J. Mol. Biol. (1987) **194**, 713–724

A Three-dimensional Model of an Anti-lysozyme Antibody

Sandra J. Smith-Gill[1]†, Charles Mainhart[1], Thomas B. Lavoie[1,3] Richard J. Feldmann[2], William Drohan[4] and Bernard R. Brooks[2]

[1]*Laboratory of Genetics, National Cancer Institute*

[2]*Division of Computer Research and Technology National Institutes of Health, Bethesda, MD 20892, U.S.A.*

[3]*Department of Zoology, University of Maryland College Park, MD 20752, U.S.A.*

[4]*Department of Molecular Biology Meloy Laboratories Revlon Health Care Group Springfield, VA 19014, U.S.A.*

(Received 21 April 1986, and in revised form 13 October 1986)

The primary amino acid structure of the lysozyme-binding antibody, HyHEL-10, as determined by amino acid and nucleotide sequencing was utilized to construct a scale model of the Fv (variable region domain of immunoglobulin) using energy-minimized torsional angles of the McPC603 Fv as a prototype template. This model was in turn used as a template for generating a computer-built set of co-ordinates, which were subjected to a total of 600 steps of Adopted Basis Newton-Raphson energy minimizations using the program CHARMM. Only minimal shifts of the backbone (root-mean-square 0·76 Å) were required to give an energetically stable structure with a favorable van der Waals' energy.

Several notable features were evident from both the scale model and the energy-minimized computer model: (1) the shape of the antibody combining region is that of a very shallow concavity approximately 20 Å × 25 Å; (2) the concavity is acidic and non-hydrophobic and is bordered by hydrophobic segments; (3) the lower portion of the combining site is dominated by a cluster of tyrosine residues over the L3 and H2 areas; (4) a somatic mutation encoded by the J region of the heavy chain (J_H) may contribute significantly to the complementarity of heavy chain H3 to the epitope on hen egg white lysozyme. In addition, the space-filling energy-minimized model revealed that residue 49L, a framework residue, was prominently exposed and accessible in the center of the combining-site concavity. The model suggests that variation in length of complementarity-determining regions may function not only to change directly the shape of the antibody combining site, but may also influence indirectly the nature of the antibody surface by changing the accessibility of residues not usually involved in antigen binding.

1. Introduction

In order to understand the molecular basis of interaction between an antibody and a protein antigen, the details of three structures must be defined: the antibody, the protein epitope and the antibody–protein complex. Hen egg white lysozyme (HEL)‡ has proven to be a valuable model antigen

† Author for correspondence at: National Institutes of Health, National Cancer Institute, Building 37, Room 2B10, Bethesda, MD 20892, U.S.A.

‡ Abbreviations used: HEL, hen egg white lysozyme: mAb, monoclonal antibody; L, light chain; H, heavy chain; V_L and V_H, variable region of light and heavy chain, respectively; CDR, complementarity determining region; L1, L2, L3 are CDR-1, -2 and -3 of the light chain; H1, H2, H3 are CDR-1, -2 and -3 of the heavy chain; J_K and J_H, J regions of light and heavy chains; D, diversity region of the heavy chain; Fv, variable region domain of immunoglobulin; Fr, framework region of immunoglobulin; cDNA, complementary DNA; HPLC, high-performance liquid chromatography; bp, base-pairs; ABNR, Adopted Basis Newton-Raphson; r.m.s., root-mean-square.

0022-2836/87/080713-12 $03.00/0

for the study of anti-protein specificity. The structure and function of HEL has been characterized to high resolution (Phillips, 1967; Diamond, 1974), and monoclonal antibodies (mAbs) specific for the native protein have proven amenable to both functional characterization (Smith-Gill *et al.*, 1982, 1984*b*, 1986; Metzgar *et al.*, 1983; Darsely & Reese, 1985*a*) and structure determination (Mariuzza *et al.*, 1983; Smith-Gill *et al.*, 1984*a*; Darsley & Reese, 1985*b*; Silverton *et al.*, 1984; Amit *et al.*, 1985, 1986; de la Paz *et al.*, 1986).

Recently, structural studies have been reported for two contrasting sets of HEL-binding antibodies. X-ray analysis of one antibody–HEL complex indicates an extensive surface area of interaction involving all six CDRs of the antibody (Amit *et al.*, 1985, 1986). To date, details of the structure of the uncomplexed antibody, which was derived from the secondary response of mice immunized with native HEL (Mariuzza *et al.*, 1983), have not been reported. In addition, models of HEL-binding peptide-generated antibodies also suggested an extensive surface area of antibody involved in contact, and an inverse correlation between overall CDR length in the antibody and epitope size in the antigen (de la Paz *et al.*, 1986). These anti-peptide antibodies all have affinities for HEL at least one order of magnitude lower than for the corresponding immunizing peptides (Darsley & Reese, 1985*a*), and also lower than the affinity of antibodies induced by the native protein (Amit *et al.*, 1986). Thus, these studies represent two different windows of the spectrum of anti-protein specificity, sampled utilizing the same protein antigen.

We have been investigating in detail (Smith-Gill *et al.*, 1982, 1984*a*,*b*, 1986; Silverton *et al.*, 1984) a group of anti-HEL mAbs representing a third view of the anti-protein immune response. These anti-bodies were derived from BALB/c mice hyperimmunized with HEL for over six weeks, then allowed to rest for four to six weeks prior to a single boost and fusion (Smith-Gill *et al.*, 1982). They should represent a sample of highly antigen-selected memory cells. Their affinities for HEL (at least 10^{-9}; M. Denton & H. A. Sheraga, personal communication) are at least two orders of magnitude higher than those of the peptide-generated (Darsley & Reese, 1985*a*) or the secondary response anti-HEL antibodies (Amit *et al.*, 1986). Two of these antibodies, HyHEL-8 and HyHEL-10, were considered to recognize overlapping but non-identical epitopes, including residues 19 and/or 21 on HEL structural domain I (Smith-Gill *et al.*, 1984*b*). These two independently derived IgG_1K antibodies both express the same V_H and V_K isotypes as the DNP-binding myeloma protein XRPC-25 (Barstad *et al.*, 1978), yet they do not bind DNP nor does XRPC-25 bind HEL (Smith-Gill *et al.*, 1984*b*). Thus, an understanding of the structures of these antibodies in relation to their respective protein or hapten specificities, as well as in comparison to the other two sets of HEL-binding antibodies discussed above, should prove valuable in our understanding of anti-protein specificity. To this end, we have determined the variable-region amino acid sequences, and have built an energy-minimized (Brooks *et al.*, 1983) model of the antibody HyHEL-10, which will be used in functional studies.

2. Materials and Methods

(a) *Amino acid sequence analysis*

HyHEL-10 protein was purified from ascites and the amino-terminal amino acid sequences determined for the heavy and light chains as described (Smith-Gill *et al.*, 1984*a*). The amino-terminal amino acid sequence of V_H was completed to residue 52† by Edman degradation using a Beckman 890c automated sequencer (Beckman Instruments, Inc., Fullerton, CA), and high-pressure liquid chromatography to analyze the phenyl-thiohydantoin derivatives (Zimmerman *et al.*, 1977). The remaining segment of V_H was determined from a cyanogen bromide (CNBr) fragment. The heavy chain was cleaved with CNBr and the resulting fragments separated on a Sephadex G-100 column equilibrated in 5 M-guanidine, 0·2 M-NH_4HCO_3 (Rudikoff *et al.*, 1980). The fraction containing the V_H peptides was completely reduced and [^{14}C]alkylated, and rechromatographed on the Sephadex G-100 column to resolve 2 V_H peptides, one spanning residues 1 to 47, and a second peptide (P_{CNBr}) spanning residues 49 to 139 (Fig. 1). Edman degradation of the latter (P_{CNBr}) gave a sequence to residue 90 that overlapped with the amino-terminal sequence. The V_H region sequence was completed by trypsin cleavage of succinylated P_{CNBr}. The cleavage mixture was sequenced by Edman degradation without further purification, and after 4 cycles a single sequence could be read from residue 76 to 110. The identification of residues in H3 was also confirmed from the nucleotide sequence of a cDNA clone of the HyHEL-10 heavy chain, as described below.

The amino acid sequence of residues 1 to 52 of the HyHEL-10 V_K was determined on the automated sequencer. The amino acid sequence of the remaining segment was deduced from the nucleotide sequence of a cDNA clone, which also confirmed the amino acid sequence from 1 to 52.

(b) *RNA purification*

Hybridoma cells for RNA production were either grown in flask culture to a density of 5×10^5 cells/ml or in solid form in the axillary region of BALB/c or CD2f1 mice to a size of 0·5 to 1·0 g to minimize the presence of necrotic tissue. Solid tumors were surgically removed and immediately stored in liquid nitrogen until processed for RNA isolation. The cells used to construct library I were maintained in culture for approximately 8 months prior to growth for RNA isolation. This library was used to generate a V_K23 cDNA clone that was used as a probe to screen library II; library II was constructed from RNA isolated from cells grown in tissue culture immediately after thawing of frozen samples of the original HyHEL-10 clone (i.e. a primary hybridoma that had been cloned three times to ensure monoclonality), to prevent introduction of somatic mutations due to *in-vitro* passage (Rudikoff *et al.*, 1981).

Total cellular RNA was isolated either by homogenization of solid tissue in 3 M-LiCl/6 M-urea (Auffray & Rougeon, 1980) or by lysis of pelleted tissue culture cells

† The Kabat (1980) amino acid numbering system is used throughout this paper.

in 5 M-guanidine monothiocyanate, 50 mM-Tris · HCl (pH 7·5), 10 mM-EDTA and 8% 2-mercaptoethanol adjusting the lysate to 3·5 M-LiCl (Cathala *et al.*, 1983). Poly(A)$^+$ RNA was isolated by poly(U)-Sepharose chromatography (Buell *et al.*, 1978).

(c) *cDNA library production*

cDNA for library I was synthesized starting with 3 μg of poly(A)$^+$ HyHEL-10 RNA by the method of Buell *et al.* (1978). The cDNA was selected after second-strand synthesis by chromatography through Sephadex G-200 (Pharmacia) column (0·7 cm × 45 cm) in 10 mM-Tris · HCl (pH 8·0), 1 mM-EDTA. Digestion with S_1 nuclease (Efstratiadis *et al.*, 1976) was followed again by Sephadex G-200 chromatography. The blunt-ended cDNA was tailed with deoxycytidine using terminal transferase (Peacock *et al.*, 1981) and annealed to dG-tailed pBR322 (Maniatis *et al.*, 1982). The annealed insert/plasmid hybrid was then used to transform competent *Escherichia coli* K12 RRI (Dagert & Ehrlich, 1979) or MM294 (Hanahan, 1983) strains. Library II was constructed starting with 100 μg of total HyHEL-10 RNA by the method of Gubler & Hoffman (1983), with the addition of 2 μg of actinomycin D/ml in the first-strand reaction, and without *E. coli* DNA ligase in the second-strand reaction. The cDNA was size selected by 6% polyacrylamide gel electrophoresis in Tris/borate/EDTA buffer (pH 8·3), to be > 600 bp. The DNA was then tailed, annealed and used to transform competent cells as above.

(d) *Library screening*

The kappa chain oligonucleotide (5′ CACGACTGAG-GCACCTCCGA 3′) was synthesized using a manual DNA synthesizer by dimer coupling (Bachem) (Ito *et al.*, 1982) and the final product was purified by reverse-phase HPLC (Spectra Physics) using a C-18 column (Waters) with a 10% to 20% acetonitrile gradient before detritylation and 0% to 20% acetonitrile gradient after detritylation.

Libraries were screened by the method of Grunstein & Hogness (1975) using either the kinased oligonucleotide probe (Wallace *et al.*, 1979) or a nick-translated cDNA probe (Rigby *et al.*, 1977). Probes for the kappa chain were either the kappa oligonucleotide or nick-translated clone K10.7D11. For the heavy chain the probe was a 142 bp *Eco*RI–*Bam*HI fragment of V_H36–60 supplied by R. I. Near and M. Gefter (Near *et al.*, 1984). The fragment was ligated to itself, nick-translated, digested with *Eco*RI and *Bam*HI, and purified by gel electrophoresis for use as a probe. Positive colonies were picked, plated and rescreened. Plasmid DNA was prepared from positive clones by an alkaline lysis protocol (Birnboim & Doly, 1979), and then analyzed by agarose gel electrophoresis for the size of inserted cDNA and the presence of diagnostic restriction enzyme digest fragments. Large-scale plasmid purification was either by an alkaline lysis protocol or by the method of Mukhopedhyay & Mandal (1983). Restriction enzyme digests were performed either in the recommended buffers or with buffer supplied by the manufacturer (International Biotechnologies, Inc.).

(e) *DNA sequencing*

DNA was sequenced by the method of Maxam & Gilbert (1980) with either 5′ protruding or flush-ended restriction fragments labeled with bacteriophage T4 polynucleotide kinase. Single end labeled fragments were isolated on polyacrylamide gels, electroeluted and purified by Elutip-d chromatography (Schleicher and Schuell).

(f) *Molecular models*

The backbone of our HyHEL-10 scale model was built with the Precision Molecular Model System (Academic Press) (Barrett, 1979) using the energy-minimized torsional angles of the McPC603 Fv (Segal *et al.*, 1974) as described (Mainhart *et al.*, 1984). There were 3 HyHEL-10 backbone segments that differed from those of McPC603 in size and required a separate modeling step. These 3 disparate segments were all located in Wu-Kabat (Kabat, 1980) hypervariable regions. Both the L1 and H3 of HyHEL-10 were 6 residues shorter than the homologous McPC603 region, and the HyHEL-10 H2 was 3 residues shorter than the McPC603 H2. The L1 of a crystallized human kappa chain dimer, REI (Epp *et al.*, 1975), has an L1 of size identical with that of HyHEL-10, and the torsional angles of REI were used to construct this segment. The H2 and H3 regions of HyHEL-10 were very short. These segments of HyHEL-10 were constructed by removing the long segments from the McPC603 model and rejoining the open ends to pre-existing beta-sheets in the Fv framework. In addition, the H2 segment of HyHEL–10 contained a substituted Pro residue at position 61. Although a transpeptide bond was retained during the modeling of HyHEL–10, the energy-minimized M603 backbone had to be slightly modified in order to accommodate the Pro-specific phi torsional angle of about −60°. This region does not appear to be a region of contact when the model is complexed with HEL (S. J. Smith-Gill *et al.*, unpublished results). The L2, L3 and H1 segments were of the same lengths as those of McPC603, and were constructed utilizing the McPC603 co-ordinates. Overall, the proposed HyHEl-10 model is more compact and lacks the 2 regions of loose structure (L1 and H3) of McPC603.†

(g) *Computer modeling and energy minimization of the HyHEL-10 Fv*

The HyHEL-10 scale model was used as a guide to construct a computer model on the PPD 1170, by breaking and rejoining the backbone of the energy-minimized McPC603 co-ordinates, as previously described (Mainhart *et al.*, 1984). The resulting computer co-ordinates of HyHEL-10 were then migrated to an Apollo DN300 computer in order to be studied using CHARMM (Brooks *et al.*, 1983) for the energy minimization and analysis. All calculations using CHARMM (Brooks *et al.*, 1983) were performed with the parameter set PARAM19 (Reiher, 1985) *in vacuo*. A distance-dependent dielectric ($\varepsilon = r$) with a shift potential termination at 7·5 Å was employed. The Adopted Basis Newton-Raphson (ABNR) algorithm (Brooks *et al.*, 1983) was used for all energy minimizations.

The purpose of the minimization was to provide an energetically reasonable structure with a large degree of overlap with the model structure; the method as applied in this study did not allow for large-scale or global changes, but rather it optimized local structure where the protein had been modeled. This was achieved through the use of harmonic constraints during minimization.

The co-ordinate files for the HyHEL-10 light and heavy chain were combined and hydrogenated (A. Brunger & M. Karplus, unpublished results) using CHARMM prior to any energy minimization. Each amino

† The co-ordinates of the model (side-chain and α-carbon) have been deposited with the Brookhaven Laboratory Data Bank.

acid of the HyHEL-10 structure was then grouped into 1 of 3 categories on the basis of the degree of its variability from the McPC603 prototype (Table 1). A residue was defined as being conservative if it was a framework position that was identical in both the McPC603 and HyHEL-10 sequence. All unmodeled CDR segments and any homologous residues in the framework regions that differed only in their specific side-chain types were classified as semi-conservative, with the exception of substitutions involving a proline residue, which were defined as variable. All modeled CDR residues plus the 2 upstream and downstream residues that were contiguous with them were classified as variable. This grouping distinction was used in applying different constraints to different regions of the structure during minimization.

The combined structure was then subjected to 4 series of ABNR minimizations utilizing CHARMM. The harmonic constraint-energy force constants used for each atom depend on the residue classification (conservative, semi-conservative or variable; Table 1) and varied for each cycle of minimization (Table 2). In the final cycle nearly all constraints were removed, including those on the unsubstituted framework regions from the McPC603 structure.

Table 1

Categories of amino acids used to define harmonic constraints, as detailed in Table 2, during energy minimization of HyHEL-10 model

Variable amino acids with all atoms (backbone and side-chain) free[a]							
Thr	14L	Asn	32L	Gly	55H	Trp	95H
Pro	15L	Leu	33L	Ser	56H	Asp	96H
Gly	16L	His	34L	Asn	60H	Gly	97H
Ser	28L	Val	51H	Pro	61H	Asp	101H
Ile	29L	Ser	52H	Ser	62H	Tyr	102H
Gly	30L	Tyr	53H	Ala	88H		
Asn	31L	Ser	54H	Asn	94H		
Semi-conservative amino acids with fixed backbone and free side-chains[b]							
Leu	4L	Glu	79L	Thr	21H	Leu	63H
Thr	10L	Thr	80L	Ser	23H	Ser	65H
Asn	17L	Phe	83L	Val	24H	Ile	67H
Ser	18L	Met	85L	Thr	25H	Ser	68H
Leu	21L	Gln	90L	Asp	27H	Ile	69H
Arg	24L	Ser	91L	Ile	29H	Thr	70H
Ser	40L	Asn	92L	Thr	30H	Lys	75H
His	41L	Ser	93L	Ser	31H	Asn	76H
Glu	42L	Trp	94L	Asp	32H	Gln	77H
Ser	43L	Tyr	96L	Tyr	33H	Tyr	78H
Arg	45L	Asp	1H	Trp	34H	Asp	81H
Lys	49L	Gln	3H	Ser	35H	Leu	82H
Tyr	50L	Gln	5H	Ile	37H	Ser	82bH
Glu	53L	Pro	9H	Lys	39H	Val	82cH
Ser	54L	Ser	10H	Phe	40H	Thr	83H
Ile	55L	Lys	13H	Asn	43H	Thr	84H
Ser	56L	Ser	15H	Tyr	47H	Thr	89H
Ile	58L	Gln	16H	Met	48H	Trp	103H
Ser	60L	Thr	17H	Tyr	50H	Gln	105H
Asn	76L	Ser	19H	Tyr	58H	Leu	108H

All residues not listed in this Table were classified as conservative; both the backbone and side-chains of these amino acids remained fixed during the first 200 steps of minimization, but were allowed to move during the final 400 steps of minimization.

[a] Both side-chains and backbones of these amino acids were allowed to move during all 600 steps of HyHEL-10 minimization.

[b] Only the side-chains of these amino acids were allowed to move during the first 200 steps of HyHEL-10 minimization, their backbones remained fixed. However, during the final 400 steps of the HyHEL-10 minimization both the side-chains and the backbones of these amino acids were allowed to move.

3. Results

(a) *Heavy chain primary structure*

The HyHEL-10 V_H region is 96 amino acid residues long (Fig. 1). In general, most V_H regions have two different lengths in H2 (Kabat, 1980) and HyHEL-10 has the short form, which is three residues shorter than that of McPC603. The HyHEL-10 V_H sequence corresponds to that of the V_H36–60 group (Near *et al.*, 1984). In H3, the D region contributes only three residues; the amino acid sequence (Asn94-Trp95-Asn96) corresponds to that encoded by the DQ52 germline gene (Kurosawa & Tonegawa, 1982). The J_H sequence corresponds to that of J_H3.

Although all other amino acid residues of H3 and J_H were unambiguous, it was not possible to determine from amino acid sequencing whether residue 101H was Ala or Asp. The identification of residue 101 was confirmed from the nucleotide sequence of a cDNA clone of the HyHEL-10 heavy chain. The sequence was determined from the *Eco*RI site in the V region and the *Bst*EII site in the C region. The nucleotide sequence confirmed that residue 101H was Asp, representing a somatic mutation from the J_H3-encoded Ala at that position. The region of H3 is slightly hydrophilic and acidic, due to the two Asp residues at positions 96 and 101. The H2 region is very hydrophobic (Fig. 2).

(b) *Light chain primary structure*

A V_K23-containing cDNA clone K.10.7D11 was isolated from library I and was used as a probe to isolate cDNA clone 10K-106 from library II. The predicted amino acid sequence was identical over the first 52 residues to the previously determined N-terminal amino acid sequence (Fig. 3). The region corresponding to amino acids 30 to 110 was verified by nucleotide sequence comparison to an independent clone derived from library II (10K-1077). The nucleotide sequences and the predicted amino acid sequences are shown in Figure 3. Both L1 and L2 have very hydrophobic sections (Fig. 2).

(c) *Scale model*

Because of the shortness of H2 and H3 modeled regions, there was little flexibility in the positioning of these segments. The points of rejoining are shown in Figure 4(a). In each case they form the β-turns of the rigidly bonded framework beta-sheets. The resulting modeled Fv is illustrated in Figure 5(a). Several characteristics were discernible from this unrefined scale model and the hydrophobicity analysis of the primary structure (Fig. 2). The

Table 2
HyHEL-10 energies during ABNR energy minimization with CHARMM

Cycles	Initial	100 [a]	200 [b]	200 [c]	400 [d]	600	
r.m.s. [f] (free)	0·0	0·335	0·519	0·519	0·671	0·766	
r.m.s. (all)	0·0	0·188	0·291	0·291	0·403	0·499	
Potential energy (kcal/mol):							
	All atoms	Free atoms only		All atoms			
Total	272,251·3	261,860·1	3233·1	700·0	4699·9	−2965·0	−4975·1
E bond	1111·8	253·4	147·4	104·6	969·5	110·8	73·4
E angle	2845·2	766·2	428·6	418·5	2497·5	861·3	657·2
E dihedral	666·1	310·9	283·7	264·4	619·7	560·1	531·5
E improper	612·7	286·3	157·5	125·3	451·6	160·1	132·5
van der Waals'	268,464·2	261,513·2	1885·0	879·3	5902·7	341·1	−279·4
Electrostatic	−1455·3	−1271·0	−1382·9	−1442·6	−5741·1	−6217·7	−6538·4
Constraint	0·0	0·0	1713·7	350·5	0·0	1219·3	448·1

See Brooks *et al.* (1983) for CHARMM and potential energy methods.

[a] 100 steps whereby only the free atoms were allowed to move, but their movement was restricted by a harmonic constraint of 2·0 × atomic mass.

[b] 100 steps similar to above except that the harmonic constraint was reduced to 1·0 × atomic mass for free atoms.

[c] The same structure after a total of 200 steps of minimization, with the energy recalculated to include all atoms.

[d] 200 steps whereby the fixed atoms were allowed to move with a harmonic constraint restriction of 2·0 × their atomic mass. The atomic constraint on free atoms was further reduced to 0·5 × atomic mass.

[e] 200 steps where all atoms were allowed to continue movement but fixed atoms were restricted with a harmonic constraint of 1·0 × atomic mass and free atoms retained the 0·5 × atomic mass constraint term.

[f] Root-mean-square difference, in Å, from the initial HyHEL-10 structure, calculated for only the free amino acids (free) or over the entire structure (all).

modeled antigen-combining site of HyHEL-10 is a large, shallow concavity, approximately 20 Å × 25 Å (1 Å = 0·1 nm), lined by acidic amino acids and bordered by hydrophobic sites. H3 at the centre of the concavity is acidic due to the aspartic acid side-chains of residues 96 and 101. Residue 32 in H1, which in the model is in close proximity to Asp94 of H3, might contribute further to the acidity of the H3 region. In contrast to the non-hydrophobic nature of H3, the borders of the concavity are predominantly hydrophobic (Fig. 2): the strongly hydrophobic L2 and the third Fr of the L chain (Fr3L) on the upper left, the hydrophobic β-turn of L1 on the lower left, and H2 on the lower right. The only exception is the upper right, bordered by H1.

Another feature of the three-dimensional combining site, which was not obvious from the primary structure but notable from the three-dimensional scale model, was a cluster of tyrosine residues surrounding the combining site (Fig. 5(b)); a total of seven tyrosine side-chains (50L, 96L, 33H, 50H, 53H, 58H and 102H) was prominently exposed.

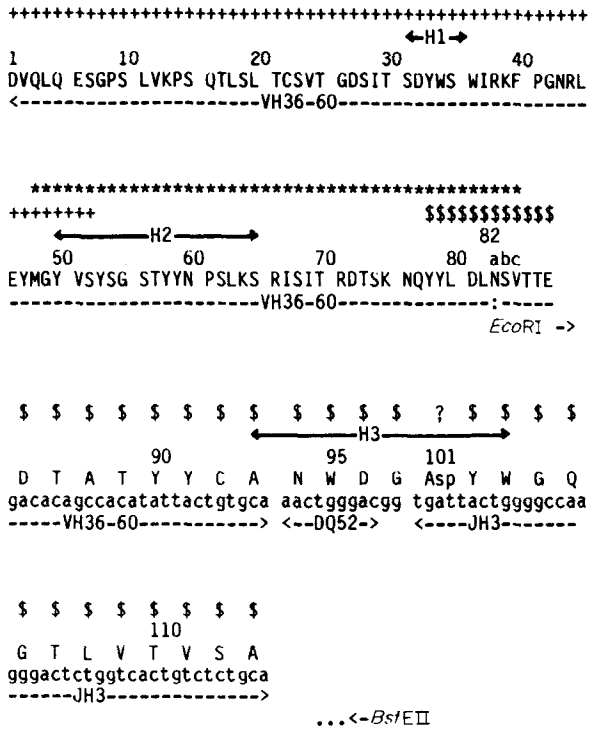

Figure 1. Complete amino acid sequence of HyHEL-10 heavy chain V regions, and nucleotide sequence of H3 of HyHEL-10 H-chain cDNA clone. Confirmed amino acid residues and the peptide fragments from which their sequences were determined are shown above the sequence: +++, amino terminal sequence; ***, cyanogen bromide fragment P_{CNBr}; $$$, trypsin cleavage of succinylated P_{CNBr}. Restriction sites used in sequencing the cDNA clone are shown below the sequences.

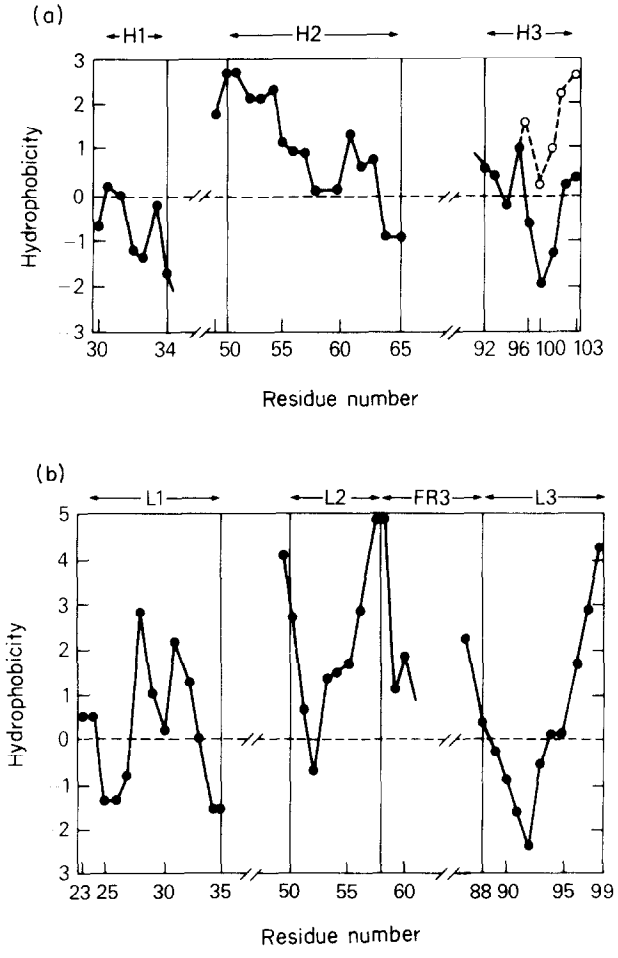

Figure 2. Hydrophobicity plots (Wolfenden *et al.*, 1981) for HyHEL-10. (a) Heavy chain hypervariable regions. The hydrophobicity plot of H3 without the Ala101 → Asp substitution is indicated by the broken line. (b) Light chain hypervariable regions.

(d) *Energy-minimized structure*

The first 100 iterations of ABNR minimizations were performed to reduce strained bond angles and illegal contacts in the immediate regions of modeling and substituted side-chains, with all conservative residues fully constrained, all semi-conservative residues moderately constrained, and all variable residues weakly constrained (Tables 1 and 2). After 20 steps, the total energy of freed atoms had fallen to 4198 kcal (1 cal = 4·184 J), and only a few high-energy contacts remained. At this point, the position of several tyrosine side-chains in H2 were shifted slightly to relieve contacts, and the minimization was continued. A total of 100 iterations yielded a structure whose total resting potential energy was 3233 kcal (Table 2). This structure was still being quite forcefully constrained to its starting co-ordinate positions with a constraint energy of 1714 kcal. The relatively low-energy structure was subjected to 100 more cycles of minimization where the constraints on variable and semi-conservative residues were reduced by nearly 50%, and total free energy fell to 700 kcal. The van der Waals' energy due to unfavorable contacts was still in excess of 800 kcal, so constraints were again reduced for all atoms for 200 more cycles. The resulting potential energy of this structure was −2965 kcal, and the root-mean-square difference from the initial HyHEL-10 structure was 0·67 Å. The energy needed to constrain this structure was still quite significant (in excess of 1000 kcal) and the van der Waals' energy was still positive (Table 2). One final series of minimizations was performed where nearly all constraints were removed during the process. A total of 200 additional iterations were completed on the structure before the energy level stabilized at around −4975 kcal. The van der Waals' energy

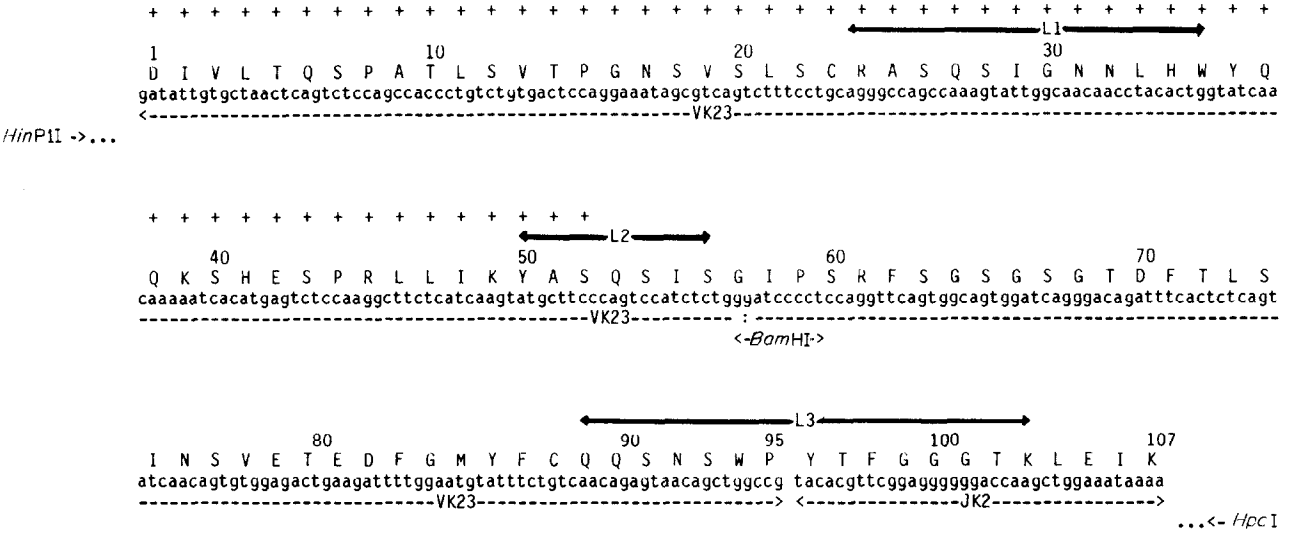

Figure 3. Complete nucleotide and amino acid sequence of HyHEL-10 light chain V region. Span of amino acid residues confirmed from the amino-terminal sequence is shown above the sequence (+ + +). Restriction sites used in sequencing the cDNA clones are shown below the sequence.

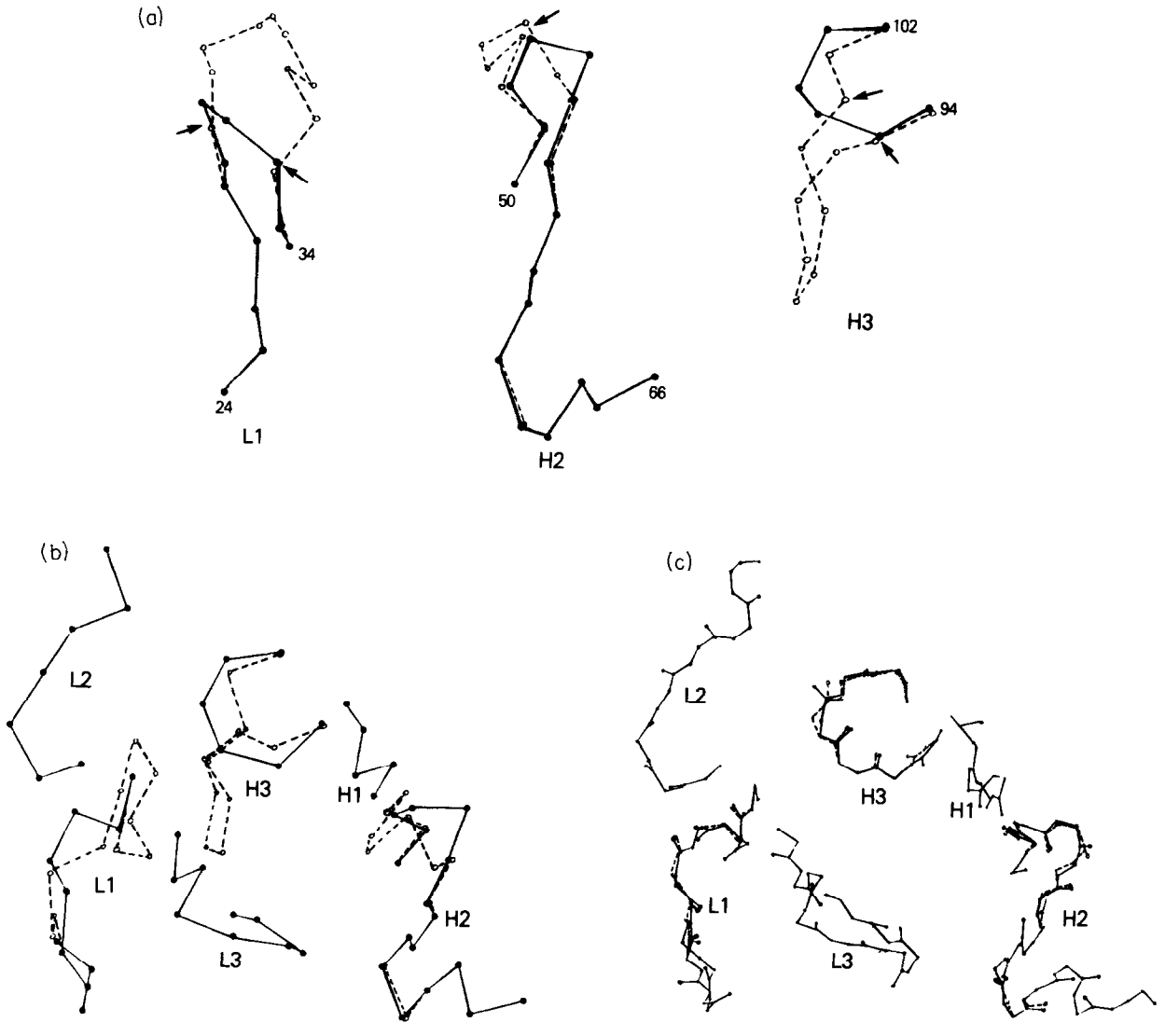

Figure 4. (a) α-Carbon backbone of modeled HyHEL-10 (————) compared with McPC603 (– – – –) from which it was modeled. Arrows show points of breaking and rejoining (Mainhart *et al.*, 1984). (b) α-Carbon backbone of modeled HyHEL-10 (————) CDR regions compared with those of McPC603 (– – – –), showing absence of any change in unmodeled regions. Orientation is approximately the same as the entire Fv shown in Fig. 5(a) to (c). (c) Peptide backbones of modeled HyHEL-10 CDRs prior to (————) and after energy minimization (— — — —). Orientation is same as in (b).

resulting from close contacting atoms was a favorable −279 kcal, and there were no illegal contacts in the structure. The total r.m.s. difference from the initial HyHEL-10 structure was 0·5 Å, and most of this could be attributed to the modeled regions, which had a r.m.s. difference of 0·77 Å (Fig. 4(b) and (c)). The energy used to constrain this structure was still around 450 kcal (Table 2) because some constraints on the conserved regions were maintained. The alpha-carbon backbone of this energy minimized structure is shown in Figure 5(c), and an overlay of this structure with that of the McPC603 from which it was modeled is shown in Figure 5(d). The shallow, concave shape, noted from the scale model, was also characteristic of the energy-minimized model (Fig. 5(c), (e) and (f)). The cluster of tyrosines over the lower portion of the combining site was also readily apparent in the computer-generated, energy-minimized model (Fig. 6). An additional prominent feature of the three-dimensional combining site was very obvious from the space-filling model. Residue 49L, a lysine, is exposed and closely juxtapositioned to the H3 region within the combining site. This is a framework residue that is not exposed in antibodies with longer H3, but in HyHEL-10 would be potentially available to interact with antigen as part of the combining site.

4. Discussion

We have built a physical scale model of the Fv regions of an anti-lysozyme antibody, and used this scale model as a template for producing an

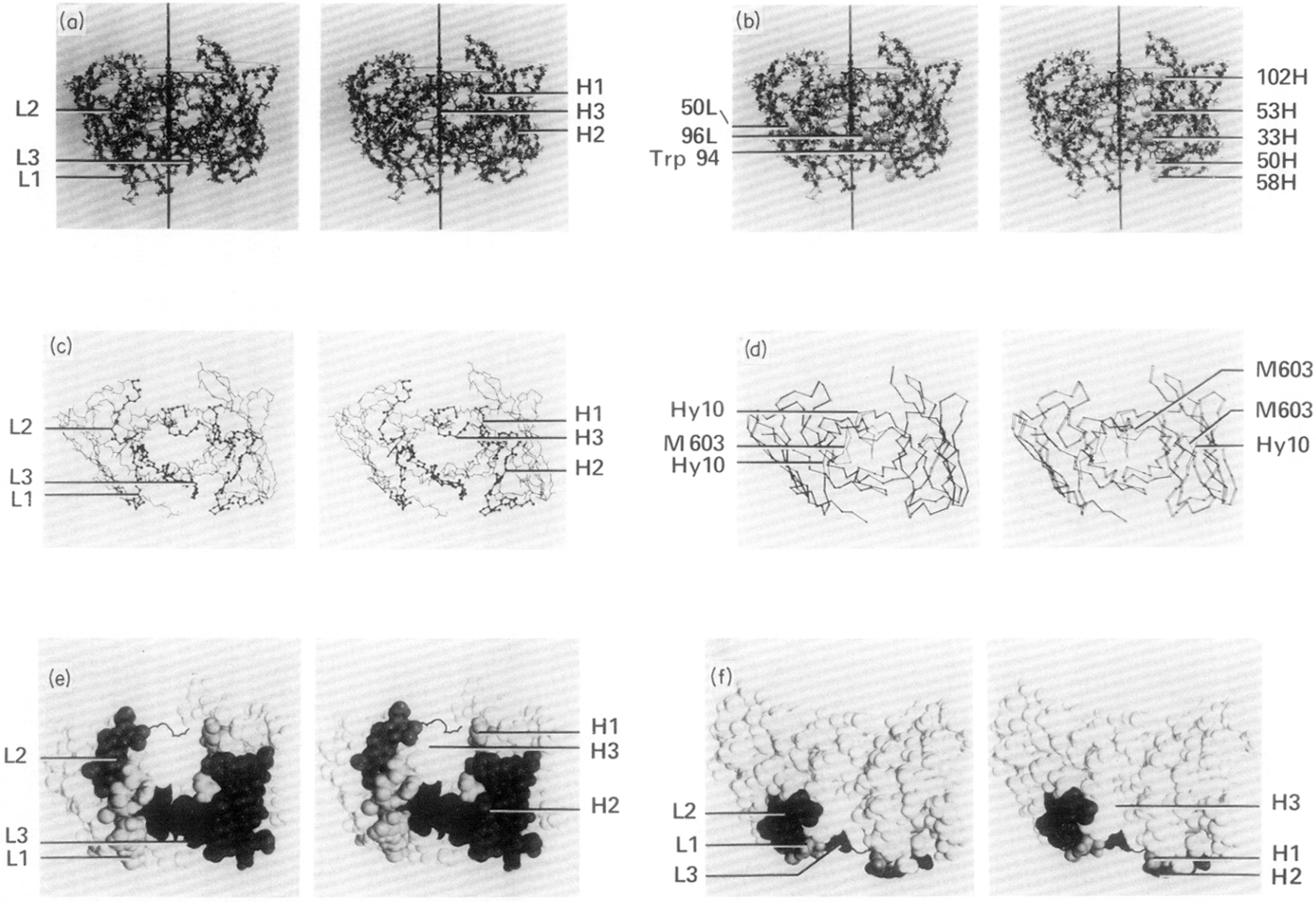
(a)
L2
L3
L1
H1
H3
H2
(b)
50L
96L
Trp 94
102H
53H
33H
50H
58H
(c)
L2
L3
L1
H1
H3
H2
(d)
Hy10
M 603
Hy10
M603
M603
Hy10
(e)
L2
L3
L1
H1
H3
H2
(f)
L2
L1
L3
H3
H1
H2

Fig. 5.

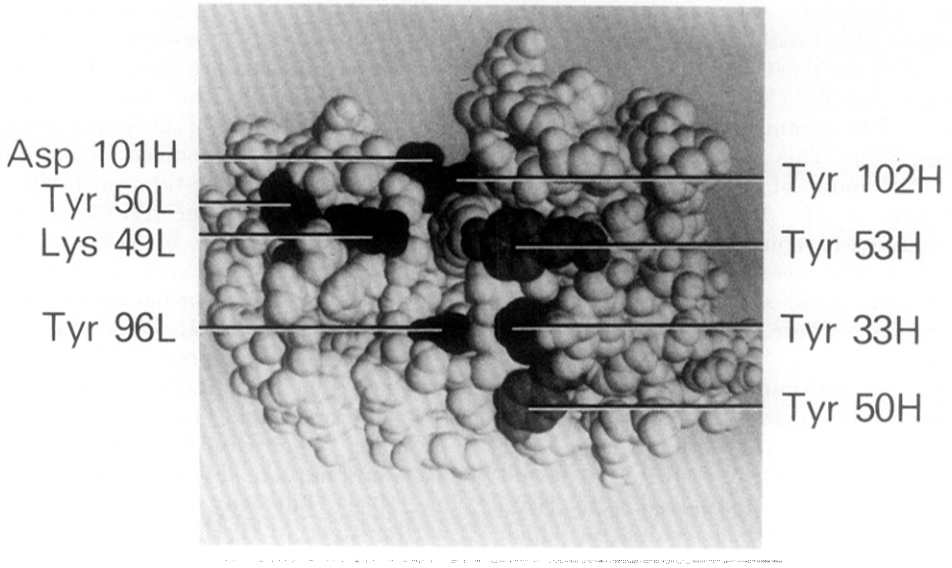

Figure 6. Detail of computer-generated energy-minimized space-filling model of HyHEL-10 Fv region, showing location of external tyrosine residues, residues Lys49L (L) and Asp101H (H).

energetically feasible structure. The modeled combining site is a shallow concavity that includes at least one prominently accessible framework residue. The energy minimization of the structure has not been carried to complete convergence, nor has the configuration space of all possible local structures been fully explored. The purpose of this model was to investigate topographical features of the combining site, and to provide an energetically stable structure that will be used to generate related structures (such as HyHEL-8 and XRPC-25) and can be subjected to functional studies.

HyHEL-10 was chosen for model building for both functional and structural reasons. Functionally, it is one of several V_H36–$60/V_K23$ antibodies that have definable specificities that are closely related in primary sequence, and for which an X-ray analysis of both the Fv and the Fv–HEL complex should be available in the future (Silverton *et al.*, 1984). Thus, the correctness of our model will ultimately be testable by comparison with the X-ray crystallographic data. Comparative study of the structure and function of the V_H36–60 antibodies with each other, as well as with peptide-generated and secondary-response antibodies specific for HEL (Amit *et al.*, 1985, 1986; de la Paz *et al.*, 1986), promises to yield valuable correlates of the structural basis of both fine specificity for a protein epitope as well as specificity for hapten *versus* protein.

Structurally, the short CDRs of both the H and L chain made HyHEL-10 particularly amenable to model building (Mainhart *et al.*, 1984). While there existed a crystal structure with an L1 of the same length that could be used to model that region of HyHEL-10 by homology, no crystal structures were available with identical lengths to serve as a model for H2 and H3. The short lengths of these CDRs, especially H3, allowed very little flexibility in the construction of the β-turns in the modeled CDR regions, essentially reducing the degrees of freedom for error relative to building a larger CDR region. The Precision Molecular Model System allowed control of the dihedral angles when building the scale model, which in turn served as a template for the computer model. As a consequence, only minimal shifts of the computer-modeled backbone were required to give a very energetically stable structure with a potential energy of -4975 kcal/mol.

Although energy minimization does not in itself ensure a correct structure (Novotny *et al.*, 1984), a recent comparison of a model of antibody D1.3 with the X-ray structure of the D1.3 Fv complexed to HEL suggests that a combination of structural homologies with conformational-energy analysis can successfully predict main-chain conformations (Chothia *et al.*, 1986). In fact, the same antibody, REI, was used to predict L1 of D1.3 correctly as we have used to model L1 of HyHEL-10. For H2 and H3, for which homologous structures were not

Figure 5. Stereodiagrams of 3-dimensional models of HyHEL-10 Fv region. All front views are with same orientation with the L-chain on the left side and the H-chain on the right. (a) Scale molecular model, with side-chains, front view. (b) Scale model, front view, with space-filling oxygen atoms added to emphasize location of exposed tyrosine side-chains in CDRs. (c) Computer-generated alpha-carbon backbone of energy-minimized model, front view, with CDRs highlighted. (d) Energy-minimized HyHEL-10 backbone overlayed on McPC603 starting structure. (e) Computer-generated space-filling model of energy-minimized structure, front view. (f) Computer-generated space-filling model of energy-minimized structure, top view.

available, our approach was to shorten the McPC603 backbone by breaking and rejoining the alpha carbon chain at positions that had deleted long loops unique to McPC603 and that retained sections of the CDR common to other antibody structures (Fig. 4(a)). This approach appears to increase the likelihood of obtaining a correct structure (Chothia *et al.*, 1986). These modeled regions were minimally constrained during minimization to allow any necessary adjustments of angles, and, in fact, small shifts did occur in the immediate vicinity where the backbone was modeled (Fig. 4(c)).

The rationale for retaining constraints on the structure rather than allowing it to converge to an energy minimum was to not allow large-scale motion that might "lock" it into a state that would preclude interactions when complexed with HEL (B. R. Brooks *et al.*, unpublished results), or which might be incompatible with side-chain substitutions associated with other antibodies (such as HyHEL-8) with which HyHEL-10 would be compared after minimization. However, even with these constraints, the physical model provided a good template for an energetically feasible structure that could represent one of the states assumed by the native molecule in solution.

The overall topography of the predicted combining site, readily apparent from both the low-resolution scale model and the energy-minimized structure, was that of a shallow concavity approximately 20 Å × 25 Å. The modeled concavity is acidic and non-hydrophobic and is bordered by hydrophobic segments. The lower portion of the combining site is dominated by a cluster of tyrosine residues over the H2 area. Thus, the scale molecular model also was very useful for analysis of most global features of the antibody combining site. In addition, in the energy-minimized space-filling model a framework residue (Lys49L) could be seen as obviously exposed and intimately associated with H3 in the center of the concavity. This could also be seen in the unrefined scale model, but its apparent accessibility was not evident until we examined the computer-generated space-filling model.

The predicted shallow, concave shape is a direct consequence of the short lengths of L1, H2 and H3 (11, 16 and 8 residues, respectively). The compactness of H3 creates the concavity on the frontal surface. The upper region of the concavity is bordered by L2 and H1 on the left and right sides, respectively (Fig. 4(a)). The lower region of the concavity is bordered by L1, L3 and H2. Because both L1 and H2 also are quite short, the overall surface of the predicted combining-site concavity is relatively shallow. The relative flatness of the predicted HyHEL-10 surface is consistent with the recent X-ray structure of anti-HEL antibody D1.3 (Amit *et al.*, 1986). We predict that the concavity of the HyHEL-10 surface will be even shallower than that of D1.3, because H3 and H2 of HyHEL-10 are both shorter than that of D1.3.

The cluster of tyrosine residues around H3 was not immediately evident from the primary structure, but was prominent in both the unrefined scale model and the energy-minimized model. A high proportion of aromatic residues has also been noted in modeled combining sites of HEL-binding anti-peptide antibodies (de la Paz *et al.*, 1986). The tyrosine residues in HyHEL-10 are contributed primarily by H2, and their clustering could effectively prevent contact of any other surrounding residues or backbone atoms from direct interaction with HEL. This may be particularly significant with respect to Trp94 in L3. In our model, Trp94L is partially masked by the overlying tyrosine residues. The L3 tryptophan has been proposed to contribute significantly to binding of at least one hapten, β-(1,6)D-galactan (Glaudemans *et al.*, 1984), but may not be accessible to interact with antigen in HyHEL-10.

The observation that the fundamental antigenic units of HEL correspond to the structural domains of HEL (Smith-Gill *et al.*, 1984*b*) predicts that the HyHEL-10 combining site would be complementary to HEL domain I. Complementarity of the antibody to HEL structural domain I would require portions of the antibody actually to fit into the hydrophobic pockets between domain I and the adjacent domain III of HEL. The strong hydrophobicity of the CDRs, especially H2, L1 and L2 (Fig. 2) are consistent with this expectation. The topography of the combining site with respect to the hydrophobicity profile is notable: the strongly hydrophobic segments of L1, L2 and H2 are all brought together on the surface of the combining site to border the hydrophilic H3. In addition, the J_H3-encoded Ala101H → Asp somatic mutation makes H3 acidic and hydrophilic, which is complementary to the basic epitope on HEL domain I. In contrast to HyHEL-10, a group of arsonate-binding Id36–60 hybridomas from both BALB/c and A/J mice also express V_H36–60 and J_H3 but do not have the Ala → Asp somatic mutation (Near *et al.*, 1984; Juszczak *et al.*, 1984), and express a basic and hydrophobic H3. Thus, even though the 36–60 H chain could assume a conformation similar to that of HyHEL-10, its hydropathic topography would be very different. This somatic mutation in HyHEL-10, therefore, very probably represents an antigen-selected substitution (McKean *et al.*, 1984), which may increase the specificity and/or affinity of HyHEL-10 for HEL, a hypothesis we are testing by site-specific mutagenesis of the expressed HyHEL-10 rearranged genes.

The apparent prominent exposure of Lys49 of the L chain and its close juxtaposition to H3 makes it likely that this residue is a contact residue (S. J. Smith-Gill *et al.*, unpublished results). It is notable that residue 49L differs between HyHEL-10 and the closely related HyHEL-8 (T. B. Lavoie, unpublished results), and may contribute to some of the L-chain-determined differences in fine specificity between these two antibodies (S. J. Smith-Gill *et al.*, unpublished results). Residue 49L

has been shown to be a contact residue in antibody D1.3 (Amit *et al.*, 1986) and has also been suggested as a potentially important contact residue in a set of related anti-idiotypic antibodies; these antibodies also have relatively short H3 and express V_K23 light chains (Stablitzky *et al.*, 1985). Although residue 49L is considered to be a relatively conserved framework residue not traditionally involved in antigen binding (Kabat, 1980), it now appears that it is important to antigen binding for several anti-protein antibodies. The apparent accessibility of Lys49L in the HyHEL-10 combining site is a consequence of the short CDRs, especially H3. We suggest that variation in length of the CDRs not only directly changes the shape of the combining site, but may also change the nature of the accessible antibody surface by exposing (or blocking) residues not usually involved in antigen binding.

We have constructed this energetically feasible model of the HyHEL-10 Fv in order to visualize a three-dimensional structure that could be one of many the molecule might take in solution. It has provided insight into the topographical and hydropathic properties of the HyHEL-10 combining site and has suggested that at least one framework residue may contribute to antigen binding. The topography of the model suggests that a J_H-encoded somatic mutation may contribute significantly to the complementarity of H3 to the epitope on HEL. The predicted topographical arrangement of the strongly hydrophobic segments also suggests that they may play an important role in complex formation by fitting into hydrophobic pockets adjacent to the basic, hydrophilic epitope on HEL. The model serves as an aid in the design and interpretation of experiments investigating the nature of specificity of this antibody for its epitope on lysozyme, rather than as a prediction of the crystal structure at atomic resolution. We are currently performing energy-minimization experiments on a complex modeled from this hypothetical Fv structure and the X-ray co-ordinates of HEL (B. R. Brooks *et al.*, unpublished results; S. J. Smith-Gill *et al.*, unpublished results). The functional correlates predicted by this model are being tested in our laboratory by chain recombination experiments and by site-specific mutagenesis of the expressed HyHEL-10 V_H and V_L. In addition, final determination of details of the Fv structure, and thus experimental test of the model itself, will come from X-ray analysis of the crystallized Fab, which is in progress (Silverton *et al.*, 1984).

We thank R. Near and M. Gefter for the V_H36–60 probe and for providing us with the 36–60 amino acid sequence prior to its publication, S. Rudikoff for help and advice with the amino acid sequencing, and M. Millison for preparation of the manuscript. We are grateful to M. Potter for many helpful suggestions throughout this project.

References

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

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

Auffray, C. & Rougeon, F. (1980). *Biochemistry*, **107**, 303–314.

Barrett, E. J. (1979). *J. Chem. Educ.* **56**, 168–169.

Barstad, P., Hubert, J., Hunkapiller, M., Goetze, A., Schilling, J., Black, B., Eaton, B., Richards, J., Weigert, M. & Hood, L. (1978). *Eur. J. Immunol.* **8**, 497–503.

Birnboim, H. C. & Doly, J. (1979). *Nucl. Acids Res.* **7**, 1513–1523.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *J. Computat. Chem.* **4**, 187–217.

Buell, G. N., Wickens, M. P., Payvar, F. & Shimke, R. T. (1978). *J. Biol. Chem.* **253**, 2471–2482.

Cathala, G., Savouret, J.-F., Mendez, B., West, B. L., Karin, M., Martial, J. A. & Baxter, J. D. (1983). *DNA*, **2**, 329.

Chothia, C., Lesk, A. M., Levitt, M., Amit, A. G., Mariuzza, R. A., Phillips, S. E. V., Phillips, R. J. & Poljak, R. J. (1986). *Science*, **233**, 755–758.

Dagert, M. & Ehrlich, S. D. (1979). *Gene*, **6**, 23–26.

Darsley, M. J. & Reese, A. R. (1985*a*). *EMBO J.* **4**, 383–392.

Darsley, M. J. & Reese, A. R. (1985*b*). *EMBO J.* **4**, 393–398.

de la Paz, P., Sutton, B. J., Darsley, M. J. & Rees, A. R. (1986). *EMBO J.* **5**, 415–425.

Diamond, R. (1974). *J. Mol. Biol.* **82**, 371–391.

Efstratiadis, A., Kafatos, F. C., Maxam, A. M. & Maniatis, T. (1976). *Cell*, **7**, 279–288.

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

Glaudemans, C. P. J., Kovac, P. & Rasmussen, K. (1984). *Biochemistry*, **23**, 6732–6736.

Grunstein, M. & Hogness, D. S. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 3961–3965.

Gubler, Y. & Hoffman, B. J. (1983). *Gene*, **25**, 263–269.

Hanahan, D. (1983). *J. Mol. Biol.* **166**, 557–580.

Ito, H., Ike, Y., Ikuta, S. & Itakura, K. (1982). *Nucl. Acids. Res.* **10**, 1755–1769.

Juszczak, E., Near, R. I., Gefter, M. L. & Margolies, M. N. (1984). *J. Immunol.* **133**, 2603–2609.

Kabat, E. A. (1980). In *Methods in Enzymology*, vol. 70 (VanVanakis, H. & Langone, J. J., eds). pp. 3–49. Academic Press, New York.

Kurosawa, Y. & Tonegawa, S. (1982). *J. Exp. Med.* **15**, 201–218.

Mainhart, C. R., Potter, M. & Feldmann, R. J. (1984). *Mol. Immunol.* **21**, 469–478.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*, p. 242, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mariuzza, R. A., Jankovic, D. L., Boulot, G., Amit, A. G., Saludjian, P., LeGuern, A., Mazie, J. C. & Poljak, R. J. (1983). *J. Mol. Biol.* **170**, 1055–1058.

Maxam, A. M. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499–560.

McKean, D., Huppi, K., Bell, M., Staudt, L. & Gerhard, W. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 3180–3184.

Metzgar, D. W., Ch'ng, L.-K., Miller, A. & Sercarz, E. E. (1983). *Eur. J. Immunol.* **14**, 87–93.

Mukhopedhyay, M. & Mandal, N. C. (1983). *Anal. Biochem.* **113**, 265–270.

Near, R. I., Juszczak, E. C., Huang, S. J., Sicari, S. A., Margolies, M. N. & Gefter, M. L. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 2167–2171.

Novotny, J., Bruccoleri, R. & Karplus, M. (1984). *J. Mol. Biol.* **177**, 787–818.

Peacock, S. L., McIver, C. M. & Monohan, J. J. (1981). *Biochim. Biophys. Acta*, **655**, 243–250.

Phillips, D. C. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **57**, 484–495.

Reiher, W. E. (1985). Dissertation, Harvard University, pp. 20–32.

Rigby, P. W. J., Diekmann, M., Rhodes, C. & Berg, P. (1977). *J. Mol. Biol.* **113**, 237–256.

Rudikoff, S., Rao, D. N., Glaudemans, C. P. J. & Potter, M. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 4270–4274.

Rudikoff, S., Giusti, A. M., Cook, W. & Scharff, M. D. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 1979–1983.

Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 4298–4308.

Silverton, E., Padlan, E. A., Davies, D. R., Smith-Gill, S. J. & Potter, M. (1984). *J. Mol. Biol.* **180**, 761–765.

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

Smith-Gill, S. J., Mainhart, C. R., Lavoie, T. B., Rudikoff, S. & Potter, M. (1984*a*). *J. Immunol.* **132**, 963–967.

Smith-Gill, S. J., Lavoie, T. B. & Mainhart, C. R. (1984*b*). *J. Immunol.* **133**, 384–393.

Smith-Gill, S. J., Hamel, P. J., Klein, M. H., Rudikoff, S. R. & Dorrington, K. J. (1986). *Mol. Immunol.* **23**, 919–926.

Stablitzky, F., Wildner, G. & Rajewsky, K. (1985). *EMBO J.* **4**, 345–350.

Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T. & Itakura, K. (1979). *Nucl. Acids Res.* **6**, 3543–3557.

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

Zimmerman, C. L., Appella, E. & Pisano, J. J. (1977). *Anal. Biochem.* **77**, 569–573.

Edited by S. Brenner