

Refined Three-Dimensional Structure of the Fab Fragment of a Murine IgG1, λ Antibody

BY T. BIZEBARD

Laboratoire de Biologie Structurale, UMR 9920, CNRS, Bâtiment 34, 91198 Gif-sur-Yvette CEDEX, France

R. DANIELS

NIMR, The Ridgeway, Mill Hill, London NW7 1AA, England

R. KAHN

LURE, Université Paris-Sud, 91405 Orsay CEDEX, France

B. GOLINELLI-PIMPANEAU

Laboratoire de Biologie Structurale, UMR 9920, CNRS, Bâtiment 34, 91198 Gif-sur-Yvette CEDEX, France

J. J. SKEHEL

NIMR, The Ridgeway, Mill Hill, London NW7 1AA, England

AND M. KNOSSOW*

Laboratoire de Biologie Structurale, UMR 9920, CNRS, Bâtiment 34, 91198 Gif-sur-Yvette CEDEX, France

(Received 14 June 1993; accepted 11 February 1994)

Abstract

We report the cDNA sequence determination and the crystal structure of the Fab fragment of a murine IgG1, λ antibody (HC19), specific for an influenza virus hemagglutinin. The HC19 Fab-fragment structure has been refined; the crystallographic *R* factor is 19.5% at 2.3 Å resolution. We have compared the conformation of HC19 complementarity determining regions (CDRs) with those of CDR loops of Fab structures available from the Protein Data Bank. These loops were chosen based on the identity of key residues, following the canonical-structure approach; four CDRs have a main-chain conformation very similar to the canonical structure that had been identified. HC19 L1 CDR adopts a conformation clearly distinct from all L1 CDRs that belong to a chain of a different class or origin; this is determined by the nature of a few residues at positions in the sequence different from those of key residues in other light chains. This canonical structure should be representative of most murine λ -class light chains, as inferred from the very high sequence homologies of these polypeptides.

Introduction

Immunoglobulins are made up of four polypeptide chains, two light (L) chains and two heavy (H) chains. Each chain consists of compact domains, approximately

110 amino acids long, connected by elongated segments. All immunoglobulin domains are built on a conserved structural framework consisting of two antiparallel β -sheets linked by a disulfide bridge. Structural studies have concentrated mostly on Fab fragments which consist of the entire light chain and the two NH₂-terminal domains of the heavy chain; these fragments are easily prepared and retain all the binding properties of the complete molecule. The sequences of the NH₂-terminal domains of both chains vary most from one immunoglobulin molecule to the other and are, therefore, labelled V_L and V_H (V, variable); strongly associated by non-covalent interactions, they constitute what has been called the Fv dimer.

Most of the antigen affinity and binding specificity of the antibody molecule comes from six loops of the polypeptide chain, of highly variable length and sequence (complementarity determining regions, CDRs), which are supported by the conserved super-secondary structure of the variable domains [for a precise definition of these loops, see Chothia *et al.* (1986)]. Three loops come from the light chain (labelled L1, L2 and L3) and three come from the heavy chain (H1, H2 and H3). Prediction of the structure of antibody combining sites is of central importance in the design of antibodies with a predefined specificity. Since this specificity depends only on limited portions of the polypeptide chain embedded in a well defined framework, its engineering appears to be an easier objective to reach than the *de novo* design of a protein.

* Author for correspondence.

As sequence information for specific antibodies is now readily available, understanding the sequence-structure relationship for the CDRs has become the focus of renewed attention; investigators have used both theoretical (Fine, Wang, Shenkin, Yarmush & Levinthal, 1986) and semi-empirical approaches (Chothia & Lesk, 1987; Martin, Cheetham & Rees, 1989) to tackle this problem. Among the latter approaches, Chothia & Lesk have examined the structural data available on Fab molecules and identified key residues in known Fab structures which, through their packing, hydrogen bonds or particular main-chain conformation, are responsible, together with the CDR length, for a given CDR main-chain conformation. An analysis of immunoglobulin sequences has shown that these key residues are found in many immunoglobulins and suggested that five of the six CDR loops (excluding H3) adopt a small number of main-chain conformations called canonical structures. Such a structural analysis has helped assessing the effect of mutations on a given combining-site structure in the engineering of an antibody in which all the CDRs of a specific murine antibody had been grafted onto a human framework (Riechmann, Clark, Waldmann & Winter, 1988).

The 'canonical-structure' approach is regularly being tested by new Fab or Fv structures that are determined directly (Steipe, Plückthun & Huber, 1992). Following this approach, we will compare the observed HC19 CDR conformations with those which could have been inferred from structures available in the Brookhaven Protein Data Bank (PDB; Bernstein *et al.*, 1977). Immunoglobulin HC19 (IgG1, λ of murine origin) light chain belongs to a class for which no structural information is presently available from the PDB. Among the light-chain CDRs, the L1 loop is the longest and its conformation is, therefore, the most problematic to predict. Indeed, the HC19 L1 loop conformation differs from those of homologous loops in PDB structures. After completion of this work, Wu & Cygler (1993) have reported the structure of the Fab fragment of another murine IgG1, λ (Se155-4) and described the conformation of its L1 loop; we compare this conformation with the one found in HC19 and assess the relevance of the new common canonical structure to those of other L1 loops of murine λ chains of known sequence.

The HC19 Fab structure has been determined in the course of a study of the molecular mechanisms that allow influenza viruses to escape neutralization by the humoral immune response. HC19 immunoglobulin neutralizes influenza virus X31 (a laboratory-adapted variant of strain Hong Kong/68/H3N2) and is specific for its hemagglutinin (HA). Together with the present determination, knowledge of the structures of X31 HA (Weis, Brünger, Skehel & Wiley, 1990) and of a mutant that escapes neutralization by HC19 (Rigolet, 1991) should allow insight into molecular-recognition events associated with infectivity neutralization and antigenic variation.

Materials and methods

Structural comparisons

The antibody residue numbering adopted here is that proposed by Chothia & Lesk (1987). Representative structural data for comparisons have been taken from the Brookhaven Protein Data Bank using the following entries: 2FB4 (Fab KOL; Marquart, Deisenhofer, Huber & Palm, 1980), 2FBJ (Fab J539; Suh *et al.*, 1986); 2HFL (Fab HyHEL5; Sheriff *et al.*, 1987); 2MCP (Fab McPC603; Segal *et al.*, 1974); 3HFM (Fab HyHEL10; Padlan, Silverton, Sheriff, Cohen, Smith-Gill & Davies, 1989); 1FDL (Fab D1.3; Fischmann *et al.*, 1991); and 7FAB (Fab' New; Saul & Poljak, 1992). The coordinates of Fab Se155-4 were kindly provided by M. Cygler (personal communication).

All superpositions of atoms were calculated using the algorithm of Kabsch (1976), as programmed by P. Alzari (personal communication).

Preparation, crystallization and molecular-replacement studies

The preparation, crystallization and initial structural determination of Fab fragment HC19 have been described in detail elsewhere (Bizebard *et al.*, 1990; Bizebard, Mauguén, Skehel & Knossow, 1991). Crystal data are: space group $P3_121$; $a = b = 98.9$, $c = 89.3$ Å; $Z = 6$.

Primary sequence determination

RNA extraction. HC19 hybridoma cells were obtained as described by Fazekas de St Groth & Scheidegger (1980). Total RNA (15–20 µg) was extracted from the cytoplasm of $\sim 10^7$ hybridoma cells prepared by cold lysis in a buffer containing 10 mM KCl, 10 mM Tris pH 8, 1 mM MgCl₂, plus 25 units of ribonuclease inhibitor (RNasin from Amersham), followed by three successive phenol extractions and two ether extractions. The nucleic acids were precipitated with 1/16th volume of 2 M NaCl and 3 volumes of ethanol, washed with 70% ethanol, dried and dissolved in 200 µl H₂O.

Polymerase chain reaction. mRNA was copied into cDNA by the following procedure: total RNA (3 µl) was made up to 20 µl in a reaction mixture containing 40 pmol of poly-dT (12–18 nucleotides long), 50 mM Tris pH 8.3, 10 mM MgCl₂, 140 mM KCl, 20 mM DTT, 1 mM of each dNTP ($N = G, C, A$ or T , Pharmacia), 10 units of RNasin and 30 units of avian myeloblastosis virus reverse transcriptase (RT, Promega). The mixture was heated to 315 K for 90 min and the reaction stopped by increasing its temperature to 353 K for 10 min.

Genes for both immunoglobulin chains were amplified with the polymerase chain reaction (PCR; Saiki *et al.*, 1988) using two oligonucleotide primers: for the light chain 5'-ATGGCCTGGAT(or C)TTCATTAT-3' (corresponding to a consensus mouse λ light-chain leader sequence) and 5'-GAA(or G)ACAG(or T)TCTGCAC(or G)GAGACAGACTCTT-3' (corre-

sponding to mouse CL λ terminal sequence); for the heavy chain 5'-ATGGCTGTCC(or T)TA(or G)GC(or G)GCTA(or G)CTCTTCTGC-3' (corresponding to a heavy-chain leader sequence) and 5'-TATGCA-AGGCTTACAACCACA-3' (corresponding to the G1 immunoglobulin hinge-region sequence). The leader sequences of mouse heavy chains are much more varied than for λ light chains: the one chosen for HC19 was inferred from immunoglobulin nucleotide sequences (Kabat, Wu, Perry, Gottesman & Foeller, 1991) similar to the partially determined HC19 mRNA sequence (for this, we used the protocol described below, except for the nature of the template).

PCR amplification was performed in 100 μ l mixture containing the cDNA-RNA products (1 μ l), 10 mM Tris pH 8.8, 50 mM KCl, 1.5 mM MgCl₂, 0.25 mM of each dNTP, 8 μ g of carageenan, 3 units of Taq (*Thermus aquaticus*) polymerase (Perkin Elmer, Cetus) and 0.2 μ M of each of the two primers. The sample mixture was overlaid with mineral oil, then subjected to 30 cycles of PCR amplification, each cycle consisting of 1 min of denaturation at 367 K, 1 min of annealing at 333 K and 2 min of elongation at 343 K plus a final cycle (333 K, 1 min; 343 K, 10 min) to ensure completion of the synthesized DNA.

The amplified product was purified by preparative agarose gel electrophoresis (1% agarose; 1 μ g ml⁻¹ ethidium bromide) and filtration (Millipore UltraFree-mc, UFC3 0HV 00) according to manufacturers instructions, ethanol precipitated, washed, dried and redissolved in 100 μ l H₂O. One tenth of this purified product was subjected to a second PCR amplification, following the same protocol, and the final DNA product was recovered as previously.

Sequencing. The cDNA was sequenced by the chain-termination method (Sanger, Nicklen & Coulson, 1977) using avian myeloblastosis virus reverse transcriptase (RT). The oligonucleotide primer was end-labelled in a reaction mixture containing 50 mM Tris pH 8, 10 mM MgCl₂, 10 mM β -mercaptoethanol, 4 μ mol γ -³²P-ATP and 10 units of T4 polynucleotide kinase (Pharmacia). The mixture was heated at 310 K for 1 h, then stopped by addition of 1.5 volumes of 4 M ammonium acetate, 1 μ l tRNA (20 mg ml⁻¹) and 3 volumes of ethanol. The labeled primer was then ethanol precipitated, washed, dried and redissolved in 2 μ l H₂O. Prior to the sequencing reaction, cDNA was denatured by incubation at room temperature with 0.4 M NaOH (15 min), then the reaction was neutralized by adding 0.4 volumes of 5 M ammonium acetate and the cDNA ethanol precipitated, washed, dried and redissolved in H₂O. 2 μ l of template solution (e.g. 0.10 pmol of denatured cDNA) and 1 μ l of labeled primer (3–5 pmol) were then mixed and heated to 373 K for 5 min. This second denaturation step was stopped by shock-freezing the mixture (193 K).

Sequencing reactions were carried out with 3 μ l of the resulting primer/template solution plus 7 μ l of a

solution 107 mM Tris pH 8.3, 21 mM MgCl₂, 114 mM KCl, 42 mM DTT, and containing 10 units of RT. This mixture was divided into four aliquots (2 μ l each) and to each one was added 1 μ l of a solution 0.15 mM ddXTP (X = G, A, C or T, Pharmacia), 1.2 mM dATP, 1.2 mM dGTP, 1.2 mM dCTP, 1.2 mM dTTP. The four mixtures were incubated at 315 K for 30 min, and terminated by addition of 5 μ l of a solution containing 80%(v/v) formamide, 50 mM Tris-borate pH 8.3, 1 mM EDTA, 0.1%(w/v) xylene cyanol, 0.1%(w/v) bromophenol blue; the mixture was heated to 353 K for 10 min and 2–3 μ l were analyzed on a 6%(w/v) acrylamide gel.

Sequence information determined using the four initial primers allowed the design of 11 additional primers to continue the sequencing and remove ambiguities. The HC19 amino-acid sequence is given in Fig. 1.

Data collection

A native data set to 2.3 Å resolution, consisting of 68 753 measurements for 21 391 unique reflections, was collected from two crystals on the multiwire area detector MARK II (Kahn, Fourme, Bosshard & Saintagne, 1986), at the D23 workstation of the LURE synchrotron (Orsay, France). The raw data were integrated using a profile-fitting algorithm (Kabsch, 1988), as implemented in the MADNES software (Messerschmidt & Pflugrath, 1987). The intensity measurements were further processed using the programs ROTAVATA and AGROVATA and reduced to structure factors by the program TRUNCATE; all of these are part of the CCP4 program suite (SERC Daresbury Laboratory, 1979).

The R_{merge} on intensities was 5.7% [$R_{\text{merge}} = \Sigma(I - \langle I \rangle) / \Sigma I$]. Of the possible unique reflections, 93.6% were measured (with 90.4% more than 2 r.m.s.d. above the background). The last resolution shell from 2.42 to 2.3 Å

GLN ALA VAL VAL THR GLN GLU SER ALA LEU THR THR SER PRO GLY GLU THR VAL
THR LEU THR CYS ARG SER SER THR GLY ALA VAL THR THR SER ASN TYR ALA ASN
TRP VAL GLN GLU LYS PRO ASP HIS LEU PHE THR GLY LEU ILE GLY GLY THR ASN
ASN ARG ALA PRO GLY VAL PRO ALA ARG PHE SER GLY SER LEU ILE GLY ASP LYS
ALA ALA LEU THR ILE THR GLY ALA GLN THR GLU ASP GLU ALA ILE TYR PHE CYS
ALA LEU TRP TYR SER ASN HIS TRP VAL PHE GLY GLY GLY THR LYS LEU THR VAL
LEU GLY GLN PRO LYS SER SER PRO SER VAL THR LEU PHE PRO PRO SER SER GLU
GLU LEU GLU THR ASN LYS ALA THR LEU VAL CYS THR ILE THR ASP PHE TYR PRO
GLY VAL VAL THR VAL ASP TRP LYS VAL ASP GLY THR PRO VAL THR GLN GLY MET
GLU THR THR GLN PRO SER LYS GLN SER ASN ASN LYS TYR MET ALA SER SER TYR
LEU THR LEU THR ALA ARG ALA TRP GLU ARG HIS SER SER TYR SER CYS GLN VAL
THR HIS GLU GLY HIS THR VAL GLU LYS SER LEU SER PRO ALA ASP CYS

(a)

GLN VAL GLN LEU LYS GLU SER GLY PRO GLY LEU VAL ALA PRO SER GLN SER LEU
SER ILE THR CYS THR VAL SER GLY PHE LEU LEU ILE SER ASN GLY VAL HIS TRP VAL
ARG GLN PRO PRO GLY LYS GLY LEU GLU TRP LEU GLY VAL ILE TRP ALA GLY GLY
ASN THR ASN TYR ASN SER ALA LEU MET SER ARG VAL SER ILE SER LYS ASP ASN
SER LYS SER GLN VAL PHE LEU LYS MET LYS SER LEU GLN THR ASP ASP THR ALA
MET TYR TYR CYS ALA ARG ASP PHE TYR ASP TYR ASP VAL PHE TYR TYR ALA MET
ASP TYR TRP GLY GLN GLY THR SER VAL THR VAL SER SER ALA LYS THR THR PRO
PRO SER VAL TYR PRO LEU ALA PRO GLY SER ALA ALA GLN THR ASN SER MET VAL
THR LEU GLY CYS LEU VAL LYS GLY TYR PHE PRO GLU PRO VAL THR VAL THR TRP
ASN SER GLY SER LEU SER SER GLY VAL HIS THR PHE PRO ALA VAL LEU GLN SER
ASP LEU TYR THR LEU SER SER SER VAL THR VAL PRO SER SER THR TRP PRO SER
GLU THR VAL THR CYS ASN VAL ALA HIS PRO ALA SER SER THR LYS VAL ASP LYS
LYS ILE VAL PRO ARG ASP CYS GLY CYS LYS PRO CYS ILE

(b)

Fig. 1. Primary sequence of HC19. (a) Light-chain sequence. (b) Partial heavy-chain sequence. CDRs are underlined.

was complete to 86.2% (75.7% above 2 r.m.s.d.) and its R_{merge} was 15.5%.

Refinement

The structure of Fab fragment HC19 was solved by molecular replacement (Bizebard, Mauguen, Skehel & Knossow, 1991). The initial model was subjected to one cycle of simulated-annealing refinement [as programmed in the *X-PLOR* software (Brünger, 1990)] to 3.5 Å resolution. The crystallographic R factor was 17.6%.

With the sequence of Fab fragment HC19, the model was modified as follows, in order to remove as much model bias as possible: all side chains that were different in the model structure from the actual sequence were truncated to $C\beta$ or $C\alpha$. We also removed all atoms of parts that are structurally very different from one antibody to another: these zones were selected by superposition of $C\alpha$ atoms of Fab fragments of known structures (the comparisons were made between single domains to eliminate quaternary structure differences); we removed those portions that were significantly different in several structures (deviation above 1.5 Å) or were impossible to align because of sequence insertions or deletions. The six CDRs were thus eliminated, as well as the connecting regions between V and C domains and a few other loops which do not belong to the combining site.

The resulting model consisted of 2533 non-H protein atoms (as compared to 3261 in the final structure); it was subjected alternately to cycles of conjugate-gradient refinement – first up to 2.8 Å resolution, then extended to 2.3 Å – using the program *PROLSQ* (Hendrickson, 1985) and to stages of interactive model building on an Evans & Sutherland graphics terminal using the program *FRODO* (Jones, 1978) and weighted $2F_o - F_c$ maps (Read, 1986).

As the refinement progressed to an R factor of 31.8%, all missing non-H atoms could be approximately placed into the electron density; the model was then subjected to one cycle of the simulated-annealing protocol of the program *X-PLOR* (final R factor 24.5%), which led to significant improvements in the quality of many ill defined regions of the electron density.

Alternate conjugate-gradient refinement and model building were continued and new parameters were introduced: individual temperature factors, then water molecules. The latter were introduced only after the R factor had reached 23.7% (at 2.3 Å resolution). Solvent molecules were placed into regions of the $F_o - F_c$ difference electron-density map which were more than 4 r.m.s.d. of this map above the mean, in stereochemically plausible positions (*i.e.* at less than 3.4 Å from at least one hydrogen-bond donor/acceptor and making no unfavorable contacts with any protein atom). They were retained only if their temperature factor did not increase above 45 Å² upon further refinement.

Table 1. *Refinement statistics for Fab fragment HC19*

No. of atoms	3261	
No. of added water molecules	91	
<i>R</i> factor (%) (7–2.3 Å)		
for 20 393 reflections with <i>F</i> > 2σ	19.5	
for all 20 603 reflections in that resolution range	19.7	
Average temperature factor (Å ²)	24	
Geometry of the model		
	R.m.s. deviation from ideal values	Target value
Bond distances (Å)	0.012	0.02
Angle distances (Å)	0.043	0.04
Planar 1–4 distances (Å)	0.046	0.05
Planes (Å)	0.013	0.02
Chiral volumes (Å ³)	0.14	0.15
Single-torsion contacts (Å)	0.17	0.3
Multiple-torsion contacts (Å)	0.25	0.3
Hydrogen-bond contacts (Å)	0.16	0.3
ω torsion angles (°)	2.3	3
χ torsion angles (°)	20.7	15
Aromatic torsion angles (°)	17.2	20

The final model characteristics are summarized in Table 1. The model shows clear electron density for most of the molecule, including the six CDRs. The only exceptions are in loops remote from the combining site, the first two residues of the light chain and the last three and four residues of the light and heavy chains, respectively.

Comparison of the observed structure with a model built from canonical structures

We examined the results of an empirical approach in the modelling of an antibody combining site, in the particular case of Fab fragment HC19. This methodology has already been discussed (Chothia *et al.*, 1986) and follows the observation that the six loops constituting the antibody combining site are supported by a conserved framework structure. We summarize here the successive steps in this procedure:

Firstly, two parent framework structures were chosen from the Protein Data Bank (for the V_L and the V_H domains, respectively), in order to maximize the sequence homology with the predicted antibody. A structure was chosen among the two Fab fragments from which the variable domains were extracted, and the relative orientation of the two variable domains in the model was taken from this structure.

Secondly, main-chain conformations of the CDRs (excluding H3) were inferred from 'canonical structures' (Chothia & Lesk, 1987).

Finally, these loops were grafted onto the modelled framework, with some local refinement to adjust the geometries of the different pieces [using Hermans' regularization routine *REFINE* (Hermans & McQueen, 1974) and the program *PROLSQ* (Hendrickson, 1985)]. As for the combining-site side-chain atoms, we limited our approach as follows: an HC19 side chain was included

in our model only if there was in the CDRs of known structure one potential model with both the same side chain at the same position and the same loop main-chain conformation as that predicted for HC19.

Results

Sequence

Examination of the electron density showed one contradiction with the predicted amino-acid sequence. Residue L94 was sequenced as Tyr; this constituted the only deviation in the V_L gene from the germline sequence. In the electron-density maps, this side chain, although exposed to the solvent, was clearly defined and non-aromatic; it was, therefore, modelled as an Asn – the normal germline residue – and this side chain fitted smoothly into the electron density. The primary sequence was checked by direct mRNA sequencing, thereby avoiding the PCR protocol during which an error may have been introduced; this revealed a single nucleotide heterogeneity (TAC or AAC, *i.e.* Tyr or Asn). The electron density, checked again in the final stages of the refinement, removes the ambiguity between Asn and Tyr, and an Asn is included in our model at position L94. Since the L94 residue is 3.7 Å from a crystal contact, it is not impossible that the L94 Asn was selected from a heterogeneous population in the crystallization process. An alternative explanation is that a mutation occurred in the HC19 cell line after the antibody was prepared but prior to sequencing.

Structure description

The final model contains 3261 protein atoms and 91 solvent atoms. The average coordinate error, as calculated from a Luzzati plot (Luzzati, 1952) is 0.31 Å. The Ramachandran plot for HC19 is shown Fig. 2: almost all residues are found within the boundaries determined by Ramachandran, Venkatachalam & Krimm (1966). The few exceptions found include residues located in poorly defined electron-density areas. Only two significant deviations are found in well defined electron-density areas: Thr L51 and Ser L93; these are discussed below.

Tertiary and quaternary structure. The globular domains of the refined HC19 structure display the typical immunoglobulin-fold super-secondary structure. In the case of the variable domains, structural comparisons of HC19 with other Fab fragments of known structure have been performed by superposing $C\alpha$ atoms belonging to the core β -strand region [for a definition of this core region, see Chothia & Lesk (1987)]. The root-mean-square coordinate differences (r.m.s.d.) were in the range usually observed for other comparisons of Fab fragments (Chothia & Lesk, 1987): see Table 2. The fact that the Fab D1.3 V_H domain is coded by the same gene as the HC19 V_H domain probably accounts for the very low r.m.s.d. value in this particular case. The

higher deviation observed for V_L domains (except for Se155–4), as compared to V_H , is due to a loop of the V_L domain (residues L67 to L69) contacting L1 in the HC19 structure and whose conformation is affected by the conformational change observed for L1 in HC19 (see below); when residues of this loop are not included in the comparison, the deviation between HC19 and other V_L cores is 0.9 Å, which is similar to values observed for V_H cores.

The quaternary structure of Fab fragment HC19 also has the typical shape observed for other Fab fragments. The elbow angle (the angle between the two pseudo-twofold axes of the V and C domains) is 146°. Values for other Fab fragments range from 120 to 180°. A remarkable feature of the displacement which best superposes the two constant domains is the translation along the pseudo-twofold axis: its value distinguishes between κ light chains (values ranging from 1.6 to 1.7 Å, up to 2.7 Å when the heavy chain is of the α class) and λ light chains (from 3.5 to 3.9 Å; for HC19, 3.15 Å). This slight structural difference is accounted for by the nature of the buried side chains at the CL:CH1 interface: almost half the light-chain residues at that interface are different in κ and λ light chains, and a small adjustment is needed to accommodate them. This feature has been reviewed (Padlan, Cohen & Davies, 1986), using a more limited repertoire of Fab-fragment structures than is presently available. In the case of HC19, we reach conclusions identical to theirs.

Conformation of the combining site. In the following description of the HC19 combining site, we restrict ourselves to the conformation of each individual loop main chain and of those side chains which have been shown

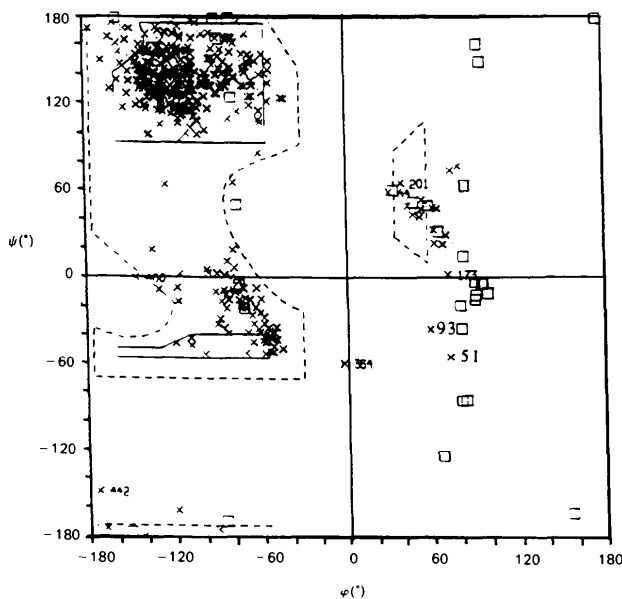


Fig. 2. Ramachandran plot of Fab fragment HC19. Glycine residues are represented as squares, other residues as crosses; outliers are labelled.

Table 2. *Sequence homologies and structural deviations of Fab fragment HC19 variable domains with other Fab fragments*

D1.3, Kol, HyHEL5, HyHEL10 and J539 coordinates were taken from the Protein Data Bank. Sel155-4 coordinates were kindly provided by the authors (M. Cygler, personal communication).

	Fab fragment	Primary sequence homology (without CDRs) (%)	R.m.s.d.* (Å)
VL	D1.3	44	1.59
	KOL	51	1.52
	HyHEL5	45	1.50
	HyHEL10	42	1.60
	J539	43	1.58
	Sel155-4	97	0.52
VH	D1.3	89	0.34
	KOL	54	0.88
	HyHEL5	42	0.97
	HyHEL10	57	0.80
	J539	55	0.86
	Sel155-4	38	1.05

* On core C α atoms.

to be important for the loop structure; we compare them with the corresponding features in other Fab fragments of known structure. For a complete description – or prediction – of the antibody combining site, two other points should be considered: the conformation of all side chains and the relative position of the CDR loops. These are addressed in the next section.

Light chain. Because of its class (murine λ), the HC19 light chain displays a low sequence homology to all other light chains available from the PDB (for the V_L domain alone this is never higher than 51%, a figure reached for Fab KOL, CDRs excluded). Despite this low homology, the key residues which have been postulated to determine the CDR main-chain conformation in the canonical structures are identical or similar to those found in the HC19 light chain.

In particular, when the size of the CDR is short (as is the case for L2 and L3) it can be anticipated that, because of the limited number of possible tight turns that the main chain can adopt, these loops will have in HC19 a conformation similar to the equivalent loops in other Fab-fragment structures. We will first describe the conformation of these short loops and then analyze that of L1.

L2. L2 is the shortest of the six CDRs (three residues); the main chain adopts a nearly identical conformation in all known structures, a tight three-residue turn. It is also the case for HC19 (r.m.s.d. on the main-chain atoms of the CDR and of the two neighboring residues: 0.09 Å with Fab KOL; 0.21 Å with Fab McPC603). The CDR L2 electron density is unambiguous (Fig. 3a). In HC19, L2 consists of the three residues Gly L50, Thr L51 and Asn L52. The main-chain dihedral angles of residue L51 lie outside the boundaries postulated by Ramachandran ($\varphi_{51} = 72^\circ$; $\psi_{51} = -54^\circ$). Such a dihedral angle pair has

already been found in two other Fab-fragment CDR L2 loops at the same position: in Fab KOL and Fab D1.3 with L51 being Asp and Thr, respectively. The turn conformation adopted by L2 in HC19 has been included by Chothia & Lesk (1987) in their β -turn classification, with any amino acid at position 2.

L3. As observed in Fab fragments of the λ -class, CDR L3 residues L91, L92, L95 and L96 are an extension of the β -sheet framework. The other two residues (Ser L93

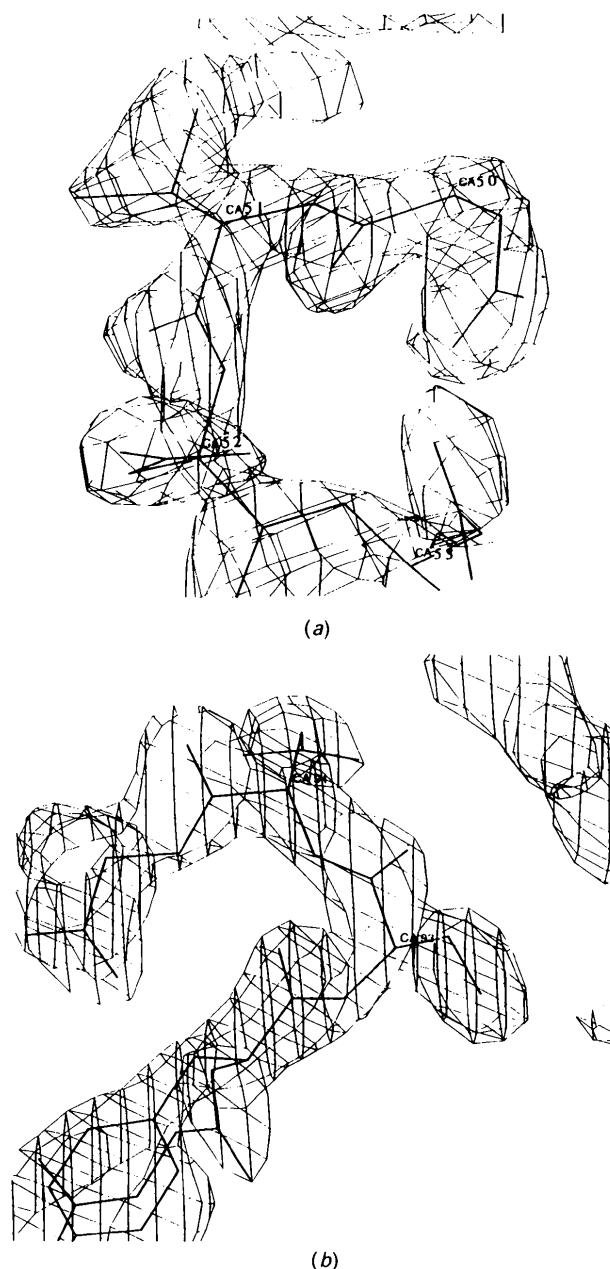


Fig. 3. Diagrams of a $2F_o - F_c$ electron-density map of Fab fragment HC19 showing (a) CDR L2; the map is contoured at the 2 r.m.s.d. level. (b) CDR L3; the map is contoured at the 1.5 r.m.s.d. level.

and Leu L94) adopt a classic γ -turn conformation with a hydrogen bond between the CO of residue L92 and NH of residue L94. The electron density is clear and the chain tracing is unambiguous (Fig. 3b).

Classic γ -turns are often associated with the loop-end of a β -hairpin, as is the case here; the φ, ψ torsion angles of their first residue are usually around ($75^\circ, -65^\circ$); this is outside the boundaries postulated by Ramachandran, although in most of the γ -turns this residue is not a glycine (Milner-White, Ross, Ismail, Khaled & Poet, 1988). Thus, the main-chain conformation of HC19 CDR L3 is not new in protein structures. In particular, the HC19 L3 conformation resembles Fab' New L3 (r.m.s.d. for main-chain atoms: 0.77 Å), which is of the same length; other Fab-fragment L3 loops have different lengths and/or main-chain conformations.

L1. L1 (residues 26–32) is longer than other light-chain CDRs (in HC19, it is ten residues long) and its main chain has clearly distinct conformations in λ and κ light chains. In the three light chains of human λ class of

known structure [Fab KOL, Fab' New and Bence-Jones protein Rhe (Furey, Wang, Yoo & Sax, 1983)], main-chain conformations are very similar, although Fab' New L1 is one residue longer than the other two; the main distinctive feature of that conformation is the anchoring of the aliphatic side chain of residue Ile L30 into the protein core.

The length of the L1 loop is the same in most λ mouse light chains and in Fab' New (Kabat, Wu, Perry, Gottesman & Foeller, 1991). The residue in λ mouse light chains equivalent to human λ Ile L30 is Val; this residue is buried in the protein core. Despite this conserved feature, the main-chain atom superposition of HC19 L1 with human λ light chains L1 gives a very high r.m.s.d. (e.g. 4.1 Å with Fab' New). The HC19 L1 conformation is also very different from all κ L1 conformations (Fig. 4). On the other hand, the same superposition with Fab Se155–4 yields an r.m.s.d. of 0.3 Å: the two L1 conformations are identical.

The position of residue L30 differs in human and mouse λ chains: the C α atoms are 3.5 Å apart after superposition of the V_L frameworks. This is due to the low sequence homology of mouse and human λ chains which results in differences for the residues that contact L30. In human λ light-chain structures, Ile L30 contacts Gly L25, Val (or Ile) L33 and Ala L71; in HC19, Val L30 has the following neighbors: the side chains of Leu L66 (L66 is a Lys in human λ chains), Ala L71 and Leu L90 (this is an Ala or a Ser in human λ chains of known structure) and the main chain of L70. Wu & Cygler (1993) have also noticed that if the L1 CDR in Fab Se155–4 were to adopt the conformation it has in Fab' New, the Ser L25 hydroxyl group would be buried in a hydrophobic pocket, whereas it is exposed in Se155–4 and HC19. Other factors that prevent HC19 L1 from adopting a conformation close to the one observed in human Fab fragments with a λ light chain become apparent when one tries to graft the CDR L1 of Fab' New onto the HC19 framework (Fig. 5). This results in two groups of sterically unfavorable interactions on both sides of the anchoring residue L30. First, the side chains of grafted residues L30, 31, 32 make unfavorable steric interactions with the side chains of residues L90 and L91, which are identical in HC19 and Se155–4, and with

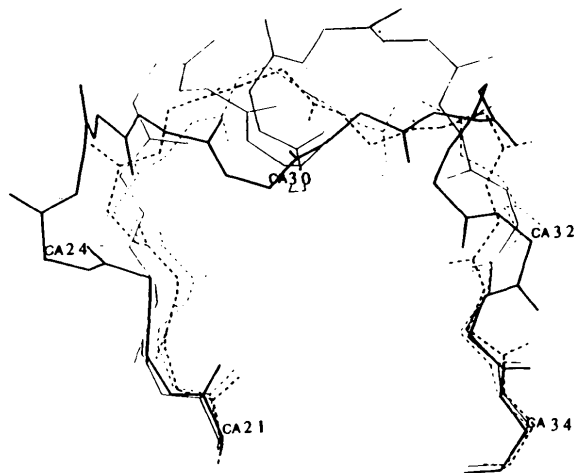


Fig. 4. Comparison of CDR L1 (residues 24–32) main-chain conformation in HC19 and representative Fab fragments. Framework main-chain atoms immediately adjacent to the loop are also displayed. HC19 is drawn in thick lines, Fab' New in thin lines, Fab J539 (κ light chain, L1 canonical structure 1) in dashed thick lines, Fab HyHEL10 (κ light chain, L1 canonical structure 2) in dashed thin lines. Canonical structures are from Chothia *et al.* (1989).

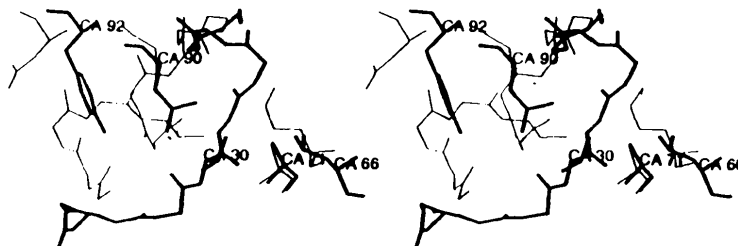


Fig. 5. Stereoview of the CDR L1 of Fab' New grafted onto the HC19 framework: main-chain atoms (plus the side chain of Val L30) are shown as thin lines. The conformation in HC19 is displayed in thick lines, as well as the residues in HC19 that contact the L1 loop or the grafted one; the residues in Fab' New at equivalent positions in the structure are shown as thin lines.

the main chain of CDR L2. Secondly, the main chain of grafted residues L28–29 collides with the side chain of Tyr L92 (in Fab Se155–4, instead of the germline Tyr, residue L92 is a cysteine; that cysteine would also collide with the Fab' New L1 loop when grafted on Se155–4); in human λ chains of known structure, residue L92 is Asp or Asn whose side chain points in a direction different from that of the tyrosine or cysteine in mouse λ chains. In HC19 or Se155–4, the conformation of L1 avoids contact of residue L92 with L28 or L29. The collisions of the grafted Fab' New L1 CDR with L90 and L92 could not be alleviated by minor main-chain modifications. Thus, the presence in mouse λ chains of the bulky buried L90 and L92 residues which were exposed and/or small in human λ light chains correlates with the shift in the position of the anchoring residue L30 and prevents the Fab' New L1 conformation from being adopted.

Heavy chain. The V_H domain of Fab HC19 displays a high sequence homology with the corresponding domain of Fab fragment D1.3, whose structure is known (sequence homology without CDRs: 89%). The length and nature of key residues of CDRs H1 and H2 are conserved from D1.3 to HC19, and their main-chain conformations are very similar; these conformations are also similar to those of equivalent loops in other Fab fragments [e.g. for H1, the r.m.s.d. (on main-chain atoms) is 0.32 Å with D1.3, but also 1.06 Å with HyHEL5; for H2, r.m.s.d. 0.13 Å with D1.3, but also 1.07 Å with HyHEL10]. This is in very good agreement with the 'canonical-structure' hypothesis.

With CDR H3, the problem encountered in all structure comparisons is met again in the case of HC19. This loop – coded by the V–J–D gene junction – is of highly variable length and sequence and its conformation differs

from one Fab to another. In addition, this is very often a long loop where the main chain adopts a rather loose conformation. In the case of HC19, CDR H3 is a 12-residue loop (Fig. 6); it is stabilized at its base by a conserved salt bridge between Arg H94 and Asp H101 and three internal hydrogen bonds. The remainder of H3 loops out of the framework; its stabilization depends almost exclusively on the packing of its six aromatic side chains, which form two clusters protruding on each side of the loop, both of them engaging many contacts, mainly with aromatic residues of other CDRs.

Comparison of the observed structure with a model based on canonical structures

Here we compare results of the empirical approach described in *Materials and methods* and of our crystallographic analysis. We first chose two parent framework structures available from the PDB, based on their sequence homology to the V_H and V_L domains. Highest homologies were found with Fab D1.3 for the V_H domain and with Fab KOL for the V_L domain (respective values: 89 and 51%). The resulting r.m.s.d. on framework $C\alpha$ atoms between the observed and the model structures was 0.76 Å, a normal value for this type of superposition (Chothia & Lesk, 1987).

The next step in this process concerns the overall structure of the HC19 combining site. We have seen in the previous section that CDR loops L2, L3, H1 and H2 are satisfactorily accounted for by canonical structures. Following the empirical approach, we grafted the four CDRs onto the model framework (V_L Fab KOL/ V_H Fab D1.3); loop L2 came from Fab McPC603, L3 from Fab' New, and H1 and H2 from Fab D1.3 (the

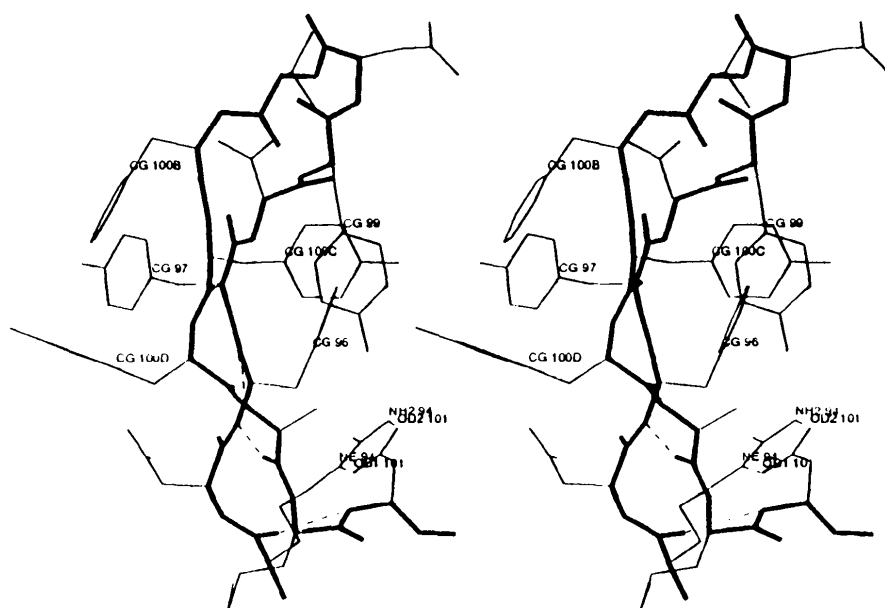


Fig. 6. Stereoview of CDR H3 of Fab fragment HC19. The main chain is shown as thick lines. The salt bridge and the three internal hydrogen bonds that stabilize the loop at its base are in dashed lines. Amino acids in the two clusters of aromatic residues protruding on each side of the loop are labelled.

Table 3. HC19 CDRs and canonical structures

CDR	Length of loop in HC19	Key residues	Canonical structure	Parent structure chosen for prediction
L2	3	48I, 64G	No. 1*	McPC603
L3	6	None	V λ type†	New
H1	7	26G, 27F, 29L, 34V, 94R	No. 1*	D1.3
H2	3	55G	No. 1*	D1.3

* Chothia *et al.* (1989).

† Chothia & Lesk (1987).

Table 4. R.m.s. deviations on main-chain atoms between observed and predicted CDRs of Fab fragment HC19

CDR	R.m.s.d. 1*	R.m.s.d. 2†
L1	4.0	4.7
L2	0.81	1.32
L3	0.77	1.01
H1	0.32	0.85
H2	0.13	0.58
L2 + L3 + H1 + H2‡	—	0.97
L2 + L3 + H1 + H2§	—	1.10

* CDRs individually superposed.

† Framework C α atoms superposed, CDRs compared.

‡ On main-chain atoms (92 atoms).

§ Predicted side chains included (133 atoms).

canonical-structure approach is summarized in Table 3). The methodology we followed allows us to determine the conformation of only seven side chains out of a total of 16 in these loops. Results of a superposition of the resulting atoms with the observed structure are shown in Table 4. It is worth noting that the overall superposition of the loops gives an r.m.s.d. with the observed structure which is only slightly higher than when each loop is considered individually; the relative position of the loops is, therefore, correct. As for the side chains, there is no incorrect conformation and the r.m.s.d. for these atoms is acceptable.

Discussion

The canonical-structure approach yields a partial model, excluding in particular H3, and would, therefore, fall short of enabling one to dock the HC19 paratope on its X31 HA epitope. The overall accuracy of this HC19 partial model is nevertheless notable: the r.m.s.d. between observed and predicted loops is 1.10 Å, which implies that both the main-chain conformation and the relative orientation of the predicted CDRs were correct. This error is comparable to the r.m.s.d. between crystallographically determined atomic positions in the common core of highly homologous protein structures (Chothia & Lesk, 1986). Several authors have quoted only slightly lower or even comparable deviations between identical and independently determined protein structures (Bhat,

Bentley, Fischmann, Boulot & Poljak, 1990; Stanfield, Fieser, Lerner & Wilson, 1990). This result is obtained despite the contacts that the highly variable H3 loop makes with the rest of the combining site and when no particularly accurate model could be found for the framework of the V_L domain (highest sequence homology to V_L domains available from the PDB: 51%). Two factors can account for the accuracy of this partial model:

(a) Three of the four modelled loops have very short sequences outside the β -sheet framework (two to three residues); their conformation is therefore heavily constrained. The fourth loop (H1) model is taken from Fab D1.3, where all residues critical for the observed conformation are conserved in HC19; in such cases it has been observed that main-chain atoms have an r.m.s.d. lower than 0.5 Å (Chothia *et al.*, 1992).

(b) The variable dimer of Fab KOL was chosen to orient V_H with respect to V_L, based on the sequence homologies and because it gave the best result in our molecular-replacement studies (Bizebard, Mauguen, Skehel & Knossow, 1991): in the KOL V dimer the relative position of the two domains is very close to that in HC19; had the other option been taken, to use the D1.3 V dimer for orienting the variable domains, the agreement would be slightly worse (r.m.s.d. = 1.1 Å, as compared to 0.97 Å when Fab KOL is chosen as the parent structure). In this context it should be mentioned that the displacements that best superpose the variable domains in published Fab structures are very close to each other (average value of the rotation angle = 170.7°, standard deviation = 3.4°). In the case of NC41, which has been reported to be an outlier in that distribution (rotation angle = 177°), deviations between observed and modelled CDRs of up to 3 Å were obtained (Tulip, Varghese, Laver, Webster & Colman, 1992); this is significantly higher than observed here.

The L1 loop is noticeably different from its counterparts in all structures available from the PDB. HC19 L1 length and key residues differ from those of the equivalent loops in κ -chain models. On the other hand, the L1 loops of human λ light chains – which all have very similar conformations – could be considered as potential models for HC19 because of the conservation of a key hydrophobic residue at position 30. Among them, we selected Fab' New because its L1 loop has the same length as HC19 L1. Although residues responsible for the observed L1 conformation in human λ light chains are of the same type as in HC19, most are not conserved (25 Gly \rightarrow Ser, 30 Ile \rightarrow Val, 33 Val \rightarrow Ala, 29 Asp \rightarrow Ala, 71 Ala conserved). More importantly though, residues which are exposed or small in human λ chains become large and/or buried in HC19 and interact with L1; this accounts for the observed difference between HC19 and human λ loops.

In HC19 CDR L1, a glycine at position L28 adopts a φ, ψ dihedral angle pair which is unfavorable for all other amino acids; a change of this residue would induce some

strain. All other side chains but one are exposed to the solvent; it is, therefore, likely that almost any mutation of these L1 residues could be accommodated by the same main-chain conformation. The only exception is Val L30: this residue and its contacting neighbors (Leu L66, Ala L71 and Leu L90) should, therefore, be key residues responsible for the observed conformation. These are very well conserved in mouse λ sequences (Kabat, Wu, Perry, Gottesman & Foeller, 1991). The HC19 L1 main-chain conformation is therefore likely to be representative of most mouse λ light-chain L1 loops. The fact that the L1 conformations of the two independently determined Fab fragments with mouse λ light chains (HC19 and Se155-4) are so similar provides additional support to this conclusion.*

We thank Alan Douglas for ascites fluid preparation. This work was supported in part by a grant from the EEC (SC1-474).

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1GIG, R1GIGSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: GR0265).

References

- BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.
- BHAT, T. N., BENTLEY, G. A., FISCHMANN, T. O., BOULOT, G. & POLIAK, R. J. (1990). *Nature (London)*, **347**, 483–485.
- BIZEBARD, T., MAUGUEN, Y., PETEK, F., RIGOLET, P., SKEHEL, J. J. & KNOSSOW, M. (1990). *J. Mol. Biol.* **216**, 513–514.
- BIZEBARD, T., MAUGUEN, Y., SKEHEL, J. J. & KNOSSOW, M. (1991). *Acta Cryst.* **B47**, 549–555.
- BRÜNGER, A. T. (1990). *X-PLOR Manual*. Version 2.1. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale Univ., New Haven, CT, USA.
- CHOTHIA, C. & LESK, A. M. (1986). *EMBO J.* **5**, 823–826.
- CHOTHIA, C. & LESK, A. M. (1987). *J. Mol. Biol.* **196**, 901–917.
- CHOTHIA, C., LESK, A. M., GHERARDI, E., TOMLINSON, I. M., WALTER, G., MARKS, J. D., LLEWELYN, M. B. & WINTER, G. (1992). *J. Mol. Biol.* **227**, 799–817.
- CHOTHIA, C., LESK, A. M., LEVITT, M., AMIT, A. G., MARIUZZA, R. A., PHILLIPS, S. E. V. & POLIAK, R. J. (1986). *Science*, **233**, 755–758.
- 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. & POLIAK, R. J. (1989). *Nature (London)*, **342**, 877–883.
- FAZEKAS DE ST GROTH, S. & SCHEIDEGGER, D. (1980). *J. Immunol. Methods*, **35**, 1–21.
- FINE, R. M., WANG, H., SHENKIN, P. S., YARMUSH, D. L. & LEVINthal, C. (1986). *Proteins*, **1**, 342–362.
- FISCHMANN, T. O., BENTLEY, G. A., BHAT, T. N., BOULOT, G., MARIUZZA, R. A., PHILLIPS, S. E. V., TELLO, D. & POLIAK, R. J. (1991). *J. Biol. Chem.* **266**, 12915–12920.
- FUREY, W., WANG, B. C., YOO, C. S. & SAX, M. (1983). *J. Mol. Biol.* **167**, 661–692.
- HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
- HERMANS, J. JR & McQUEEN, J. E. (1974). *Acta Cryst.* **A30**, 730–736.
- JONES, A. T. (1978). *J. Appl. Cryst.* **11**, 268–272.
- KABAT, E. A., WU, T. T., PERRY, H. M., GOTTESMAN, K. S. & FOELLER, C. (1991). *Sequences of Proteins of Immunological Interest*. US Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, MD, USA.
- KABSCH, W. (1976). *Acta Cryst.* **A32**, 922–923.
- KABSCH, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
- KAHN, R., FOURME, R., BOSSHARD, R. & SAINTAGNE, V. (1986). *Nucl. Instrum. Methods Phys. Res. Ser. A* **246**, 596–603.
- LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
- MARQUART, M., DEISENHOFER, J., HUBER, R. & PALM, W. (1980). *J. Mol. Biol.* **141**, 369–391.
- MARTIN, A. C. R., CHEETHAM, J. C. & REES, A. R. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 9268–9272.
- MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
- MILNER-WHITE, J. E., ROSS, B. M., ISMAIL, R., KHALED, B. M. & POET, R. (1988). *J. Mol. Biol.* **204**, 777–782.
- PADLAN, E. A., COHEN, G. H. & DAVIES, D. R. (1986). *Mol. Immunol.* **9**(23), 951–960.
- 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.
- RAMACHANDRAN, G. N., VENKATACHALAM, C. M. & KRIMM, S. (1966). *Biophys. J.* **6**, 849–872.
- READ, R. J. (1986). *Acta Cryst.* **A42**, 140–149.
- RIECHMANN, L., CLARK, M., WALDMANN, H. & WINTER, G. (1988). *Nature (London)*, **332**, 323–327.
- RIGOLET, P. (1991). PhD thesis, Univ. Paris XI, France.
- SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARIF, S. J., HIGACHI, R., HOM, G. T., MULLIS, K. B. & ERLICH, H. A. (1988). *Science*, **239**, 487–491.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
- SAUL, F. A. & POLIAK, R. J. (1992). *Proteins*, **14**, 363–371.
- SEGAL, D. M., PADLAN, E. A., COHEN, G. H., RUDIKOFF, S., POTTER, M. & DAVIES, D. R. (1974). *Proc. Natl Acad. Sci. USA*, **71**, 4298–4302.
- SERC Daresbury Laboratory (1979). *CCP4. A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington WA4 4AD, England.
- 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.
- STANFIELD, R. L., FIESER, T. M., LERNER, R. A. & WILSON, I. A. (1990). *Science*, **248**, 712–719.
- STEIPE, B., PLÜCKTHUN, A. & HUBER, R. (1992). *J. Mol. Biol.* **225**, 739–753.
- SUH, S. W., BHAT, T. N., NAIRA, M. A., COHEN, G. H., RUDIKOFF, S., RAO, D. N. & DAVIES, D. R. (1986). *Proteins Struct. Funct. Genet.* **1**, 74–80.
- TULIP, W. R., VARGHESE, J. N., LAVER, W. G., WEBSTER, R. G. & COLMAN, P. M. (1992). *J. Mol. Biol.* **227**, 122–148.
- WEIS, W. I., BRÜNGER, A. T., SKEHEL, J. J. & WILEY, D. C. (1990). *J. Mol. Biol.* **212**, 737–761.
- WU, S. & CYGLER, M. (1993). *J. Mol. Biol.* **229**, 597–601.