The EMBO Journal vol.5 no.2 pp.415–425, 1986

Modelling of the combining sites of three anti-lysozyme monoclonal antibodies and of the complex between one of the antibodies and its epitope

Patricia de la Paz, Brian J.Sutton, Michael J.Darsley and Anthony R.Rees

Laboratory of Molecular Biophysics, University of Oxford, The Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK

Communicated by Sir David Phillips

Models of the antigen combining sites of three monoclonal antibodies, which recognise different but overlapping epitopes within the 'loop' region of hen egg lysozyme (HEL), have been generated from the cDNA sequences of their Fv regions (the VL and VH domains) and the known crystal structures of immunoglobulin fragments. The α-carbon backbone of the structurally conserved framework region has been derived from the IgG myeloma protein NEW, and models for the hypervariable loop regions have been selected on the basis of length and maximum sequence homology. The model structures have been refined by energy minimisation. Both the size and chemical nature of the predicted combining site models correlate broadly with the epitope boundaries previously determined by affinity studies. A model of the complex formed between one antibody and the corresponding lysozyme epitope is described, and contact residues are identified for subsequent testing by oligonucleotide-directed site-specific mutagenesis.

Key words: antibody structure/protein epitope/computer graphics/hypervariable regions/lysozyme

Introduction

Structural studies of the binding of small molecules and haptens to myeloma proteins, in particular the results of X-ray crystallographic and magnetic resonance studies (reviewed by Davies and Metzger, 1983), have shown how antibody molecules may bind to discrete haptenic determinants. Much less is known, however, about the way in which antibodies recognise and bind to an antigenic surface such as that presented by a foreign protein molecule. Each antibody binds only to a specific region of the surface, known as the antigenic determinant or epitope, but it is becoming clear that frequently the whole of the protein surface may in fact consist of a continuum of overlapping epitopes (reviewed by Benjamin *et al.*, 1984). Nevertheless, certain epitopes appear to dominate the antibody response to a protein in a given host and there has recently been considerable interest in the attempt to find correlates of antigenicity. The possible importance of local flexibility of the protein has received particular attention, following the unexpectedly high frequency with which antibodies raised to peptide fragments of proteins are found to cross-react with the native protein. Indeed, a correlation has been found between the cross-reactivity of anti-peptide antibodies and the mobility of the region of the native (haemerythrin) structure to which they bind (Tainer *et al.*, 1984) and, for certain other proteins, a correlation between mobility and antigenicity with respect to anti-protein antibodies has been found (Westhof *et al.*, 1984). However, it has also been demonstrated that accessibility correlates well with antigenicity (Novotny *et al.*, 1985; Thornton *et al.*, 1985) and this may be the primary correlate. In fact the preliminary results from the only crystallographic analysis of a complex between a protein antigen and a monoclonal Fab (antigen binding) fragment reveal that the region of lysozyme recognised is not one of above average mobility (Amit *et al.*, 1985).

We have undertaken a systematic study of the anti-peptide and anti-protein antibody response to a defined protein epitope, in order to investigate in molecular detail the nature of the interactions, the basis of the peptide/protein cross-reactivity and the genetic aspects of the diversity of the immune response to a protein determinant. The approach we have adopted is as follows: (i) raise monoclonal anti-peptide and anti-protein antibodies against a single antigenic region; (ii) define the precise boundaries of the epitopes biochemically; (iii) determine the antibody sequences: (iv) model and compare the structures of each of the antibody combining sites and their complexes with antigen; (v) test the models by site-directed mutagenesis of those residues predicted to be important followed by analysis of the altered binding properties of the new antibodies. We believe this protocol will allow us to test and progressively refine the models of the complexes and, as a result, advance our understanding of the basis of the specificity of antigen-antibody interactions.

A series of monoclonal antibodies has been raised against the 'loop' antigenic region of hen egg lysozyme (HEL) which comprises residues 57–84 (Arnon and Sela, 1969). This region contains a stretch of polypeptide chain that has a higher than average mobility as defined by the main chain atomic temperature factors (Artymiuk *et al.*, 1979). Five antibodies have been raised against the peptide (conjugated to bovine serum albumin) which cross-react with native HEL (Darsley and Rees, 1985a) while two antibodies were raised against the native protein that are also specific for the loop region (Ryan *et al.*, unpublished).

The boundaries of the epitopes that each anti-peptide antibody recognises have been determined (Darsley and Rees, 1985a) and the five antibodies (all IgG class and termed Gloop 1–5) fall into three groups: Gloop 1 and 2, Gloop 3 and 4 and Gloop 5. Each group recognises distinct but overlapping epitopes within the loop region of HEL. These specificity groups were subsequently found to correlate exactly with the germ-line V-region genes from which they are derived, when the cDNA sequences were determined (Darsley and Rees, 1985b).

In this paper we describe the modelling component of the analysis and consider the three-dimensional structures of the combining sites of these antibodies. A modelling procedure is presented based upon knowledge of the sequence and a data base of known crystal structures of immunoglobulin fragments. This procedure is applied to the three Gloop antibodies 2, 4 and 5, each of which recognises a slightly different epitope within the loop region of HEL.

Modelling of unknown structures based upon homologous known structures has been performed within other families of proteins, such as the serine proteases (Greer, 1981) and, indeed, for antibody molecules (Padlan *et al.*, 1976; Mainhart *et al.*, 1984). The antigen binding Fv region of the antibody molecule

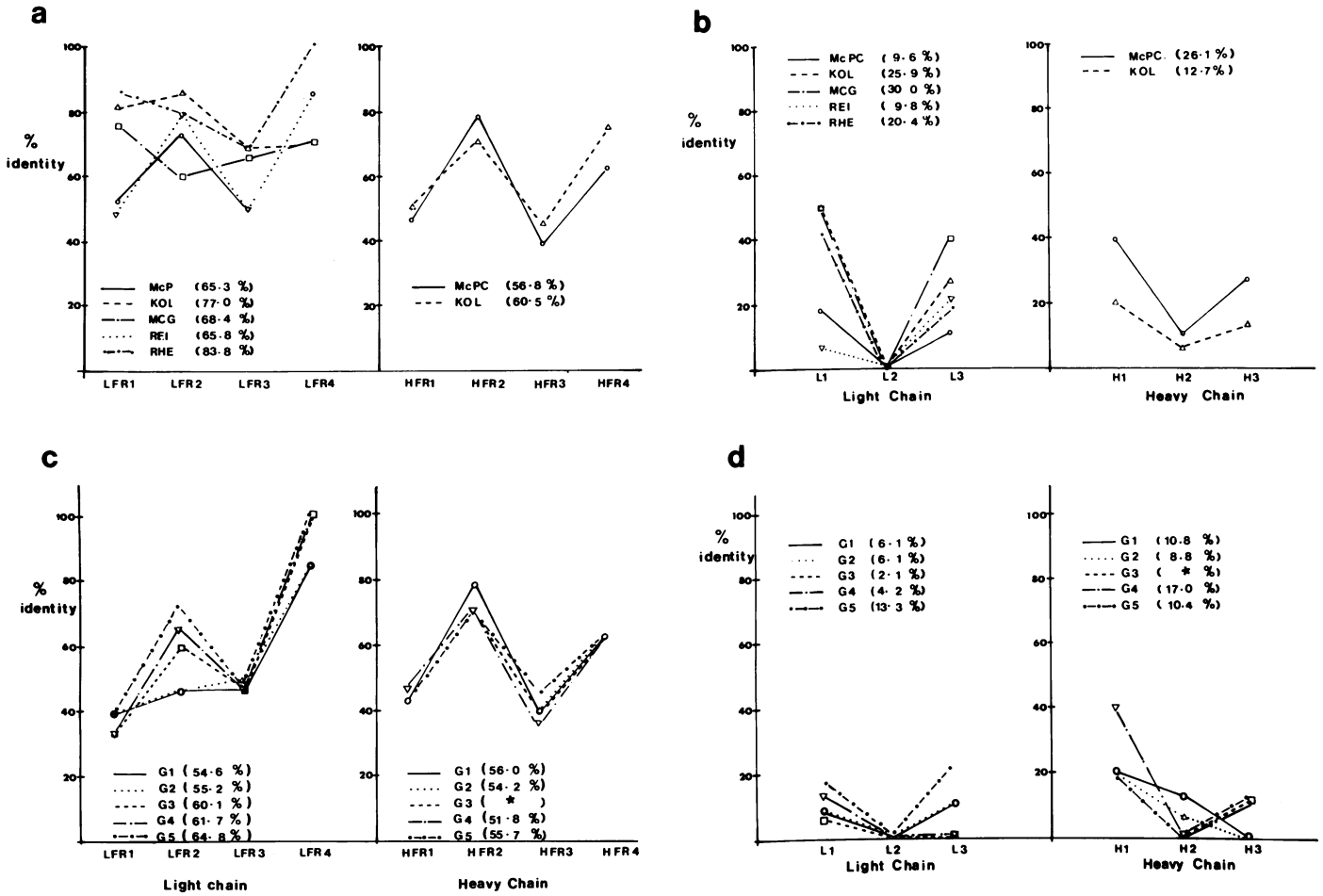

Fig. 1. Percentage sequence identity with Fv NEW for: **(a)** framework regions of database structures; **(b)** hypervariable regions of database structures; **(c)** framework regions of Gloop antibodies; **(d)** hypervariable regions of Gloop antibodies. Note: numbers in brackets represent mean sequence identity across a given framework or hypervariable region. * indicates incomplete sequence data.

is a particularly favourable system insofar as it consists of a highly conserved framework including a β-barrel structure formed between the β-sheets of the VL and VH domains (Novotny *et al.*, 1983). However, the specificity of the antibody is determined by loops of different length and hypervariable sequence that connect the strands of β-sheet, and our modelling of these complementarity-determining regions or CDRs (three in each chain: L1–3 and H1–3) is based upon a comparison of the known CDR structures, considering, in particular, length and sequence homology.

Having defined the boundaries of the epitopes and generated models of the combining sites of the antibodies that recognise them, the relative sizes and chemical nature of the three pairs of complementary structures are compared. A model of the complex formed between the anti-peptide antibody Gloop 2 and the native HEL structure with which it cross-reacts is described. This model has provided the starting point for studies now in progress to alter the antibody combining site by oligonucleotide-directed site-specific mutagenesis. While the results of these studies will initially provide a test of the modelling, we hope that, when supported by detailed X-ray structures of Fab fragments and Fab–antigen complexes, they may permit the design of further alterations that will modulate specificity towards a pre-determined sequence or structure.

Results

Database of crystal structures

Table Ia–c lists the database of crystal structures used in the present model building studies. This includes three Fab fragments, a light chain dimer and two VL dimers. Five of the six structures are independent crystal structure determinations: only RHE was analysed by molecular replacement of a pre-existing model (REI) into the observed electron density.

On the basis of the distinction between conserved framework regions and hypervariable loop regions we have accordingly divided the model building study into two parts.

Modelling of framework (FR) regions

In modelling the FR regions it is possible to exploit the high degree of sequence and structural homology among antibodies of known structure. Figure 1a shows the light and heavy chain sequence identity between the FR regions of Fv NEW and each of the other structures in the database. This reveals an average sequence identity of ~72% for the L chains and 67% for the H chains. This degree of identity is correlated with good structural homology as shown in Table II in which we compare the percentage sequence identity relative to NEW with the structural homology for the VL and VH FR regions. This emphasises that, even among antibodies from different species, there is strong con-

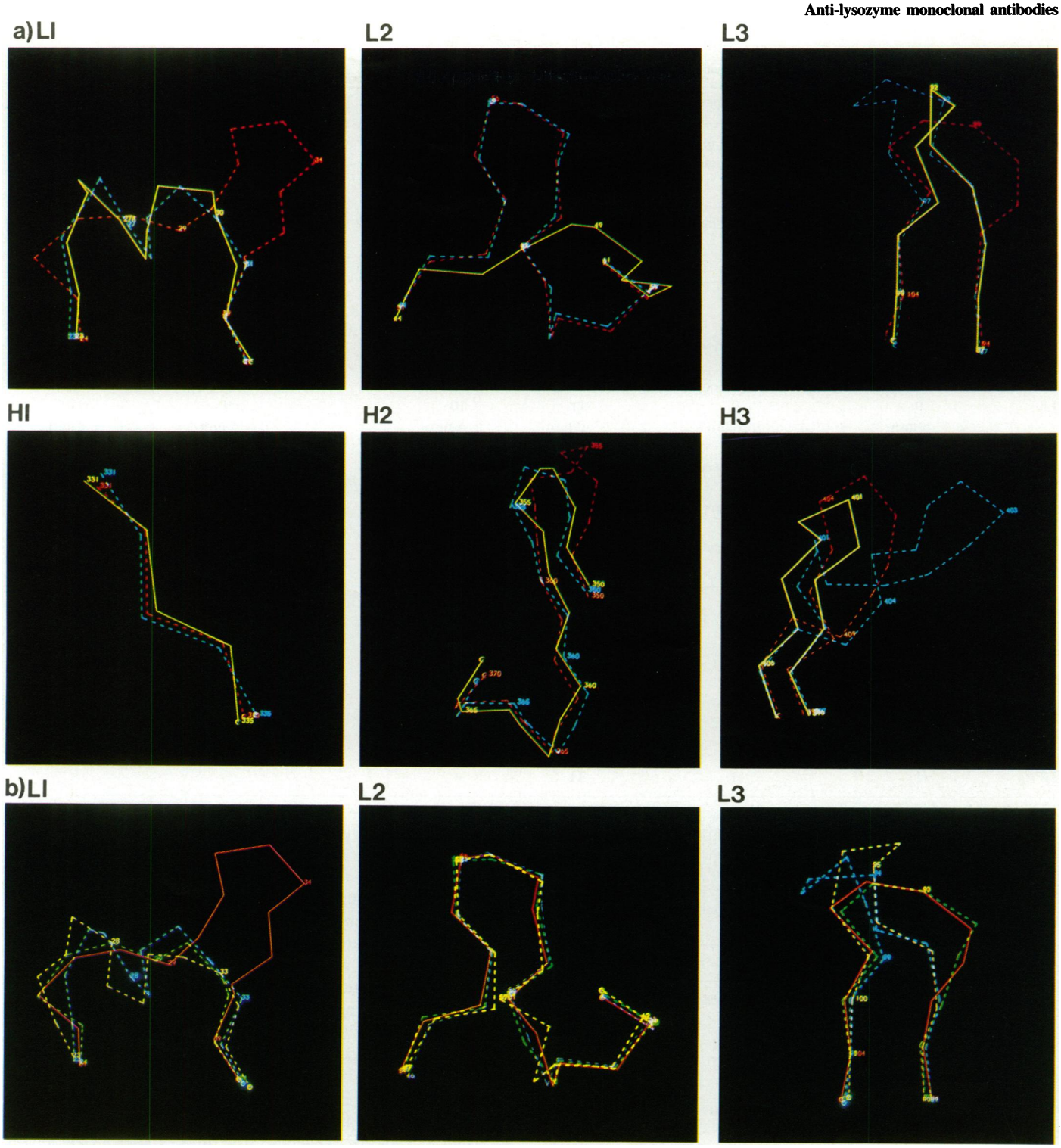

Fig. 2. **(a)** Comparison of α-backbone conformations of the CDRs from Fab fragments in the database structures: NEW = yellow; MCPC603 = red; KOL = blue. **(b)** Comparison of α-carbon backbone conformations of the CDRs from light chain dimers in the database structures: MCPC603 = red (for reference); REI = green; RHE = blue; MCG = yellow.

servation of FR structure. In view of this close agreement the model chosen for the FR regions of the Gloop antibodies was that of the human Fab NEW rather than the murine Fab from MCPC603 since the former has been defined at higher resolution.

Fab NEW displays ~60% sequence identity with the L chains of the Gloop antibodies and 54% with the H chains (Figure 1c). This conservation of sequence is reflected in the conformational constraints imposed by the β-barrel structure, but the precise orientation taken up by the VL and VH domains will be dictated by the nature of those residues at the VL:VH interface. Indeed, different modes of domain association have been observed between VL domains in VL dimer RHE (Furey *et al.*, 1983) and the light chain dimer LOC (Chang *et al.*, 1985). However, in Table III we list the VL:VH interactions common to the three database structures and the predicted Gloop models, and what is immediately obvious is that, in most instances, the contact

Table I. Properties of immunoglobulin fragments with known crystal structures

	NEW	MCPC603	KOL	MCG	REI	RHE
a. Details of immunoglobulin fragments referred to in this study						
Source	Human	Mouse	Human	Mouse	Human	Human
Composition	Fab	Fab	Fab	L chain dimer	VL dimer	VL dimer
Class	λ, γ1	κ, α	λ, γ1	λ	κ	λ
Bound hapten	Vit Kl-OH and others	Phosphorylcholine		Various aromatic compounds		
Resolution	2.0 Å	2.7 Å	1.9 Å	2.3 Å	2.0 Å	1.6 Å
Reference	1	2	3	4	5	6
b. Lengths of hypervariable (CDR) loops						
L1	14	17	13	14	11	13
	α-helix	Random coil	α-helix	α-helix	Short random coil	α-helix
L2	5	7	7	7	7	7
	Random coil	'S' loop	'S' loop	'S' loop	'S' loop	'S' loop
L3	9	9	11	10	9	11
	β-turn	β-turn	β-turn	β-turn	β-turn	β-turn
H1	5	5	5	–	–	–
	β-sheet	β-sheet	β-sheet			
H2	16	19	17	–	–	–
	β-turn	β-turn	β-turn			
H3	9	11	17	–	–	–
	β-turn	β-turn	Random coil covering site			
c. Lengths of framework (FR) regions						
LFR1	22	23	22	22	23	22
LFR2	15	15	15	15	15	15
LFR3	27	32	32	32	32	32
LFR4	11	11	11	11	11	11
HFR1	30	30	30	–	–	–
HFR2	14	14	14	–	–	–
HFR3	32	32	32	–	–	–
HFR4	11	11	11	–	–	–

Note: CDR and FR regions defined according to the method of Kabat *et al.* (1976).
References: 1. Saul *et al.* (1975); 2. Segal *et al.* (1974); 3. Marquart *et al.* (1980); 4. Edmundson *et al.* (1974); 5. Epp *et al.* (1975); 6. Furey *et al.* (1983).

Table II. Comparison of sequence identity and structural homology in framework regions of database structures versus Fv NEW

	MCPC		KOL		MCG	REI	RHE
	VL	VH	VL	VH	VL	VL	VL
Amino acid sequence identity %	65	57	77	61	68	66	84
Structural homology %	62	71	75	50	65	60	75
RMS deviation in Å	0.85	0.82	0.67	0.85	0.76	0.81	0.61

In computing the structural homology the percentage is a measure of those α-carbon atoms which could be superimposed to within <1.5 Å (Rossman and Argus, 1975). The RMS deviation is computed for those atoms within this limit.

residue is conserved. Where substitutions do occur in the Gloop sequences, similar substitutions are seen to be accepted in the database structures.

Modelling of hypervariable (HV) regions

HV regions are expected to be difficult to model because of their varying length (Table Ic) and lower sequence homology (Figure 1b and d). From Figure 2, however, in which the structural homology between different HV loops is compared, it is clear that while certain loops are of variable length and conformation (L1, L3 and H3), others are remarkably constant in structure. The latter include two regions with residues participating in β-structure (H1 and H2) and a region which forms a hydrogen-bonded arch between two β-strands (L2).

Models for the hypervariable surface of each of the Gloop antibodies were obtained by selection of those CDRs in the database that were of similar length to and displayed maximum sequence homology with the particular Gloop CDR. The models chosen are listed in Table IV. Database structures of appropriate length could be found for the majority of the three sets of CDRs (Gloop 2, 4 and 5) with a sequence homology to known structures of between 40 and 80%. Even in those instances where the homology was low (e.g. H2), the observed structural homology in different crystal structures was good. Only in modelling L1 and H3 was it necessary to introduce deletions relative to existing structures.

Energy minimisation

The preliminary models of the Gloop antibodies were subjected to 64 cycles of energy minimisation by the method of Levitt (1974), in which the potential energy of each structure was minimised by shifting the atomic positions in the direction of steep-

Table III. A summary of those interactions across the VL:VH interface common to the database structures and the Gloop antibodies

NEW		MCPC		KOL		G2		G4		G5	
V_H	V_L	V_H	V_L	V_H	V_L	V_H	V_L	V_H	V_L	V_H	V_L
V37	F99										
Q39	Q37										
Q39	Y86										
R43	Y86	K	Y	K	Y	Q	Y	Q	Y	Q	Y
L45	P43						I				
L45	Y86										
L45	F99										
W47	S93		Y		Q		Y		V		Y
W47	L94		P		S		P		P		P
W47	R95		L		Y		L		W		W
Y94	A42		P	F	A	F	T	F	T		S
W107	Y35						L				
W107	A42		P						T		S
W107	P43								I		
W107	F99										

When a residue substitution with respect to the NEW structure has occurred in either the H or L domain the new residue is indicated in the table.

est descent of the energy. The principal aim of this procedure was to improve the stereochemistry of the models by relaxing strained bond lengths and angles and relieving unfavourable contacts between non-bonded atoms. In fact the main contribution to the energy change was due to relief of non-bonded interactions, which contributed 69%, 70% and 79% of the total energy change for the Gloop 2, 4 and 5 models, respectively. However, the final conformations are close to those of the initial models: the root mean square shifts calculated for all the atoms are 0.23 Å, 0.27 Å and 0.28 Å for Gloop 2, 4 and 5, respectively. The range of convergence of this procedure is clearly small and other methods must be employed to achieve sampling of more remote conformations that may be accessible to the structures (see Discussion).

For comparison, the identical energy minimisation procedure was applied to the Fv NEW structure. The total energy change (2828 units), though less than that for the Gloop models (3683, 3993 and 5691 units for Gloop 2, 4 and 5, respectively) was still considerable and was similarly dominated by the non-bonded interaction term (58% of total energy change). The root mean square displacement of atomic positions was 0.20 Å.

Correlation of size of combining site with size of epitope

In Figure 3 we show a superposition of the α-carbon backbones of the energy refined models of Gloop 2, 4 and 5, viewed perpendicular to the axis of the Fv β-barrel. Loops L2, L3, H1 and H2 have identical lengths and very similar conformations in the three predicted structures while L1 and H3 vary substantially in length in a manner that correlates with the size of the epitope as deduced from binding studies (see Figure 6 in Darsley and Rees, 1985a). Gloop 2, for instance, which recognises the largest epitope, is characterised by a short L1 and H3 (11 and four residues, respectively), generating a relatively flat combining site surface with an area of ~ 375 Å^2. Gloop 4, by contrast, which recognises the smallest epitope, has significantly longer L1 and H3 loops (16 and eight residues, respectively), whose effect is to raise the walls of the combining site to generate a more enclosed cavity whose area is ~ 300 Å^2. Gloop 5, whose epitope is intermediate in size between Gloop 2 and 4, has a long L1 loop and a short H3 loop (17 and five residues, respectively).

These observations indicate that changes in the size of the combining site can be brought about by changes in the lengths of the CDRs, particularly L1 and H3. Furthermore, for CDRs of identical length, the size of the combining site can be modulated by changes in the volumes of the constituent side chains (see below).

Identification of possible contact residues

In Figure 4a – f we compare the chemical characteristics of the combining sites of the three Gloop antibodies, viewed down the Fv barrel axis. In our consideration of possible contact residues for the 'loop' epitope, the following were specifically excluded: (i) those residues whose side chain character is conserved across antibodies of different specificity (Padlan, 1977) — restricted mainly to buried positions within the VL and VH domains; (ii) those residues which, from inspection of the models, lie on the outer surface of the Fv domain distant from the central binding cavity. These include the three N-terminal residues of L1, the five C-terminal residues of L2 (excluding position 55) and the four C-terminal residues of H2.

Of the remaining residues, few were seen to be conserved across all the Gloop sequences. This is consistent with the biochemical data which show that although these antibodies were raised to a common antigenic determinant, they may recognise the 'loop' structure in significantly different ways.

Between antibodies of the same specificity group, however, certain residues are conserved at positions known to be hypervariable hot-spots and these are candidates for contact residues (see below). Also, there are residues which are partially buried in the VL:VH interface but which nevertheless vary between antibodies of different specificity groups (e.g. L1 position 34; L3 position 96; H2 position 35; H2 position 50 and H3 position 101). Some of these are charged residues which, away from the screening effects of solvent, could contribute significantly to the energy of antibody – antigen interaction.

Another striking feature of all the Gloop antibody combining sites is the high concentration of aromatic residues, in particular tyrosine (see Figure 4). This has also been found in several other antibody combining sites (see Discussion).

Correlation between the nature of the combining sites and that of the epitopes

In attempting to correlate the stereochemical properties of the combining sites with those of the epitopes, the following features common to the Gloop antibodies were noted. (i) Antibody binding to HEL has no effect on enzyme activity and therefore the mode of interaction must not preclude access to the active site cleft. (ii) All five antibodies are sensitive to the introduction of a positive charge at position 77 of the 'loop' epitope. This suggests that the contact residue on the combining site is itself a positive charge.

Gloop 2 and Gloop 5. These will be considered together since they display broadly similar specificity for the HEL epitope. Of the positively charged residues within the two combining sites, only that at position 56 (H2) is conserved (Lys 56 in Gloop 2; Arg 56 in Gloop 5). This is therefore proposed as the most probable contact residue for Asn 77 of HEL.

Another similarity between the combining sites is the presence of a negatively charged residue on H3: Glu 95 in Gloop 2 and Asp 101 in Gloop 5 (see Figure 4a and c). The possible significance of this residue will be discussed after presenting the model of the Gloop 2-HEL complex (see below).

The two antibodies differ in two ways. First, Gloop 5 recog-

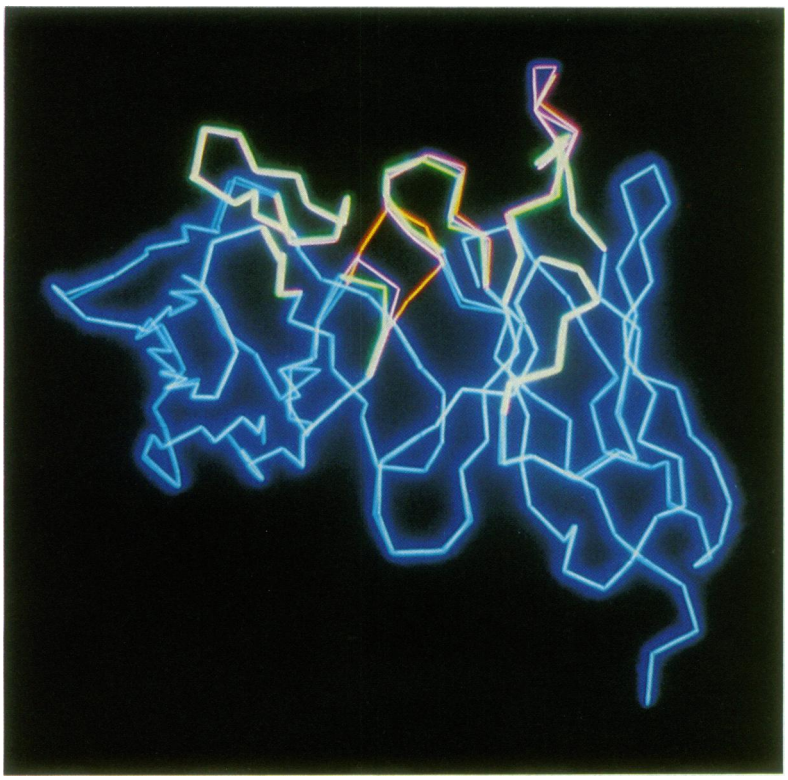

Fig. 3. Superposition of the predicted α-carbon backbone conformations for the Fv fragments of Gloop 2, Gloop 4 and Gloop 5, viewed perpendicular to the axis of the Fv β-barrel: Gloop 2 = green; Gloop 4 = red; Gloop 5 = pink.

nises a slightly smaller epitope than Gloop 2 and has a correspondingly smaller combining site as a result of differences in the lengths of L1 and H3. The greater length of L1 in Gloop 5 may be an important factor in the occlusion of certain residues of HEL that are accessible to Gloop 2. In addition, for those CDRs that are of the same length in the two antibodies, side chain volumes at particular residue positions are substantially larger in Gloop 5. For example:

on L2 — Ala 50 replaced by Trp;
on L3 — Leu 96 replaced by Trp;
on H1 — Gly 33 replaced by Trp;
Thr 35 replaced by Arg.

Second, Gloop 2 is sensitive to the substitution of arginine by lysine at position 68 in HEL while Gloop 5 is unaffected. Arg 68 may well interact with a negative charge on the combining site surface, and there are two solvent-accessible glutamic acid residues in Gloop 2, at positions 27A on L1 and 95 on H3, that are absent in Gloop 5.

Gloop 4. This antibody differs from both Gloop 2 and Gloop 5 in that it recognises a smaller epitope comprised of predominantly hydrophobic and uncharged hydrophilic residues. Unlike Gloop 2 and Gloop 5 it has no accessible negatively charged residue in the combining site. The only two solvent accessible charges, both of them lysines, are found at positions 50 on L2 and 58 on H2. The L2 lysine is, however, at the periphery of the combining site, shielded from the central cavity by the ridge formed by the residues of the long L1 and H3 CDRs, and may not be involved in binding to the small epitope. The H2 lysine may therefore be responsible for the sensitivity to the substitution at position 77 in HEL, and even occurs at a very similar location to that of the conserved positively charged residue at position 56 in H2 of Gloop 2 and Gloop 5 (see Figure 4b).

A preliminary model of the Gloop 2–HEL complex

A preliminary 'docking' of the Gloop 2 combining site to the epitope on the surface of HEL has been carried out. The initial orientation was established by matching charged residues and then adjusted, maintaining both molecules as rigid bodies, so as to optimise their stereochemical complementarity and at the same time ensure that the model was consistent with the known boundaries of the epitope. The model in shown in Figure 5.

The main features of the interaction are given below. (i) Arg 68 (HEL) interacts with Glu 27A (L1). In addition, the model places Arg 68 close to two other hydrogen bonding residues, Gln 27 (L1) and Ser 30 (L1). It is known that Gloop 2 (and Gloop 1) alone can distinguish between Arg and Lys at position 68 of HEL and it is possible that if all three antibody residues are required for the interaction this may account for the observed discrimination. (ii) Asn 77 (HEL) interacts with Lys 56 (H2). (iii) The active site residues of HEL remain fully accessible to solvent in the complex. (iv) Residues Asn 65 and Asn 74 (HEL), whose influences on Gloop antibody specificity have hitherto not been monitored, are found near the central region of the combining site, interacting mainly with hydrogen bonding residues such as Tyr 58 (H2) and Tyr 94 (L3). In addition, Leu 75 (HEL) interacts with Pro 52 (H2). Changes within the 'loop' epitope at these residue positions would therefore be expected to affect binding to Gloop 2. (v) A number of main chain atoms of HEL are seen to interact with the combining site surface, particularly at positions between residues 65 and 78. (vi) The model places Arg 73 of HEL very close to the edge of the combining site surface of Gloop 2, close to residue Glu 95 in H3. Arg 73 was excluded from the epitope recognised by Gloop 2 because of insensitivity to the substitution Arg 73 → Lys 73 (Darsley and Rees, 1985a), but at this location, such a substitution might easily

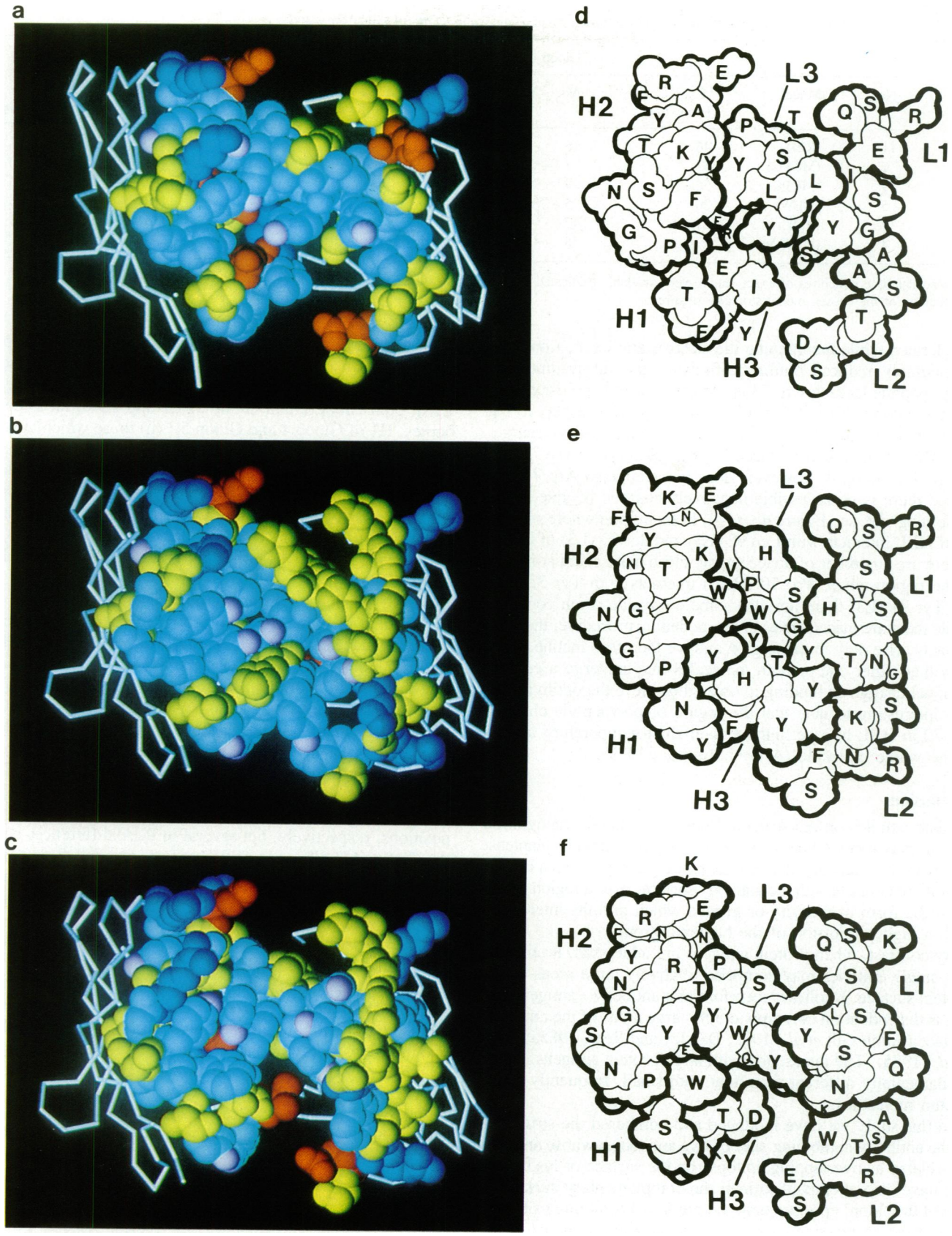

Fig. 4. Predicted structures after energy minimisation of the Fv fragments of **(a)** Gloop 2, **(b)** Gloop 4 and **(c)** Gloop 5, viewed down the axis of the Fv β-barrel. The framework regions are shown as an α-carbon backbone while the hypervariable regions are in an all-atom space-filling representation. Colour key: hydrophilic uncharged residues = yellow; hydrophobic uncharged residues = light blue; positively charged residues = dark blue; negatively charged residues = red; tyrosine -OH groups = lilac. **(d–f)** Key to the CDRs and their constituent residues for the modelled structures. The boundaries of the CDRs are indicated by bold lines.

Table IV. A summary of those database CDR structures chosen for modelling the CDRs of Gloop 2, 4 and 5

	Gloop 2			Gloop 4			Gloop 5		
	No.	Model	% Homology	No.	Model	% Homology	No.	Model	% Homology
L1	11	(REI)	46	16	(MCPC-17)*	41	17	(MCPC)	77
L2	7	(MCPC)	57	7	(MCPC)	43	7	(MCPC)	86
L3	9	(MCPC)	56	9	(REI)	33	9	(REI)	67
H1	5	(NEW)	20	5	(NEW)	40	5	(NEW)	60
H2	17	(KOL)	31	17	(KOL)	19	17	(KOL)	13
H3	4	(NEW-9)	*	8	(NEW-9)	*	5	(NEW-9)	*

When no database structures of exact length were available (indicated by *), the structure nearest in length was chosen. Only for H3 was a significant modification of the chosen model structure required.

be tolerated while maintaining the electrostatic interaction. The considerably reduced binding affinity of the sub-fragment of 'loop' peptide lacking Arg 73 is consistent with the existence of this interaction. (Gloop 5 behaves similarly with respect to the substitution and sub-fragment binding, and has a negative charge, Asp 101 in H3, similarly placed for interaction with Arg 73, see Figure 4c. Gloop 4, however, does not recognise Arg 73 and indeed there is no accessible negatively charged residue in the combining site, see Figure 4b.) (vii) In the region where residue 67 of HEL interacts with Leu 92 (L3) and Ser 93 (L3) of Gloop 2 there are a number of unacceptably close main chain contacts. In the region where Pro 70 (HEL) interacts with Tyr 32 (L1) and Tyr 91 (L3) there are also some close side chain contacts. While there are inadequacies in the modelling procedure, the poor contacts arise only in this region and we therefore incline to the view that either one or both molecules may undergo a conformational change on binding in order to achieve the optimum fit. It is interesting to note that the region of polypeptide chain at Pro 70 in HEL has the highest main chain temperature factors of the whole molecule (Artymiuk *et al.*, 1979).

Discussion

To date, studies on the structural basis of antibody recognition have largely focussed on hapten–antibody interactions. Haptens, however, are small and derive their binding energy from a limited number of contacts with the antibody surface, over a region which takes the form of a cavity or groove which permits interaction with a high proportion of the haptenic group.

By contrast, when protein antigens are addressed, it emerges that single epitopes may comprise relatively large areas of the protein surface. It follows therefore that antibody–antigen contact is dependent on correspondingly large areas of the antibody surface (Smith-Gill *et al.*, 1982; Darsley and Rees, 1985a; Amit *et al.*, 1985). The interaction with large protein antigens is thus fundamentally different from that with small, frequently rigid, hapten molecules.

In this study we have modelled and compared the structures of the antibody combining sites of five antibodies whose antigens are related as overlapping epitopes on the surface of lysozyme. We have also attempted to map the complementary surface of one of the 'loop' epitopes onto the predicted combining site surface of the corresponding antibody, to generate a model of the complex.

Modelling of the combining sites

Our comparative analysis of the conformations of hypervariable loop regions in the known crystal structures has allowed us to place each CDR of the Gloop antibodies into one of three groups: (i) those which are of the same length as, and exhibit appreciable sequence homology with, existing structures in the database (e.g. L1 of both Gloop 2 and Gloop 5; L2 and L3 of all three antibodies; H1 of Gloop 4 and Gloop 5); (ii) those which are of the same length as CDRs which, in the majority of observed structures, are of constant length and backbone conformation despite significant differences in sequence (e.g. H1 of Gloop 2 and H2 in all three antibodies); (iii) those which contain deletions relative to existing models (e.g. 1-residue deletion: L1 and H3 of Gloop 4; 4-residue deletion: H3 of Gloop 5; 5-residue deletion: H3 of Gloop 2).

Clearly we have greater confidence in the models of those CDRs that fall in the first two groups. Figure 2 illustrates that there are indeed remarkable similarities between the backbone conformations of certain CDRs (H1, H2 and L2) despite significant differences in amino acid sequence, and in some instances complete lack of sequence homology (Figure 1b). Further support for the modelling of all three light chain CDRs upon homologous structures is provided by the crystal structure determinations of two other VL dimers, AU (Fehlhammer *et al.*, 1975) and ROY (Colman *et al.*, 1977), not included in the database for this study. AU and ROY differ in sequence from REI at only 16 and 18 positions, respectively, but several of these differences occur in the CDRs. AU differs from REI at two, one and three positions in L1, L2 and L3, respectively, and ROY differs at three, two and four positions, respectively. All three proteins have virtually identical backbone conformations.

Where deletions must be made, the correctness with which the new structure can be modelled will depend on the length of the loop. The three H3 CDRs are all short (four, five and eight residues) and will therefore have a restricted range of conformations available to them. A comparative study of modelling deletions and insertions in homologous loop structures has been undertaken by others (e.g. for the serine proteases – Greer, 1981) with some degree of success although the conclusions of this study must be tempered by those of Read *et al.* (1984) who, in evaluation of comparative model building of trypsin, expressed less optimism about the successful prediction of variable loop structures.

An obvious improvement in the present study would be to increase the size of the database of CDR conformations. This would not necessarily demand more antibody crystal structures since the CDR loops form part of a much larger family of interconnecting loops, found between β-strands in β-sheet-containing proteins, an analysis of which is being undertaken by others (Sibanda and Thornton, 1985).

A further limitation of the present study is that the energy mini-

misation procedure used does not allow for large departures from the initial conformation. One approach to allow sampling of a wider range of conformations that may be accessible to the CDR structures would be to compute a dynamic simulation. In addition, more sophisticated potential energy functions could be incorporated when generating the final conformation, in order to take explicit account of terms due, for instance, to hydrogen bonds and electrostatic effects.

Despite these reservations, we are confident that the models are sufficiently correct to allow useful comparisons to be made between the five Gloop antibody combining sites and also to provide a basis for our protein engineering studies (see below), the results of which will, in turn, provide a test of the initial modelling.

Nature of combining sites of antibodies raised to the same antigenic region

It is perhaps surprising that there are few obvious similarities between the five anti-loop antibodies considering that they recognise overlapping determinants. However, there are common features. First, they share certain solvent-accessible positively charged residues, one of which (Lys 56 on H2) we have already suggested may provide an important contact for Asn 77 in the Gloop 2-HEL complex (see Darsley and Rees, 1985a for arguments relating to contacts at this epitope position). In Gloop 5, which recognises a broadly similar epitope, the positively charged residue at this position is conserved (Arg 56 in H2). In Gloop 4, which also recognises Asn 77 but within a much smaller epitope, there is a positive charge in a similar position (Lys 58 in H2). [The only other positive charge in the Gloop 4 combining site (Lys 50 in L2) is distant from the central cavity and shielded by residues of the L1 and H3 CDRs.]

The second common feature is that all three antibodies possess a high proportion of aromatic residues in the contact regions of their combining sites. Since the majority of the 'loop' peptide residues are either hydrophobic or neutral hydrophilic, this might be an expected complementarity. In fact, this preponderance of aromatic residues is a feature common to other antigen binding fragments, both at the VL:VH interface and in the combining site itself. A cluster of invariant aromatic residues at the VL:VH interface has been described (Novotny and Haber, 1985; Burley and Petsko, 1985). Here, their presence may contribute to structural stabilisation, but in the combining site, the amphipathic character of Tyr residues in particular may be important for providing both general hydrophobic interactions and more specific antigen binding through hydrogen bonds.

When the differences between the Gloop antibodies are considered, the comparative modelling procedure allows further determinants of specificity to be identified. Thus Gloop 2, the only antibody whose epitope includes Arg 68, contains two negatively charged residues, Glu 27A on L1 and Glu 95 on H3, that are absent from all the other combining sites. In the model of the complex, Glu 27A interacts with Arg 68 and Glu 95 may interact with Arg 73 as discussed above. Gloop 5 may also recognise Arg 73 through the similarly located Asp 101 in H3. The combining site of Gloop 4 in contrast does not recognise Arg 73 and there is no negatively charged residue present.

There is a considerable variation in the lengths of certain hypervariable loops which correlates well with the size of the epitope. Thus Gloop 2, which recognises the largest epitope, has the shortest L1 and H3 CDRs, while Gloop 4, which recognises the smallest epitope, has the longest L1 and H3 CDRs. The epitope recognised by Gloop 5 is intermediate in size, and this antibody has a long L1 and a short H3 CDR. Inspection of the models shows that shorter CDRs, in particular L1 and H3, can result in an opening-up of the combining site while longer loops can cause partial occlusion of its solvent-accessible surface. In this context it is interesting to note that all the myeloma proteins in the database which bind small molecules have longer H3 loops than any of the Gloop antibodies (Tables I and IV); Fab KOL is an extreme example of this with an almost completely occluded binding cavity. Furthermore, the lengths of L1 and H3 in an anti-HEL antibody whose structure has been recently described (Amit *et al.*, 1985), and which recognises a large surface region of HEL, are more similar to those of the Gloop antibodies.

Thus, the gross architecture of the combining site appears to be determined by variations in the length of the CDRs. Moreover, for CDRs of the same or similar length, changes in the volume of the combining site can be achieved by substitution of side chains with different steric properties. This was demonstrated by comparing Gloop 2 and Gloop 5, which recognise epitopes of similar chemical character but different size.

Modelling of the Gloop 2–HEL complex

By 'docking' the model of the Gloop 2 antibody combining site with its associated epitope on native HEL it has been possible to generate a model of the complex that is consistent with the experimental binding data. The striking manner in which a matching of charged residues led to an acceptable relative orientation of the two surfaces, later refined by considering surface complementarity, may reflect the importance of long range electrostatic interactions in initial contact between antigen and antibody.

The modelling of the complex also revealed the following features: (i) the large surface of interaction (~20 Å × 15 Å), with all CDRs involved in antigen contact; (ii) the high proportion of contacts involving main chain atoms of the 'loop' epitope. This may be a reflection of the importance of backbone conformation and not merely side chain stereochemistry, and perhaps also the requirement for accessibility or protrusion of an epitope in the recognition of antigenic structures (Novotny *et al.*, 1985; Thornton *et al.*, 1985); (iii) the existence of additional residues within the epitope that are probably contact residues from Gloop 2 (e.g. Asn 65, Asn 74 and Leu 75); (iv) the requirement for conformational change in either the epitope or the Gloop 2 combining site on formation of the complex. Although no conformational changes have been detected on binding haptens to Fab combining sites, the haptens are small and in fact changes in some CDRs have been observed on binding site-filling ligands to the light chain dimer MCG (Edmundson *et al.*, 1984).

The Gloop antibodies were originally raised to the 'loop' peptide fragment conjugated to a carrier molecule (BSA), and not to the native HEL molecule, but all five antibodies cross-react with the 'loop' region of native HEL. We chose to model the complex with HEL in order to understand how anti-peptide antibodies might interact with the native protein. The isolated 'loop' peptide undoubtedly has greater flexibility than it does as part of the native molecule, and may bind to Gloop 2 in a slightly different conformation than that of native HEL. Indeed, Gloop 2 binds HEL with a somewhat lower affinity than the 'loop' peptide (Darsley and Rees, 1985a). The requirement for conformational adjustment in the modelling complex may be a reflection of this possibility, and the high main chain temperature factors observed for the region of HEL in question suggest that such a change might be accommodated by the structure.

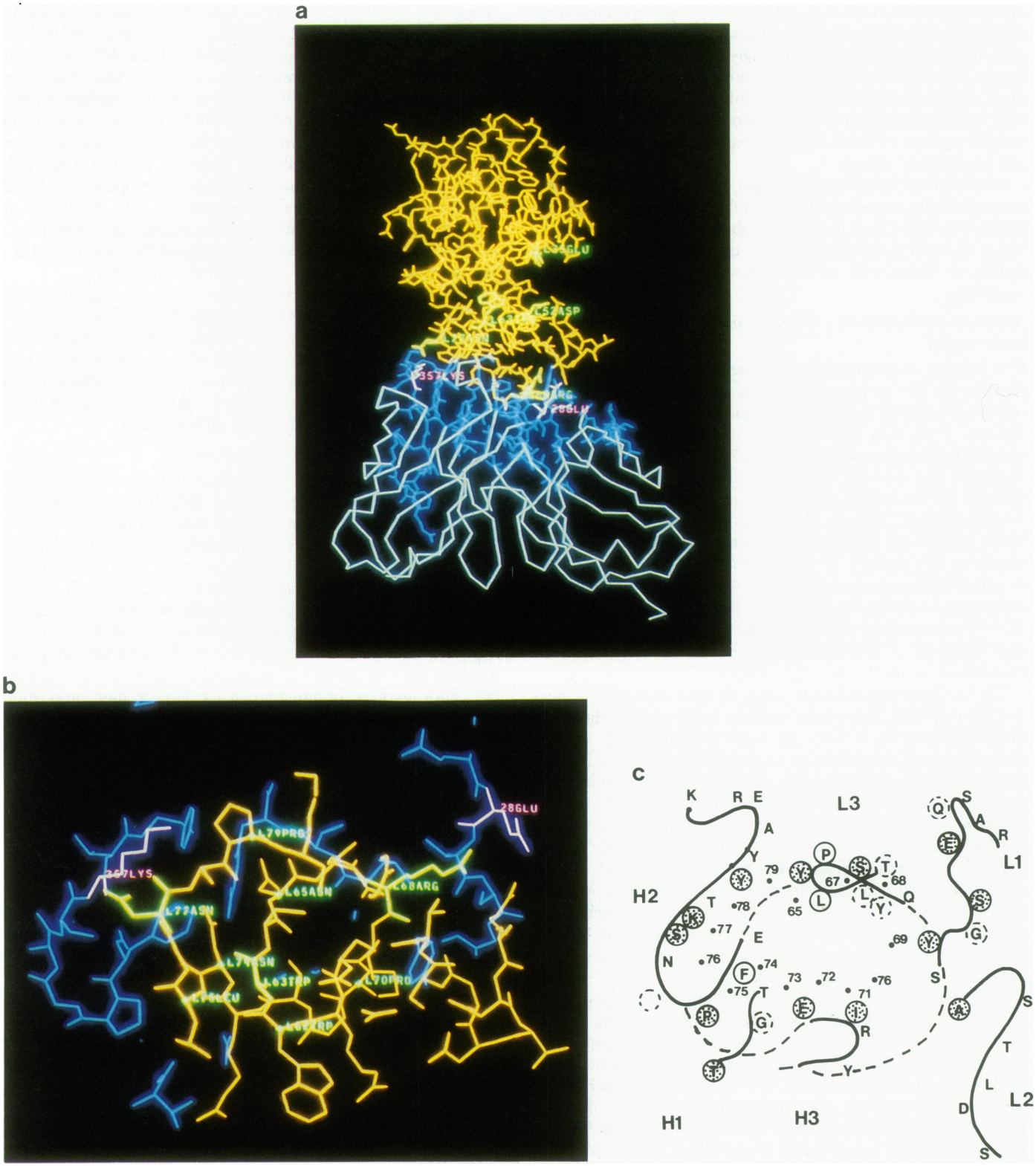

Fig. 5. **(a)** Preliminary model of the complex formed between the predicted Fv fragment of Gloop 2 and HEL. All-atom representations are shown for both the HEL molecule (yellow) and the Gloop 2 CDRs (blue). The Gloop 2 Fv α-carbon backbone is shown in white. The following HEL residues are highlighted in green: Glu 35, Trp 62, Trp 63 from the active site; Arg 68 and Asn 77 from the 'loop' epitope. The following Gloop 2 residues are shown in pink: Glu 28 (L1) and Lys 57 (H2). **(b)** A cross section through the contact surface between HEL and Gloop 2 viewed down the axis of the Fv β-barrel. Additional residues of HEL that participate in the interaction with the Gloop 2 surface are indicated in this view, in green. **(c)** Key to **(b)** above showing schematically the six CDRs [not all of which appear in **(b)**] and identifying the following residues: ● positions of α-carbon atoms of 'loop' contact residues; ⊗ predicted Gloop 2 contact residues; ○ residues conserved across all five Gloop antibodies; ◌ residues conserved between Gloop antibodies of the same specificity group.

Conclusions

In this study we have described an approach to predict and compare the structures of anti-protein antibody combining sites, using five antibodies generated against a series of overlapping epitopes on the surface of HEL as a model system. The boundaries of these epitopes have been defined, and also we have been able to model the three representative antibodies Gloop 2, 4 and 5. This has allowed us to compare the nature of the combining sites of the antibodies and correlate them with their epitopes, a knowledge of whose structure is an essential requirement in this approach. For one antibody, Gloop 2, we have proposed a model of the complex with native HEL.

The overall conclusions are: (i) that CDR length, in particular that of L1 and H3, is an important factor in determining the size and architecture of the combining site and that this effect is modulated by differences in residue volume; (ii) that an extensive region of the surface of the combining site is involved in the interaction including residues from all CDRs; (iii) that electrostatic interactions may be important in establishing the initial orientation between antibody and antigen, followed by adjustment to maximise stereochemical complementarity, perhaps strongly influenced by the high proportion of aromatic residues in the combining sites; (iv) that extensive contacts are made with the polypeptide backbone of the epitope; (v) that conformational changes may be necessary in either antibody or antigen, or both, in order to optimise interactions when an anti-peptide antibody combines with the native protein antigen.

These conclusions are part observation, part prediction. To test the prediction element we are engaged in studies of the Fab fragments and Fab-HEL complexes by X-ray crystallography and n.m.r. Since the acquisition of structural information by X-ray methods is a time-consuming process we have opted to carry out a parallel strategy (outlined in detail by Rees and de la Paz, 1986) in which the cloned antibody H and L genes are subjected to site-directed mutagenesis, expressed in *Xenopus* oocytes and subjected to binding analysis. The model of the Gloop 2–HEL complex has provided the basis for initial selection of residues for mutation. In this manner we hope to test the models that we have produced, and dissect the components of the specific interaction with antigen. This approach will, we hope, lead to a better understanding of the factors governing antibody affinity and specificity and permit the rational design of combining sites with predetermined specificity.

Acknowledgements

We would like to thank Professor Sir David Phillips for stimulating the development of this project in its early stages and for continued support and critical comments. We also thank Dr Michael Sternberg, Dr Garry Taylor and Dr Janet Thornton for many helpful discussions. We are also grateful to Drs Jane Burridge and Andrew Morffew of Graphics Systems Research, IBM Scientific Centre, Winchester, UK for producing the all-atom representations shown in Figure 4. We acknowledge the provision of the most recent co-ordinates for MCPC603 from Dr David Davies, NIH, Bethesda and thank the Medical Research Council and The Royal Society (BJS) for support and the Science and Engineering Research Council through grants to The Oxford Enzyme Group.

References

Amit,A.G., Mariuzza,R.A., Phillips,S.E.V. and Poljak,R.J. (1985) *Nature*, **313**, 156–158.

Arnon,R. and Sela,M. (1969) *Proc. Natl. Acad. Sci. USA*, **62**, 163–170.

Artymiuk,P.J., Blake,C.C.F., Grace,D.E.P., Oatley,S.J., Phillips,D.C. and Sternberg,M.J.E. (1979) *Nature*, **280**, 563–568.

Benjamin,D.C., Berzofsky,J.A., East,I.J., Gurd,F.R.N., Hannum.C., Leach, S.J., Margoliash,E., Michael,J.G., Miller,A., Prager,E.M., Reichlin,M., Sercars,E.E., Smith-Gill,S.J., Todd,P.E. and Wilson,A.C. (1984) *Annu. Rev. Immunol.*, **2**, 67–101.

Burley,S.J. and Petsko,G.A. (1985) *Science*, **229**, 23–28.

Colman,P.M., Schramm,H.J. and Guss,J.M. (1977) *J. Mol. Biol.*, **16**, 73–79.

Darsley,M.J. and Rees,A.R. (1985a) *EMBO J.*, **2**, 383–392.

Darsley,M.J. and Rees,A.R. (1985b) *EMBO J.*, **2**, 393–398.

Davies,D. and Metzger,H. (1983) *Annu. Rev. Immunol.*, **1**, 87–117.

Edmundson,A.B., Ely,K.R., Girling,R.L., Abota,E.E., Schiffer,M., Westholm,F.A., Fausch,M.D. and Deutsch,H.F. (1974) *Biochemistry*, **13**, 3816–3827.

Edmundson,A.B., Ely,K.R. and Herron,J.N. (1984) *Mol. Immunol.*, **21**, 561–576.

Epp,O., Latham,E., Schiffer,M., Huber,R. and Palm,W. (1975) *Biochemistry*, **14**, 4943–4952.

Fehlhammer,H., Schiffer,M., Epp,O., Colman,P.M., Lattman,E.E., Schwager,P. and Steigemann,W. (1975) *Biophys. Struct. Mech.*, **1**, 139–146.

Furey,W., Wang,B.C., Yoo,C.S. and Sax,M. (1983) *J. Mol. Biol.*, **167**, 661–692.

Greer,J. (1981) *J. Mol. Biol.*, **153**, 1027–1042.

Kabat,E.A., Wu,T.T., Bilofsky,H., Reid-Miller,M. and Perry,H. (1983) *Sequences of Proteins of Immunological Interest.* U.S. Department of Health and Human Services, NIH, Bethesda.

Levitt,M. (1974) *J. Mol. Biol.*, **82**, 393–420.

Mainhart,C.R., Potter,M. and Feldmann,R.J. (1984) *Mol. Immunol.*, **21**, 69–478.

Marquart,M., Deisenhofer,J., Huber,R. and Palm,W. (1980) *J. Mol. Biol.*, **141**, 369–391.

Novotny,J. and Haber,E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4592–4596.

Novotny,J., Bruccoleri,R., Newell,J., Murphy,D., Haber,E. and Karplus,M. (1983) *J. Biol. Chem.*, **258**, 14433–14437.

Novotny,J., Handschumacher,M., Haber,E., Bruccoleri,R.E., Carlson,W.B., Fanning,D.W., Smith,J.A. and Rose,G.D. (1985) *journal?*, in press.

Olson,A.J., Lerner,R.A. and Hendrickson,W.A. (1984) *Nature*, **314**, 127–133.

Padlan,E.A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 2551–2555.

Padlan,E.A., Davies,D.R., Pecht,I., Givol,D. and Wright,C. (1976) *Cold Spring Harbor Symp. Quant. Biol.*, **41**, 627–637.

Read,R.J., Brayer,G.D., Jurasek,L. and James,M.N.G. (1984) *Biochemistry*, **23**, 6570–6575.

Rees,A.R. and de la Paz,P. (1986) *Trends Biochem. Sci.*, in press.

Rossman,M.G. and Argos,P. (1975) *J. Biol. Chem.*, **250**, 7525–7532.

Saul,F., Amzel,L. and Poljak,R. (1978) *J. Biol. Chem.*, **253**, 585–597.

Segal,D., Padlan,E., Cohen,G., Rudikoff,S., Potter,M. and Davies,D. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 4298–4302.

Sibanda,B.L. and Thornton,J.M. (1985) *Nature*, **316**, 170–174.

Smith-Gill,S.J., Wilson,A.C., Potter,M., Feldman,R.J. and Mainhart,C.R. (1982) *J. Immunol.*, **128**, 314–322.

Tainer,J.A., Getzoff,E.D., Alexander,H., Houghten,R.A., Thornton,J.M., Edwards,M.S., Taylor,W.R. and Barlow,D.J. (1985) *journal?*, in press.

Westhof,E.; Altschuh,D., Moras,D., Bloomer,A.C., Mondragon,A., Klug,A. and van Regenmortel,M.H.V. (1984) *Nature*, **311**, 123–126.

Received on 22 November 1985