D. C. (1990) *J. Mol. Biol.* **216**, 511–512.
28. Howard, A. J., Gilliland, G. L., Finzel, B. C. & Poulos, T. L. (1987) *J. Appl. Crystallogr.* **20**, 383–387.
29. Rossmann, M. G. & Blow, D. M. (1962) *Acta Crystallogr.* **15**, 24–31.
30. Fitzgerald, P. M. D. (1988) *J. Appl. Crystallogr.* **21**, 273–278.
31. Crowther, R. A. (1972) in *The Molecular Replacement Method*, ed. Rossmann, M. G. (Gordon & Breach, New York), pp. 173–178.
32. Lattman, E. E. & Love, W. E. (1970) *Acta Crystallogr.* **B26**, 1854–1857.
33. Crowther, R. A. & Blow, D. M. (1967) *Acta Crystallogr.* **23**, 544–548.
34. Herron, J. N., He, X.-M., Mason, M. L., Voss, E. W., Jr. & Edmundson, A. B. (1991) *Proteins* **11**, 159–175.
35. Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980) *J. Mol. Biol.* **141**, 369–391.
36. Saul, F. A., Amzel, L. M. & Poljak, R. J. (1978) *J. Biol. Chem.* **253**, 585–597.
37. Jones, A. T. (1978) *J. Appl. Cryst.* **11**, 268–271.
38. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235**, 458–460.
39. Hendrickson, W. A. (1985) *Methods Enzymol.* **115**, 252–270.