J. Mol. Biol. (1991) **217**, 133–151

Structure, Function and Properties of Antibody Binding Sites

I. Saira Mian[1]†, Arthur R. Bradwell[2] and Arthur J. Olson[1]‡

[1]*Department of Molecular Biology*
Research Institute of Scripps Clinic
10666 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.

[2]*Department of Immunology*
Medical School, Birmingham University, B15 2TJ, U.K.

(Received 16 May 1990; accepted 19 July 1990)

Do antibody combining sites possess general properties that enable them to bind different antigens with varying affinities and to bind novel antigens? Here, we address this question by examining the physical and chemical characteristics most favourable for residues involved in antigen accommodation and binding. Amphipathic amino acids could readily tolerate the change of environment from hydrophilic to hydrophobic that occurs upon antibody–antigen complex formation. Residues that are large and can participate in a wide variety of van der Waals' and electrostatic interactions would permit binding to a range of antigens. Amino acids with flexible side-chains could generate a structurally plastic region, i.e. a binding site possessing the ability to mould itself around the antigen to improve complementarity of the interacting surfaces. Hence, antibodies could bind to an array of novel antigens using a limited set of residues interspersed with more unique residues to which greater binding specificity can be attributed. An individual antibody molecule could thus be cross-reactive and have the capacity to bind structurally similar ligands. The accommodation of variations in antigenic structure by modest combining site flexibility could make an important contribution to immune defence by allowing antibody binding to distinct but closely related pathogens.

Tyr and Trp most readily fulfil these catholic physicochemical requirements and thus would be expected to be common in combining sites on theoretical grounds. Experimental support for this comes from three sources, (1) the high frequency of participation by these amino acids in the antigen binding observed in six crystallographically determined antibody–antigen complexes, (2) their frequent occurrence in the putative binding regions of antibodies as determined from structural and sequence data and (3) the potential for movement of their side-chains in known antibody binding sites and model systems. The six bound antigens comprise two small different haptens, non-overlapping regions of the same large protein and a 19 amino acid residue peptide. Out of a total of 85 complementarity determining region positions, only 37 locations (plus 3 framework) are directly involved in antigen interaction. Of these, light chain residue 91 is utilized by all the complexes examined, whilst light chain 32, light chain 96 and heavy chain 33 are employed by five out of the six. The binding sites in known antibody–antigen complexes as well as the postulated combining sites in free Fab fragments show similar characteristics with regard to the types of amino acids present. The possible role of other amino acids is also assessed. Potential implications for the combining regions of class I major histocompatibility molecules and the rational design of molecules are discussed.

1. Introduction

Antibodies are powerful recognition and binding molecules that the immune system employs to eliminate foreign molecules. Antibody binding sites are formed by six hypervariable loops or complementarity determining regions (CDRs§). The CDRs,

† Present address: Sinsheimer Laboratory, Biology Department, University of California Santa Cruz, Santa Cruz, CA 95064, U.S.A.

‡ Author to whom reprint requests should be addressed.

§ Abbreviations used: CDR, complementarity determining region; MHC, major histocompatibility complex; FR, framework.

0022-2836/91/010133–19 $03.00/0

three from each of the heavy and light chain variable domains, are connected to a relatively invariant β-sheet framework (Alzari *et al.*, 1988; Davies & Metzger, 1983; Capra & Edmundson, 1977; Wu & Kabat, 1970). Early analysis of a data bank of complete and partial sequences of 415 light and 197 heavy chains demonstrated that CDRs are rich in aromatic residues (Kabat *et al.*, 1977). The combining region represents only a small part of the antibody molecule, whose overall three-dimensional structure is highly conserved. Although, the pairing of light and heavy chains can generate some antibody diversity, most of it is generated by the somatic recombination of variable region gene segments (Yancopoulos & Alt, 1986; Wysocki & Gefter, 1989). Such genetic mechanisms yield antibodies exhibiting extensive diversity in hypervariable loop sequences. This potential repertoire is estimated to be approximately 10^9 in mouse (Berek *et al.*, 1985). However, the initial repertoire that confronts an antigenic challenge is smaller than the potential repertoire, since it is restricted to the antibody specificities expressed on existing immunocompetent B cells at a point in time (Holmberg *et al.*, 1986). This available repertoire can yield an apparently unlimited repertoire of antigen binding specificities and affinities.

Although a single antibody has a unique three-dimensional structure, biophysical and biochemical evidence indicates that it is multispecific or cross-reactive (Richards *et al.*, 1975). This capacity to combine both with its inducing antigen and with antigens of similar or disparate structure augments the genetically determined antigen-binding capabilities of antibodies. The extent of molecular complementarity between determinants on the antigen molecule and amino acid residues in the combining site determines the degree of antibody specificity. Increased cross-reactivity, therefore, is at the expense of specificity and affinity.

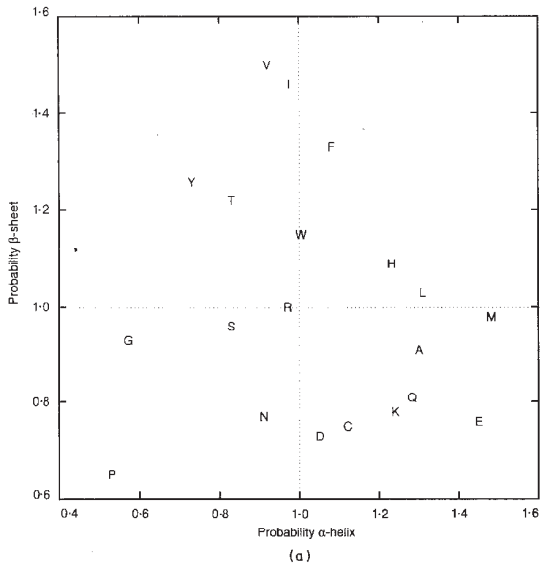
An improved understanding of both antibody cross-reactivity and binding can be obtained by a study of antibody–antigen interactions at the atomic level. The role of residues in the definition of combining site structure and interaction with antigen can be assessed as a function of the chemical and structural properties of individual amino acids. First, we examine those characteristics that appear to be of general importance in antibody–antigen interactions. This is followed by a detailed study of the binding sites in six antibody–antigen complexes and four free Fab fragments of known three-dimensional structure, and the much larger database of antibody sequences. Padlan (1990) has performed a similar, though not identical, analysis of antibody combining sites in general, and three anti-lysozyme antibody antigen complexes in particular. On the basis of their propensity to occur in the combining sites and their greater exposure relative to those in the framework regions, he has suggested that these amino acids determine specificity. Our results and their interpretation lead us to conclude that Tyr residues may play more generally important roles in binding and non-specific antibody–antigen interactions.

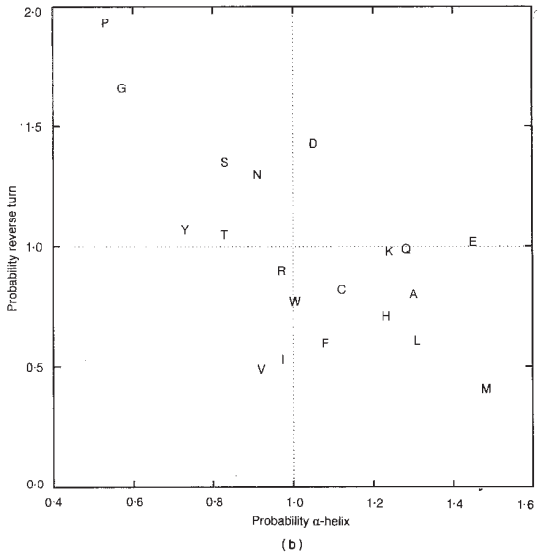
2. Physical and Chemical Properties of Amino Acids

Since antibody binding sites are formed by six hypervariable loops supported on a highly conserved β-sheet framework, there is likely to be a bias towards amino acids that are generally found in non-helical regions of proteins. Figure 1 shows the normalized frequencies of occurrence of amino acids in α-helix, β-sheet and reverse turns in 66 globular proteins comprising 31 different conformations (Levitt, 1978). In these structures, the occurrence of Pro, Gly, Tyr, Ser, Thr, Asn, Val, Arg, Ile and Trp in α-helices is less frequent than random. Leu, His, Trp, Thr, Tyr, Phe, Ile and Val have a greater than random probability of occurring in β-sheets; the same is true for Thr, Tyr, Asn, Ser, Asp, Gly and Pro in reverse turns. Arg appears to be equally tolerated in all the secondary structures elements considered. In general structural terms, Tyr and Thr seem to be the most useful non-helix forming residues, since they could be positioned in either the strand or turn regions of the hypervariable loops.

The free energy of interaction between an antibody and its antigen is a function of both enthalpy and entropy. Non-bonded forces between the interacting molecules include hydrophobic, hydrogen bond, van der Waals' and electrostatic interactions (for a review, see Fersht, 1985). In general terms, antibody combining site residues need to be as multifaceted as possible to accommodate the varied stereochemical and electronic features of the antigen. Hence, amino acids with non-polar (for example Leu, Ile and Val) and charged (for example Asp, Glu, Lys and Arg) side-chains would be of more limited usefulness than, for example, His, which is known to be capable of cross-linking sequentially distant but spatially close regions of proteins (Baker & Hubbard, 1984; I.S.M. & A.J.O., unpublished results). Similarly, the amides Asn and Gln would be generally more preferable than Asp and Glu, since the former pair are both hydrogen bond donors and acceptors whereas their charged counterparts are only acceptors.

If a positive charge is required in the antibody combining site, Arg would be more suitable than Lys because of its greater functional versatility; for example, Arg can form a larger number of hydrogen bonds than Lys. As a consequence of its planar nature and π-electron system, the terminal guanidinium group of Arg often exhibits pseudo-aromatic behaviour by participating in most of the interactions previously catalogued for true aromatic–aromatic interactions (I.S.M. & A.J.O., unpublished results). These interactions occur at the intersubunit interfaces of a number of oligomeric proteins, including viral coat proteins and a membrane protein; the photosynthetic reaction centre of *Rhodopseudomonas viridis*. The ability to form hydrogen bonds, hydrophobic interactions and





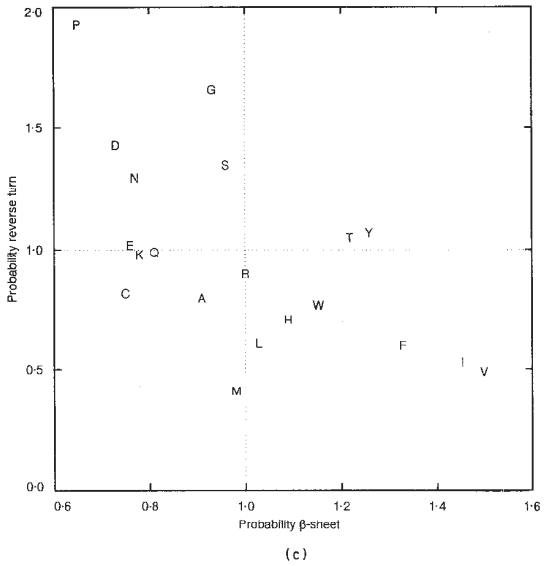


Figure 1. Scatter diagrams showing the normalized frequencies of occurrence of amino acids in α-helix, β-sheet and reverse turns in 66 globular proteins comprising 31 different conformations (Levitt, 1978). (a) Probability of forming α-helix *versus* β-sheet. (b) Probability of forming α-helix *versus* reverse turn. (c) Probability of forming β-sheet *versus* reverse turn. The values represent the ratio of the fraction of residues of each amino acid that occurred in the secondary structure element to this fraction for all residues. To eliminate a bias towards structures that were determined more than once, the values were each weighted by a factor of 1/(number of related proteins with same conformation). Normalized frequencies of 1 indicate random occurrence, whilst >1 indicate more frequent occurrence than random. The actual point is marked by the bottom left of the 1 letter amino acid code: A, Ala; R, Arg; N, Asn; D, Asp; C, Cys; Q, Gln; E, Glu; G, Gly; H, His; I, Ile; L, Leu; K, Lys; M, Met; F, Phe; S, Ser; T, Thr; W, Trp; Y, Tyr; and V, Val.

attractive electrostatic interactions between positively charged groups and aromatic rings permits Tyr and Trp to interact with structurally diverse antigens. Another functional advantage in locating Tyr and Trp in antibody combining sites is that, unlike amino acids having shorter side-chains, such as Asn and Ser, they lack the capacity to interact easily with other groups on the antibody surface but are ideally suited to interact with another molecule.

The accommodation of charged areas on the antigen need not necessitate an antibody combining site possessing amino acids of complementary charge. Analysis of Arg, Lys, Glu and Asp side-chains buried at the intermolecular interfaces of oligomeric systems indicates that oriented dipoles are usually preferred over countercharges in stabilizing these buried residues (I.S.M. & A.J.O., unpublished results). Thus, the peptide backbone and polar side-chains of hypervariable loop residues could be deployed to stabilize both negatively and positively charged regions. In some instances, this may be as effective as employment of formally charged amino acids: in cases of charge–charge interaction, the steric effects of neighbouring regions may prevent the formation of geometrically optimal ion-pairs such that the potentially available energy is not fully realized.

The non-covalent association between antibody and antigen requires the removal of water from surfaces buried by the interacting molecules. Antibody regions involved in this process should be capable of tolerating both the polar and non-polar environments that exist before complex formation and upon antigen binding, respectively. Individual residues exposed on the surface of the free antibody can become completely or partially buried in the complex. In addition to residue amphipathicity, residue size might be a factor. There is a good correlation between the surface area of amino acids and their free energies of transfer from water to an organic phase (Chothia, 1974, 1975; Gelles & Klapper, 1978). A value of 1 $Å^2$ (1 Å = 0·1 nm) of

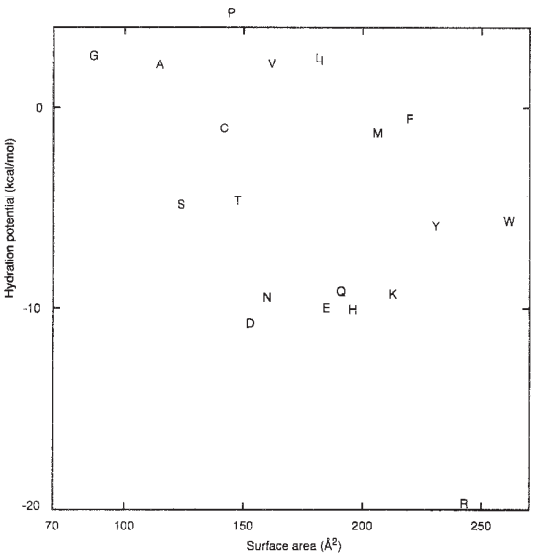


Figure 2. Comparison of the size of amino acids and the affinities of their side-chains for water. The surface area (Rose *et al.*, 1985) comprises the mean solvent accessibility for amino acid X in an ensemble of Gly-X-Gly tripeptides. The hydration potential (Wolfenden *et al.*, 1981) is the effective free energy of transfer from the vapour phase to dilute aqueous solution buffer at pH 7 of molecules having the structure R-H, where R is the side-chain of each amino acid; for P (Pro), only the surface area is indicated, since no hydration potential was evaluated. Side-chains were modelled by the following compounds: A, methane; R, methylguanidine; N, acetamide; D, acetic acid; C, methanethiol; Q, propionamide; E, propionic acid; G, H_2; H, 4-methylimidazole; I, isobutane; L, butane; K, *n*-butylamine; M, ethylmethyl sulphide; F, toluene; S, methanol; T, ethanol; W, 3-methylindole; Y, *p*-cresol; and V, propane. As a result of technical difficulties (Wolfenden *et al.*, 1981), methylguanidine (shorter than the side-chain of Arg by 2 methylene groups) was employed to estimate the value for propylguanidine; this probably leads to the hydrophilic and hydrophobic nature of Arg being over- and under-estimated, respectively.

surface area gives a hydrophobic energy of 25 cal/mol (1 cal = 4·184 J: Chothia, 1974). Whilst these van der Waals' energies may be small compared to a hydrogen bond, when summed over the entire combining site they may be important in stabilizing the complex. Figure 2 compares the affinities of amino acid side-chains for water (Wolfenden *et al.*, 1981) with the surface area of the entire amino acid (Rose *et al.*, 1985). The classical groupings into small, large, hydrophobic and hydrophilic amino acids are evident. With respect to amphipathicity, Ser, Thr, Tyr and Trp seem desirable residues to locate in antibody binding sites, since their side-chains are in the midrange of hydrogen potential values. The aromatic residues Tyr and Trp are also two of the largest and are capable of contributing significantly to the total interaction energy (Fig. 2).

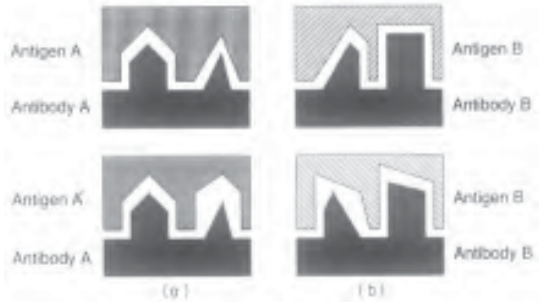


Figure 3. A diagram illustrating specificity and cross-reactivity for a given antibody. The specific binding of the antibodies A and B to antigens A and B is a function of the high degree of complementarity between their molecular surfaces in terms of shape, size and functionality. (a) Cross-reactivity may arise as a result of structural similarity of epitopes between antigens A and A′ (Richards *et al.*, 1975). A poor fit in one region may be compensated for by a good fit elsewhere. This could result in a sufficient number of short-range interactions to produce a stable antibody–antigen complex. Another cross-reacting antigen, A″, may fit the antibody combining site in a slightly different way. (b) The antibody B may accommodate the related antigens B and B′ if it is able to vary the stereochemical features of the combining site, i.e. if it is intrinsically pliable.

In an antibody–antigen complex, the stabilization energy gained from the various intermolecular forces must more than offset losses due to conformational entropy and conformational strain. The free energy of complex formation could therefore be maximized by minimizing the loss of conformational entropy upon association. It is known that a single antibody is able to combine with a spectrum of different antigens (Richards *et al.*, 1975). Although such cross-reaction may occur either because the antigens share epitopes, or because the epitopes are sufficiently similar in shape to bind the same antibody (Richards *et al.*, 1975), it could arise also if the topography of the combining site could be modulated (Fig. 3). Thus, antibodies could utilize amino acids whose side-chains were sufficiently structurally and functionally flexible to permit them to alter the stereochemical features of the combining site with minimal loss of entropy. The potential importance of side-chain motion has been further highlighted by a recent comparative study of known antibody structures and sequences (Chothia *et al.*, 1989). It has been suggested that the number of main-chain conformations of at least five of the six loops appears to be limited. The adoption of a specific backbone conformation is believed to be a reflection of only a few key conserved residues in the loop or framework of the antibody (Chothia *et al.*, 1989). This small repertoire of canonical structures would represent a reduction in the spectrum of specificity and affinity potentially available to the antibody binding site were the number of conformations proportional to the number of sequences that could be produced genetically.

Table 1

Preferred conformations of amino acid side-chains as described by their torsion parameters (Cody, 1985)

Residue	Torsion parameters (°)							
	ψ	χ^{11}	χ^{12}	χ^{21}	χ^{22}	χ^{3}	χ^{4}	χ^{5}
A	5; −19							
R	−15	60		180		180	±10	
		−60		180		180	±10	
N	11	60		0				
D	−7; ±35	−60		±5				
C	−30	−60						
		60						
Q	−20	60		180		−15		
		−60		180		25		
E	0; ±35	180		180		±15		
		−60		180		±15		
		60		180		±15		
G	0; 25							
H	−25	−60		60	−120			
		60		60	−120			
I	−15; −45	60	180	180				
		180	−60	180				
		−60	180	180				
L	−18; ±36	180		−60	180			
		−60		−60	180			
		60		60	180			
K	−20	−60		180		180	180	
M	±20	180		180		−60		
		−60		180		180		
		60		180		180		
F	±20	180		90				
		−60		90				
		60		90				
P	±10	±35		±35		±25	±20	±10
S	5	60						
T	−25	−60	180					
		60	−60					
W	−10	−60		90				
		60		90				
Y	±20	−60		90				
		180		90				
		60		90				
V	−11; −35	−60	180					

These were derived from crystallographic studies of amino acids and their derivatives. All atoms are numbered using the Greek letter designations starting with the C^{α} $\psi = (O\text{–}C\text{–}C^{\alpha}\text{–}N)$, $\chi^{11} = (N\text{–}C^{\alpha}\text{–}C^{\beta}\text{–}C^{\gamma 1})$, $\chi^{12} = (N\text{–}C^{\alpha}\text{–}C^{\beta}\text{–}C^{\gamma 2})$, $\chi^{21} = (C^{\alpha}\text{–}C^{\beta}\text{–}C^{\gamma}\text{–}C^{\delta 1})$, $\chi^{22} = (C^{\alpha}\text{–}C^{\beta}\text{–}C^{\gamma}\text{–}C^{\delta 2})$, $\chi^{3} = (C^{\beta}\text{–}C^{\gamma}\text{–}C^{\delta}\text{–}C^{\varepsilon})$, $\chi^{4} = (C^{\gamma}\text{–}C^{\delta}\text{–}C^{\varepsilon}\text{–}C^{\zeta})$. To account for the ring pucker in Pro, $\chi^{4} = (C^{\gamma}\text{–}C^{\delta}\text{–}N\text{–}C^{\alpha})$ and $\chi^{5} = (C^{\delta}\text{–}N\text{–}C^{\alpha}\text{–}C^{\beta})$. The most frequently observed values are given first.

The number of degrees of freedom of each amino acid side-chain can be approximated by examining the range and distribution of the observed conformations (Table 1). Residues with short side-chains such as Ser and Thr lose little entropy when fixed upon antigen association, since they have only one and two variable side-chain torsion angles, respectively. However, they project only a short distance from the surface of the antibody and so could not effect substantial changes in binding site topography. Large residues can elicit great changes in the surface contours of the combining site, since they can sweep out large volumes of space. Residues with the largest surface area (Fig. 2) are Trp, Arg, Tyr and Phe and of these, Arg has twice the number of variable torsion angles of the other three. Additionally, Arg is less suitable than the other aromatic residues because it is charged and therefore requires a more restrictive interaction at the interface.

Tyr represents a balance between the many different, though sometimes conflicting, desirable aspects we believe to be of general importance in antibody–antigen interactions. Thus, it would be expected to be the most common residue in antibody combining sites. Experimental evidence from antibody sequences and structures appears to verify these assumptions.

3. Combining Sites in Known Antibody–Antigen Complexes

Table 2 lists the antibody residues that bind antigen in six complexes whose structures have been determined by X-ray crystallography. The numbering scheme used follows the convention of

Table 2

Antigen binding residues observed in six crystallographically determined antibody–antigen complexes

Residue position		Antibody–antigen complex: McPC603	HyHEL-5	HyHEL-10	Fab D1.3	Fab 4-4-20	B13I2	Percentage composition
A. *Light chain*								
CDR1	27D	.	.	.	.	H	.	S24, H20, Y16, G14
	28	.	.	.	.	.	D	D17, N16, S16, Y9
	30†	.	.	G (30)	H	.	D	S25, N23, V14, L10
	31†	.	N (30)	N (31)	.	.	.	S33, N28, T18, H5
	32†	.	Y (31)	N (32)	Y	Y	Y	Y68, N6, S4, W3
	34	.	Y (33)	.	.	R	.	A30, N24, H20, S9
FR2	49	.	.	.	Y	.	.	Y83, G9, F3, S2
CDR2	50†	.	D (49)	Y (50)	Y	.	.	G19, D15, K11, R10
	53	.	.	Q (53)	.	.	.	S38, K18, T14, S11
CDR3	91†	D (97)	W (90)	S (91)	F	S	G	W27, Y22, S16, G12
	92†	.	G (91)	N (92)	W	.	.	N20, Y18, S15, D12
	93†	.	R (92)	.	S	.	.	S36, E14, H10, Y9
	94†	Y (100)	.	.	.	.	V	S17, N16, V12, L12
	95†	.	P (94‡)	.	.	.	.	P63, H10, L8, S6
	96†	L (102)	.	Y (96)	.	W	P	W19, Y18, L18, R10
B. *Heavy chain*								
FR1	30†	.	.	T (30)	T	.	.	S48, T42, K5, G1
CDR1	31†	.	.	S (31)	G	.	R	S39, D34, N7, R6
	32†	.	.	D (32)	Y	.	.	Y60, F17, S6, T6
	33	Y (33)	W (33)	Y (33)	.	W	A	Y40, G24, W16, A5
	35	E (35)	E (35)	.	.	.	.	H26, N21, S17, E15
FR2	47	.	W (47)	.	.	.	.	W94, Y3, L<1, F<1
CDR2	50	.	E (50)	Y (50)	.	.	.	Y19, E12, A12, R12
	51	.	.	.	.	.	I	I84, S9, F1, R1, V1
	52	R (52)	.	S (52)	W	.	S	N27, R16, S13, D12
	52A†	.	.	.	.	.	S	P54, N19, S9, L4
	53†	.	.	Y (53)	G	.	G	N29, G21, S11, A10
	54†	.	S (55)	S (54)	D	.	.	S27, N26, G23, D16
	55†	.	.	.	.	.	S	G56, Y19, S15, D3
	56	.	S (57)	S (56)	.	.	Y	S25, T24, G16, Y10
	57	.	T (58)	.	.	.	.	T74, I12, K3, S2
	58	.	N (59)	Y (58)	.	.	F	N20, Y19, L115, E13
CDR3	95	N (101)	G (99)	W (98)	.	.	Y	D28, G18, S10, Y10
	96	.	.	.	R	.	.	Y26, G12, R9, D6
	97	.	Y (101)	.	D	.	.	Y32, G14, D6, L6
	98	.	.	.	Y	.	.	Y32, G19, V7, S6
	99	.	.	.	R	.	P	G26, Y21, S16, E5
	100	.	.	.	.	.	F	S26, Y12, F12, R7
	100A	.	.	.	.	Y	.	S23, F12, G10, Y8
	100B	.	.	.	.	Y	.	Y29, F12, D11, S11
	100I	W (107)	.	.	.	.	.	W35, Y34, A10, M6

McPC603 (Satow *et al.*, 1986), HyHEL-5 (Sheriff *et al.*, 1987), HyHEL-10 (Padlan *et al.*, 1989), Fab D1.3 (Amit *et al.*, 1986), Fab 4-4-20 (Herron *et al.*, 1989) and B13I2 (Stanfield *et al.*, 1990). Residue positions, framework (FR) and complementarity determining regions (CDR) are from Kabat *et al.* (1987). The sequential residue numbers of structures in the Protein Data Bank (Bernstein *et al.*, 1977): McPC603 (File 2MCP), HyHEL-5 (1HFL) and HyHEL-10 (3HFM), are indicated in parentheses. At each of the positions known to bind antigen, the column headed Percentage composition indicates the 4 most common amino acids and their frequencies as calculated from the summary tables in Kabat *et al.* (1987). For example, one of the fluorescein binding residues in Fab 4-4-20 is His L27D (located in CDR1). Examination of the sequence database at this location indicates that Ser, His, Tyr and Gly are found in 74% of all sequences (24%, 20%, 16% and 14%, respectively) with the remaining 26% comprising the other 16 amino acids (each is <14% of the total). Asn, Asp, Gln and Glu are underestimated because positions given as Asx or Glx by Kabat *et al.* (1987) (i.e. where there are uncertainties in sequence data), whilst being included in the total number of sequences, are not incorporated into totals for these 4 amino acids. Although positions L95 and L96 are listed separately, examination of the crystal structures of McPC603, HyHEL-5 and HyHEL-10 indicates that Leu L96 (102), Pro L95 (94) and Tyr L96 (96) are structurally equivalent (see also Fig. 5). In HyHEL-10, Ser L93 (93) and Trp L94 (94) are listed as being partially buried by the antigen lysozyme (Padlan *et al.*, 1989). In B13I2, Leu L27D, Cys H32, Trp H47, Gly H50, Ser H96, Ser H97 and Asp H98 are buried by, but not in van der Waals' contact with, the antigen (Stanfield *et al.*, 1990).

Also marked are residues in the hypervariable and framework regions that a comparative study of known antibody structures and sequences (Chothia *et al.*, 1989) has suggested as being important (†) and mainly responsible (‡) for generating the observed main-chain conformations of 5 of the 6 hypervariable loops (predictions for CDR3 of the heavy chain were not made). Amongst all the antigen binding residues in these complexes, only Pro L95 (94) of HyHEL-5 belongs to the group of key residues. Other positions that are mainly responsible for producing the canonical structures occur at the following positions. Light chain: 2 (1/4, i.e. out of a total of 4 classes in which this is important for the observed conformation, only in canonical structure number 1 is this a residue that is a primary determinant of the loop main chain conformation), 25 (1/4), 29 (1/4), 33 (1/4), 48 (1/1), 64 (1/1), 71 (1/4), 90 ([1, 2, 3]/3), 95 ([1, 3]/3). Heavy chain: 26 (1/2), 27 (1/2), 29 (1/2), 34 (1/2), 52a (2/4), 55 ([1, 4]/4), 71 ([2, 3, 4]/4).

Kabat *et al.* (1987) with L (light) and H (heavy) prefixing the position number (the sequential residue numbers are given in parentheses where available). The complexes are McPC603 (Satow *et al.*, 1986: murine myeloma IgA-κ), HyHEL-5 (Sheriff *et al.*, 1987: BALB/c murine monoclonal antibody, IgG1-κ), HyHEL-10 (Padlan *et al.*, 1989: BALB/c murine monoclonal antibody, IgG1-κ), Fab D1.3 (Amit *et al.*, 1986: BALB/c murine monoclonal antibody, IgG1-κ), Fab 4-4-20 (Herron *et al.*, 1989: BALB/c murine monoclonal antibody, IgG2a-κ) and Fab B13I2 (Stanfield *et al.*, 1990: murine monoclonal antibody, IgG1-κ). McPC603 and Fab 4-4-20 bind small haptens (phosphocholine and fluorescein, respectively), B13I2 binds a synthetic 19 amino acid residue peptide (a homologue of the C-helix of myohaemerythrin) and the remainder bind different regions of the same large protein antigen (lysozyme). Because of the limited number of X-ray structures of antibody–antigen complexes, Table 2 includes the frequency of different amino acids obtained from the more extensive antibody sequence database (Kabat *et al.*, 1987). These potential binding positions are obtained by projecting the information from the structurally determined interface residues to those same locations in the larger sequence database. For example, position L91 binds antigen in all the structurally determined complexes considered and at this location in other antibody sequences, the four most common amino acids are Trp, Tyr, Ser and Gly.

In spite of differences in the stereochemical features of the antigen (the anti-lysozyme antibodies bind to *different* regions of the protein), there appears to be an overall bias in the types of amino acid found in the combining sites. At both the known and potential binding positions, Tyr, Trp, Ser and Asn are the most common residues that interact with antigen. As outlined in the previous section, these four, particularly Tyr and Trp, possess structural and functional characteristics that are highly desirable for antibody binding sites. Details of the combining region of HyHEL-5 and McPC603 (Fig. 4) demonstrate the extensive binding repertoire and versatility of Tyr and Trp in accommodating both small and large ligands. In the remaining complexes (HyHEL-10, Fab D1.3, Fab 4-4-20 and Fab B13I2), these residues also employ both the aromatic ring system and hydrogen-bonding atoms in interacting with the hapten or antigen (Padlan *et al.*, 1989; Amit *et al.*, 1986; Herron *et al.*, 1989; Stanfield *et al.*, 1990).

Table 2 indicates that the concentration of Tyr is highest in the heavy chain. This may account for the possession of antigen-binding affinities by complete variable heavy domains (Ward *et al.*, 1989). CDR3 of the heavy chain contains 19 amino acid positions, including a frequently inserted eight residue loop (Kabat *et al.*, 1987). Tyr comprises 25% of the total residues; in 11 of the positions it is the most common residue and many individual sequences contain consecutive Tyr residues, including five in two sequences (human heavy chain CDR3 of proteins OU and WOL; Kabat *et al.*, 1987). Chothia *et al.* (1989) were unable to make predictions about the conformation of this loop. Another observation from the sequence data (Kabat *et al.*, 1987) is that Tyr and Trp frequently alternate with small amino acids such as Gly, Ala and Ser. This pattern of residues might allow maximum mobility of Tyr and Trp. In HyHEL-5, the internally directed residue, Met L33(32), is located between Tyr L32(31) and Tyr L34(33) and permits the side-chains of the flanking aromatics to be exposed and potentially mobile. This Met is one of the residues responsible for stabilizing canonical structure 1 of the light chain CDR1 (Chothia *et al.*, 1989).

The number of locations involved directly or indirectly in antigen binding that are additional to those observed in the structures examined should be relatively small. Nuclear magnetic resonance spectroscopy studies of three different antibodies raised against the same 14 amino acid residue peptide antigen (Anglister & Zilber, 1990) indicate that the aromatic residues that interact with antigen are all situated at positions given in Table 2 (His L31, Tyr L32, Phe L96, Tyr H32, Trp H50, Trp H100A). The same is true for all but one position (Lys H100c) believed to be potentially involved in hapten binding in the preliminary crystal structure of Fab 36-71, an anti-arsonate monoclonal antibody (Rose *et al.*, 1990). The other presumed binding positions are Phe L32, Phe L50, Phe L96, Asn H35, Trp H47, Tyr H50 and Glu H96. Figure 5 shows all residues observed to bind antigen in McPC603, HyHEL-5 and HyHEL-10, and two potential supplemental positions, H59 and H94. H59 is implicated by studies on anti-morphine (Miller & Glasel, 1989) and anti-common acute lymphocytic leukaemia antigen (Kudo *et al.*, 1985) antibodies. In addition, this is the only position that is highly conserved amongst all six CDRs; Kabat *et al.* (1987) indicated that 96% of all antibody sequences have Tyr at H59. The importance of H94 is supported by the observation that changes in affinity of variant anti-digoxin antibodies could be accounted for by a mutation at this position (Panka *et al.*, 1988).

As the number of crystallographically determined complexes with greater diversity in the nature of antigen increases, the types of amino acid found to bind antigen should primarily remain limited to those already observed in McPC603, HyHEL-5, HyHEL-10 Fab D1.3, Fab 4-4-20 and Fab B13I2. There may be minor individual variations to accommodate properties peculiar to the inducing antigen. Furthermore, of all possible positions in the six hypervariable loops (Kabat *et al.* (1987) defined a total of 85 CDR positions), the location of such residues should largely be confined to a subset of those indicated in Table 2 (note that binding residues occur at framework as well as CDR positions). Whilst some locations are used in only one antibody–antigen structure, the most universal binding positions seem to be L91 (employed in all the complexes examined) and L32/L96/H33 (used

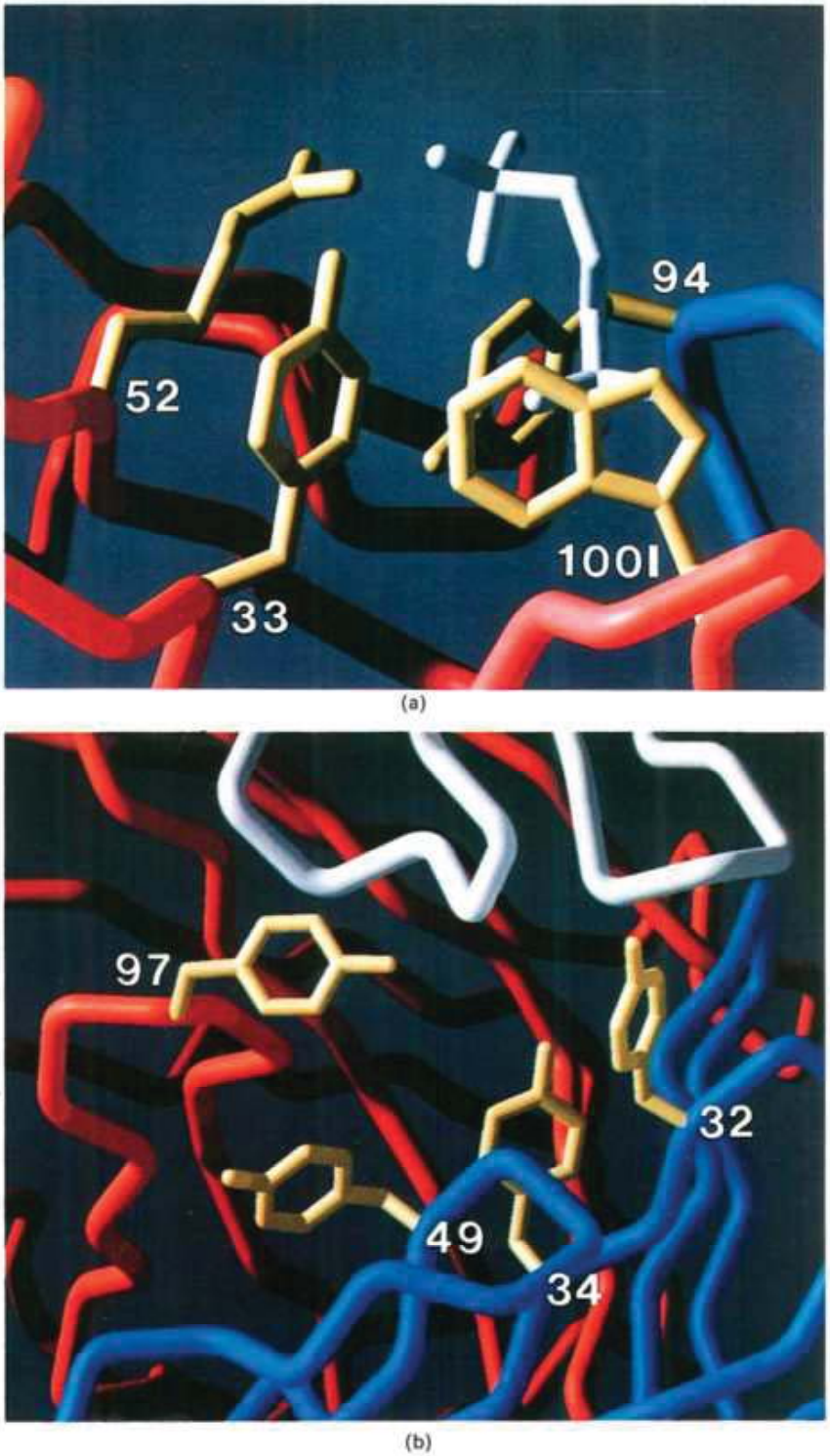


Figure 4. Details of the interactions between antigen (white) and Tyr, Trp and Arg binding site residues in the Fab monomer fragments from the complexes of (a) McPC603 and (b) HyHEL-5 (other binding amino acids present in Table 2 are not shown). The C^{α} backbone of the light (blue) and heavy chains (red) are represented as smooth tubes with Tyr, Trp and Arg side-chains in yellow. (a) The positively charged quaternary ammonium group of phosphocholine is stabilized by the π electron cloud of Trp H100I (107) and Tyr L94 (100). Tyr H33 (33) and Arg H52 (52) form hydrogen bonds with the terminal phosphate group; the partially positively charged hydrogens on the aromatic ring systems assist in stabilizing the negatively charged phosphate group. (b) Four Tyr residues, L32 (31), L34 (33), L49 (48) and H97 (101), that interact with antigen in at least 1 of the 3 anti-lysozyme antibodies listed in Table 2. The images were composed using GRAMPS (O'Donnell & Olson, 1981) and GRANNY (Connolly & Olson, 1985) and subsequently rendered with MCS (Connolly, 1985).

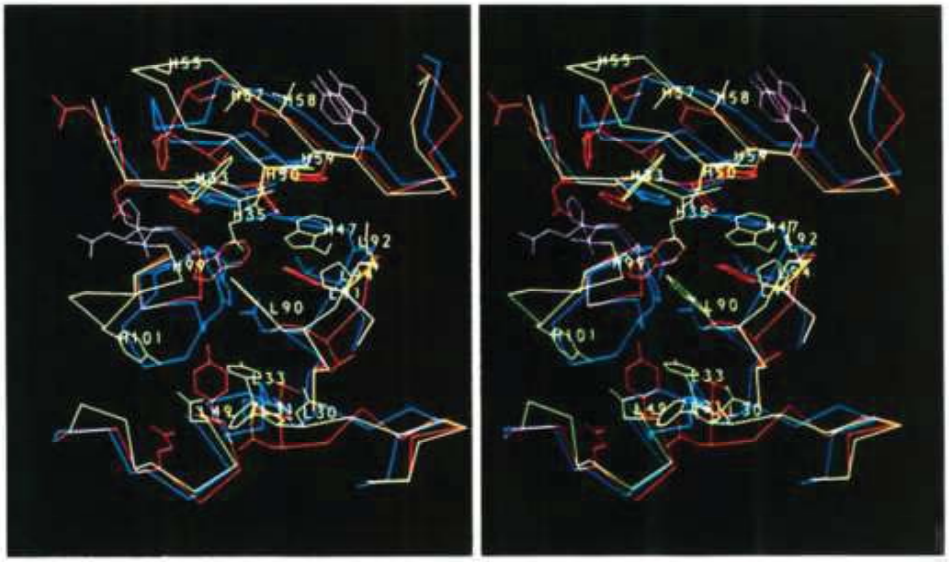


Figure 5. The combining site and antigen binding residues in 3 antibody–antigen complexes whose structures are available in the Brookhaven Data Base (Bernstein *et al.*, 1977): HyHEL-5 (yellow), McPC603 (blue) and HyHEL-10 (red). The 3 Fab fragments have been superimposed using the variable framework regions. Only the C^{α} backbone of the 6 CDR loops, the flanking FR positions, side-chains of amino acids given in Table 2 and the *sequential* residue numbers for binding position in HyHEL-5 are shown. Indicated in magenta are 2 locations that may be directly or indirectly involved in antigen binding in other antibodies: positions H59 (HyHEL-y, Tyr60; HyHEL-10, Tyr59; McPC603, Tyr62) and H94 (HyHEL-5, His98; HyHEL-10, Asn97; McPC603, Arg100). The view is looking down onto the combining site.

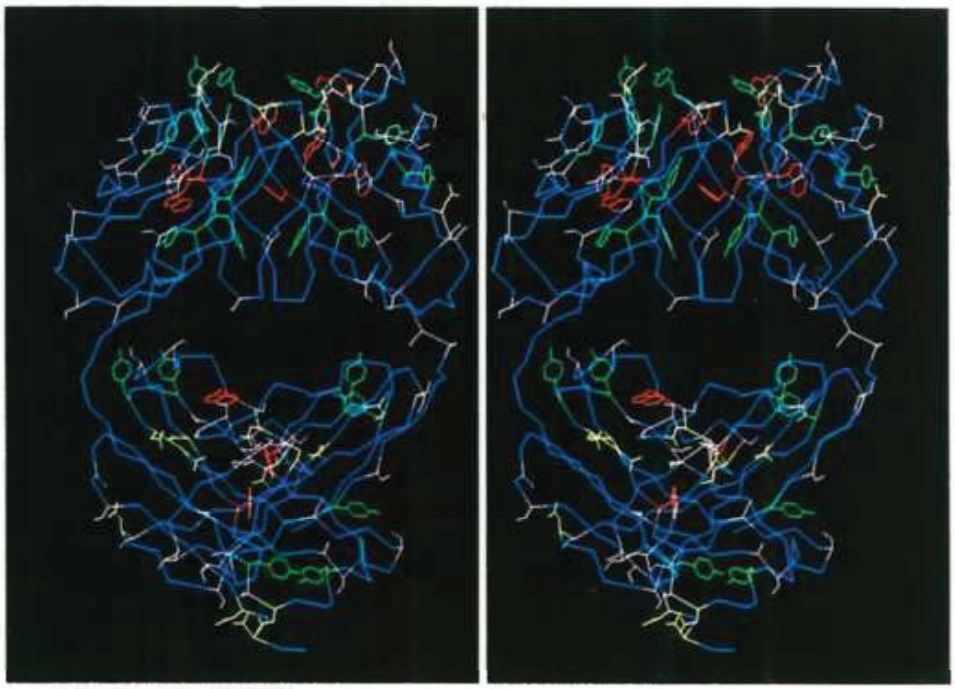


Figure 6. C^{α} carbon backbone of the Fab fragment from an antigen–antibody complex (HyHEL-5) showing the location of all Tyr (green), Trp (red), Ser (magenta) and Asn (yellow) residues. The heavy chain is on the right hand side and the antigen binding site is at the top of the image. The crystal structures of 2 other antigen–antibody complexes and 4 free fab fragments (see section 4) show similar distributors for these 4 amino acids.

by 5 out of 6 complexes; see also Fig. 5). These positions, L91 in particular, represent good sites for use in the redesign of antibody combining sites for specific functions.

4. Combining Sites and Amino Acid Distribution in Fab Fragments

From Table 2, it can be seen that the amino acids most likely to be found in antibody binding sites are Tyr, Trp, Ser and Asn. Figures 6 and 7 show the distribution of these residues in the entire Fab fragments from three of the complexes, McPC603, HyHEL-5, HyHEL-10, and four free Fab fragments whose combining sites are less well characterized; galactan-binding Fab J539 (Suh *et al.*, 1986: murine, IgA-κ), Kol (Matsushima *et al.*, 1978: human myeloma IgG1-λ), anti-*p*-azobenzenearsonate Fab R19.9 (Lascombe *et al.*, 1989: murine, IgG2b-κ) and vitamin K_1OH cross-reacting immunoglobulin New (Saul *et al.*, 1978: human myeloma IgG1-λ). In Fab R19.9, a large number of Tyr residues in the CDR regions have their side-chains pointing towards the solvent, and Lascombe *et al.* (1989) surmise that they may have an important function in antigen binding. The crystal structure of unliganded anti-single-stranded DNA Fab BV04-01 (monoclonal antibody isolated from mice with an autoimmune syndrome similar to human systemic lupus erythematosus) indicates that the putative combining site contains Trp, Tyr, His, Arg and Lys (Herron *et al.*, unpublished results).

The antigens for the crystallographically determined Fab molecules are quite diverse, ranging from saccharides through nucleic acids, peptides, large protein, small neutral organic and small charged inorganic molecules. Although there are no lipid-binding antibodies amongst them, the wide variation in the nature of the antigen suggests that the conclusions derived from an analysis of their cognate Fab structures should be applicable to those that bind lipids. Thus, using Table 2 and the additional locations given in the previous section, it should be possible to identify residues that are most likely to bind antigen in these Fab fragments and in other antibodies for which only sequence data are available.

Table 3 expands the information present in Table 2 by comparing the frequency of amino acids in different parts of antibody structures in relation to the probability of their binding antigen. In reading the Table from left to right, attention is gradually focused on the binding site, such that the changing concentrations of different amino acids become more apparent. An asymmetry in the location of residues across the antibody is clear. For example, some amino acids that are common in the CDR regions are uncommon in the known and potential binding positions and *vice versa*. Trp, Asn and Tyr exhibit a skewed distribution (see Fig. 7), such that the greatest accumulation occurs at or near the vicinity of the antibody binding site. In the case of Trp and Tyr, there is a marked preference

Table 3

Percentage composition of amino acids in different parts of antibody structures with vertebrate average shown for comparison (the preferred choice since the antibody sequence data are from several species)

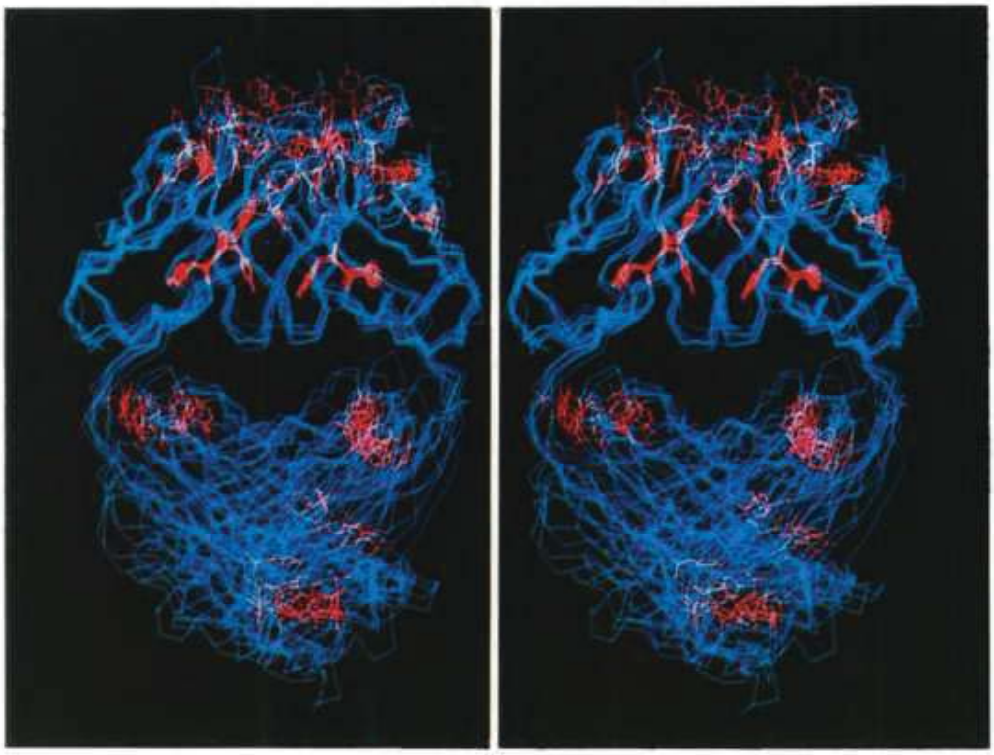
Amino acid	Vertebrate average	Variable regions	CDR regions	Potential binding positions	Known binding positions
A	7·2	6·44	7·06	3·50	1·1
R	4·9	3·55	3·80	2·84	5·7
N	4·1	2·55	7·44	11·17	6·8
D	5·4	3·81	5·09	6·14	8·0
C	2·5	2·03	0·13	0·16	0·0
Q	3·9	5·68	4·30	0·57	1·1
E	6·8	3·58	2·38	2·42	3·4
G	7·6	9·40	7·01	8·68	8·0
H	2·2	0·77	2·39	3·92	2·3
I	5·0	3·98	3·41	3·43	1·1
L	8·7	7·79	5·09	2·43	1·1
K	6·5	4·42	4·10	2·59	0·0
M	2·3	1·76	2·07	0·23	0·0
F	4·1	2·96	2·53	2·28	3·4
P	4·9	4·45	2·81	3·68	3·4
S	7·0	13·84	17·27	14·88	13·6
T	5·6	8·66	6·05	5·90	4·6
W	1·3	2·10	1·94	5·52	10·2
Y	3·2	5·31	11·08	17·30	25·0
V	6·8	6·90	4·05	2·34	1·1
Amino acids in database	300,000	110,816	27,007	15,054	88

The vertebrate average is from Doolittle (1986) and the variable and CDR region data from Kabat *et al.* (1987) summary tables. The columns headed Potential binding positions and Known binding positions show amino acid frequencies restricted to the known antigen binding positions of the complexes given in Table 2. Thus, Tyr comprises 25·0% (22 out of 88) of the observed binding residues and 17·30% (2605 out of 15,054) of the amino acids that occur at these positions in the sequences of the other antibodies examined. Asn, Asp, Gln and Glu are slightly underestimated in columns 2, 3 and 4 because of uncertainties in sequence data.

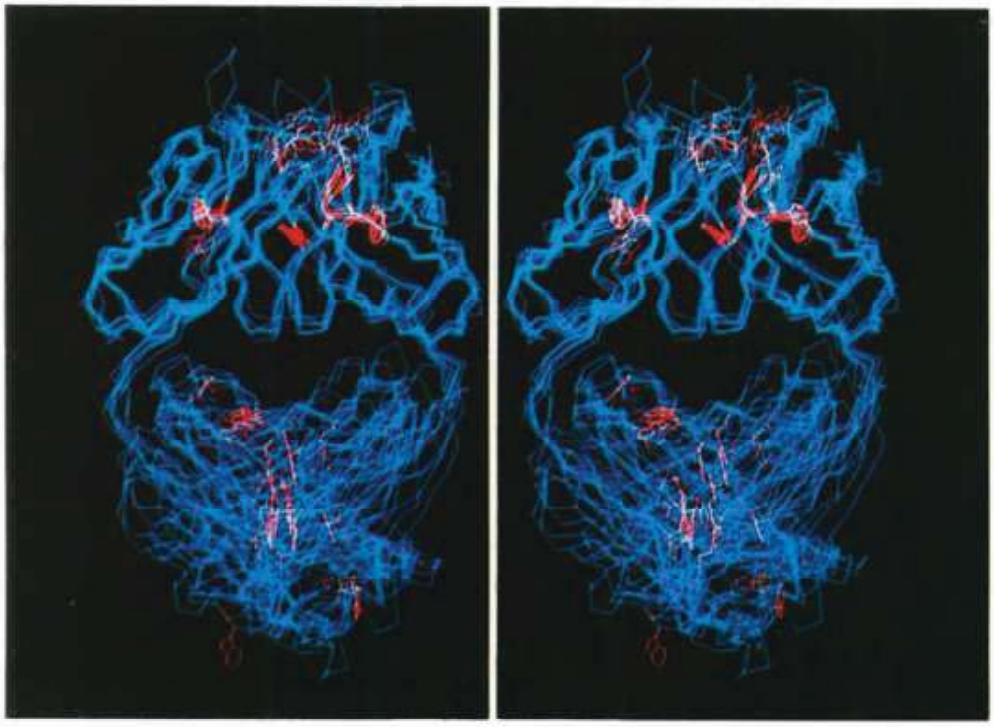
relative not only to the CDRs, but also the variable regions and proteins in general. Whilst Ser may be a common antibody combining site residue numerically, it is also often found in the β-strand regions of the CDR loops distal to the binding site. The common occurrence of Ser and Asn may be rationalized on structural grounds. Figure 1 indicates that these residues are preferentially found in β-sheet and reverse turns, the structural components of the antibody combining site.

5. Conformational Flexibility of Binding Site Residues

Figure 7 demonstrates the importance of Tyr and Trp in the binding sites of antibodies. Section 2 postulated that, since Tyr and Trp are large and each has only two rotatable side-chain torsion angles, they would seem particularly parsimonious residues to effect variation in the surface contours of

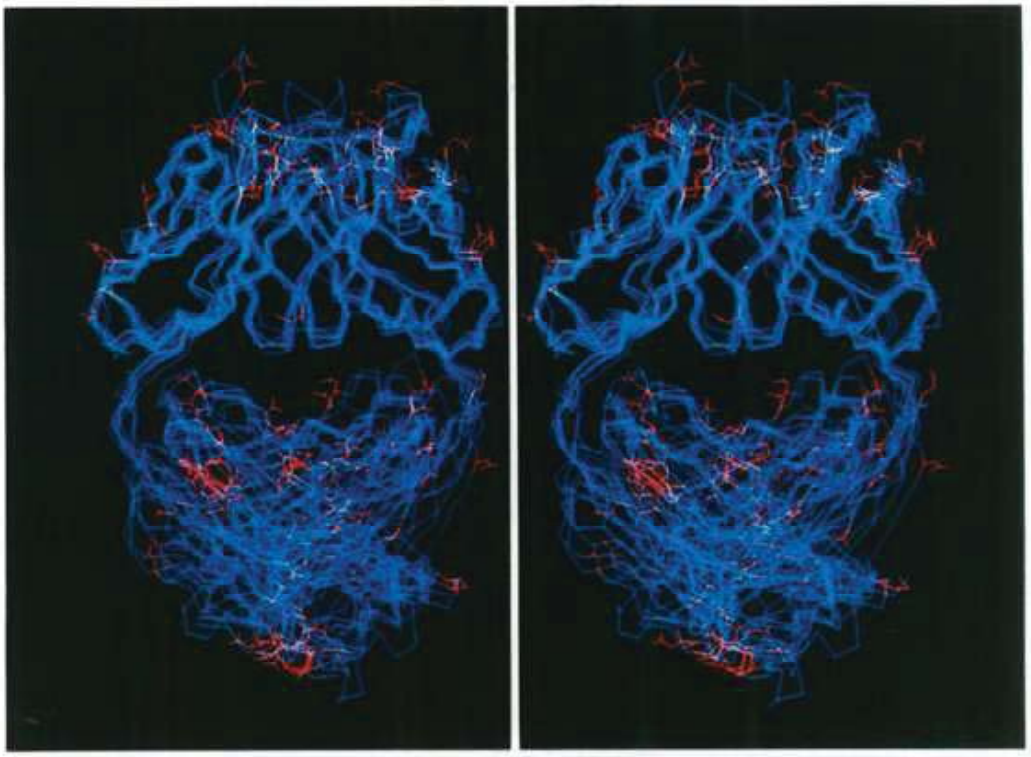


(a)

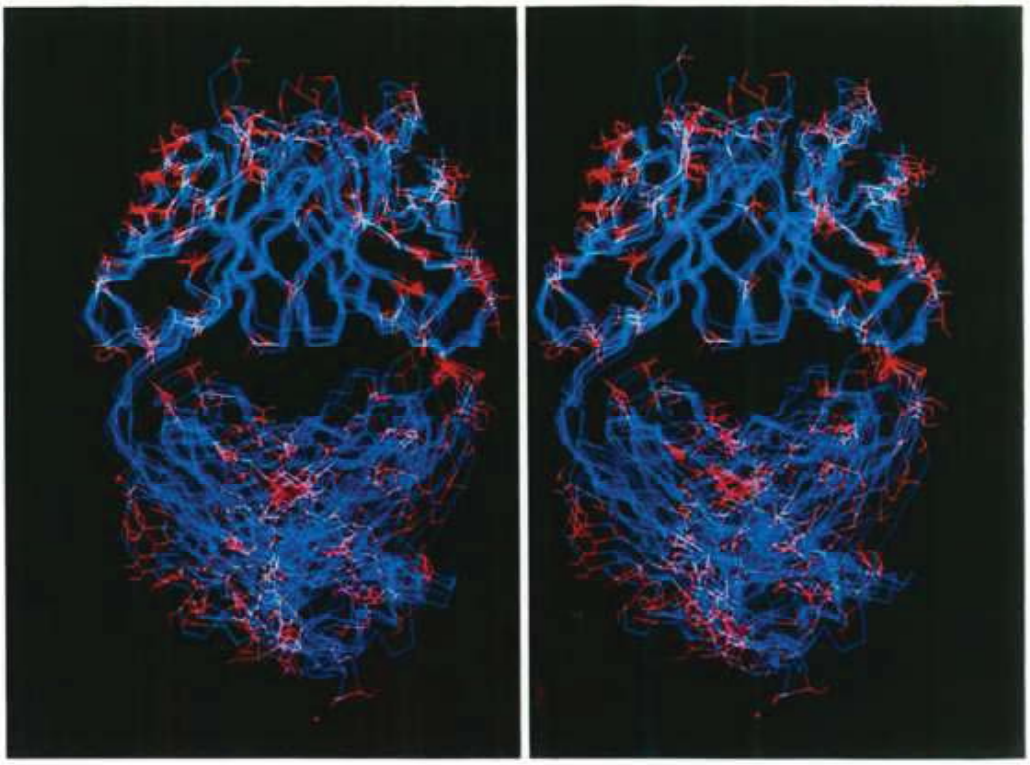


(b)

Fig. 7.

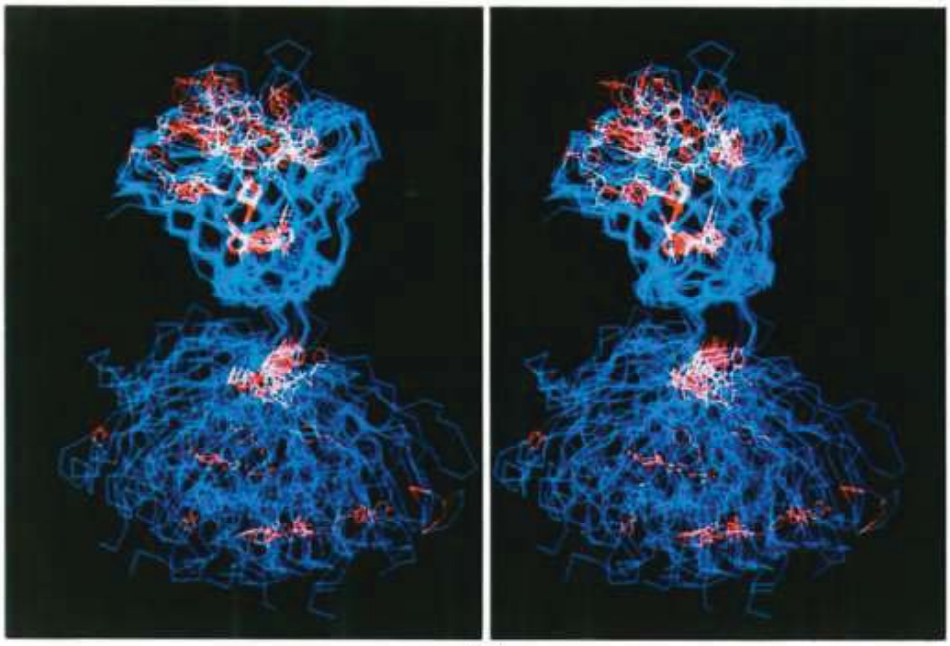


(c)



(d)

Fig. 7.



(e)

Figure 7. The location of all (a) Tyr, (b) Trp, (c) Asn, and (d) Ser residues in 7 Fab fragments whose Brookhaven Data Base (Bernstein *et al.*, 1977) codes are HyHEL-5 (1HFL), HyHEL-10 (3HFM), McPC603 (2MCP), J539 (1FBJ), Kol (2FB4), R19.9 (1F19) and New (3FAB). (e) An orthogonal view of (a). Only the C^{α} backbone and appropriate side-chains are shown. The molecules have been superimposed using only the framework amino acids of the variable regions (note how differences in elbow angle result in different conformations for the constant regions). The heavy chain is on the left-hand side and the antigen binding site is at the top of the picture. The highly conserved Tyr and Trp residues in the core of the variable region are Trp L35 (99% of all antibodies in the sequence database presented by Kabat *et al.*, 1987), Tyr L36 (78%), Tyr L86 (99%), Tyr L87 (74%), Trp H36 (99%), Trp H47 (94%), Tyr H90 (98%), Tyr H91 (81%) and Trp H103 (97%). Note (1) L87 and H91 precede Cys residues that form intrachain disulphides, (2) H47 is an antigen binding position and (3) L86, H91, L36 and H103 are at the junction of the heavy and light chains with the last 2 interacting with each other. In the constant regions, the highly conserved Tyr and Trp residues are Tyr L140, Trp L148, Tyr L173, Tyr H147, Trp H157 and Tyr H185. Note that the side-chains of L140, L173 and H185 point up towards the variable region and may be involved in motion between the 2 domains.

the antigen combining site (especially with regard to entropic cost), without alterations in backbone conformation. If necessary, however, additional main-chain movement could augment this feature. Such a mechanism could allow a variety of slightly differently shaped antigens to be accommodated and might thus optimize the antibody repertoire against evolving pathogens. X-ray crystallographic analysis of engineered protease has shown that structural plasticity, a combination of alternate side-chain conformations and binding site flexibility, permits both large and small substrates to be well accommodated in an enzyme binding site (Bone *et al.*, 1989). In the case of antibodies, the only direct experimental evidence showing amino acid movements to accommodate antigen interaction is that based upon the 2·8 Å crystal structures of free and complexed B1312 (Stanfield *et al.*, 1990). Significant movements in the side-chain of Phe H100 and the main chain of Pro H99 and Tyr H100B were observed between the native and complexed forms of the antibody.

Additionally, there is circumstantial and indirect evidence to support the idea that Tyr and Trp impart to the antibody binding site an ability to vary its geometry. Figure 8 illustrates how Tyr and Trp residues at structurally equivalent positions in the hypervariable loops of seven Fab fragments can have alternative conformations. Side-chain movement of these residues could change the contours of the binding site significantly beyond that which could be achieved by other sorts of residue; any main-chain movements could augment such a capability. Figure 9 shows a cluster of four Tyr residues from HyHEL-5 and the results of modelling their potential side-chain movement in the absence of bound antigen. Figure 9(b) illustrates how changes in side-chain conformation alone could considerably alter the surface topography of the antibody binding site. Preliminary data from a molecular dynamics simulation of this Fab monomer in solvent and with counterions indicate that the range of motions modelled is easily accessible to these residues (I.S.M., U. C. Singh & A.J.O., unpublished results). In spite of differences in specificity and amino acid sequence in the variable domains, the α-carbon backbones of unliganded Fab BV04-01 and hapten bound Fab 4-4-20 are remarkably similar

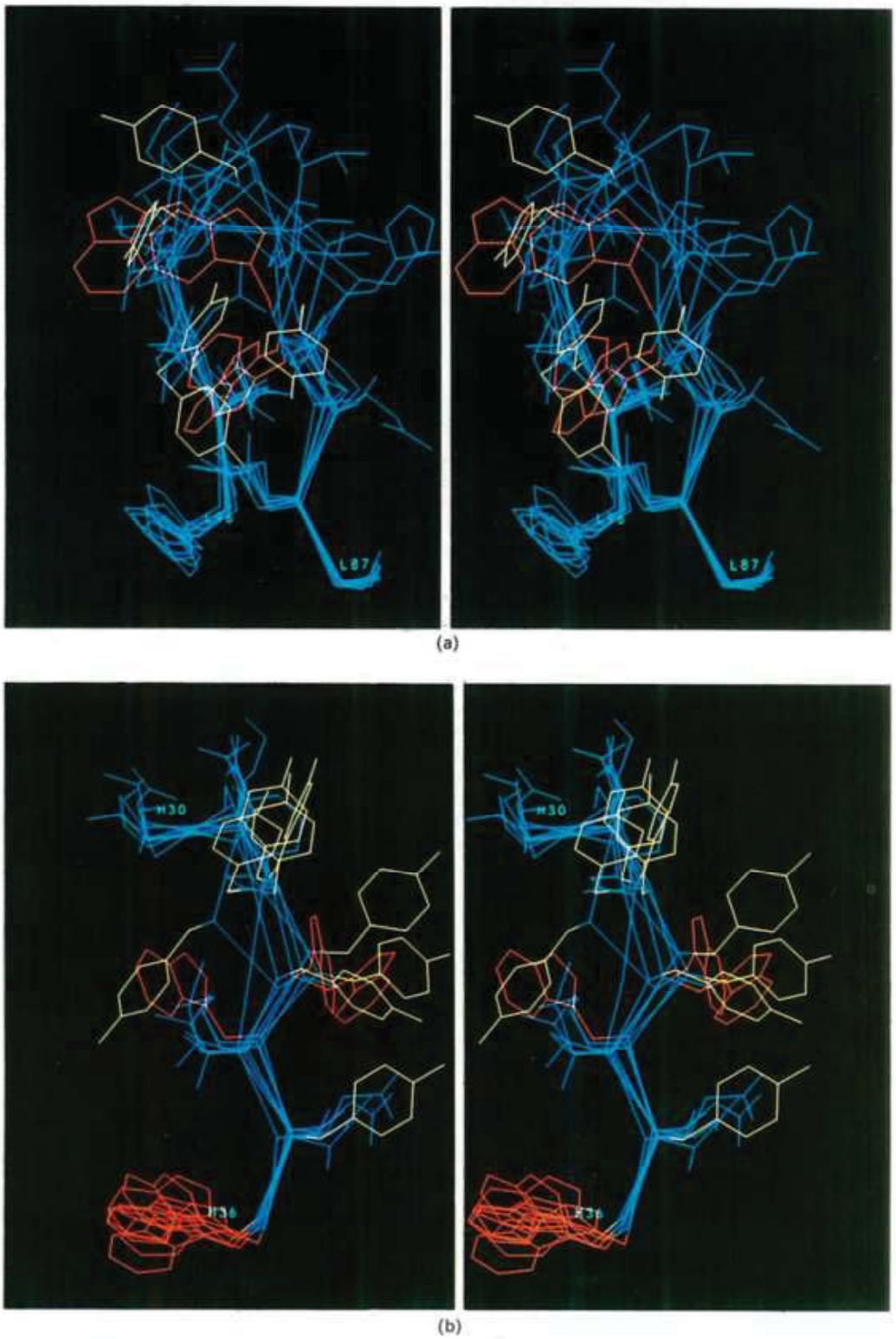
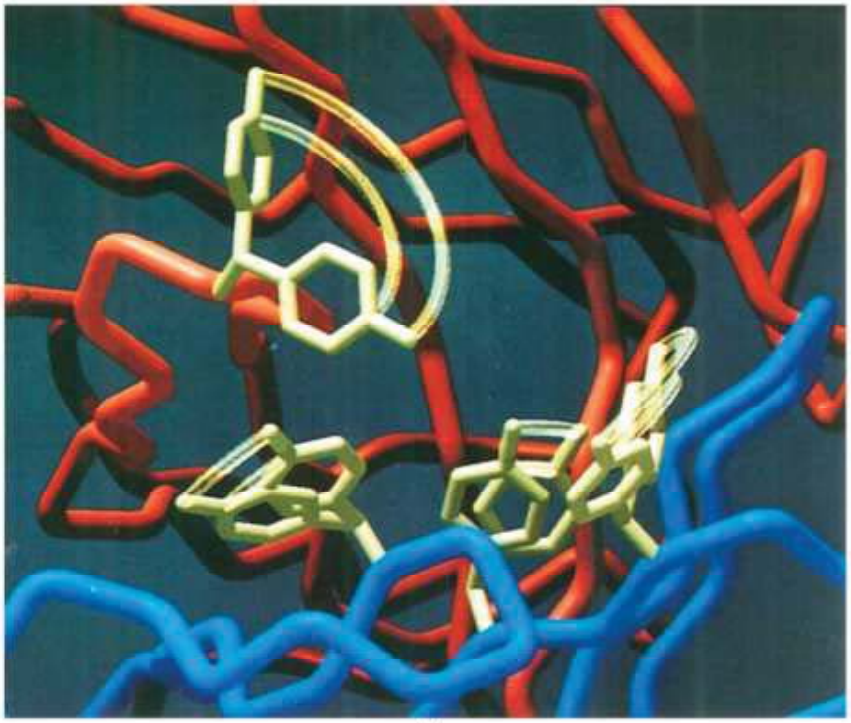
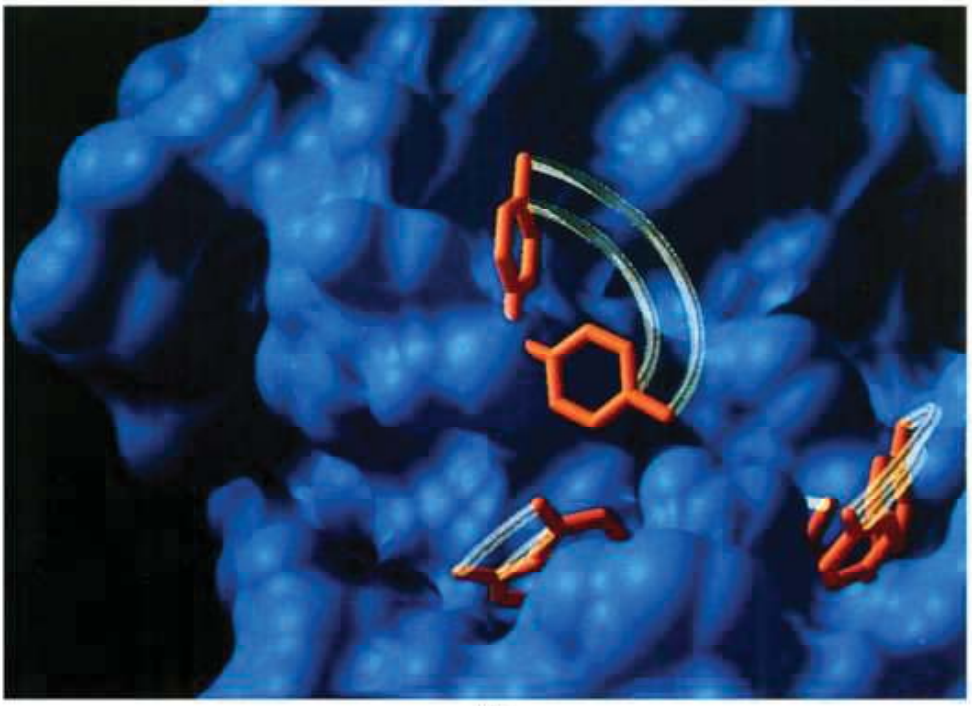


Figure 8. Two of the 6 hypervariable loops from the Fab fragments illustrated in Fig. 7, showing the C^{α} backbone and Tyr (yellow) and Trp (red) side-chains. (a) and (b) Light chains 3 and heavy chain 1 loops, respectively, which comprise the CDR and flanking FR residues. The structures were superimposed using the C^{α} carbon atoms of the first and last 2 residues whose sequential residue numbers in the HyHEL-5 structures are indicated. The other 4 hypervariable loops (not shown) also display the ability of Tyr and Trp residues at structurally equivalent positions to adopt different side-chain conformations.



(a)



(b)

Figure 9. (a) Potential movements of the 4 Tyr side-chains displayed in Fig. 4(b) in a hypothetical free HyHEL-5 molecule. These are based on the HyHEL-5/lysozyme complex in which the antigen has been removed. The arcs indicate the limits of rigid body rotations of C^{ζ} and O^{η} around χ^1 (with minor movements of χ^2 in some instances) such that no side-chain atom approaches closer than 3 Å of an atom from another residue. χ^1 rotations for these and other Tyr and Trp residues are: binding positions Tyr L32 (31), 115°; Tyr L34 (33), 25°; Trp L91 (90), 85°; Trp H33 (33), 100°; Trp H47 (47), 30°; Tyr H97 (109), 130°; non-binding Tyr L49 (50, invariant), 35°; Tyr H27 (27), 10°; Tyr H32 (32), 20°; Tyr H59 (60, invariant), 90°. (b) A molecular surface on the antibody (blue) indicates the scale of the modelled Tyr side-chain movements relative to the size of the entire binding site region. It can be seen how small, possibly co-ordinated, movements of these residues could alter the surface topography of the antigen binding site.

(Herron *et al.*, 1989); this suggests the importance of side-chain conformation.

In crystallographic studies on the binding of fluorescein and rhodamine to a model system consisting of an Mcg light chain (Bence–Jones) dimer, hapten-induced changes were observed in side-chains of key Tyr and Phe residues (Edmundson *et al.*, 1984); the polypeptide backbone exhibited lesser changes (Edmundson *et al.*, 1984). These aromatic residues appeared to produce a "conformational tuning" of the Mcg active site to improve ligand complementarity (Edmundson *et al.*, 1984). Thermodynamic measurements on the binding of fluorescein by three monoclonal anti-fluorescyl antibodies (4-4-20, 20-19-1 and 20-20-3) suggest hapten-induced conformational changes in the antigen binding region (Herron *et al.*, 1986). Nuclear magnetic resonance spectroscopy experiments on an anti-lysozyme humanized mouse F_v possessing the same CDRs as Fab D1.3 indicate perturbation of amide proton chemical shifts in six out of 12 Tyr upon antigen binding (L. Reichmann & P. E. Wright, personal communication). Although the effects due to the new framework are unknown, the results are suggestive of changes in the environments of these Tyr residues. The relevance of light chain dimers to real antibodies is unclear and the results from the thermodynamics measurements and nuclear magnetic resonance experiments could arise from very small changes, such as are almost always found on association. In addition, there is the possibility of conformational changes within the protein antigen itself. However, when taken in conjunction with the other observations, the arguments for conformational flexibility of Tyr and Trp residues in antibody binding sites is strong.

6. Antibody Affinity

In addition to their role in permitting the binding of differently shaped molecules, Tyr and Trp may contribute significantly to antibody affinity. In some known structures, Arg seems to play a similar role because it bears some physical and chemical resemblances to these aromatics (see section 3). One example of the importance of Tyr, Trp and Arg in antibody affinity involves Fab 4-4-20, which binds fluorescein with an association constant of $3{\cdot}4 \times 10^{10}$ M in aqueous solution (Herron *et al.*, 1989). In the crystal structure of the complex, Tyr, Trp and Arg comprise six out of the eight amino acids involved in hapten binding (Table 2) and are believed to make major contributions to the binding energy (Herron *et al.*, 1989). Studies on the active site structure and antigen binding properties of 4-4-20 and other cross-reactive anti-fluorescein monoclonal antibodies implicate a specific Arg (L34) in the increased affinity of 4-4-20 (Bedzyk *et al.*, 1990).

A different system displaying unusually high affinity is the non-covalent interaction between streptavidin and biotin (dissociation constant $K_d = 10^{-15}$ M; Green, 1975). This forms the basis of many diagnostic assays requiring the formation of irreversible and specific links between macromolecules. In the three-dimensional structure of the biotin–streptavidin complex, five of the residues lining the biotin-binding site are aromatic (Weber *et al.*, 1989). Trp79, Trp92, Trp108 and Trp120 form a hydrophobic binding site by packing around the tetrahydrothiophene ring whilst Tyr43 is situated to hydrogen bond with the biotin ureido oxygen (Weber *et al.*, 1989). In a study of germ-line affinity and germ-line variable-region genes in the primary B cell response (Chua *et al.*, 1987), the affinity of anti-5-dimethylaminonaphthalene-1-sulphonyl IgM antibodies was proposed to be primarily associated with CDR3 regions, CDR1 and CDR2 playing lesser roles. An apparent preference for Tyr codons in the D segment was noted (Chua *et al.*, 1987) and, as shown in Table 2, the concentration of Tyr is highest in the CDR3 regions of the heavy chain. Semi-quantitative estimates of Gibbs free energy changes accompanying complex formations of McPC603, HyHEL-5 and Fab D1.3 show that Tyr, Trp and Arg residues in the binding site make important contributions to the binding energy (Novotny *et al.*, 1989).

While Tyr, Trp and Arg appear to augment binding, they do not by themselves guarantee high affinity, since this is a function of the degree of complementarity between the interacting molecules. The combining site residues need to be positioned precisely in order to take full advantage of electrostatic, hydrogen bonding and van der Waals' interactions. This may not always be possible, as for example in the binding of saccharide antigens. Crystallographic studies of protein–carbohydrate complexes (Vyas *et al.*, 1988; Quiocho, 1988) indicates that the primary factors involved in substrate specificity and binding are intricate networks of cooperative and bidentate hydrogen bonds to both polar side-chain protein residues and solvent molecules. Further stabilization occurs *via* van der Waals' interactions with Tyr and Trp aromatic rings. In a saccharide-binding antibody, therefore, generation of the correct ligand site geometry with relatively large numbers of precisely oriented hydrogen bond donor and acceptor groups would necessitate the presence of residues other than Arg, Trp and Tyr.

7. A Model for Antibody Diversity and Specificity

In our model, the presence of Tyr and Trp produces an antibody binding site that behaves as a malleable template responsible for more general binding. Such a combining site can accommodate antigens that have never been encountered in the course of evolution. Successive generations could attempt to "fine tune" and/or improve specificity towards the novel antigen by modulating the nature and/or location of other residues that would play a greater role in providing the discriminatory power

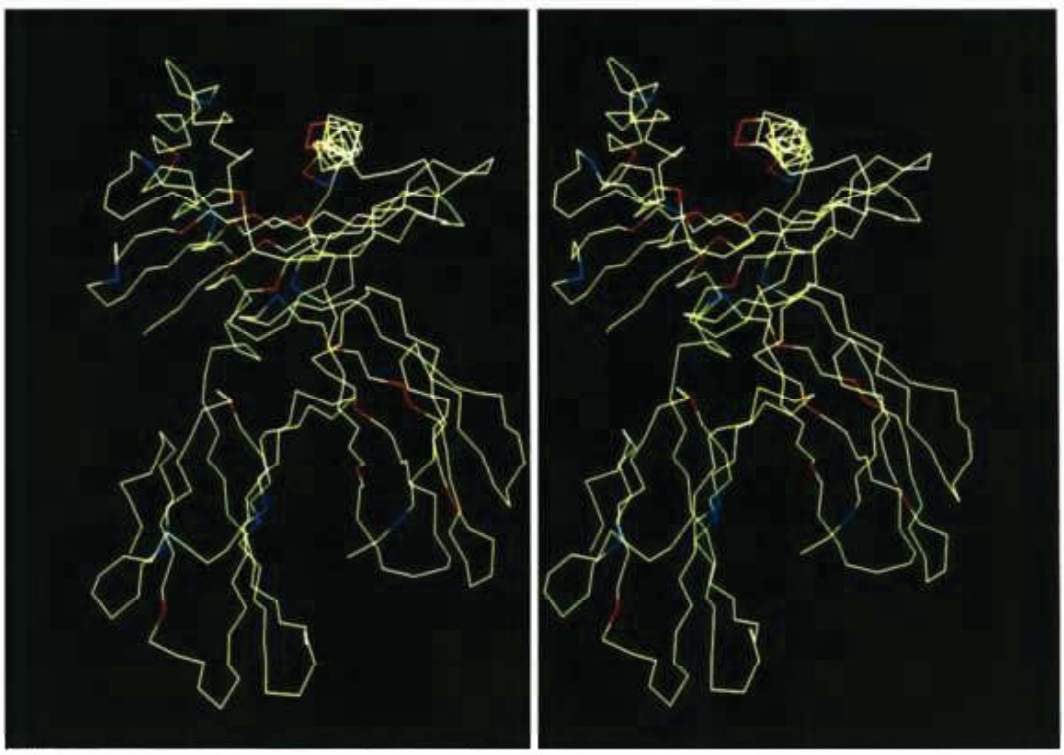


Figure 10. C^{α} skeleton of the human class I histocompatibility antigen A2 (PDB file 1HLA) showing the location of Tyr (red) and Trp (blue) residues. The binding site for processed foreign antigens is the large groove formed by the 2 helices supported by a β-sheet platform.

exhibited by antibodies. The concomitant improvement in shape and electrostatic complementarity, and by implication ligand–antibody binding, would arise from the improved apposition of hydrophobic groups and juxtaposition of hydrogen bond donating/accepting atoms. Unique combinations with other residues would give rise to specificity. The presence of Tyr and Trp allows individual antibodies to be multispecific, i.e. the potential to bind a variety of structurally unrelated ligands.

Although the precise nature of the specificity-determining residues is dependent on the physicochemical properties of the hapten/antigen itself, this role may be partially fulfilled by Ser and Asn, since Table 2 indicates that they are also common in antibody binding sites. His may play a part similar to Ser and Asn, because of the hydrogen bond donor and acceptor functionality of the imidazole ring. Such a model does not preclude residues having multiple roles, i.e. providing both specificity and general binding. Those occurring at the "key" positions defined by Chothia *et al.* (1989) would additionally determine the conformations of the hypervariable regions.

There are a number of lines of experimental support for this model. Firstly, from two-dimensional nuclear magnetic resonance studies, which compare the amino acid sequences, predicted structures and observed antibody–antigen interactions in complexes of two antipeptide antibodies (Levy *et al.*, 1989). Tyr, Trp, His and Phe residues in the combining site were believed to create a general hydrophobic pocket, whilst other residues altered the shape and polarity to fit the specific antigen (Levy *et al.*, 1989). Secondly, crystallographic studies on the binding of opioid peptides to the Mcg light chain dimer have indicated malleable binding sites (Edmundson *et al.*, 1987). Finally, the increase in antibody binding affinity during maturation of the immune response to the hapten 2-phenyl-5-oxazolone has been shown to be correlated with a non-random increase in the total number of mutations of framework and CDR residues (Berek & Milstein, 1987).

8. Class I Major Histocompatibility Molecules

In the present model for antibody combining sites, the structural and functional features of Trp and especially Tyr permit them to form a pliable binding pocket with other residues largely modulating the precise shape and polarity to fit the specific antigen. This may represent a general mechanism whereby molecules of defined sequence can evolve sites that accommodate and bind a broad array of ligands with differing sequence and conformation. The number of class I major histocompatibility molecules is smaller than the potential number of foreign antigens with which they must complex. The crystal structure of HLA-A2 (Bjorkman *et al.*, 1987) indicates that Tyr represents half of the residues pointing into the antigen recognition site (see Fig. 10) and that they are highly or

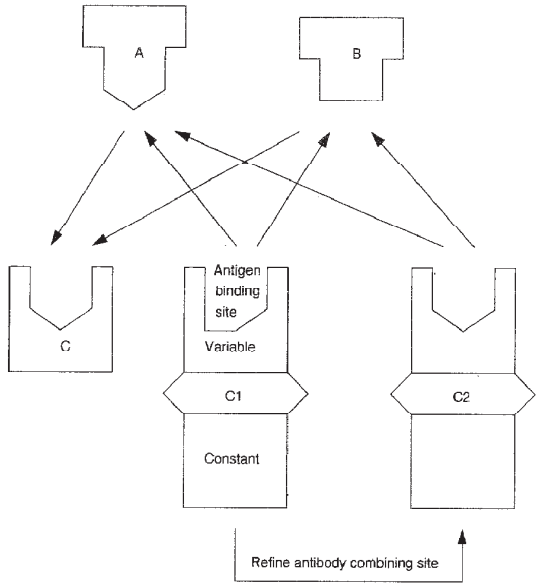


Figure 11. A diagram showing the potential use of antibodies as probes for macromolecular structure and redesign of molecules (the various objects are not drawn to scale). A, B and C may be of similar or different size; antibody C2 either may be one that has been independently isolated or is the product of refinement of C1. In the 1st instance, the combining site of C1 and C2 are complementary to the region of A and B to which they bind. For the 2nd case, A and B of known structure bind C, whose 3-dimensional form is unknown. Here, antibodies C1 and C2 bind A and B, with the combining site of C2 being a more accurate template of the binding site of C.

completely conserved (Bjorkman *et al.*, 1987). These residues may assist in enabling the HLA-A2 molecule to bind many different peptides by imparting conformational pliability. Townsend *et al.* (1989) have shown that MHC molecules fold up around the peptides they present to T-cell receptors, and that the nature of the peptide has significant consequences for the ultimate function and conformation of the MHC molecule (Townsend *et al.*, 1989). The precise mechanism of peptide binding may exhibit some differences since, in contrast to antibodies that can bind to free virus or soluble antigen, HLA-A2 must interact with both an antigen and a T-cell receptor (for a review of T-cell receptors, see Davis & Bjorkman, 1988).

9. Antibody Binding Sites as Structural Probes

Recent advances in the *rapid* production of *large* numbers of *different* Fab fragments in bacteria (Huse *et al.*, 1989) and the ability to model the structure of hypervariable loops of given amino acid sequence with reasonable success (see Thornton, 1990) should allow for greater exploitation of antibody specificity and diversity. The present results have identified the physicochemical characteristics of amino acids that may play roles in antigen binding and the most likely location of these residues in the combining site. It should be easier, therefore to utilize antibodies as probes of structure and in the redesign of molecules by screening the library of Fab fragments for those possessing the desired activity (Fig. 11). The affinity and specificity of candidate molecules could be improved further by making more detailed changes to the combining site region or generating chimaeric antibodies by crossing their heavy and light chains.

When deployed as tools for structural analysis, evaluation of the three-dimensional structure of the antibody binding site would yield a surface region complementary in nature to the combining area of the system of interest. As an aid to more rational drug design, consider a set of ligands of known structure that interact with a target whose sequence, structure or binding site properties may be unknown. It should prove possible to isolate and identify antibodies that reproduce the spectrum of specificity and affinity against the ligands displayed by the natural system. Modelling and subsequent refinement of the antibody binding site would produce a region having stereochemical features compatible with the combining site of interest. This facsimile of the desired region could be employed to optimize existing ligands and design novel molecules, and the antibody molecule itself may be a more tractable assay system for hypothesis testing than the real target.

We thank E. S. Golub, D. S. Goodsell and B. Toyonaga for comments on the manuscript. This work was supported by a grant from the Cancer Research Campaign, England (A.R.B) and NIH Programme Project grant no. GM38794-01 (A.J.O).

References

Alzari, P. M., Lascombe, M.-B. & Poljak, R. J. (1988). *Annu. Rev. Immunol.* **6**, 555–580.

Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. (1986). *Science*, **233**, 747–753.

Anglister, J. & Zilber, B. (1990). *Biochemistry*, **29**, 921–928.

Baker, E. N. & Hubbard, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

Bedzyk, W. D., Herron, J. N., Edmundson, A. B. & Voss, E. W., Jr (1990). *J. Biol. Chem.* **265**, 133–138.

Berek, C. & Milstein, C. (1987) *Immunol. Rev.* **96**, 23–41.

Berek, C., Griffiths, G. M. & Milstein, C. (1985). *Nature (London)*, **316**, 412–418.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennet, W. S., Strominger, J. L. & Wiley, D. C. (1987). *Nature (London)*, **329**, 512–518.

Bone, R., Silen, J. L. & Agard, D. A. (1989). *Nature (London)*, **339**, 191–195.

Capra, D. J. & Edmundson, A. B. (1977). *Sci. Amer.* **236**, 50–59.

Chothia, C. (1974). *Nature (London)*, **248**, 338–339.

Chothia, C. (1975). *Nature (London)*, **254**, 304–308.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. (1989). *Nature (London),* **342**, 877–883.

Chua, M.-M., Goodgal, S. H. & Karush, F. (1987). *J. Immunol.* **138**, 1281–1288.

Cody, V. (1985). In *The Chemistry and Biochemistry of Amino Acids* (Barrett, G. C., ed.), pp. 625–653, Chapman and Hall, London.

Connolly, M. L. (1985). *J. Mol. Graph.* **3**, 19–24.

Connolly, M. L. & Olson, A. J. (1985). *Comput. Chem.* **9**, 1–6.

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

Davis, M. M. & Bjorkman, P. J. (1988). *Nature (London),* **334**, 395–402.

Doolittle, R. F. (1986). In *Of Urfs and Orfs: A Primer on How to Analyse Derived Amino Acid Sequences,* p. 25, University Science Books, Mill Valley, CA.

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

Edmundson, A. B., Ely, K. R., Herron, J. N. & Cheson, B. D. (1987). *Mol. Immunol.* **24**, 915–935.

Fersht, A. R. (1985). In *Enzyme Structure and Function* pp. 293–310, W. H. Freeman and Company, New York.

Gelles, J. & Klapper, M. H. (1978). *Biochim. Biophys. Acta,* **533**, 465–477.

Green, N. M. (1975). *Advan. Protein Chem.* **29**, 85.

Herron, J. N., Kranz, D. M., Jameson, D. M. & Voss, E. W., Jr (1986). *Biochemistry,* **25**, 4602–4609.

Herron, J. N., He, X.-M., Mason, M. L., Voss, W. E., Jr & Edmundson, A. B. (1989). *Proteins: Struct. Funct. Genet.* **5**, 271–280.

Holmberg, D., Freitas, A. A., Portnoi, D., Jacquemart, F., Avrameas, S. & Coutinho. A. (1986). *Immunol. Rev.* **93**, 147–169.

Huse, W. D., Sastry, L., Iverson, S. A., Kang, A. S., Alting-Mees, M., Burton, D. R., Benkovic, S. J. & Lerner, R. A. (1989). *Science,* **246**, 1275–1281.

Kabat, E. A., Wu, T. T. & Bilofsky, H. (1977). *J. Biol. Chem.* **19**, 6609–6616.

Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987). In *Sequences of Proteins of Immunological Interest,* US Department of Health and Human Services, Washington, DC.

Kudo, A., Nishimura, Y. & Watanabe, T. (1985). *J. Immunol.* **135**, 642–645.

Lascombe, M.-B., Alzari, P. M., Boulot, G., Saludjian, P., Tougard, P., Berek, C., Haba, S., Rosen, E. M., Nisonoff, A. & Poljak, R. J. (1989). *Proc. Nat. Acad. Sci., U.S.A.* **86**, 607–611.

Levitt, M. (1978). *Biochemistry,* **20**, 4277–4285.

Levy, R., Assulin, O., Scherf, T., Levitt, M. & Anglister, J. (1989). *Biochemistry,* **28**, 7168–7175.

Matsushima, M., Marquart, M., Jones, T. A., Colman, P. M., Bartels, K., Huber, R. & Palm, W. (1978). *J. Mol. Biol.* **121**, 441–459.

Miller, A. & Glasel, J. (1989). *J. Mol. Biol.* **209**, 763–778.

Novotny, J., Bruccoleri, R. E. & Saul, F. A. (1989). *Biochemistry,* **28**, 4735–4749.

O'Donnell, T. J. & Olson, A. J. (1981). *Comput. Graph.* **15**, 133–142.

Padlan, E. A. (1990). *Proteins: Struct. Funct. Genet.* **7**, 112–124.

Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J. & Davies, D. R. (1989). *Proc. Nat. Acad. Sci., U.S.A.* **86**, 5938–5942.

Panka, D. J., Mudgett-Hunter, M., Parks, D. R., Peterson, L. L., Herzenberg, L. A., Haber, E. & Margolies, M. N. (1988). *Proc. Nat. Acad. Sci., U.S.A.* **85**, 3080–3084.

Quiocho, F. (1988). *Curr. Top. Microbiol. Immunol.* **139**, 135–148.

Richards, F. F., Konigsberg, W. H., Rosenstein, R. W. & Varga, J. M. (1975). *Science,* **187**, 130–130.

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985). *Science,* **229**, 834–838.

Rose, D. R., Strong, R. K., Margolies, M. N., Gefter, M. L. & Petsko, G. A. (1990). *Proc. Nat. Acad. Sci., U.S.A.* **87**, 338–342.

Saul, F. A., Amzel, M. & Poljak, R. J. (1978). *J. Biol. Chem.* **235**, 585–597.

Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). *J. Mol. Biol.* **190**, 593–604.

Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. & Davies, D. R. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8075–8079.

Stanfield, R. L., Fieser, T. M., Lerner, R. A. & Wilson, I. A. (1990). *Science,* **248**, 712–719.

Suh, S. W., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., Rudikoff, S. & Davies, D. R. (1986). *Proteins: Struct. Funct. Genet.* **1**, 74–80.

Thornton, J. M. (1990). *Nature (London),* **343**, 411–412.

Townsend, A., Öhlén, C., Bastin, J., Ljunggren, H.-G., Foster, L. & Kärre, K. (1989). *Nature (London),* **340**, 443–448.

Vyas, N. K., Vyas, M. N. & Quiocho, F. (1988). *Science,* **242**, 1290–1295.

Ward, E. S., Detlef, G., Griffiths, A. D., Jones, P. T. & Winter, G. (1989). *Nature (London),* **341**, 544–546.

Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. & Salemme, F. R. (1989). *Science,* **243**, 85–88.

Wolfenden, R., Andersson, L., Cullis, P. M. & Southgate, C. C. B. (1981). *Biochemistry,* **20**, 849–855.

Wu, T. T. & Kabat, E. A. (1970). *J. Expt. Med.* **132**, 211–250.

Wysocki, L. J. & Gefter, M. L. (1989). *Annu. Rev. Biochem.* **58**, 509–531.

Yancopolous, G. D. & Alt, F. W. (1986). *Annu. Rev. Immunol.* **4**, 339–368.

Edited by F. E. Cohen