J. Mol. Biol. (1991) **221**, 239–256

2·9 Å Resolution Structure of an Anti-dinitrophenyl-spin-label Monoclonal Antibody Fab Fragment with Bound Hapten

Axel T. Brünger, Daniel J. Leahy†, Thomas R. Hynes and Robert O. Fox‡

The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry Yale University, New Haven, CT 06511, U.S.A.

(Received 9 November 1990; accepted 29 January 1991)

The crystal structure of the Fab fragment of the murine monoclonal anti-dinitrophenyl-spin-label antibody AN02 complexed with its hapten has been solved at 2·9 Å resolution using a novel molecular replacement method. Prior to translation searches, a large number of the most likely rotation function solutions were subjected to a rigid body refinement against the linear correlation coefficient between intensities of observed and calculated structure factors. First, the overall orientation of the search model and then the orientations and positions of the four Fab domains (V_H, V_L, C_H1 and C_L) were refined. This procedure clearly identified the correct orientation of the search model. The refined search model was then subjected to translation searches which unambiguously determined the enantiomer and position in the unit cell of the crystal. The successful search model was the refined 2·5 Å crystal structure of the Fab fragment of HyHel-5 from which non-matching residues in the variable domains had been removed. HyHel-5 is a murine monoclonal antibody whose heavy and light chains are of the same subclass ($\gamma 1$, κ, respectively) as AN02. After molecular replacement the structure of the AN02 Fab has been refined using simulated annealing in combination with model building and conjugate gradient refinement to a current crystallographic *R*-factor of 19·5% for 12,129 unique reflections between 8·0 and 2·9 Å. The root-mean-square (r.m.s.) deviation from ideal bond lengths is 0·014 Å, and the r.m.s. deviation from ideal bond angles is 3·1°. The electron density reveals the hapten sitting in a pocket formed by the loops of the complementarity determining region. The dinitrophenyl ring of the hapten is sandwiched between the indole rings of Trp96 of the heavy-chain and Trp91 of the light-chain. The positioning of the hapten and general features of the combining site are in good agreement with the results of earlier nuclear magnetic resonance experiments.

Keywords: antibody; fab fragment; crystal structure; molecular replacement; PC-refinement

1. Introduction

X-ray structural analyses of several immunoglobulin Fab fragments have revealed that the variable domains of Fab fragments have a characteristic folding pattern (Suh *et al.*, 1986; Satow *et al.*, 1986; Poljak *et al.*, 1974; Marquart *et al.*, 1980; Sheriff *et al.*, 1987; Epp *et al.*, 1975). This characteristic "immunoglobulin" fold consists of two antiparallel β-sheets whose strands are connected by a conserved topology of loops. Several genetic mechanisms combine to concentrate a high level of sequence diversity in three loops of both the heavy and the light-chain variable domains. These "hypervariable" or "complementarity determining region" (CDR§) loops all occur on the same exposed surface of the immunoglobulin molecule and form the antigen binding site.

Recent analyses of the known sequences and

† Present address: Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168th Street, New York, NY 10032, U.S.A.

‡ Author to whom all correspondence should be addressed.

§ Abbreviations used: CDR, complementarity determining region; CPU, central processing unit; DNP, dinitrophenyl; Fab, antigen binding fragment of an antibody; H1, heavy chain CDR1; H2, heavy chain CDR2; H3, heavy chain CDR3; L1, light chain CDR1; L2, light chain CDR2; L3, light chain CDR3; n.m.r., nuclear magnetic resonance; NOE, nuclear Overhauser effect; *PC*, Patterson correlation coefficient, that is the standard linear correlation coefficient between $|E_{obs}|^2$ and $|E_{model}|^2$; r.m.s., root-mean-square; SA, simulated annealing; S/N, signal to noise; σ, standard deviation; TEMPO, 2,2,6,6,-tetramethyl-1-piperindinyloxy; V_L, variable domain, light chain (residues 1 to 106); C_L, constant domain, light chain (residues 107 to 212); V_H, variable domain, heavy chain (residues 1 to 116); C_H1, constant domain, heavy chain (residues 117 to 217).

0022-2836/91/170239-18 $03.00/0

structures of antibodies indicate that the main-chain conformations of at least five of the six hypervariable loops which form the antigen binding site may be limited to a small number of discrete structural classes (Chothia & Lesk, 1987). The existence of these "canonical" structures has enabled the accurate modeling of hypervariable loop conformation for several antibodies prior to the determination of their crystal structure (Chothia *et al.*, 1989). As more structural data on immunoglobulins become available, it is likely that the modeling of combining site structure for most antibodies will be reduced to understanding hypervariable loop side-chain conformations and the main-chain conformation of the third hypervariable region of the heavy chain (Martin *et al.*, 1989). In addition to providing insight into the relationship between amino acid structure and function, the ability to accurately model antibody-combining site structure would have direct practical applications. Understanding more precisely the chemical features of antigen–antibody interactions would aid in designing antibodies with improved properties for use as drugs or diagnostic reagents. Efforts to engineer enzymatic function into antibody-combining sites would also be benefited.

AN02 is a member of a family of 12 murine monoclonal anti-dinitrophenyl-spin-label antibodies. cDNAs encoding the heavy and light chains of each of these antibodies have been cloned and sequenced (Leahy *et al.*, 1988*a*). An expression system has been adapted to enable production of recombinant antibodies, and several site-specific mutants of AN02 have been constructed (T. P. Theriault & H. M. McConnell, unpublished results; Leahy, 1988). This family of antibodies was created to exploit the effects of the paramagnetic hapten on the antibody nuclear magnetic resonance (n.m.r.) spectra to study the properties of the hapten binding site. Due to the broadening effect of the spin label on the resonance signals of nearby protons, the difference of the n.m.r. spectra of the anti-spin-label antibodies with and without bound hapten contains only those signals from protons within approximately 17 Å (1 Å = 0·1 nm) of the hapten. That is, one can obtain n.m.r. spectra which consist principally of only those resonance signals emanating from protons in the hapten binding site.

Growth of the hybridoma cells in media containing either perdeuterated or partially deuterated amino acids results in the production of antibody protein with specific amino acid types deuterated. In addition to allowing assignment of resonance signals to specific amino acid types, deuteration affords considerable simplification of n.m.r. spectra by eliminating splitting in the case of partial deuteration and eliminating entire resonance signals in the case of perdeuteration. Many distances and distance constraints between various protons in the AN02-combining site have been measured from n.m.r. spectra of deuterated samples (Anglister *et al.*, 1984*a*,*b*, 1985). The principal difficulty in fully interpreting the observed n.m.r. spectra is the inability to make unequivocal assignments of resonance signals to specific amino acid residues. The expression system has been developed to help solve this problem through the analyses of the n.m.r. spectra of single site mutants.

We began a crystallographic analysis of the AN02 Fab to provide a basis for comparison with and interpretation of existing n.m.r. data. Tentative assignments of resonance signals to specific amino acid protons have been made based on the crystal structure and nuclear Overhauser effect (NOE) spectra (Theriault *et al.*, 1991). Comparison of the crystal structure and models of the antibody based upon n.m.r. data will help demonstrate the strengths and weaknesses of each technique and may reveal differences between the antibody in solution and in the crystal. Structural studies of genetically engineered single site mutants will allow assessments of the structural and functional role of specific amino acids. Crystals of the AN02 Fab without its hapten have also been produced and characterized (Leahy *et al.*, 1988*b*).

The structural basis of affinity maturation during somatic mutation can be examined by construction of mutant antibodies with sequences "intermediate" between the germline and final sequences. Candidate genes encoding the germline variable regions for the AN02 heavy and light chains have been cloned and sequenced to provide a basis for such experiments (Theriault *et al.*, 1990). Chothia & Lesk (1987) have identified several framework residues believed to be important in the determination of hypervariable loop structure. n.m.r. and X-ray structural analysis of AN02 mutants at these particular residues could directly address the specific roles played by these amino acids in determining combining site structure and function.

We report here the 2·9 Å resolution crystal structure of the Fab fragment of the AN02 antibody that was solved using novel molecular replacement techniques. Conventional molecular replacement (Hoppe, 1957; Rossmann & Blow, 1962) is a two-step procedure, first rotating (rotation function) and then translating (translation function) the known structure of a homologous search model in the unit cell of the target crystal in order to obtain a maximum correlation between diffraction data calculated from the model and observed diffraction data. As suggested by Brünger (1990*a*), molecular replacement can be generalized by subjecting the most likely rotation function solutions to a rigid body refinement of the overall orientation of the search model as well as of the orientations and positions of the four Fab domains (V_H, V_L, C_H1 and C_L) prior to translation searches. These refinements resulted in significant changes in the interdomain angles of the Fab and proved essential in identifying both the correct orientation as well as the position of the search model in the unit cell of the crystal. The application of generalized molecular replacement to AN02 proves its usefulness to crystal structure determinations with high symmetry.

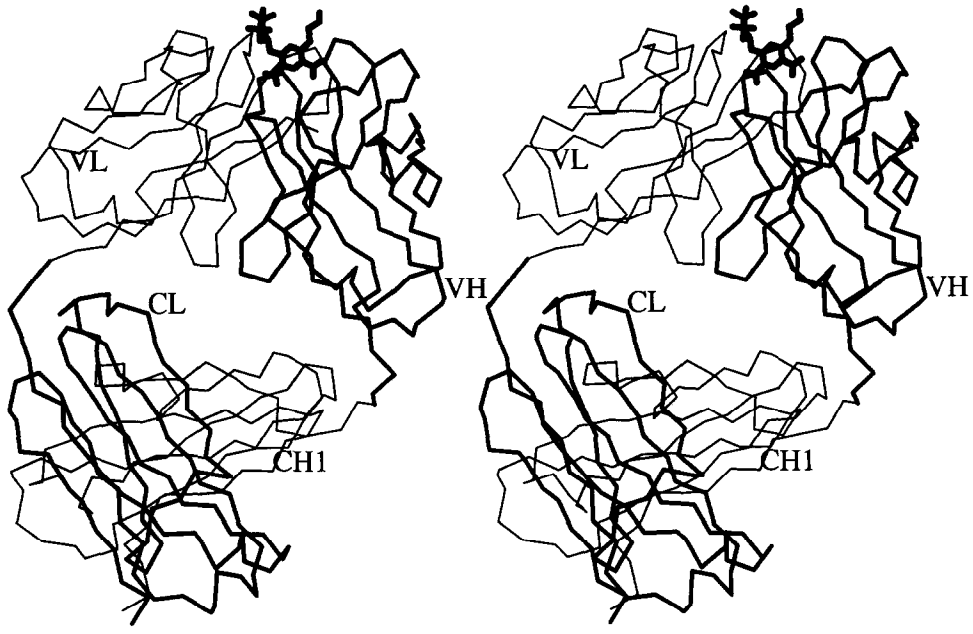

Figure 1. C^{α} backbone plot showing the 4 domains of the AN02 structure (thin and medium lines) and the hapten molecule (thick lines).

2. Materials and Methods

(a) *Definitions*

The elbow angle is defined as the angle between the pseudo-2-fold axes of symmetry of the V_L–V_H and C_L–C_H1 domain pairs. The location of the four domains and the putative location of the hapten molecule is shown in Fig. 1. All sequence numbers in this paper refer to the numbering scheme of Kabat *et al.* (1987). Table 1 shows how the Kabat numbering scheme relates to a purely sequential numbering scheme for AN02.

Table 1
A conversion table for the AN02 sequential and Kabat et al., *(1987) numbering systems*

Kabat *et al.*	Sequential
A. *Heavy chain*	
1–35	1–35
35A	36
36–82	37–83
82A, 82B, 82C	84–86
83–97	87–101
100K	102
101–130	103–132
133–154	133–154
156–157	155–156
162–169	157–164
171–180	165–174
183–196	175–188
198–200	189–191
202–206	192–196
208–223	197–212
226–229	213–216
235	217
B. *Light chain*	
1–27	1–27
29–95	28–94
95A	95
96–214	95–214

(b) *Preparation of crystals*

The purification, proteolytic cleavage and crystallization of the antibody protein were carried out as described (Leahy *et al.*, 1988*b*). To prepare the crystals a 20 mg/ml solution of the Fab in PBS (150 mM-NaCl, 10 mM-NaH_2PO_4, pH 7·2, 0·02% (w/v) NaN_3) was saturated with hapten, centrifuged 20 min at 13,000 revs/min in a microfuge, and stored as 400 μl samples in glass n.m.r. tubes at 4 °C. Crystals could be seen after 1 to 2 weeks and continued to grow for several months reaching sizes up to 2 mm × 2 mm × 3 mm.

(c) *Diffraction data*

Data were collected on film at the Stanford Synchrotron Radiation Laboratory using radiation of wavelength 1·54 Å. Two crystals were aligned with the *c*-axis parallel to the spindle axis and 3° oscillation photographs were taken. To collect data from the "missing cone", another crystal was aligned with the *c*-axis parallel to the incident radiation and 1·5° oscillation photographs were taken. Data were reduced using the software developed at the Macromolecular Supercomputer Facility at Purdue University for their Cyber 205 supercomputer (Rossmann, 1985). A total of 70,064 reflections were measured which resulted in 12,775 unique reflections to a 2·9 Å *d*-spacing. These data represent 90% of the theoretically available data. The R_{merge} value for data from all crystals was approximately 14%. The refined cell dimensions were $a = b = 73{\cdot}23$ Å, $c = 373{\cdot}8$ Å, $\alpha = \beta = 90°$ and $\gamma = 120°$. The space group was determined to be either one of the enantiomeric pair $P6_122$, $P6_522$. Final solution of the structure identified the space group as $P6_522$.

(d) *Search model*

The search model included the Fab portion of the HyHel-5 structure (Sheriff *et al.*, 1987) with temperature factors set to twice the values of the original HyHel-5 structure. This was done in order to amplify the effect of

the *B*-factors, i.e. less well determined regions contribute less to the Patterson map than with the original *B*-factors. In some of the rotation functions (see the legend to Table 2), certain residues and/or side-chain atoms were left out in regions where sequence differences exist between AN02 and HyHel-5. The sequence identity between AN02 and HyHel-5 is 83% (V_L), 41% (V_H), 100% (C_H1) and 100% (C_L).

(e) *Computational techniques*

All calculations (molecular replacement, *PC*-refinement and *R*-factor refinement of rigid groups as well as individual atoms) were carried out with version 2.1 of the program X-PLOR (Brünger, 1990*b*). X-PLOR is available upon request from A.T.B. Display and fitting of electron density maps was carried out with FRODO (Jones, 1978). Main-chain geometry was analyzed with the program GEOM (G. Cohen, personal communication). Least-squares superpositions of antibody domains were carried out using the program TOSS (Hendrickson, 1979).

3. Course of Structure Solution and Refinement

(a) *Rotation search*

The direct space Patterson search method of Huber (1985) was employed for rotation searches with the HyHel-5 search model against the AN02 data. The crystal Patterson map was computed from the observed intensities by fast Fourier transformation on a 1 Å grid. The model Patterson map was computed by placing the HyHel-5 search model into a 120 Å × 120 Å × 150 Å orthorhombic box with the largest extent of the model approximately parallel to the *z*-direction, evaluating the structure factors, and fast Fourier transformation of the squared amplitudes on a 1 Å grid. The box size was chosen such that it was about twice as large as the model in each direction in order to avoid overlap of self Patterson vectors. Model Patterson vectors were selected according to length and magnitude (Table 2). The selected Patterson vectors were rotated using the Eulerian angles $(\theta_1, \theta_2, \theta_3)$ as defined by Rossmann & Blow (1962). The product correlation of the rotated model Patterson vectors with the crystal Patterson map was computed by linear eight-point interpolation (Nordman, 1980). The orientations of the search models were sampled by using the pseudo-orthogonal Eulerian angles introduced by Lattman (1972):

$$\begin{aligned} \theta_+ &= \theta_1 + \theta_3 \\ \theta_- &= \theta_1 - \theta_3 \\ \theta_2 &= \theta_2. \end{aligned} \tag{1}$$

To ensure proper sampling of the rotation function, the sampling interval Δ for θ_2 should be less than $\arcsin(d/3r)$, where d is the high-resolution limit of the Patterson maps and r is the maximum Patterson vector length of the search model. Two different sampling intervals were tested (Table 2), both of which satisfy the sampling criterion. Following Lattman (1972), the variable interval for θ_+ is given by $\Delta/\cos(\theta_2/2)$ and the variable interval for θ_- is given by $\Delta/\sin(\theta_2/2)$. As the AN02 crystal symmetry is $P6_122$ or $P6_522$, the rotation search could be restricted to the asymmetric unit $\theta_+ = 0 \ldots 720°$, $\theta_2 = 0 \ldots 90°$, $\theta_- = 0 \ldots 60°$ (Rao *et al.*, 1980). All sampled grid points were sorted with respect to the rotation function value and the peak search described by Brünger (1990*a*) was carried out.

Table 2

Conventional rotation searches

Patterson vector selection									Peak height[c]		
Resolution range (Å)	Length[d] (Å)	P_{min}[e]	No.[f]	Atom selection[g]	Δ[a] (°)	θ_1 (°)	θ_2 (°)	θ_3[b] (°)	Highest	Next	S/N
		Correct orientation				46·9	62·9	359·1			
15–5	4–24	3·0	1778	Subset	2·5	55·8	52·5	344·5	2·7	2·5	1·1
15–4	4–24	3·0	3294	Subset	2·5	48·5	57·5	351·4	3·0	2·6	1·2
15–4	4–24	3·0	3294	Subset	1·0	48·7	57·0	350·0	3·1	2·6	1·2
15–3·5	4–24	3·0	5811	Subset	2·5	48·5	57·5	348·5	3·1	2·6	1·2
15–3	4–24	3·0	7906	Subset	2·5	48·6	57·5	351·3	3·1	2·7	1·1
15–4	4–40	2·5	9355	Subset	2·5	48·5	57·5	348·5	3·2	3·2	1·0
15–4	4–18	2·0	4007	Subset	2·5	49·9	57·5	349·9	2·8	2·4	1·2
15–4	4–24	3·0	3276	All	2·5	75·7	50·0	340·2	3·0	2·9	1·0
15–4	4–24	3·0	4548	C_L+C_H	2·5	251·6	77·5	203·6	2·5	2·4	1·0
8–3·5	4–45	3·0	13,505	C_L+C_H	2·5	47·7	60·0	352·7	4·0	3·9	1·0

[a] Sampling interval of the θ_2 angle.

[b] Eulerian angles specifying the orientations corresponding to the highest peak for each rotation function.

[c] Peak heights are measured in σ units above the mean. S/N is defined as the ratio of the 2 peak heights.

[d] This specifies a selection of the model Patterson vectors based on their length.

[e] This specifies a selection of the model Patterson vectors based on the value of the corresponding point in the Patterson map: all Patterson vectors are selected with values greater than P_{min} (in σ units above the mean of the Patterson map).

[f] This is the actual number of selected Patterson vectors.

[g] Subset means that the following atoms were omitted from the HyHel-5 search model: residues 38 to 46, 89 to 96 of the light chain, residues 24 to 34, 53 to 75, 61 to 68, 74 to 78, 84 to 94, 99 to 105, and all side-chain atoms beyond the β positions in the V_H domain.

Table 2 shows the results of rotation functions using various resolution ranges, Patterson vector selections, atom selections and sampling intervals. The values of the 130 highest peaks of the rotation function at 15 to 4 Å resolution are shown in Figure 2(a). As we show below, the correct orientation of the HyHel-5 search model is given by ($\theta_1 = 46{\cdot}9°$, $\theta_2 = 62{\cdot}9°$, $\theta_3 = 359{\cdot}1°$). The rotation functions with a subset of the HyHel-5 atoms predicted an orientation of about ($\theta_1 = 49°$, $\theta_2 = 58°$, $\theta_3 = 351°$), which is within 8° of the correct orientation. Although this orientation was relatively consistent among different resolution ranges, the low signal-to-noise (S/N) ratio and the failure of the translation search (see below) using this orientation did not make it trustworthy. Usage of all atoms of the HyHel-5 model did not produce the correct orientation. Usage of only the constant domains C_L, C_H1 as proposed by Cygler & Anderson (1988) produced the correct orientation for a particular choice of parameters but with a very low S/N ratio (Table 2). These results did not depend on the particular type of rotation function employed. For example, reciprocal-space rotation searches (Crowther, 1972) were inconclusive (not shown).

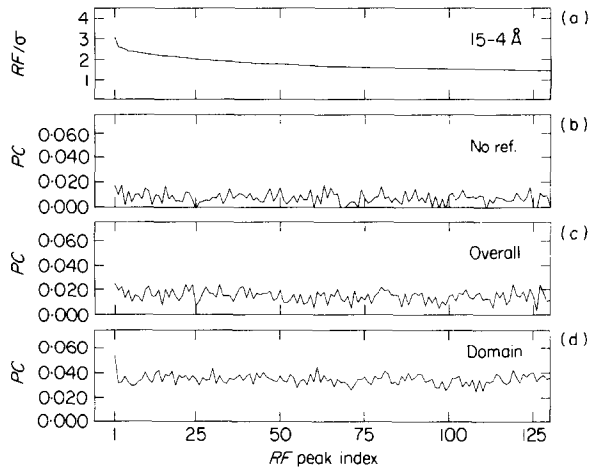

Figure 2. Rotation function and *PC*-refinement using the HyHel-5 search model against the AN02 diffraction data. (a) The values of the 130 highest peaks of the rotation function (second row in Table 2) are shown in σ units above the mean sorted by value; (b) correlation coefficients *PC* before refinement; (c) after refinement of the overall orientation; and (d) after refinement of the orientations and positions of the 4 Fab domains (V_H, V_L, C_H1 and C_L) at 15 to 4 Å resolution.

(b) PC *refinement*

The difficulties we encountered with the conventional rotation functions prompted us to try *PC*-refinement of the orientations and positions of the four Fab domains (Brünger, 1990*a*). Briefly, this consists of minimization of a target function:

$$E_{tot}(\Omega, \Omega_i, t_i) = (1 - PC(\Omega, \Omega_i, t_i)), \qquad (2)$$

where Ω is the overall orientation of the search model, and Ω_i and t_i are the individual orientations and positions of the four Fab domains, respectively. *PC* is the standard linear correlation coefficient:

$$PC(\Omega, \Omega_i, t_i) = \frac{\langle |E_{obs}|^2 |E_m(\Omega, \Omega_i, t_i)|^2 - \langle |E_{obs}|^2 \rangle \langle |E_m(\Omega, \Omega_i, t_i)|^2 \rangle \rangle}{\sqrt{\langle |E_{obs}|^4 - \langle |E_{obs}|^2 \rangle^2 \rangle \langle |E_m(\Omega, \Omega_i, t_i)|^4 - \langle |E_m(\Omega, \Omega_i, t_i)|^2 \rangle^2 \rangle}}. \qquad (3)$$

The symbols $\langle\rangle$ denote an averaging over the set of observed reflections expanded to P_1. E_{obs} denote the normalized observed structure factors and $E_m(\Omega, \Omega_i, t_i)$ denote the normalized structure factors of the search model placed in a triclinic unit cell identical in geometry to that of the crystal. The major difference to crystallographic *R*-factor refinement is that *PC*-refinement is carried out in *P*1, that is without crystallographic symmetry. Crystallographic symmetry cannot be applied since the molecule has not been positioned correctly in the unit cell. Hauptman (1982) showed that *PC* is a measure of phase error, and its use as a target for *PC*-refinement has been justified empirically by Brünger (1990*a*). More thorough studies of this point are in progress.

PC-refinements of the HyHel-5 search model were carried out where the search model was oriented according to each of the 130 selected peaks of the rotation function in Figure 2(a). It should be noted that a large number of the *PC*-refinements are actually carried out for incorrect orientations of the search model; only the *PC*-refinements starting close to the correct orientation are expected to yield a relatively large correlation coefficient after *PC*-refinement. The *PC*-refinements consisted of 15 steps of conjugate gradient minimization of E_{tot} using the method of Powell (1977) at 15 to 4 Å resolution for the overall orientation Ω of the molecule (Fig. 2(c)), followed by 50 conjugate gradient steps for the orientations and positions (Ω_i, t_i) of the four domains V_H, V_L, C_H1 and C_L (Fig. 2(d)). Peak number 1 is clearly the most significant peak in Figure 2(d). This result convinced us that this peak, indeed, corresponds to the correct orientation of the HyHel-5 search model in the AN02 crystal. However, it should be noted that *PC*-refinement produced a change of the overall orientation of about 8° (Table 3). This was the key for the success of the subsequent translation searches.

Figure 2(d) also illustrates the importance of *PC*-refinement of the orientations Ω_i and the positions t_i of the four Fab domains. The *PC* values prior to *PC*-refinement (Fig. 2(c)) or the *PC* values after refinement of the overall orientation (Fig. 2(c)) do not discriminate the correct peak among error peaks. It is only after *PC*-refinement of Ω_i and t_i that the first peak clearly shows a higher correlation

Table 3
Shifts of individual domains produced by PC-*refinement*

Domain	θ_1 (°)	θ_2 (°)	θ_3 (°)	Δx (Å)	Δy (Å)	Δz (Å)
$V_H+V_L+C_H1+C_L$	8·0	1·5	2·5	0	0	0
V_L	2·6	−1·7	5·5	−0·2	0·3	−1·4
C_L	−4·5	−0·6	0·6	0·5	−0·8	0·6
V_H	−2·5	−0·1	0·5	−0·5	0·4	−0·6
C_H1	−0·8	−1·2	−0·5	−0·4	0·2	0·3

Listed are the Eulerian angles and the shift vectors that represent the rotations and translations produced by rigid-body *PC*-refinement of the search model with the initial orientation given by ($\theta_1 = 48·5°$, $\theta_2 = 57·5°$, $\theta_3 = 351·4°$) (parameters of this rotation function are described in the second row of Table 2). The orientation after *PC*-refinement was ($\theta_1 = 46·9°$, $\theta_2 = 62·9°$, $\theta_3 = 359·1°$). *PC*-refinement consisted of 40 conjugate gradient steps with the overall orientation of the search model ($V_H+V_L+C_H1+C_L$). This was followed by 50 conjugate gradient steps of the orientations and positions of the 4 Fab domains (V_H, V_L, C_H1 and C_L).

compared to all other peaks. The individual shifts of the four domains are up to 5·5° and 0·8 Å (Table 3) resulting in a 4° change of the elbow angle. The *PC*-refinements for all 130 orientations consumed a total of 60 hours of CONVEX-C210 CPU time.

(c) *Translation search*

Translation searches using the method of Fujinaga & Read (1987) were carried out with the correctly oriented and *PC*-refined HyHel-5 search model. Briefly, the standard linear correlation coefficient between the normalized observed structure factors (E_{obs}) and the normalized calculated structure factors (E_{calc}) was computed:

$$TF(xyz) = \frac{\langle |E_{obs}|^2 |E_{calc}(xyz)|^2 - \langle |E_{obs}|^2 \rangle \langle |E_{calc}(xyz)|^2 \rangle \rangle}{\sqrt{\langle |E_{obs}|^4 - \langle |E_{obs}|^2 \rangle^2 \rangle \langle |E_{calc}(xyz)|^4 - \langle |E_{calc}(xyz)|^2 \rangle^2 \rangle}} \quad (4)$$

where the symbols $\langle\rangle$ denote an averaging over the set of observed reflections. Equation (4) is identical to equation (3) except E_m is replaced by $E_{calc}(xyz)$, where the latter denotes the normalized calculated structure factors of the search model and all of its symmetry mates. Data were included between 15 and 4 Å resolution. The sampling interval was set to about 1·33, that is, one-third of the high-resolution limit. The three-dimensional translation search was restricted to the asymmetric unit of the translation search ($0 \le x \le 1$, $0 \le y \le 1$, $0 \le z \le 0·5$ in fractional units). The CPU time for one translation search was 17 hours on a CONVEX-C210.

Translation searches were carried out with and without prior *PC*-refinement and in both enantiomeric space groups $P6_122$ and $P6_522$ (Table 4). Since the maximum correlation coefficients achieved with space group $P6_522$ was more than twice as large as that with space group $P6_122$ we concluded that $P6_522$ is the correct enantiomer. Key to the success of the translation search was the overall orientation *PC*-refinement. In fact, the translation search failed when orienting the HyHel-5 search model according to ($\theta_1 = 48·5°$, $\theta_2 = 57·5°$, $\theta_3 = 351·4°$), which is the orientation prior to *PC*-refinement. Sampling of the rotation function on a finer grid did not improve the situation. As noted by Brünger (1990*a*), this clearly demonstrates the superiority of the *PC* correlation coefficient compared to conventional rotation functions.

The *PC*-refinement of the orientations and positions of the four Fab domains further increased the S/N ratio of the translation search (Table 4). The position of the HyHel-5 search model predicted by the translation search was $x = 0·06$, $y = 0·44$, $z = 0·435$ in fractional units. Sections of the translation search through this position are shown in Figure 3. No packing overlaps between symmetry-related molecules were present for this position, as measured by the packing function (Brünger, 1990*a*) and visual inspection. The Fab molecules cover about 34% of the total volume of the unit cell.

(d) *Rigid-body* R-*factor refinement*

The initial *R*-factor of the properly placed and *PC*-refined HyHel-5 search model was 47% at 8 to 2·9 Å resolution and the correlation coefficient was

Table 4
Translation searches

Model	Enantiomer	TF_{max}[a]	Peak height[b]		
			Correct[c]	Error	S/N
No refinement	$P6_522$	0·11	4·6	5·3	0·9
Overall orientation ref.	$P6_522$	0·27	15·3	9·5	1·6
Domain refinement	$P6_522$	0·43	20·7	11·1	1·8
Domain refinement	$P6_122$	0·19			

[a] Maximum correlation coefficient obtained by the translation search.
[b] Peak heights are measured in σ units above the mean. S/N is defined as the ratio of the peak heights.
[c] The correct position of the model is given by $x = 0·06$, $y = 0·44$, $z = 0·435$ in fractional units.

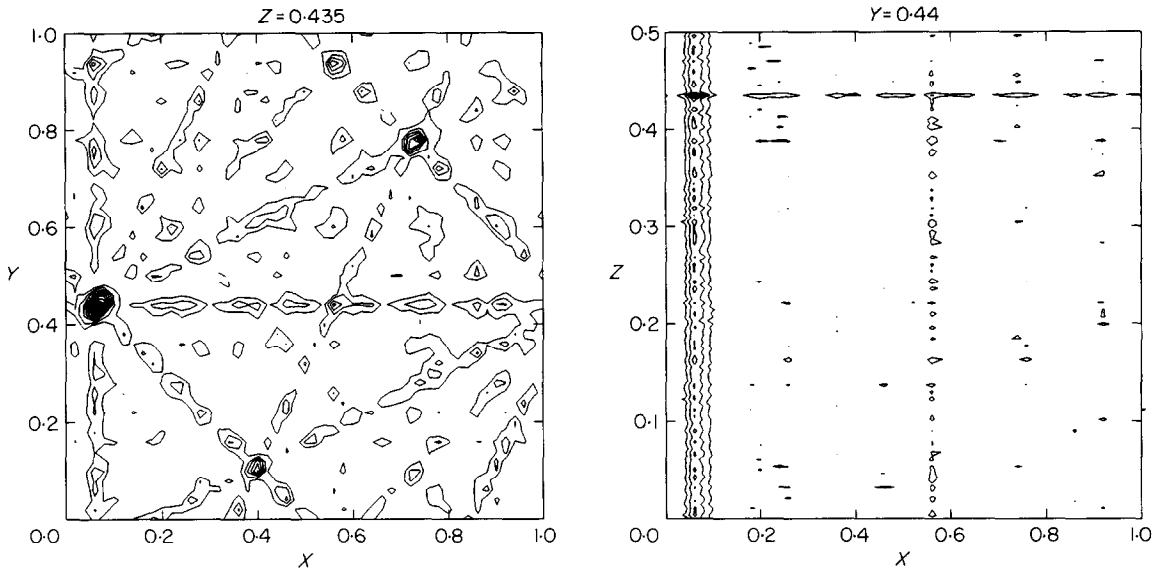

Figure 3. Contour plots of the translation function for the correctly oriented HyHel-5 search model after *PC*-refinement of the orientations and positions of the 4 Fab domains at 15 to 4 Å resolution. Shown are x–y and x–z slices through the maximum of the translation search located at ($x = 0{\cdot}06$, $y = 0{\cdot}44$, $z = 0{\cdot}435$). The sampling interval was approximately 1·3 Å, that is, one-third of the high-resolution limit. The mean and σ of the x–y slice is 0·087 and 0·024, respectively. The mean and σ of the x–z slice is 0·084 and 0·031, respectively. The lowest contours of the translation function are drawn at 1σ above the mean.

0·43 at 15 to 4 Å resolution. Rigid-body *R*-factor refinement was carried out by running 40 steps of conjugate gradient minimization with the overall orientation and position of the HyHel-5 search model, followed by 50 steps with the orientations and positions of the four Fab domains at 8 to 2·9 Å resolution. The *R*-factor dropped to 42% at 8 to 2·9 Å resolution. The shifts of the orientations and positions are shown in Table 5. The change in elbow angle during the rigid-body refinement is 3·5°, which reduces the elbow angle to 153·6°. The difference in elbow angle before and after rigid-body refinement is purely due to the extension of the diffraction data to 2·9 Å resolution and is not related to a problem with *PC*-refinement. In fact, *PC*-refinement at 15 to 2·9 Å resolution yields an elbow angle of 153·7° (not shown).

Table 5

Shifts of individual domains produced by rigid-body R-*factor refinement*

Domain	θ_1 (°)	θ_2 (°)	θ_3 (°)	Δx (Å)	Δy (Å)	Δz (Å)
$V_H+V_L+C_H1+C_L$	0·3	0·5	−0·1	0	0	0
V_L	−1·5	0	−0·4	−0·1	0	0·1
C_L	−0·2	−0·5	−0·4	−0·1	0	−0·2
V_H	3·3	1·0	3·5	0·4	0	0
C_H1	−2·7	1·5	0·2	0·2	−0·1	0·1

Listed are the Eulerian angles and the shift vectors that represent the rotations and translations produced by rigid-body *R*-factor refinement of the *PC*-refined search model with the initial position given by the translation search ($x = 0{\cdot}06$, $y = 0{\cdot}44$, $z = 0{\cdot}435$ in fractional units). The refinement consisted of 40 conjugate gradient steps with the overall orientation and position of the search model ($V_H+V_L+C_H1+C_L$). This was followed by 50 conjugate gradient steps of the orientations and positions of the 4 Fab domains (V_H, V_L, C_H1 and C_L).

(e) *First round of SA-refinement*

The rigid-body refined search model with the HyHel-5 sequence was the starting point for the first round of crystallographic refinement using the simulated annealing (SA) technique by Brünger *et al.* (1987). A slow-cooling protocol (Brünger *et al.*, 1990*a*) was employed that consisted of 80 steps C^{α}-restrained conjugate gradient minimization, followed by molecular dynamics starting at 3000 K and ending at 300 K over a period of 2·7 picoseconds, followed by 120 steps of conjugate gradient minimization. After this first round of SA-refinement, the *R*-factor dropped to 23·1% at 8 to 2·9 Å resolution with root-mean-square (r.m.s.) deviations of bond lengths and bond angles from ideality of 0·021 Å and 4·3°, respectively.

(f) *Subsequent refinement rounds and placement of the hapten molecule*

Following the initial SA-refinement, $(2|F_{obs}|-|F_{calc}|)e^{i\phi_{calc}}$ difference maps were computed, side-chains were substituted, inserted, or deleted on the computer graphics according to the sequence of AN02, and the hypervariable loops were inspected for evidence of the hapten. A model of the hapten had not been included in either the molecular replacement or the first round of SA-refinement. By inspection of the map, a likely position for the hapten was identified. This positioning placed the hapten in a pocket formed by the hypervariable loops with the dinitrophenyl (DNP) ring of the

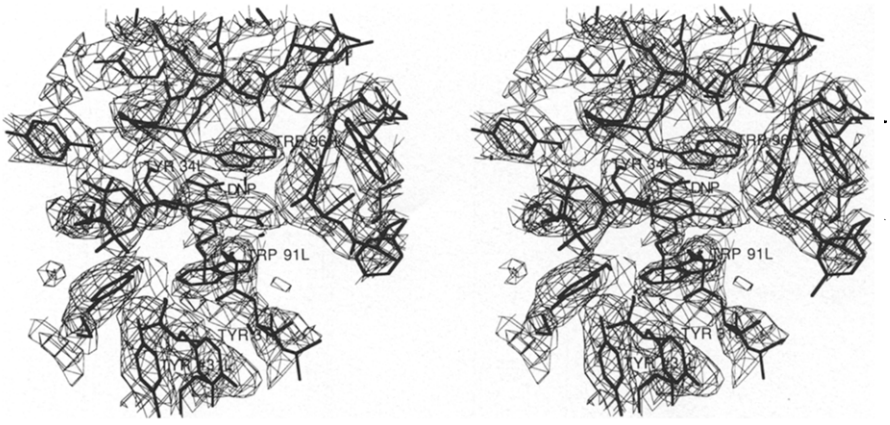

Figure 4. A stereo pair showing electron density (light lines) and the final atomic model (heavy lines) for the binding site region. The electron density map was calculated following an SA run with the binding site omitted as described in the text ("annealed" omit maps). The hapten and adjacent tryptophan and tyrosine residues are labeled. The sign (#) refers to a symmetry-related residue involved in a crystal lattice contact.

hapten sandwiched between the indole rings of Trp91 L and Trp96 H. Placement of a hapten model required rebuilding of heavy-chain CDR 3 residues 94 to 100. The SA-refinement had placed the indole ring of Trp96 H into the putative DNP ring density, moved the backbone atoms for this residue into what appeared to be the indole density, and arranged nearby side-chains into what appeared to be the correct main-chain density. The rebuilt model, now including the hapten, was submitted to another course of simulated annealing as described above, during which the *R*-factor dropped to 22·1% with an r.m.s. deviation from ideal of 0·020 Å for bond lengths and 4·2° for bond angles.

At this point several rounds of model building followed by conjugate gradient minimization were performed. The program GEOM (G. Cohen, personal communication) was used to identify residues with bad main-chain geometry, and these regions were omitted from map calculations when calculating new maps for subsequent cycles of rebuilding. Throughout these refinements an overall *B*-factor of 15 Å^2 was maintained for all atoms. These cycles of rebuilding resulted in an *R*-factor of 21·8% with an r.m.s. deviation of 0·016 Å for bonds and 3·5 Å for angles. Another course of simulated annealing was carried out as described above but with tightened force constants (Weis *et al.*, 1990) to see if the overall geometry could be improved. Following this annealing, the *R*-factor and r.m.s. bond length deviation remained at 21·8% and 0·016 Å, respectively, but the r.m.s. deviation in bond angles improved to 3·3°. The model was then subjected to three more cycles of rebuilding and conjugate minimization of both atomic positions as well as *B*-factors. *B*-factors were restrained where the target value for the deviations of *B*-factors for bonded atoms was set to 1·5 Å^2 and that for *B*-factors for atoms related by bond angles was set to 2·0 Å^2 (Brünger, 1990*b*). After the final rebuilding, alternating cycles of minimization and restrained *B*-factor refinement were performed until the *R*-factor had converged.

To verify the correctness of the combining site region, a slow-cooling SA-refinement was performed as described above with the binding site residues 89 to 96 L and 93 to 102 H, and the hapten omitted from the calculations. The free ends of the polypeptide chain were anchored by harmonic restraints with a force constant of 10 kcal/(mol Å^2) (1 cal = 4·184 J) to prevent distortions due to the removal of the binding site residues. Following this SA-refinement, an omit map was calculated from which the hapten and surrounding residues were omitted. From other work we have concluded that this "annealed" omit map minimizes model bias in the phase computation (unpublished results). The binding site region of this map and the current model is shown in Figure 4. The unambiguous presence of the hapten and surrounding residues in this map convinces us that our model of the binding site region is correct.

The current *R*-factor is 19·5% for reflections between 8·0 and 2·9 Å resolution, with an r.m.s. deviation of 0·014 Å for bond lengths and 3·1° for bond angles. No solvent molecules have been added to the model.

4. Results

A representative region of the electron density map is shown in Figure 5. The quality of the map is, in general, quite high and, with the exception of two disordered regions, allowed confident placement of main and side-chain atoms. For the vast majority of residues backbone carbonyl density was clearly evident. This carbonyl density provides confidence that the ϕ and ψ angles are generally correct. A Ramachandran plot is shown in Figure 6. Of the

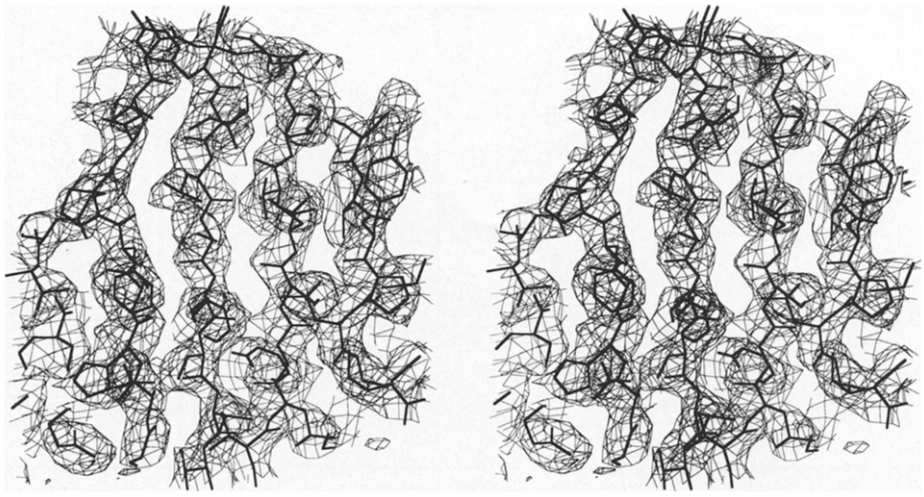

Figure 5. Stereo pair of strands 1, 2, 3 and 4 of the 4b sheet of the C_H1 domain. Small regions at the top left and right corners of this view were removed for clarity. The electron density was contoured at 1σ.

eight non-glycine residues with ϕ, ψ values significantly distant from sterically "allowed" regions, five (Ser128 H, Ala129 H, Asp229 H, Tyr94 L and Glu213 L) occur in disordered regions of the map. The remaining three (Thr51 L, Thr69 L and Ser15 H) occur in tight turns with stabilizing hydrogen bond networks and have ϕ, ψ values consistent with type II′ turns as described by Richardson (1981). The current R-value for the AN02 Fab structure is 19·5%. The r.m.s. deviations from ideal values for the following parameters are: bond lengths, 0·014 Å, bond angles, 3·1°, dihedral angles, 29·0° and improper angles, 1·5° (see Weis *et al.* (1990) for a description of this parameter). Side-chain and main-chain temperature factors are shown in Figure 7. A Luzzati (1952) plot is shown in Figure 8 and reveals an estimated r.m.s. error in co-ordinate position of 0·3 Å. While the Luzzati method may not be rigorously applicable to proteins, it has provided results consistent with other methods of estimating error in protein crystal structures (Kuriyan *et al.*, 1987).

The secondary structure and domain organization of the AN02 Fab are typical for immunoglobulin Fab fragments. Figure 9 compares the C^α trace of AN02 and HyHel-5; also shown is the placement of the hapten in AN02. The V_L and V_H domains of AN02 are related by a pseudo-dyad axis of 173·9° with a translation of −0·06 Å along this axis. The comparable axis in HyHel-5, the search model, consists of a 172·4° rotation and a 0·12 Å translation. These values are consistent with the values for other antibodies: 165·9° to 172·6° of rotation and −0·9 to 0·8 Å of translation (Padlan *et al.*, 1989). The C_H1 (murine $\gamma1$) and C_L (murine κ) domains of AN02 are related by a pseudo-dyad axis of 167·0° with a translation of −1·7 Å. These values are virtually identical to the values (166·6° and −1·6 Å for rotation and translation, respectively) reported for identical C_H1 and C_L domains in the HyHel-10 Fab (Padlan *et al.*, 1989). The elbow angle, defined as the angle between the C_H1/C_L and V_H/V_L pseudo-dyad axes, is 153·6°.

Table 6 shows the results of superpositions of AN02 backbone atoms (N, C^α, C and O) with homologous atoms from the structures of Fabs J539 (Suh *et al.*, 1986), McPC603 (Satow *et al.*, 1986), NEW (Poljak *et al.*, 1974), KOL (Marquart *et al.*, 1980), HyHel-5 (Sheriff *et al.*, 1987), and the Bence–Jones protein REI (Epp *et al.*, 1975). The percentage of shared residues is also shown. The superpositions were carried out using the program TOSS (Hendrickson, 1979), and the specific residues used for each superposition are given in the Table legend. The V_L domain of AN02 is structurally most similar to the V_L domains of REI and HyHel-5. The AN02 V_H domain is structurally most similar to the V_H domains of McPC603 and J539, but the match is not as close as the best matched V_L domains.

The results of comparisons of the AN02 C_L and C_H1 domains show r.m.s. differences of approximately 0·8 Å (Table 6). The largest deviations between domains invariably occur in the loops connecting the framework β-strands. During the first round of SA-refinement Pro200 in C_H1 converted from *trans* to *cis*. Examination of $(2|F_{obs}|-|F_{calc}|)e^{i\phi_{calc}}$ and $(|F_{obs}|-|F_{calc}|)e^{i\phi_{calc}}$ maps from which this region had been omitted from the map calculations indicate that this residue is most likely *cis* in AN02, unlike the corresponding residue in the identical HyHel-5 C_H1 domain.

Table 6 also lists the results of comparisons of the AN02 CDR loops with corresponding loops from other antibody structures. Comparisons were only made between loops of identical size. All backbone atoms were again used in the comparisons and the loops were superimposed both alone and in the context of the framework residues. The results of the comparisons are generally consistent with the

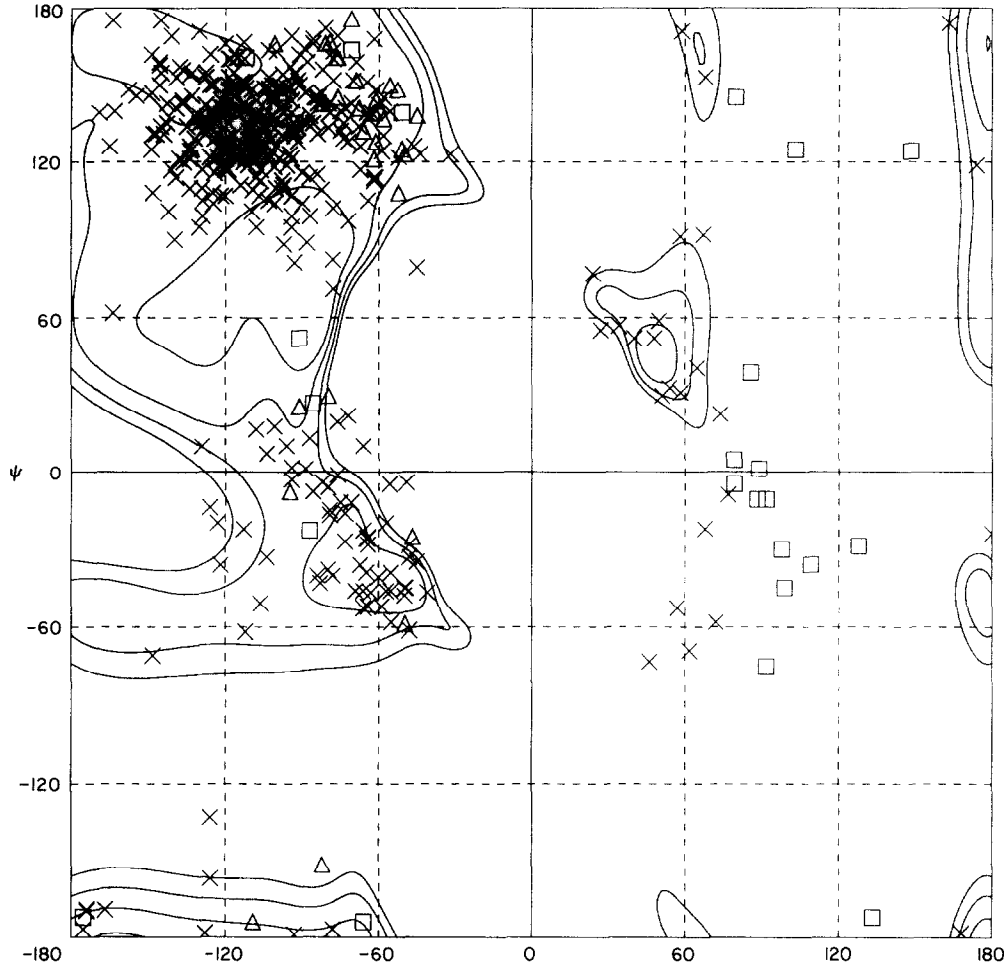

Figure 6. Ramachandran (ϕ, ψ) plots for the AN02 Fab. The contours shown represent the allowed (ϕ, ψ) values of Ramachandran & Sasisekharan (1968) for non-glycine residues. The boxed crosses represent glycine residues. The triangles represent proline residues.

Table 6

Results of superpositions with AN02 backbone atoms

Antibody	V_L	$\%V_L$	V_H	$\%V_H$	C_L	C_H1	C_H1/C_L	L1	L1/F	L2	L2/F	H2	H2/F
REI	0·64	61								0·26	0·56		
HyHel-5	0·65	83	1·13	41	0·75	0·68	0·76	0·47	0·99	0·89	1·2		
J539	0·83	73	0·99	48	0·82			0·81	1·06	0·96	1·72		
McPC603	0·76	60	1·04	42	0·82					0·95	1·09		
KOL	1·01	48	1·17	47									
NEW	1·50	46	1·19	57								0·48	2·19

r.m.s. differences in atomic position (in Å) following superposition of backbone (C^α, C, O and N) atoms of AN02 and Fabs J539 (Suh *et al.*, 1986), McPC603 (Satow *et al.*, 1986) NEW (Poljak *et al.*, 1974), KOL (Marquart *et al.*, 1980), HyHel-5 (Sheriff *et al.*, 1987), and the Bence–Jones protein REI (Epp *et al.*, 1975). The superpositions were carried out using the program TOSS (Hendrickson, 1979). The V_L and V_H domains were superimposed with the hypervariable loops deleted. Also shown are the percentage of matching residues between the AN02 and other antibody-variable domains. For the V_L regions the residues 2 to 28, 33 to 49, 53 to 90 and 98 to 107 were superimposed for the McPC603, J539, HyHel-5 and REI Fabs, and then residues 15 to 25, 33 to 49, 62 to 90 and 98 to 107 were superimposed for the KOL and NEW Fabs. For all V_H regions the residues 2 to 27, 34 to 49, 56 to 95 and 103 to 114 were superimposed. For the C_L domains, all residues were used, while all residues except 126 to 135 (which were disordered in AN02) were used for the C_H1 superpositions. The C_L/C_H1 superposition consisted of simultaneous superposition of the C_L and C_H1 domains. For L1 residues 26 to 32 were superimposed. For L2 residues 50 to 52 were superimposed. For H2 residues 50 to 55 were superimposed. L1/F, L2/F and H2/F refer to superpositions in which the specified CDR loop was superimposed in the context of its variable domain. All numbers refer to the numbering scheme of Kabat *et al.* (1987).

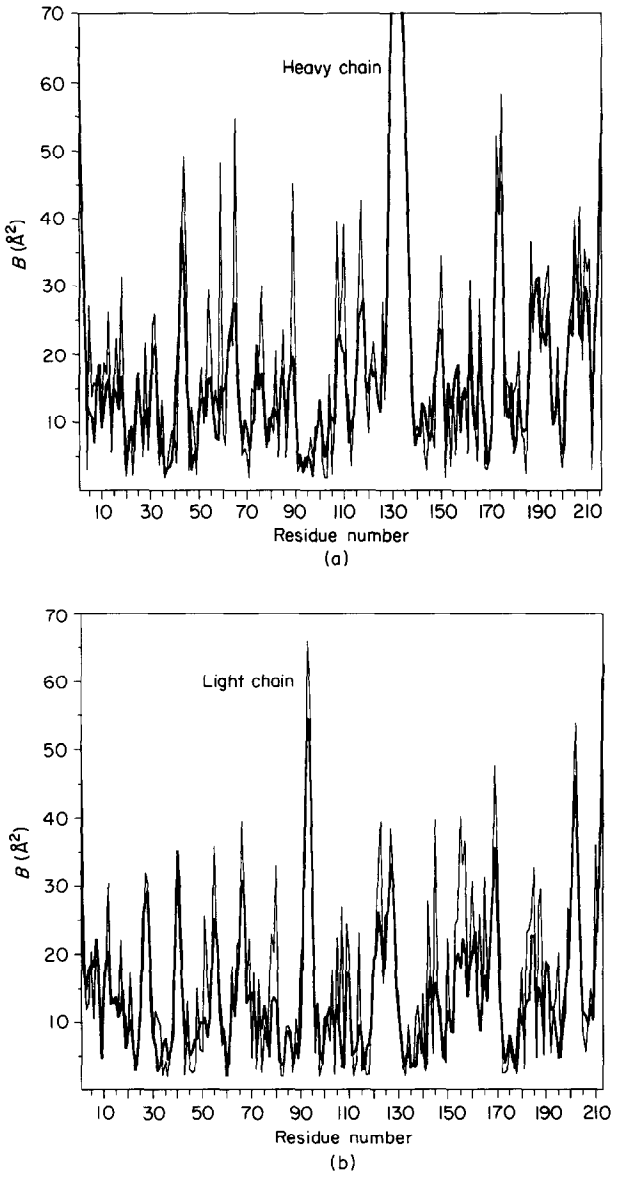

Figure 7. Average main-chain (C, C^{α}, O, N) (thick lines) and side-chain (thin lines) temperature factors *versus* residue number for the heavy (b) and light (a) chains. Glycine "side-chain" temperature factors were set to the main-chain value for that residue. Average values >70 Å are not shown.

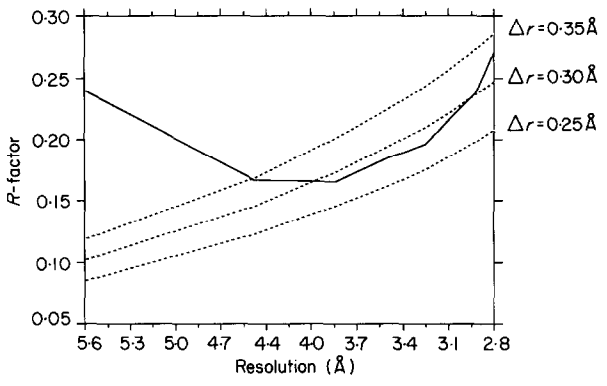

Figure 8. The crystallographic *R*-factor is plotted *versus* resolution for the AN02 Fab (continuous line) superimposed on theoretical curves (Luzzati, 1952) corresponding to different levels of co-ordinate error δr.

scheme of canonical loop structures developed by Chothia *et al.* (1989), and the following discussion refers to their classification system. The length and sequence of L1 from AN02 places it in canonical class 1, and it superimposes well with L1s from HyHel-5 and J539, other members of this class. The deviation between the L1 loops from AN02 and J539 is larger than that between AN02 and HyHel-5 because the peptide bonds following the residue at position 26L are modeled in the opposite orientations (i.e. the ϕ, ψ angles are reversed, corresponding to tight turns of either type I (AN02 and HyHel-5) or I′ (J539) (Richardson, 1981)). It should be noted, however, that the resolution of the AN02 structure is not high enough to allow a completely unambiguous determination of the peptide bond orientation in these loop regions. L2 from AN02 superimposes almost exactly on L2 from REI. The L2 loops from HyHel-5 and McPC603 have essentially the same main-chain conformation as the L2s in REI and AN02 with the exception that the peptide bond following the residue at position 50 is modeled with an opposite orientation. The main-chain conformation of the L2 from J539, however, deviates somewhat from the pattern in REI, AN02, HyHel-5 and McPC603. The deviation is principally due to different dihedral angles about the residues at positions 49 and 50. The cause and significance of this slight deviation are difficult to assess. The AN02 L3 loop is one residue longer than the L3 loop in any of the previously solved structures. This extra residue is characteristic of a small subclass of V_L regions, and the AN02 L3 loop is thus likely to represent a new class of L3 loops. The AN02 L3 loop cannot currently be compared with other L3 loops due to two disordered residues within it, Tyr94 L and Pro95 L, whose conformation is uncertain.

The AN02 H1 loop is one residue longer than the H1 loop in any of the previously solved Fab structures. The extra residue of the AN02 H1 loop is characteristic of certain subclasses of V_H regions and not specific to AN02. The AN02 H1 loop is thus likely to represent the first example of a new (although small) canonical class of H1 loops. Like other H1 loops, a glycine at position 26 produces a sharp turn with ϕ, ψ values (77°, 9°) outside the range allowed for non-glycine residues. Also, like other H1 loops, a residue at position 29 (Ile29 in AN02) is buried deeply in the framework residues, in the case of AN02, against the side-chains of residues 24, 35 and 76 of the heavy chain. The AN02 H1 loop extends through residue 33 with the β-stranded framework beginning again at position 34. The residue at position 34 in AN02 is in a structurally homologous position to the residue at position 33 in the previous, shorter H1 loops. The sequence and length of the AN02 H2 loop places it in canonical class 1, and it superimposes well with H2 of Fab NEW, another member of this class.

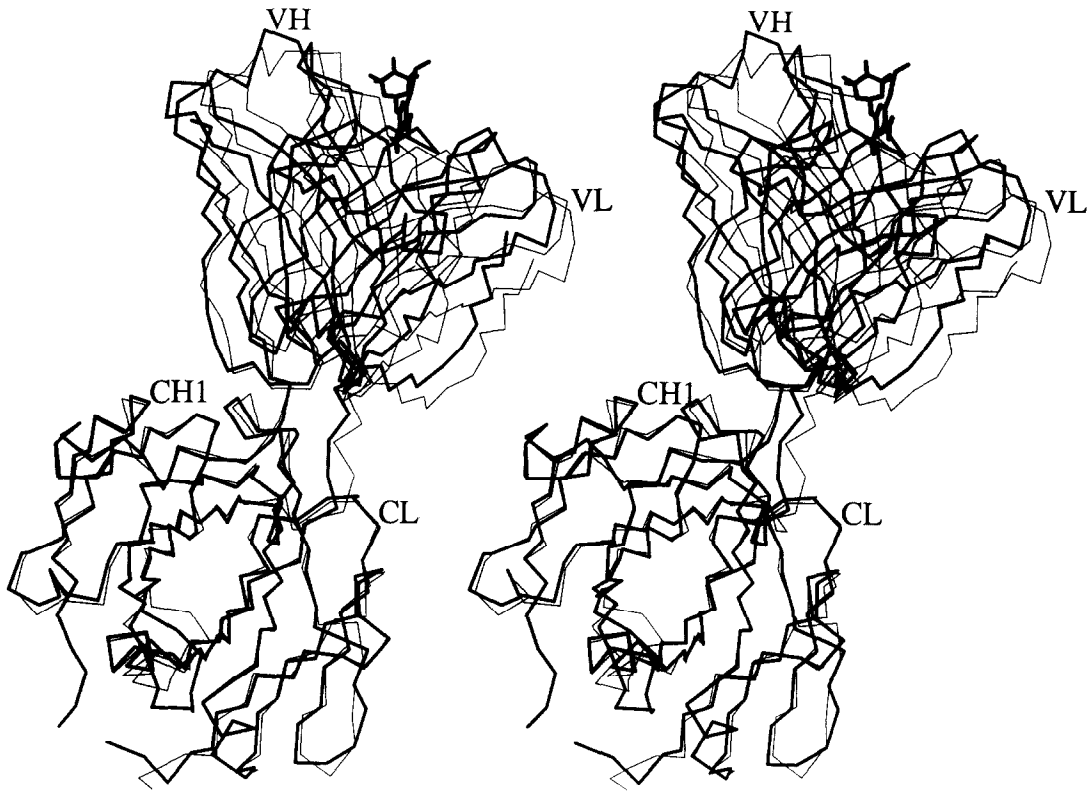

Figure 9. Comparison of AN02 and HyHel-5: shown is the refined AN02 C^{α} backbone trace (thick lines), the bound hapten (thick lines) of the refined AN02 structure, and the C^{α} trace of the HyHel-5 structure (thin lines).

When superimposed in the context of the framework, however, significant deviations between AN02 H2 and NEW H2 occur. Figure 10(a) shows a superposition of the isolated H2 loops of AN02 and NEW, and Figure 10(b) shows the orientation of the same H2 loops when superimposed in the context of the framework residues. Examination of the superimposed structures reveals that, despite similar main-chain traces for the isolated loops, the H2 loops in AN02 and NEW are canted at different angles relative to the framework residues. The axis about which the AN02 and NEW H2 loops rotate relative to one another is formed by the residues at positions 51 and 56. Two residues that appear important for the different H2 loop orientations are at positions 52 and 71 of the heavy chain. AN02 has a serine at position 52, whose side-chain makes a hydrogen bond with the backbone N at position 56. NEW has a phenylalanine residue at position 52, which not only cannot make this hydrogen bond, but also whose side-chain is oriented in the opposite direction. A hydrogen bond that is part of the β-sheet is formed between the backbone N at position 53 and the backbone O at position 33 in AN02. This hydrogen bond is not formed in NEW, and the region where this hydrogen bond would occur in NEW is occupied by the Phe52 side-chain. The absence of this hydrogen bond in NEW allows the H2 loop in NEW to assume a different orientation relative to the framework than found in AN02. Position 71 in AN02 is an arginine residue that makes a contact with the backbone O at position 53 (and also residues in H1) but would interfere with the orientation of H2 found in NEW. NEW has a much smaller residue at this position, valine, which can accommodate the course of the H2 loop found in NEW. Chothia *et al.* (1989) have identified position 71 as important for determining H2 loop conformation. The electron density for the AN02

Table 7

V_H residues which contain atoms within 4 Å of atoms in V_L residues are shown for the AN02 and HyHel-5 Fabs

AN02		HyHel-5	
Light	Heavy	Light	Heavy
		1D	61E
36Y	97P, 98L, 103W	36Y	103W
38Q	39Q	38Q	39Q, 91Y
41G	105Q	42T	91Y
43S	91F, 104G, 103W	43S	91Y, 104G, 103W
44P	103W	44P	91Y, 103W, 44L
46L	98L, 101A	46R	101D, 95G, 96N
		50D	97Y
87Y	45L, 39Q, 43N	87Y	43G, 44L, 39Q
		91W	35E
94Y	59Y, 64R, 47W, 58R	94N	60H, 61E
95(A)P	60N, 61P, 47W	95P	47W
96I	47W		
98F	45L	98F	45L
100V	43N, 44K	99G	44G

The residues are numbered according to Kabat *et al.* (1987) and the standard single letter amino acid designations are used.

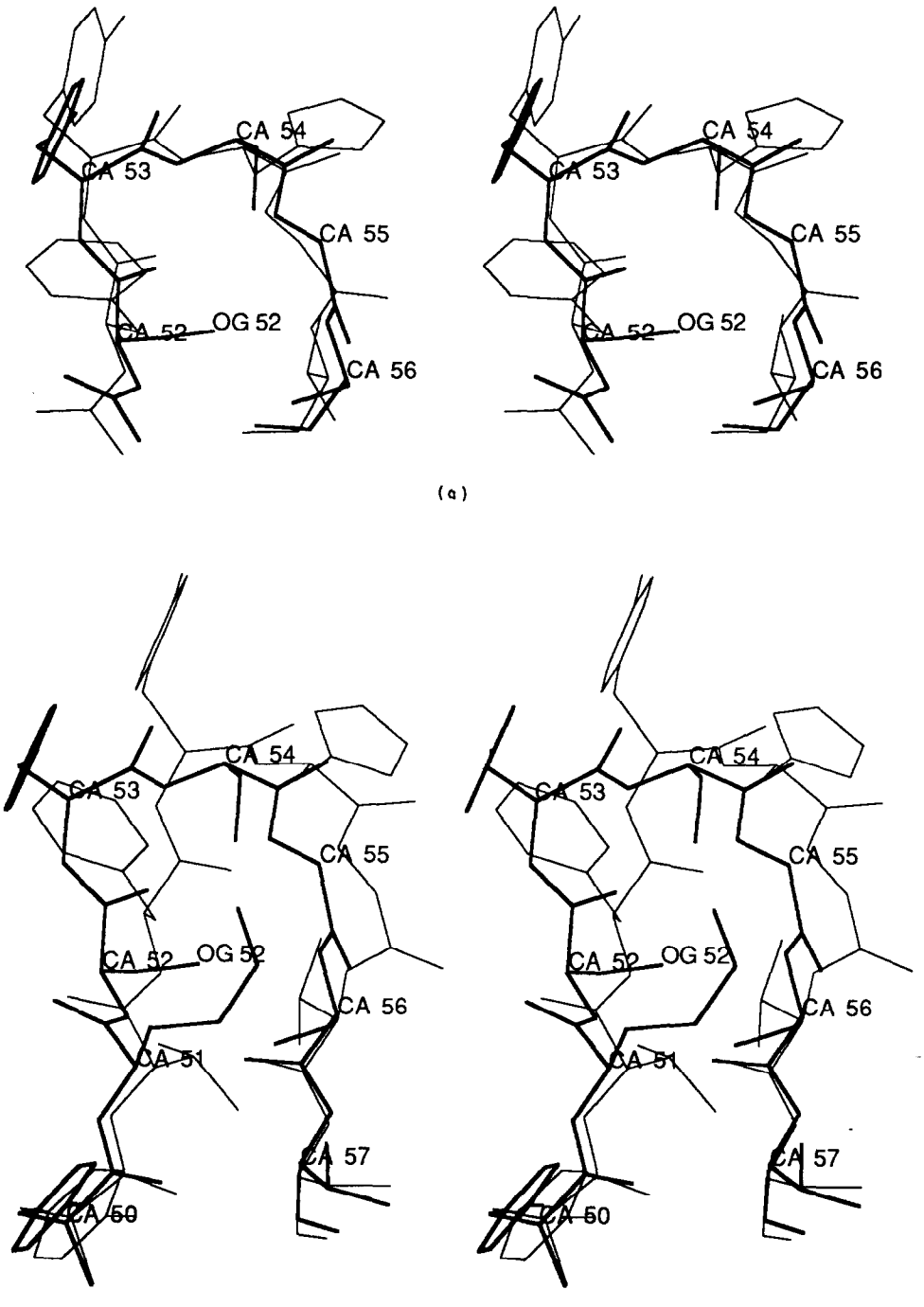

Figure 10. Stereo pairs showing the CDR2 loop of the heavy chains of AN02 and Fab NEW (Poljak *et al.*, 1974) following superposition using the program TOSS (Hendrickson, 1979). In (a) only the CDR2 residues (52 to 55) were aligned, while in (b) the CDR2 residues were aligned in the context of the V_H domain non-CDR residues. In both (a) and (b) the heavy lines represent AN02 and the light lines represent NEW. The numbering is according to Kabat *et al.* (1987).

H2 loop is well defined with *B*-factors ranging from 8·6 to 23·5 Å^2 for main-chain atoms in residues 53 to 56. The *B*-factors for NEW were not available in the Brookhaven Data Bank.

Table 7 lists the residues involved in the V_L/V_H contacts for both AN02 and HyHel-5. The contacts for HyHel-5 are shown for comparison, because while the AN02 and HyHel-5 V_L domains are highly similar (83% of the residues match exactly), the V_H domains are relatively different (41% of the residues match). Table 7 clearly shows that the highly similar V_L regions make analogous contacts to dissimilar V_H regions. As noted earlier, the rotations and translations relating V_L and V_H are very similar for AN02 and HyHel-5.

Several residues in the vicinity of the binding site are involved in lattice contacts and may lead to differences between the binding site structure in the

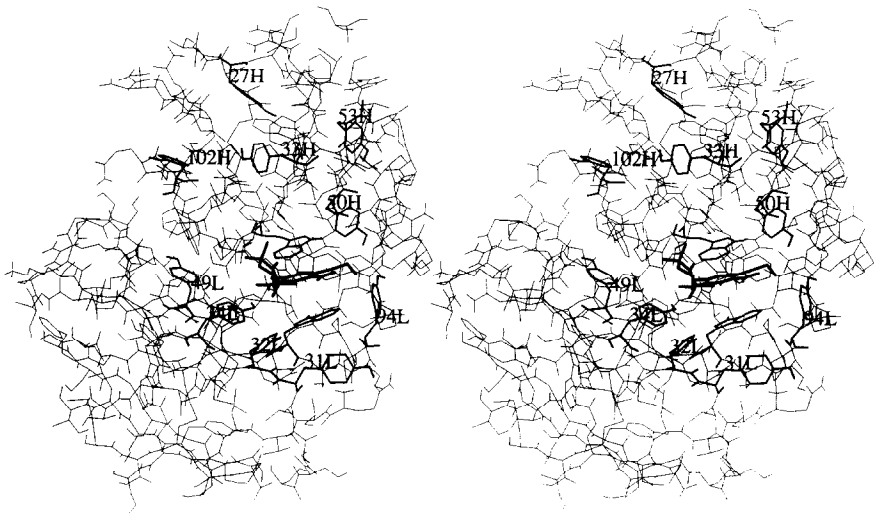

Figure 11. Stereo pair showing the AN02 combining site. Shown are all model elements within 20 Å of the DNP hapten. The 2 tryptophan residues stack on the DNP. The DNP and tyrosine residues on the surface of the binding site are labeled. The numbering is according to Kabat *et al.* (1987); H and L refer to the heavy and light chains, respectively. The DNP hapten and the tyrosine residues are highlighted by thick lines.

crystal and in solution. Of particular note, Tyr31 L stacks on Tyr31 L from a symmetry-related molecule with the planes of the aromatic rings parallel to one another. The aromatic ring of Tyr31 L forms a perpendicular contact with the indole of Trp91 L. The opposite side of the Trp91 L stacks directly on the DNP ring of the hapten (see Fig. 11). The exact influence of the lattice contacts on the binding site structure is difficult to assess directly but may complicate comparison of n.m.r. and X-ray results.

The hapten is bound in a pocket formed principally by L3 and H3 and to a lesser extent by L1 and L2. The pocket is approximately 6 Å wide, 12 Å long and 8 Å deep. The DNP ring of the hapten is sandwiched between the indole side-chains of Trp91 L and Trp96 H as shown in Figures 11 and 4. The DNP ring appears to stack more directly over the five-membered ring of Trp96 H with the six-membered ring of Trp96 H extended between the DNP nitro groups. The indole side-chain of Trp91 L has an orientation opposite to that of Trp96 H. The five-membered ring of the Trp91 L indole extends out between the two nitro groups with the six-membered ring more directly over the DNP ring. In addition to the two tryptophan residues, five other residues come into close contact (within 4 Å) with the hapten. Asp50 L contacts the middle of the 2,2,6,6-tetramethyl-1-piperindinyloxy (TEMPO) ring. Tyr32 L and Pro97 H contact methyl groups on opposite sides of the TEMPO moiety. Tyr34 L and Gln89 L both make contacts with the 2-nitro group of the DNP ring. The ethylene diamine group at position 5 of the DNP ring is oriented away from the pocket and into the solvent. For immunization, the hapten was coupled to a carrier protein, KLH, *via* an amine link at position 5 and, thus, the orientation of this side of the hapten towards the solvent and away from the antibody was not unexpected. While the electron density for the hapten is quite good, the isotropic temperature factors for the hapten are higher than the temperature factors for neighboring residues. The average *B*-factor for the side-chain atoms of Trp91 L and Trp96 H are 18·2 Å^2 and 11·8 Å^2, respectively, while the average *B*-factors for the DNP and TEMPO rings are 22·9 Å^2 and 37·4 Å^2, respectively.

Throughout refinement, two regions displayed consistently uninterpretable electron density, and, ultimately, high *B*-factors. One of these regions, residues Pro126 to Asn135 H (heavy chain, see Fig. 7), consists of an exposed portion of the C_H1 domain. This region also displayed significantly higher *B*-factors in the HyHel-5 search model. The other disordered residues, Tyr94 L and Pro95 L, form the carboxyl-terminal end of L3 and are thus of greater interest. Examination of $(2|F_{obs}|-|F_{calc}|)e^{i\phi_{calc}}$ and $(|F_{obs}|-|F_{calc}|)e^{i\phi_{calc}}$ "annealed" omit maps from which these and surrounding residues had been omitted from SA-refinement failed to reveal convincing electron density throughout all stages of refinement. As the refinement progressed, however, likely density for the main-chain atoms became more apparent. Following several attempts to rebuild these residues, their ϕ, ψ values were built with the most sterically favored values (Cantor & Schimmel, 1980), consistent with the observed electron density, and left unmodified after conjugate gradient minimization. The *B*-factors for these two residues are greater than 55 Å^2.

As shown in Course of Structure Solution and Refinement (section 3), *PC*-refinement of the overall orientation of the search model and of the orientations and positions of the four Fab domains was essential to solve the structure. This immediately poses the question of how different the interdomain angles of the search model could have been from the

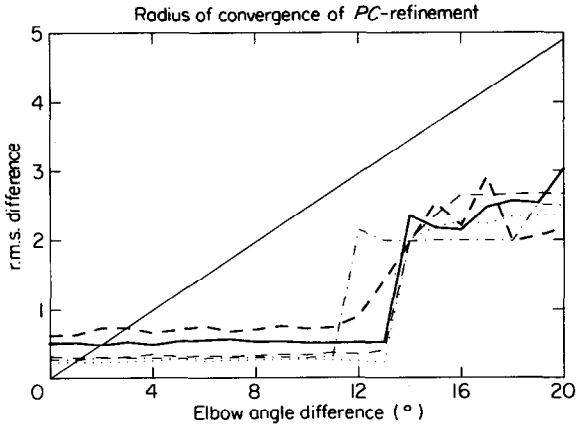

Figure 12. Atomic r.m.s. differences for (N, C, C^{α}) backbone atoms before (continuous line) and after *PC*-refinement as a function of the elbow-angle difference between the artificially modified (see the text) AN02 structure and the correct AN02 structure. The *PC*-refinements were carried out at 15 to 5 (— — —), 15 to 4·5 (——), 15 to 4 (— — —), 15 to 3·5 (—·—·—), and 15 to 3 Å (·····) resolution.

crystal structure in order to still solve the structure. Another way to phrase this question is to ask about the radius of convergence of *PC*-refinement. A detailed assessment of this question is in progress. Here we simply present the radius of convergence of *PC*-refinement as a function of the elbow angle by carrying out *PC*-refinements of the artificially modified AN02 structure. The elbow angle of the AN02 structure was made smaller by rotating the C_L, C_H1 dimer around the axis that intersects the C^{α} positions of residues 106 of the light chain and 116 of the heavy chain, while keeping the V_H, V_L dimer stationary. Figure 12 shows the r.m.s. differences of the modified AN02 structure before and after *PC*-refinement from the initial AN02 structure. It appears that *PC*-refinement is capable of undoing the artificial elbow angle change up to 13°. Thus, the search model could have had maximally a 13° difference in elbow angle from the crystal structure for a successful solution by generalized molecular replacement.

5. Discussion

The solution of the AN02 Fab crystal structure demonstrates the usefulness of generalized molecular replacement (Brünger, 1990*a*). The success of *PC*-refinement at establishing the correct overall orientation of the search model and at finding the correct orientations and positions of the four Fab domains was instrumental in identifying the correct solution in subsequent translation searches. While we cannot completely exclude the possibility that the structure could have been solved solely by conventional rotation and translation functions, our results suggest that it would have been a very difficult case. This also applies to the use of the C_H1 and C_L dimer as a search model presumably because of the high crystal symmetry. *PC*-refinement of the orientations and positions of the Fab domains has made the solution of the AN02 Fab an almost straightforward process. The Patterson correlation refinement will expand the solution of crystal structures by the molecular replacement method. The solution of multidomain structures, such as Fabs, whose domains have flexible orientations relative to one another will be especially benefited. In fact, several other structures have been solved by generalized molecular replacement (Brünger, 1991; Brünger *et al.*, 1990*b*).

The crystal structure of the AN02 Fab confirms the result of earlier n.m.r. experiments. Both methods reveal eight to 10 (depending on how broadly the "combining site" is defined) exposed tyrosine residues in the combining site (Anglister *et al.*, 1984*b*; Fig. 11). Examination of antibody sequences reveals a large number of tyrosine residues to be a general feature of antibody CDR regions (tyrosine is the most commonly occurring residue in heavy chain CDRs and second only to serine in overall CDR occurrence (Kabat *et al.*, 1987)), suggesting that many tyrosine residues may be advantageous to overall antigen recognition. In the current structure, the exposed tyrosine residues may have facilitated recognition of the hapten through interactions with the carrier protein (KLH) to which the hapten was coupled for immunization. Several exposed tyrosine residues (and other aromatic residues) are also a prominent feature of the proposed HLA antigen binding site (Bjorkman *et al.*, 1987).

Germline-variable region genes from which the AN02 heavy and light chains are likely to have been derived have been cloned and sequenced (Theriault *et al.*, 1990). Table 8 lists the residues which are different in the AN02 germline and somatic sequences. Examination of these differences reveals that the majority of the changed residues are distant from the antigen binding site and that none of the changed residues are in direct contact with the hapten. Tyr31 L (originally a serine in the germline sequence) stacks perpendicularly on Trp91

Table 8

A listing of residues that differ between the AN02 germline and somatic sequences (Theriault et al., *1990)*

Residue	Chain	Germline	AN02
31	Light	Ser	Tyr
100	Light	Ala	Val
18	Heavy	Leu	Gln
51	Heavy	Ile	Met
58	Heavy	Ser	Arg
64	Heavy	Lys	Arg
82A	Heavy	Gln	Lys
91	Heavy	Tyr	Phe
108	Heavy	Leu	Gln
110	Heavy	Thr	Ser
113	Heavy	Ala	Glu

The number is according to Kabat *et al.* (1987).

L (Trp91 L stacks on the DNP ring of the hapten) and may thus be important in determining a binding site conformation favorable to hapten binding (see Fig. 4). As noted earlier, Tyr31 L forms a crystal lattice contact which complicates assessment of its role in the binding site in solution. Arg58 H and Arg64 H (originally serine and lysine in the germline sequence, respectively) are on the surface of the protein and approximately 7 and 10 Å away, respectively, from the hapten at closest approach. It does not appear that changing either of these residues would affect hapten binding. Changes at these positions, however, could have been important for improving the affinity of AN02 for the carrier protein to which the hapten was coupled for immunization. Phe91 H (originally a tyrosine in the germline sequence) is buried in the framework region at the V_L/V_H interface (see Table 8). This change does not appear to greatly affect the V_L/V_H interaction (and thus indirectly the binding site) because Phe91 H and its neighboring residues in AN02 have virtually identical conformations to Tyr91 H and its neighboring residues in HyHel-5. Met51 H (originally isoleucine in the germline sequence) may affect the conformation of the H2 loop (53 to 56 H) (see Fig. 10(a)) by virtue of its proximity, but, as noted earlier, other residues seem more important for determining the H2 loop conformation and the H2 loop does not contact the hapten in AN02. The remaining six differences between the germline and somatic sequences all occur far removed from the combining site or the V_L/V_H interface and do not seem likely to influence the combining site region. This assessment of the germline to somatic changes in AN02 is essentially the same as that described by Theriault *et al.* (1990), who made their evaluation based on a calculated model of AN02.

Anti-DNP antibodies have long been observed to cause a red shift in the ultraviolet spectrum of DNP following DNP binding to antibody (Little & Eisen, 1967). This shift has been interpreted as resulting from a charge transfer interaction between DNP and antibody tryptophan residues. Such a red shift is seen following hapten binding to AN02, and a nuclear magnetization transfer was observed between a diamagnetic DNP hapten and two distinct tryptophan residues, one each on the heavy and light chains. The light chain tryptophan was previously identified as Trp91 L (Anglister *et al.*, 1987). The crystal structure shows the DNP ring stacking between Trp91 L and Trp96 H consistent with the earlier observations. Anglister *et al.* (1987) specifically assigned a NOE to an interaction between proton 5 on Trp91 L and proton 6 of a diamagnetic hapten. As the Fab used in the NOE study contained tryptophan deuterated so that only positions 2 and 5 on the indole side-chain remained protonated, proton 5 is the only proton which could have given rise to the observed NOE according to the crystal structure. Dwek *et al.* (1977) also found stacking of a DNP hapten with antibody tryptophan residues using magnetic resonance techniques. While the current resolution (2·9 Å) of this structure does not allow precise statements about the indole-DNP-indole stacking, it is clear that the indole rings are shifted off the center of the DNP stacking axis, as expected for non-zero overlap between charge donor and acceptor (Mullikan & Person, 1969) and seen, for example, in the trinitrobenzene-anthracene (Brown *et al.*, 1964) and trinitrobenzene-indole complexes (Hanson, 1964). A more detailed comparison of the AN02 crystal structure and n.m.r. results presented in this volume (Theriault *et al.*, 1991).

The structure of the AN02 CDR loops are generally consistent with the scheme of canonical loop structures developed by Chothia *et al.* (1989). The orientation of the peptide bond between residues in tight turns appears variable in otherwise similar loops, but does not greatly alter the overall trace of the loops' backbone atoms. This variability may result from the ambiguity in the electron density of highly mobile regions in structures at moderate resolution. The different orientations of the AN02 and NEW H2 regions relative to the framework residues despite highly similar isolated structures indicate that attention should be paid to the orientation of canonical loop structures relative to the framework when modeling unknown structures.

This work was supported by the American Cancer Society (R.O.F., grant NP 631), the Office of Naval Research (H. M. McConnell grant N00014-90-J-1407), the NSF funded Pittsburgh Supercomputer Center (A.T.B., grant DMB890008P). Calculations were also carried out on the CONVEX C-220 of the Yale Center for Structural Biology. We thank W. A. Hendrickson, W. I. Weis and H. M. McConnell for helpful discussions and W. A. Hendrickson for generous use of the computing facilities at Columbia. We thank J. Johnson and M. Rossmann for access to the film processing facility at Purdue University and for helpful advice. D.J.L. is a fellow of the Helen Hay Whitney Foundation.

References

Anglister, J., Frey, T. & McConnell, H. M. (1984*a*). Magnetic resonance of a monoclonal anti-spin-label antibody. *Biochemistry*, **23**, 1138–1142.

Anglister, J., Frey, T. & McConnell, H. M. (1984*b*). Distances of tyrosine residues from a spin-label hapten in the combining site of a specific monoclonal antibody. *Biochemistry*, **23**, 5372–5375.

Anglister, J., Frey, T. & McConnell, H. M. (1985). NMR technique for assessing contributions of heavy and light chains to an antibody combining site. *Nature (London)*, **315**, 65–67.

Anglister, J., Bond, M. W., Frey, T., Leahy, D. J., Levitt, M., McConnell, H. M., Rule, G. S., Tomasello, J. & Whittaker, M. (1987). Contribution of tryptophan residues to the combining site of a monoclonal anti-dinitrophenyl spin-label antibody. *Biochemistry*, **26**, 6058–6064.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature (London)*, **329**, 512–518.

Brown, D. S., Wallwork, S. C. & Wilson, A. (1964). Molecular complexes exhibiting polarizing bonding. IV. The crystal structure of the anthracene-s-trinitro benzene complex. *Acta Crystallogr.* **17**, 168–176.

Brünger, A. T. (1990*a*). Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Crystallogr. sect. A*, **46**, 46–57.

Brünger, A. T. (1990*b*). *X-PLOR Manual*, Version 2.1, Yale University, New Haven, CT, U.S.A.

Brünger, A. T. (1991). Solution of a Fab (26-10)/Digoxin complex by generalized molecular replacement. *Acta Crystallogr. sect. A*, **47**, 195–204.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic *R* factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Brünger, A. T., Krukowski, A. & Erickson, J. (1990*a*). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. sect. A*, **46**, 585–593.

Brünger, A. T., Milburn, M. V., Tong, L., DeVos, A. M., Jancarik, J., Yamaizumi, Z., Nishimura, S., Ohtsuka, E. & Kim, S.-H. (1990*b*). Crystal structure of an active form of *ras* protein, a complex of GTP analog and c-H-*ras* P21 catalytic domain. *Proc. Nat. Acad. Sci., U.S.A.* **87**, 4849–4853.

Cantor, C. R. & Schimmel, P. R. (1980). Results of potential energy calculations. In *Biophysical Chemistry*, part I, pp. 269–272, W. H. Freeman & Co., San Francisco.

Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917.

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). Conformations of immunoglobulin hypervariable regions. *Nature (London)*, **342**, 877–883.

Crowther, R. A. (1972). The fast rotation function. In *The Molecular Replacement Method* (Rossmann, M. G., ed.), Int. Sci. Rev. Ser., no. 13, pp. 173–178, Gordon & Breach, New York.

Cygler, M. & Anderson, W. F. (1988). Application of the molecular replacement method to multidomain proteins. 1. Determination of the orientation of an immunoglobulin Fab fragment. *Acta Crystallogr. sect. A*, **44**, 38–45.

Dwek, R. A., Wain-Hobson, S., Dower, S., Gettins, P., Sutton, B., Perkins, S. J. & Givol, D. (1977). Structure of an antibody combining site by magnetic resonance. *Nature (London)*, **266**, 31–37.

Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. (1975). The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2·0-Å resolution. *Biochemistry*, **14**, 4943–4952.

Fujinaga, M. & Read, R. J. (1987). Experiences with a new translation-function program. *J. Appl. Crystallogr.* **20**, 517–521.

Hanson, A. W. (1964). The crystal structure of the 1 : 1 complexes of skatole and indole with *s*-trinitrobenzene. *Acta Crystallogr.* **17**, 559–564.

Hauptman, H. (1982). On integrating the techniques of direct methods and isomorphous replacement. I. The theoretical basis. *Acta Crystallogr. sect. A*, **38**, 289–294.

Hendrickson, W. A. (1979). Transformations to optimize the superposition of similar structures. *Acta Crystallogr. sect. A*, **35**, 158–163.

Hoppe, W. (1957). Die 'Faltmolekülmethode'—eine neue Methode zur Bestimmung der Kristallostruktur bei ganz oder Teilweise bekannter Molekülstruktur. *Acta Crystallogr.* **10**, 750–751.

Huber, R. (1985). Experience with the application of Patterson search techniques. In *Molecular Replacement, Proceedings of the Daresbury Study Weekend, Daresbury, February 1985*, pp. 58–61, Science and Engineering Research Council, The Librarian, Daresbury Laboratory, Daresbury, United Kingdom.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.

Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987). *Sequences of Proteins of Immunological Interest*, 4th edit., National Institutes of Health, Bethesda, MD.

Kuriyan, J., Karplus, M. & Petsko, G. A. (1987). Estimation of uncertainties in X-ray refinement results by use of perturbed structures. *Proteins: Struct. Funct. Genet.* **2**, 1–12.

Lattman, E. E. (1972). Optimal sampling of the rotation function. *Acta Crystallogr. sect. B*, **28**, 1065–1068.

Leahy, D. J. (1988). Construction of a system to study the antibody combining site by NMR. Ph.D. thesis, Stanford University.

Leahy, D. J., Rule, G. S., Whittaker, M. M. & McConnell, H. M. (1988*a*). Sequences of 12 monoclonal anti-dinitrophenyl spin-label antibodies for NMR studies. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 3661–3665.

Leahy, D. J., Hynes, T. R., McConnell, H. M. & Fox, R. O. (1988*b*). Crystallization of an anti-tempo-dinitrophenyl monoclonal antibody Fab fragment with and without bound hapten. *J. Mol. Biol.* **203**, 829–830.

Little, J. R. & Eisen, H. N. (1967). Evidence for tryptophan in the active sites of antibodies to poly-nitrobenzenes. *Biochemistry*, **6**, 3119–3125.

Luzzati, P. V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr.* **5**, 802–810.

Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980). Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen binding fragment at 3·0 Å and 1·9 Å resolution. *J. Mol. Biol.* **141**, 369–391.

Martin, A. C. R., Cheetham, J. C. & Rees, A. R. (1989). Modeling antibody hypervariable loops: a combined algorithm. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 9268–9272.

Mullikan, R. S. & Person, W. B. (1969). *Molecular Complexes*, pp. 53–58. Wiley-Interscience, New York.

Nordman, C. E. (1980). Procedures for detection and idealization of non-crystallographic symmetry with application to phase refinement of the satellite tobacco necrosis virus structure. *Acta Crystallogr. sect. A*, **36**, 747–754.

Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J. & Davies, D. R. (1989). Structure of an antibody–antigen complex: crystal structure of the HyHel-10 Fab-lysozyme complex. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 5938–5942.

Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1974). The three-dimensional structure of the Fab-fragment of a human myeloma immunoglobulin at 2·0-Å resolution. *Proc. Nat. Acad. Sci., U.S.A.* **71**, 3440–3444.

Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Math. Progr.* **12**, 241–254.

Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23**, 283–437.

Rao, S. N., Jih, J.-H. & Hartsuck, J. A. (1980). Rotation-function space groups. *Acta Crystallogr. sect. A*, **36**, 878–884.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Rossmann, M. G. (1985). Determining the intensity of Bragg reflections from oscillation photographs. *Methods Enzymol.* **114**, 237–280.

Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24–31.

Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2·7 Å. *J. Mol. Biol.* **190**, 593–604.

Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. & Davies, D. R. (1987). The three-dimensional structure of an antibody–antigen complex. *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8075–8079.

Suh, S.-W., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., Rudikoff, S. & Davies, D. R. (1986). The galactan-binding immunoglobulin Fab J539: an X-ray diffraction study at 2·6-Å resolution. *Proteins: Struct. Funct. Genet.* **1**, 74–80.

Theriault, T., Rule, G. S. & McConnell, H. M. (1990). Specificities of germ line antibodies. In *Protein Structure and Engineering* (Jardetzky, O. & Holbrook, R., eds), pp. 367–376, Plenum Press, New York and London.

Theriault, T. P., Leahy, D. J., Levitt, M., McConnell, H. M. & Rule, G. S. (1991). Structural and kinetic studies of the Fab fragment of a monoclonal anti-spin label antibody by nuclear magnetic resonance. *J. Mol. Biol.* **221**, 257–270.

Weis, W. I., Brünger, A. T., Skehel, J. J. & Wiley, D. C. (1990). Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **212**, 737–761.

Edited by R. Huber