PROTEINS: Structure, Function, and Genetics 7:112–124 (1990)

On the Nature of Antibody Combining Sites: Unusual Structural Features That May Confer on These Sites an Enhanced Capacity for Binding Ligands

Eduardo A. Padlan
Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892.

ABSTRACT **A detailed analysis of the structural aspects of antibody–antigen interactions has been made possible by the availability of X-ray structures for three complexes of antilysozyme Fabs to lysozyme (reviewed by Davies et al.: J. Biol. Chem. 263:10541–10544, 1988.) Examination of the antigen-contacting residues in the three antilysozyme Fabs reveals the occurrence of a large number of aromatics, particularly tyrosines, and the absence of apolar, aliphatic residues. Calculation of the frequency of occurrence of the various amino acid types reveals that tyrosines are three times, and histidines and asparagines eight times, more likely to be found in the complementarity-determining regions than in the framework of the variable domains. Analysis of the solvent accessibility of the residues in Fvs (the modules containing variable domains of the light and heavy chains) of known three-dimensional structure indicates that tyrosines and tryptophans are more exposed when they occur in the complementarity-determining regions than when in the framework. Furthermore, many more of the asparagines in the complementarity-determining regions than in the framework are buried. These asparagines appear to have a structural role in that they hydrogen-bond through their side chains to other side chains and, even more so, to the protein backbone. The stabilizing effect of the asparagines, plus the rigidity of the framework, may serve to allow the greater exposure of the aromatic residues to solvent. In view of the greater potential contribution of aromatic side chains to the total binding energy, these results suggest that antibody combining sites have structural features that make them especially suited for interacting with ligands.**

Key words: amino acid propensities, aromatic residues, conformational entropy, exposure patterns, interaction energy

INTRODUCTION

A large body of evidence supports the view that the CDRs,* which had been identified by sequence comparisons,[1,2] constitute the antibody combining site. In particular, there has been direct visualization by X-ray diffraction of the involvement of the CDRs in the specific interaction with antigen. While the pronounced sequence variability in these regions provides a ready explanation for the diversity of antigen-binding specificities, it is not altogether obvious why antigen binding should primarily involve the CDRs.

Evidence is likewise accumulating that all accessible areas of a macromolecule can potentially be bound by antibody[3,4] so that, with the possible exception of immunodominant sites, no special features would appear to be required for antigenicity. If the CDRs are to participate in strong interactions with the many possible antigenic structures, they must possess structural features that make them especially suited for binding ligands.

Recently, detailed three-dimensional structures of several antibody–antigen complexes have been reported.[5–8] Three of these complexes are with the same antigen, hen egg-white lysozyme. The three antilysozyme antibodies, D1.3,[5] HyHEL-5,[7] and HyHEL-10,[8] bind to essentially nonoverlapping epitopes, and their combining sites are quite different. The nature of the interaction between each combining site and its antigenic determinant was also found to be different in these three complexes.[4]

The residues in the CDRs of D1.3, HyHEL-5, and

Received July 17, 1989; accepted October 27, 1989.

Address reprint requests to Dr. Eduardo A. Padlan, Building 2, Room 323, National Institutes of Health, Bethesda, MD 20892.

Abbreviations used: CDR, complementarity-determining region; CH1, the first constant domain of the heavy chain; CL: CH1, the module containing the constant domain of the light chain and CH1; Fab, the antigen-binding fragment containing the light chain and the first two domains of the heavy chain; Fc, the fragment containing the C-terminal constant domains of the heavy chains; Fv, the module containing VL and VH; PDB, Protein Data Bank; VL and VH, variable domains of the light and heavy chains, respectively.

TABLE I. CDR Residues in Contact With Lysozyme*

	D1.3	HyHEL-5	HyHEL-10
ALA			
ARG	1	1	
ASN		2	2+1†
ASP	1	1	1
CYS			
GLN			1
GLU		2	
GLY	2	2	1
HIS	1		
ILE			
LEU			
LYS			
MET			
PHE	1†		
PRO		1	
SER	1	2	4+1†
THR		1†	
TRP	2	2	1
TYR	5	2	6
VAL			
Totals	13+1†	15+1†	16+2†

*Adapted from Amit et al.,[5] Sheriff et al.,[7] and Padlan et al.[8]
†Residues that contact the antigen but only through their backbone atoms.

HyHEL-10, which contact the lysozyme,[5,7,8] are listed in Table I. Charged amino acids are found among these residues, and some of those are involved in the interaction with the antigen.[7,8] Also, a disproportionately large number of tryptophans and tyrosines have been found in the combining sites of D1.3[5] and HyHEL-10,[8] and the aromatic side chains of many of them were found to be involved in antigen binding. Interestingly, one group of amino acids is found to be entirely missing from this list of contacting residues (Table I): No alanines, valines, isoleucines, or leucines are found to be in direct contact with the antigen. The side chains of these aliphatic residues have no polar groups that could allow them to participate in ionic interactions. On the other hand, tryptophans and tyrosines have large, aromatic side chains that could allow these amino acids greater surface interactions with ligand, as well as allow them to participate in hydrogen bonding. In view of these and other known differences in the potential contribution of the various amino acids to binding interactions, the unequal representation of amino acid types in contact with the antigen that is found in the antilysozyme–lysozyme complexes suggests that the CDRs and the combining sites that they form indeed possess unusual features that may confer on these sites an enhanced capacity for binding ligands.

In this paper, an attempt is made to gain some insight into these special structural features. First, the distribution of amino acids in variable domains is examined to explore the possibility that certain residues may have a propensity for being in the CDRs versus the framework. Next, the combining site residues in seven Fabs of known three-dimensional structure are examined and the various amino acid types are compared for their accessibility and potential involvement in antigen binding. The results of these surveys are then correlated with the potential contribution of the different amino acids to antigen-binding affinity and specificity.

Statistical analyses of immunoglobulin sequences have been done by others, most notably by Kabat and coworkers[1,2,9,10] some of their results will be compared below with those from the present analysis. A preliminary account of this study was presented at the Immunology of Lysozyme Workshop held June 13–15, 1988, at the National Institutes of Health, Bethesda, Maryland.

MATERIALS AND METHODS

Sequence Data

The source of the sequence data analyzed here was the tabulation of immunoglobulin sequences by Kabat et al.[11] To avoid errors that could result from the inclusion of partially determined sequences, only those sequences in which all of the residues have been identified were included in the survey. Furthermore, it was required that all of the residues have been unequivocably identified, specifically, that there were no undetermined amidation states (e.g., Asx or Glx) in the sequence. The collection of immunoglobulin sequences is more extensive for the mouse and probably represents a better sampling of its antibody repertoire than for any other animal; accordingly, the present analysis is confined to sequences of mouse origin. The sequences used were all different, although there were many instances in which two or more sequences differed from each other at only a few positions (sometimes just one); nonetheless, these nearly identical sequences were included as separate entries in the survey. In 14 of the sequences surveyed, pyrrolidone carboxylic acids occur as the amino-terminal residue; these are included in the tabulations of glutamines. The mouse sequences that were included in this survey are given in Appendix A.

Calculation of the Propensity of Each Amino Acid Type To Be Found in the CDRs

The frequency of occurrence of each amino acid type in the CDRs (or in the framework) was computed as the ratio of the number of occurrences of the particular amino acid in the CDRs (or in the framework) to the total number of occurrences (in the CDRs and in the framework). The propensity of a particular amino acid residue to be found in the CDRs, for example, is then the ratio of the frequency of its occurrence in the CDRs to that in the frame-

work. The CDR and framework boundaries that are used here are those of Kabat et al.[11]

Atomic Coordinates

Refined coordinates for seven crystallographically determined Fab structures were available to the author. These are for Fab NEW[12] from the Protein Data Bank[13] (PDB File 3FAB); Fab KOL[14] (PDB File 1FB4); Fab McPC603[15] (PDB File 1MCP); Fab HyHEL-5[7] (PDB File 2HFL); and Fab HyHEL-10[8] (PDB File 3HFM). In addition, coordinates for Fab *D1.3*[5] were made available by Prof. Roberto J. Poljak (Pasteur Institute), and highly refined coordinates for Fab J539 have recently become available (Bhat, Padlan and Davies, to be published). Atomic coordinates for human IgG1 Fc (PDB File 1FC1) were from the refinement of Deisenhofer.[16] The pyrrolidone carboxylic acids at the amino-termini of the heavy chains of HyHEL-5 and NEW and of the light chain of NEW were included in the tabulations as glutamines.

Calculation of the Solvent Accessibilities of the Amino Acid Residues in the Immunoglobulin Domains

The solvent accessibility of individual residues was assessed using program MS of Connolly[17] and programs developed by Sheriff et al.[18] A radius of 1.7 Å was assumed for the solvent probe; standard van der Waals' radii[19] were used. The fractional accessibility of the side chain, defined as the amount of surface area that is accessible to the solvent probe in the three-dimensional structure divided by the surface area that would be exposed if the residue (X) were in an isolated Gly-X-Gly tripeptide with the same backbone configuration as the corresponding tripeptide in the structure,[18,20,21] was computed for each residue.

Discrepancies could arise in the solvent exposures of residues in the Fvs and CL:CH1 domains as a consequence of the variation in the relative disposition of these modules in Fabs. Accordingly, the accessibility values for those residues were computed in the context of isolated Fvs or CL:CH1 modules. HyHEL-5 and HyHEL-10 have the same sequence in CL and CH1, and D1.3 differs from these two in CH1 at only two positions. Consequently, the CL:CH1 structures of these proteins are very similar. Also, the primary structures of J539 and McPC603, and of KOL and NEW, are identical in their CL and CH1 domains, and their three-dimensional structures are very similar. Therefore, exposure values were computed only for the CL:CH1 modules of HyHEL-10, J539, and KOL. The values for the residues in Fc were computed for the intact fragment, but only the accessibilities for the first (as designated by Deisenhofer[16] and as identified in PDB File 1FC1) chain were included in the analysis.

Calculation of the Solvent Accessibilities of the Amino Acid Residues in Other Protein Structures

Fractional accessibilities were also computed for the various amino acid types as they occur in other protein structures. Only water-soluble protein structures that had been determined at $\leq$1.8 Å resolution and that had been refined to a crystallographic R value of $\leq$0.200 were included in this survey. None of the immunoglobulin structures currently available satisfies these criteria. The atomic coordinates for the following 50 highly refined protein structures (with their code names in parentheses) were obtained from the Protein Data Bank[13]: penicillopepsin (2APP), rhizopuspepsin (2APR), actinidin (2ACT), wheat-germ agglutinin (3WGA), α-lytic protease (2ALP), carboxypeptidase A (5CPA), α-chymotrypsin (5CHA), chymotrypsinogen A (2CGA), tuna cytochrome C (4CYT), rice cytochrome C (1CCR), cytochrome C′ (2CCY), cytochrome C2 (3C2C), cytochrome C3 (2CDV), cytochrome C551 (451C), cytochrome P450 (2CPP), *Lactobacillus casei* dihydrofolate reductase (3DFR), *Escherichia coli* dihydrofolate reductase (4DFR), erythrocruorin (1ECD), flavodoxin (4FXN), hemoglobin (2HHB), insulin (1INS), T4 lysozyme (2LZM), human lysozyme (1LZ1), L7/L12 50S ribosomal protein (1CTF), myoglobin (1MBD), myohemerythrin (2MHR), ovomucoid (2OVO), papain (9PAP), phospholipase A2 (1BP2), plastocyanin (1PCY), *S. griseus* proteinase A (2SGA), *Streptomyces griseus* proteinase B (3SGB), rubredoxin (5RXN), thermolysin (3TLN), β-trypsin (1TPO), bovine trypsin inhibitor (5PTI), trypsinogen (2TGA), ubiquitin (1UBQ), azurin (2AZA), elastase (3EST), erabutoxin (3EBX), glutathione reductase (3GRS), pepsinogen (1PSG), proteinase K (2PRK), pseudoazurin (1PAZ), ribonuclease A (7RSA), subtilisin (1CSE), tonin (1TON), Trp repressor (2WRP), and *S. griseus* trypsin (1SGT).

The calculation of amino acid accessibility for these proteins followed the same procedure described above for the Fvs. Prior to the calculation of surfaces, solvent molecules were removed, while prosthetic groups were retained. When more than one conformational isomer was found for an amino acid residue, only the one with the higher occupancy (or only the first one if they had the same occupancy) was kept. Surfaces were computed for the quaternary structure that is found in solution; however, values for only one monomer (the first in the PDB entry) were used in the analyses. Complete details of the calculations can be obtained from the author.

RESULTS

Distribution of Amino Acids in Mouse Immunoglobulin Domains

The frequencies of occurrence of the various amino acid types in the variable and constant do -

TABLE II. Distribution of Amino Acids in Mouse Immunoglobulin Domains and in Other Proteins

	L-chain variable		H-chain variable		L and H constant*		
	No.	Percent occurrence	No.	Percent occurrence	No.	Percent occurrence	Other proteins† (% occurrence)
ALA	655	6.05	818	6.48	282	4.87	8.88
ARG	338	3.12	475	3.76	148	2.56	4.70
ASN	280	2.59	425	3.37	255	4.40	4.34
ASP	407	3.76	553	4.38	219	3.78	5.49
CYS	200	1.85	214	1.70	161	2.78	2.81
GLN	622	5.75	613	4.86	204	3.52	3.90
GLU	404	3.73	490	3.88	308	5.32	6.11
GLY	1,019	9.42	1,195	9.47	309	5.34	7.76
HIS	149	1.38	88	0.70	144	2.49	2.04
ILE	544	5.03	372	2.95	220	3.80	4.60
LEU	904	8.35	871	6.90	455	7.86	7.44
LYS	496	4.58	714	5.66	346	5.97	6.95
MET	147	1.36	235	1.86	73	1.26	1.69
PHE	386	3.57	395	3.13	191	3.30	3.47
PRO	512	4.73	382	3.03	473	8.17	4.56
SER	1,433	13.24	1,570	12.44	662	11.43	7.08
THR	993	9.18	1,030	8.16	584	10.08	5.98
TRP	180	1.66	389	3.08	109	1.88	1.12
TYR	536	4.95	972	7.70	151	2.61	3.47
VAL	615	5.68	819	6.49	497	8.58	6.84

*The immunoglobulin sequences are from the compilation of Kabat et al.[11] Only those sequences in which all of the residues have been unequivocably identified were included in this survey.
†Data are from the analysis of M.H. Klapper[22] of 207 unrelated polypeptide sequences.

mains of mouse light and heavy chains are presented in Table II. The frequencies of occurrence of the amino acids in 207 unrelated proteins, compiled by Klapper,[22] are included in Table II. The 207 proteins studied by Klapper will be taken here as representing all proteins in general. By and large, the values for the immunoglobulin domains parallel those for the 207 proteins, with a few exceptions.

Table II shows that histidines are found less frequently in the immunoglobulin variable domains than in the 207 unrelated protein sequences, especially in the heavy chains. On the other hand, more tryptophans and more tyrosines are found in the immunoglobulin variable domains, also especially in the heavy chains.

Table II also shows that serines and threonines constitute more than 21% of the residues in the immunoglobulin variable domains and in the constant domains versus 13% in the 207 proteins. Since threonines are frequently found in β-pleated sheets,[23] this difference may simply reflect the predominantly β-sheet-nature of immunoglobulin structures.[24,25]

The frequencies of occurrence of the amino acids in the CDRs and in the frameworks are presented in Table III. There are four instances in which an amino acid is found at least three times more frequently in either the CDRs or in the framework. Asparagine, histidine, and tyrosine show a propensity for being in CDRs, while cysteine shows a propensity for being in the framework. Asparagines and histidines are eight times, and tyrosines are three times, more likely to be found in CDRs, whereas cysteines are 50 times more likely to be found in the framework. The cysteines in the framework are those that form the intradomain disulfide bridges that play an important role in immunoglobulin structure.[24,25]

The propensity of histidines and asparagines for being in CDRs had been noted earlier by Kabat et al.[9] Using much less sequence data, those authors found that histidines and asparagines were about two times more likely to be found in the CDRs. Tyrosines were found just as likely to be in CDRs as in the framework in that earlier survey.

The apolar, aliphatic amino acids Ala, Val, Ile, and Leu represent 26.6% of framework residues (Table III), a value that is essentially the same as the 25.1% found in the constant domains and the 27.8% found in the 207 proteins (Table II). In contrast, these amino acids represent only 16.6% of the CDR residues (Table III).

Pattern of Exposure of the Amino Acids

The accessibilities of the various amino acid types, as they occur in the CDRs of the seven Fab structures considered here, are presented in Table IV. The accessibilities of the amino acids when in the framework are presented in Table V. The accessibilities in the CL:CH1 modules and in the Fc are presented in Table VI. The accessibilities in the 50 water-soluble proteins, whose structures are known to high accuracy, are presented in Table VII.

The accessibility values shown in Table VII for

TABLE III. Distribution of Amino Acids in Mouse Variable Domains*

	CDRs		Frameworks			
	No. in the CDRs	Percent occurrence	No. in the frameworks	Percent occurrence	Propensity (P)	1/P
ALA	333	5.29	1,140	6.65	0.80	1.26
ARG	237	3.76	576	3.36	1.12	0.89
ASN	524	8.32	181	1.06	7.88	0.13
ASP	364	5.78	596	3.48	1.66	0.60
CYS	3	0.05	411	2.40	0.02	50.33
GLN	216	3.43	1,019	5.94	0.58	1.73
GLU	152	2.41	742	4.33	0.56	1.79
GLY	425	6.75	1,789	10.44	0.65	1.55
HIS	178	2.83	59	0.34	8.21	0.12
ILE	191	3.03	725	4.23	0.72	1.39
LEU	280	4.45	1,495	8.72	0.51	1.96
LYS	277	4.40	933	5.44	0.81	1.24
MET	143	2.27	239	1.39	1.63	0.61
PHE	212	3.37	569	3.32	1.01	0.99
PRO	219	3.48	675	3.94	0.88	1.13
SER	926	14.70	2,077	12.12	1.21	0.82
THR	416	6.61	1,607	9.37	0.70	1.42
TRP	137	2.18	432	2.52	0.86	1.16
TYR	827	13.13	681	3.97	3.31	0.30
VAL	238	3.78	1,196	6.98	0.54	1.85
Total	6,298	100.0	17,142	100.0		

*The immunoglobulin sequences are from the compilation of Kabat et al.[11] Only those sequences in which residues 1–107 in the light chain and 1–113 in the heavy chain have been unequivocably identified were included in this survey.

TABLE IV. Exposures for KOL, NEW, McPC603, J539, D1.3, HyHEL-5, and HyHEL-10 CDRs

	Exposure*						Average exposure	S.D.
	Bu	mB	pB	mE	Ex	Totals		
ALA	11	3	3	0	6	23	0.351	0.396
ARG	2	2	2	4	5	15	0.593	0.295
ASN	11	2	7	6	5	31	0.417	0.325
ASP	4	2	8	9	7	30	0.588	0.288
CYS	2	1	0	0	0	3	0.090	0.127
GLN	8	1	1	1	3	14	0.296	0.382
GLU	4	1	2	2	0	9	0.330	0.255
GLY	7	0	0	0	27	34	0.794	0.404
HIS	3	1	3	2	4	13	0.519	0.338
ILE	11	2	3	1	0	17	0.189	0.214
LEU	10	3	6	1	0	20	0.224	0.211
LYS	0	0	2	4	8	14	0.766	0.171
MET	5	0	0	1	0	6	0.113	0.253
PHE	3	2	2	2	0	9	0.342	0.262
PRO	3	2	1	4	2	12	0.491	0.330
SER	8	3	8	19	32	70	0.704	0.295
THR	2	4	11	7	4	28	0.558	0.229
TRP	2	3	5	2	0	12	0.415	0.200
TYR	5	11	13	11	3	43	0.480	0.231
VAL	8	4	0	1	0	13	0.187	0.204
Total	109	47	77	77	106	416	0.504	0.349

*An amino acid residue whose side chain has a fractional accessibility value between 0.00 and 0.20 is designated as completely buried (Bu), between 0.20 and 0.40 as mostly buried (mB), between 0.40 and 0.60 as partly buried (pB), between 0.60 and 0.80 as mostly exposed (mE), and at least 0.80 as completely exposed (Ex). In the special case of glycine, the residue is designated as completely exposed if its α-carbon is accessible to solvent; otherwise, it is designated as completely buried.

proteins in general parallel those obtained by others.[26,27] Thus hydrophobic residues are for the most part found buried in the interior of the proteins, while polar residues, especially those with charged side chains, are found mostly exposed to solvent. Specifically, Ile, Leu, Met, Phe, Trp, and Val are predominantly buried in these 50 proteins. Cysteines, which frequently form disulfide bridges, are found also to be mostly buried.

The accessibility values for the immunoglobulin

TABLE V. Exposures for KOL, NEW, McPC603, J539, D1.3, HyHEL-5, and HyHEL-10 Frameworks

	Exposure*						Average	
	Bu	mB	pB	mE	Ex	Total	exposure	S.D.
ALA	27	1	5	7	25	65	0.498	0.437
ARG	12	12	5	9	7	45	0.445	0.280
ASN	1	0	2	11	4	18	0.682	0.175
ASP	14	2	6	8	6	36	0.454	0.316
CYS	28	0	0	0	0	28	0.000	0.000
GLN	14	14	7	12	22	69	0.537	0.350
GLU	5	0	3	21	18	47	0.706	0.263
GLY	33	0	0	0	90	123	0.732	0.443
HIS	1	0	0	0	3	4	0.713	0.310
ILE	31	5	4	1	7	48	0.225	0.323
LEU	72	7	2	13	7	101	0.195	0.312
LYS	1	0	6	19	27	53	0.769	0.165
MET	13	1	2	1	0	17	0.125	0.224
PHE	30	3	3	2	0	38	0.128	0.181
PRO	7	0	10	24	13	54	0.623	0.282
SER	1	7	9	31	108	156	0.839	0.187
THR	10	5	19	34	37	105	0.653	0.265
TRP	28	1	0	0	0	29	0.045	0.060
TYR	38	6	4	0	0	48	0.109	0.156
VAL	42	16	7	5	10	80	0.281	0.314
Total	408	80	94	198	384	1,164	0.502	0.394

*See footnote to Table IV.

TABLE VI. Exposures for J539, KOL, and HyHEL-10 CL:CH1s and for Human IgG1 Fc

	Exposure*						Average	
	Bu	mB	pB	mE	Ex	Totals	exposure	S.D.
ALA	10	4	9	8	13	44	0.540	0.347
ARG	0	1	9	4	6	20	0.660	0.216
ASN	2	4	9	10	19	44	0.694	0.258
ASP	0	1	5	14	15	35	0.752	0.191
CYS	18	0	1	2	2	23	0.162	0.304
GLN	3	0	9	10	9	31	0.639	0.256
GLU	2	2	6	19	10	39	0.660	0.227
GLY	3	0	0	0	37	40	0.925	0.263
HIS	7	1	6	2	2	18	0.386	0.314
ILE	13	2	3	1	1	20	0.220	0.255
LEU	28	12	7	5	2	54	0.238	0.244
LYS	1	2	15	19	19	56	0.693	0.214
MET	3	2	1	1	1	8	0.340	0.279
PHE	16	5	3	0	0	24	0.154	0.177
PRO	17	2	16	17	16	68	0.525	0.340
SER	24	10	10	23	44	111	0.601	0.370
THR	9	11	15	16	26	77	0.594	0.285
TRP	10	1	1	1	0	13	0.132	0.219
TYR	13	7	8	2	0	30	0.253	0.192
VAL	48	11	7	8	3	77	0.212	0.267
Total	227	78	140	162	225	832	0.509	0.354

*See footnote to Table IV.

variable and constant domains (Table VI), the latter despite being obtained from only four structures, closely parallel those for the 50 water-soluble (Table VII). This implies that immunoglobulins are not unusual in their general structure.

In the immunoglobulin variable domains, however, the pattern of exposure for certain amino acids is different when they are in CDRs or in the framework. For example, tryptophans, when in the framework, are almost all completely buried (average exposure = 0.045, S.D. = 0.060) (Table V); when in the CDRs, the tryptophans are more exposed (average exposure = 0.415, S.D. = 0.200; Table IV). This difference is shown graphically in Figure 1. Similarly, tyrosines, when in the framework, are mostly buried (average exposure = 0.109, S.D. = 0.156; Table V); when in the CDRs, the tyrosines are more exposed (average exposure 0.480, S.D. = 0.231; Table IV). This difference is shown graphically in Figure 2.

TABLE VII. Exposures for 50 Highly-Refined, Water-Soluble Proteins*

	Exposure†						Average	
	Bu	mB	pB	mE	Ex	Totals	exposure	S.D.
ALA	309	79	97	117	199	801	0.430	0.381
ARG	21	42	83	80	50	276	0.571	0.235
ASN	49	52	99	123	145	468	0.610	0.281
ASP	69	53	79	113	151	465	0.590	0.315
CYS	174	46	26	12	1	259	0.157	0.206
GLN	39	47	61	106	91	344	0.579	0.275
GLU	30	33	72	132	127	394	0.648	0.255
GLY	291	0	0	0	635	926	0.686	0.464
HIS	63	39	41	23	25	191	0.375	0.287
ILE	273	85	62	22	8	450	0.186	0.233
LEU	372	121	70	43	22	628	0.206	0.255
LYS	10	29	91	194	187	511	0.703	0.204
MET	88	20	17	13	6	144	0.209	0.265
PHE	177	70	42	8	8	305	0.198	0.225
PRO	81	32	54	86	109	362	0.552	0.334
SER	156	74	79	145	293	747	0.586	0.355
THR	140	68	123	145	135	611	0.507	0.323
TRP	68	44	14	7	3	136	0.221	0.209
TYR	106	113	74	48	20	361	0.356	0.249
VAL	385	101	84	61	36	667	0.230	0.285
Total	2,915	1,157	1,268	1,478	2,251	9,069	0.462	0.364

*Resolution, ≤1.8Å; R value, ≤0.200.
†See footnote to Table IV.

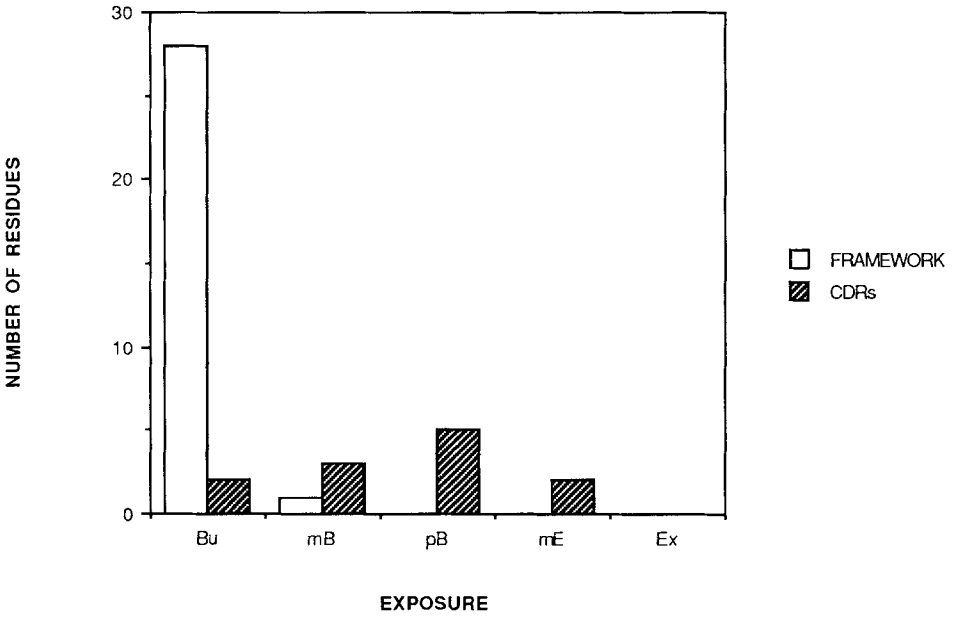

Fig. 1. The pattern of exposure for the tryptophans in the Fvs of proteins D1.3, HyHEL-5, HyHEL-10, McPC603, J539, KOL, and NEW, when in the framework (open bars) and when in the CDRs (hatched bars). See footnote to Table 4 for definitions of the exposure designations.

The exposure pattern for Asn likewise depends on where it is located in the structure. When in the framework, the asparagines are mostly exposed with little variation (average exposure = 0.682, S.D. = 0.175; Table V); when in the CDRs, more of these residues are buried (Table IV). Of the 31 asparagines in the CDRs of the seven Fabs studied here, 24 have their side chains potentially involved in hydrogen-bonding interactions; of these, 17 are potentially hydrogen bonded to main chain atoms.

The pattern of exposure for Glu also shows that this amino acid is slightly more buried in the CDRs; however, the spread in the exposure values is much greater (Tables IV, V). Of the nine glutamic acids

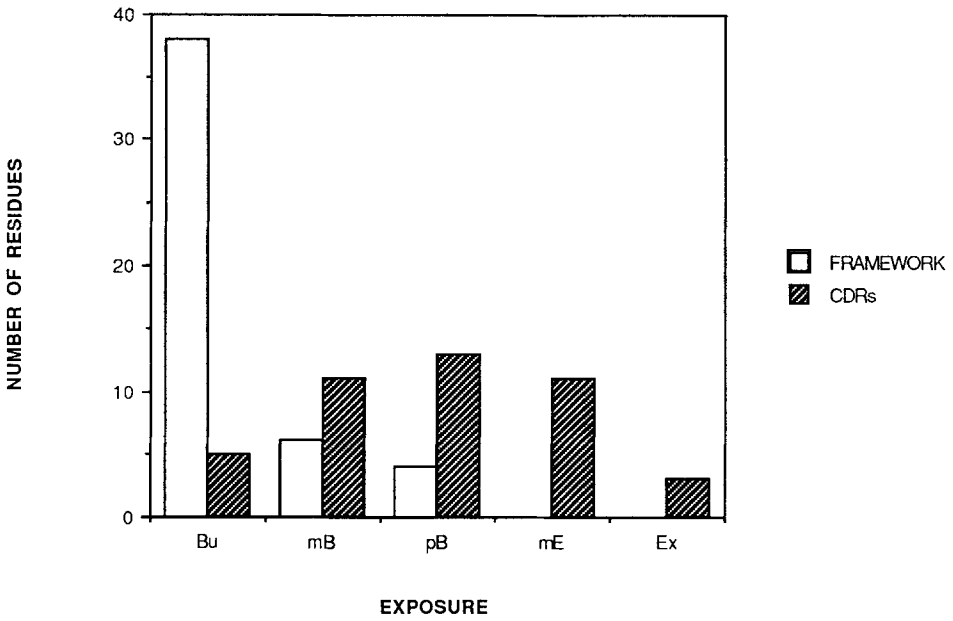

Fig. 2. The pattern of exposure for the tyrosines in the Fvs of proteins D1.3, HyHEL-5, HyHEL-10, McPC603, J539, KOL, and NEW, when in the framework (open bars) and when in the CDRs (hatched bars). See footnote to Table 4 for definitions of the exposure designations.

that are present in the CDRs of the seven Fabs, four are completely buried and one is mostly buried (Table IV). The carboxylates of all five are within hydrogen-bonding distance of other polar atoms, and one forms a salt link with the side chain of an arginine from the opposite chain. One of the buried glutamic acid residues is at the bottom of the hapten-binding cavity of McPC603 and probably plays a role in the binding to the specific ligand.[28,29] Two of the buried glutamic acids are in the combining site of HyHEL-5 and have been found to be important in neutralizing positive charges on the antigen.[7]

The differences in the exposure patterns for Asn and Glu are not as pronounced as those found for Trp and Tyr. Whereas the differences in the average exposures for CDR Trp and Tyr are 6.1 and 2.4 times the standard deviations when in the framework, respectively, the differences in the average exposures of the framework and CDR Asn and Glu are only about 1.5 S.D.

DISCUSSION

The energy of interaction between proteins derives from enthalpic and entropic factors. Positive contributions come from ion pairs and hydrogen bonds, the nonspecific van der Waals' forces, and the effects of sequestering hydrophobic groups from water.[30] For binding to occur, the negative effect of entropy loss, resulting from the immobilization of the molecules and of polypeptide segments and amino acid side chains upon complex formation, must be overcome.

The close shape complementarity of the interacting surfaces, as observed in antibody–antigen complexes,[4–7] results in the virtual exclusion of water from the interface. The removal of this surface water could have a significant effect on the interaction, for example, by altering the dielectric constant in the interface, by changing the pattern of interaction between polar groups, and so forth.

The various amino acids differ in their potential contribution to the binding energy. All residues contribute via van der Waals' interactions and the hydrophobic effect. The residues with polar side chains can contribute further in that they can participate in hydrogen bonds and, for those that are charged, in salt bridges.

The residues with aromatic side chains can contribute substantially to the energy of binding. They present large surface areas for interaction with ligand and can form hydrogen bonds either through their aromatic rings[31] or, for some, through polar atoms in the side chain. Another factor that contributes significantly to the importance of aromatic side chains in ligand binding is the fusion of many atoms in the ring systems. This prior immobilization obviates the loss of conformational entropy that could occur upon complex formation. Since the loss of conformational entropy is estimated to be equivalent to 0.6 kcal at 300 K for every side chain torsion that is fixed,[32] the utilization of aromatic side chains in the binding would lead to a more substantial contribution to the energy of interaction than if an aliphatic side chain, with the same number of atoms, were used instead. Of course, some side chains on the surface, through interactions with neighboring struc-

tures, may already be effectively immobilized even prior to complex formation.

The apolar, aliphatic side chains Ala, Val, Ile and Leu are incapable of ionic interactions and can contribute to the binding only through the weaker van der Waals' forces and through the hydrophobic effect. Furthermore, in view of the many rotational degrees of freedom in their side chains that could be frozen upon complex formation, the presence of aliphatic residues in the contact can have a significant negative contribution to the binding.

In view of the above arguments, it is not surprising that more aromatic residues and fewer apolar, aliphatic residues are found in the combining sites of antibodies.

The greater exposure of tryptophans and tyrosines in the CDRs means that they would be more available for interaction with ligand. Indeed, in the D1.3, HyHEL-5, and HyHEL-10 complexes with lysozyme, tryptophans and tyrosines are responsible for 155 of the 302 interatomic contacts between these antibodies and the antigen. Some of these contacts involve hydrogen bonds; most are van der Waals' interactions.[5,7,8]

Some of the tryptophans and tyrosines in the framework are conserved among the variable domain sequences. For example, Trp 35 in the light chain and Trp 36 and Trp 103 in the heavy chain (numbering according to Kabat et al.[11]) are present in all of the variable domain sequences and in the seven Fab structures included in this survey. Furthermore, Tyr 36 and Tyr 86 in the light chain and Tyr 90 in the heavy chain are present in all seven Fab structures. The 29 framework tryptophans in the seven Fabs are found in only five positions, while the 12 CDR tryptophans are found in eight positions. Similarly, the 48 framework tyrosines are found in only 11 positions in the seven Fabs, while the 43 CDR tyrosines are scattered among 21 positions. Nevertheless, in the tabulations and in Figures 1 and 2, the conserved tryptophans and tyrosines are counted as individual occurrences. Conclusions drawn from the differences observed here in the exposure patterns of these residues will not be affected by whether the conserved residues are counted as one or as individual occurrences. Following the argument of Kabat et al.[9], the scattering of the tryptophans and tyrosines among many positions in the CDRs leads one to conclude that these amino acids are specificity determining. The greater exposure of these residues in the combining sites supports this conclusion.

Histidines are found rarely (only 1%) in immunoglobulin variable domains, but they occur eight times more frequently in CDRs than in the framework. This suggests that this residue also plays an important role in antigen binding. The side chain of histidine has a number of properties[33] that can contribute in a variety of ways to ligand-binding affinity and specificity. It is the sixth largest and has an aromatic, imidazole ring that is readily protonated, resulting in a positive charge. Furthermore, the nitrogens in the ring can act as proton donors and acceptors in hydrogen bonding.

However, the available structural data do not allow definitive statements to be made regarding the role of histidines in antigen binding. The exposure pattern for the histidines in the CDRs reveals that this amino acid has no preference for being buried or exposed in the seven Fabs studied here (Table IV). In these seven Fabs, there is only one example of a histidine being involved in ligand binding; in D1.3 (Table I), one histidine is involved in five interatomic contacts with the lysozyme, including one hydrogen bond.[5] More examples of the involvement of histidines in ligand binding must become available before one can make a full assessment of the contribution of this residue to antigen-binding affinity and specificity.

The occurrence of eight times more asparagines in the CDRs compared with in the framework cannot be rationalized in terms of their direct contribution to binding energy, since these residues, when in the CDRs, are more buried and are thus less available for binding. Instead, these asparagines appear to be playing a more ancillary, i.e., structural, role in view of their participation in many hydrogen bonds, especially to main chain atoms. Hydrogen bonding to the polypeptide backbone by asparagine may serve to buttress local structures. Structural roles for Asn have been noted before, e.g., in β-turns[34,35] and in helices.[36]

Although the glutamic acid residues are found to be less exposed when in the CDRs, a structural role for this amino acid is not obvious. On the contrary, the involvement in ligand binding of three of the five buried glutamic acids suggests a specificity-determining role for this residue.

The exposure to water of hydrophobic side chains is energetically expensive, and their tendency would be to turn inward. The greater accessibility of the aromatic side chains in the combining sites of the seven Fabs studied here, and probably of combining sites in general, must be the result of a strong force that had overcome this tendency. Several factors come to mind. First, there is a strong framework, whose structural elements are conserved in antibodies,[24,25,37–41] that provides a strong scaffolding for the CDRs. Second, the interactions between CDR and framework elements and between the CDRs[37–42] serve to provide additional support for the CDR structures. Third, certain amino acids in the CDRs undoubtedly serve a structural rather than a specificity-determining role[9,41]; an example of this would be the asparagines, as discussed above.

It must be pointed out that some of the CDRs considered here were from crystal structures in which they were in contact with antigen, as in the case of

D1.3, HyHEL-5, and HyHEL-10, or in contact with another molecule in the lattice, as in the case of J539.[43,44] What the exposure patterns would be in the uncomplexed Fabs can only be the subject of speculation.

Of particular relevance is the possibility of conformational change occurring upon complex formation; such a change could significantly affect the exposure patterns of the amino acids. However, there are indications that no major conformational changes have occurred in at least some of these CDR structures on binding antigen. For example, the backbone trace of the CDRs of HyHEL-10 VL is essentially the same as that for REI[45] and D1.3 VL, all three proteins having the same number of amino acids in those regions. Furthermore, the first and second CDRs of the heavy chains of HyHEL-10 and D1.3 have the same lengths and, again, have essentially the same backbone structures.[8] Since the REI structure is uncomplexed and D1.3 is in an entirely different complex with lysozyme, the similarity in the backbone structures of these CDRs strongly suggests that these regions are by and large unaffected by the ligand state of the molecule. This implies, further, that the disposition of the side chains emanating from these CDRs, i.e., whether they are pointing inward or outward to solvent, will not be greatly affected by whether the Fab is complexed.

Furthermore, for the other three Fabs, McPC603, NEW, and KOL, the combining sites in the crystal structure are freely exposed to solvent. For two of these, McPC603 and NEW, the accessibility of the combining sites in the crystal was utilized for the formation of hapten complexes by diffusion.[46,47] McPC603 has many aromatics in the CDRs that are exposed to solvent. Another Fab, whose structure is known and that is uncomplexed, also has many aromatic side chains exposed to solvent.[48] All of these observations suggest that the exposure patterns observed in the seven Fab structures considered here are probably representative of those occurring in Fabs in general.

CONCLUSIONS

It is the natural tendency of water-soluble proteins to fold in such a manner as to bring its hydrophobic side chains together to form an oily core and leave its polar side chains exposed to solvent.[30] This distribution of polar and apolar amino acids in a protein structure has been confirmed by the earliest crystal structure of a protein[49] and by subsequent studies, as indicated by the exposure values in Table VII. There are instances, however, when structural stability appears to have been sacrificed for the sake of biological function.

For example, charged side chains, which are normally exposed to solvent, are sometimes found in the interior. In some of these cases, an obvious functional purpose is served. These include the buried aspartic acid in the catalytic triad of the serine proteases, which plays an important role in the mechanism of action of these enzymes[50]; the burying of the propionates of the heme group in cytochrome C, which exposes a hydrophobic edge of the porphyrin for possible interaction with another molecule[51]; and others.

The presence of large, aromatic side chains on the surface of antibody combining sites appears to be another instance in which a diminution of structural stability is permitted in order to achieve a biological purpose; the purpose, in this case, is enhanced capacity for binding ligands. The large, aromatic side chains are normally found in the interior of proteins, where their hydrophobicity contributes to the oily core; when directed outward and made available for ligand interaction, these side chains can contribute significantly to the binding energy. In view of the relative importance of aromatic amino acids in binding interactions, it is not surprising that these residues figure prominently also in the interface between closely interacting immunoglobulin domains, for example, in the contact between the variable domains[37,52,53] and in the contact between the CH3 domains of IgG Fc.[16]

Another family of proteins, whose members display the ability to bind to a large number of different structures, are the major histocompatibility complex (MHC) antigens,[54,55] which bind peptides. The putative binding groove of a class I MHC antigen, whose structure is available,[56] was found to be lined with many aromatic side chains. The structural arguments presented above for the antibody combining sites are, in all likelihood, valid also for the binding sites of the MHC antigens.

ACKNOWLEDGMENTS

I thank Drs. David R. Davies and Elvin A. Kabat for comments and discussions.

REFERENCES

1. Wu, T.T., Kabat, E.A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J. Exp. Med. 132:211–250, 1970.
2. Kabat, E.A., Wu, T.T. Attempts to locate complementarity-determining residues in the variable positions of light and heavy chains. Ann. N.Y. Acad. Sci. 190:382–393, 1971.
3. Benjamin, D.C., Berzofsky, J.A., East, I.J., Gurd, F.R.N., Hannum, C., Leach, S.J., Margoliash, E., Michael, J.G., Miller, A., Prager, E.M., Reichlin, M., Sercarz, E.E., Smith-Gill, S.J., Todd, P.E., Wilson, A.C. The antigenic structure of proteins: A reappraisal. Annu. Rev. Immunol. 2:67–101, 1984.
4. Davies, D.R., Sheriff, S., Padlan, E.A. Antibody–antigen complexes. J. Biol. Chem. 263:10541–10544, 1988.
5. Amit, A.G., Mariuzza, R.A., Phillips, S.E.V., Poljak, R.J. Three-dimensional structure of an antigen–antibody complex at 2.8 Å resolution. Science 233:747–753, 1986.
6. Colman, P.M., Laver, W.G., Varghese, J.N., Baker, A.T., Tulloch, P.A., Air, G.M., Webster, R.G. Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. Nature 326:358–363, 1987.
7. Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H.,

Smith-Gill, S.J., Finzel, B.C., Davies, D.R. Three-dimensional structure of an antibody–antigen complex. Proc. Natl. Acad. Sci. USA 84:8075–8079, 1987.
8. Padlan, E.A., Silverton, E.W., Sheriff, S., Cohen, G.H., Smith-Gill, S.J., Davies, D.R. Structure of an antibody–antigen complex: Crystal structure of the HyHEL-10 Fab–lysozyme complex. Proc. Natl. Acad. Sci. USA 86:5938–5942, 1989.
9. Kabat, E.A., Wu, T.T., Bilofsky, H. Unusual distribution of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. J. Biol. Chem. 252:6609–6616, 1977.
10. Kabat, E.A. The structural basis of antibody complementarity. *Adv. Protein* Chem. 32:1–75, 1978.
11. Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M., Gottesman, K. (eds.). "Sequences of Proteins of Immunological Interest," 4th ed. Bethesda, MD: National Institutes of Health, 1987.
12. Saul, F.A., Amzel, L.M., Poljak, R.J. Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin New at 2.0 A resolution. J. Biol. Chem. 253:585–597, 1978.
13. 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. The Protein Data Bank. A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
14. Marquart, M., Deisenhofer, J., Huber, R., Palm, W. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. J. Mol. Biol. 141: 369–391, 1980.
15. Satow, Y., Cohen, G.H., Padlan, E.A., Davies, D.R. Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. J. Mol. Biol. 190:593–604, 1986.
16. Deisenhofer, J. Crystallographic refinement and atomic models of a human Fc fragment and its complex with Fragment B of Protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å resolution. Biochemistry 20:2361–2370, 1981.
17. Connolly, M.L. Analytical molecular surface calculation. J. Appl. Crystallogr. 16:548–558, 1983.
18. Sheriff, S., Hendrickson, W.A., Stenkamp, R.E., Sieker, L.C., Jensen, L.H. Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrin. Proc. Natl. Acad. Sci. USA 82:1104–1107, 1985.
19. Case, D.A., Karplus, M. Dynamics of ligand binding to heme proteins. J. Mol. Biol. 132:343–368, 1979.
20. Shrake, A., Rupley, J.A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J. Mol. Biol. 79:351–371, 1973.
21. Padlan, E.A., Davies, D.R. A model of the Fc of immunoglobulin E. Mol. Immunol. 23:1063–1075, 1986.
22. Klapper, M.H. The independent distribution of amino acid near neighbor pairs into polypeptides. Biochem. Biophys. Res. Commun. 78:1018–1024, 1977.
23. Chou, P.Y., Fasman, G.D. Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry 13:211–222, 1974.
24. Poljak, R.J., Amzel, L.M., Avey, H.P., Chen, B.L., Phizackerley, R.P., Saul, F. Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8-Å resolution. Proc. Natl. Acad. Sci. USA 70:3305–3310, 1973.
25. Schiffer, M., Girling, R.L., Ely, K.R., Edmundson, A.B. Structure of a lambda-type Bence-Jones protein at 3.5-Å resolution. Biochemistry 12:4620–4631, 1973.
26. Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H., Zehfus, M.H. Hydrophobicity of amino acid residues in globular proteins. Science 229:834–838, 1985.
27. Miller, S., Janin, J., Lesk, A.M., Chothia, C. Interior and surface of monomeric proteins. J. Mol. Biol. 196:641–656, 1987.
28. Segal, D.M., Padlan, E.A., Cohen, G.H., Rudikoff, S., Potter, M., Davies, D.R. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen-binding site. Proc. Natl. Acad. Sci. USA 71:4298–4302, 1974.
29. Padlan, E.A., Cohen, G.H., Davies, D.R. On the specificity of antibody/antigen interactions: Phosphocholine binding to McPC603 and the correlation of three-dimensional structure and sequence data. Ann. Inst. Pasteur/Immunol. 136C:271–276, 1985.
30. Kauzmann, W. Some factors in the interpretation of protein denaturation. Adv. Protein Chem. 14:1–63, 1959.
31. Levitt, M., Perutz, M.F. Aromatic rings act as hydrogen bond acceptors. J. Mol. Biol. 201:751–754, 1988.
32. Novotny, J., Bruccoleri, R.E., Saul, F.A. On the attribution of binding energy in antigen–antibody complexes McPC 603, D1.3, and HyHEL-5. Biochemistry 28:4735–4749, 1989.
33. Creighton, T.E. "PROTEINS Structures and Molecular Principles." New York: W.H. Freeman, 1984.
34. Richardson, J.S. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34:167–339, 1981.
35. Wilmot, C.M., Thornton, J.M. Analysis and prediction of the different types of beta-turn in proteins. J. Mol. Biol. 203:221–232, 1988.
36. Richardson, J.S., Richardson, D.C. Amino acid preferences for specific locations at the ends of alpha helices. Science 240:1648–1652, 1988.
37. Davies, D.R., Padlan, E.A., Segal, D.M. Immunoglobulin structures at high resolution. In: "Contemporary Topics in Molecular Immunology," Vol. 4. Inman, F.P., Mandy, W.J., eds. New York: Plenum Press, 1975: 127–155.
38. Padlan, E.A. Structural basis for the specificity of antibody–antigen reactions and structural mechanisms for the diversification of antigen binding specificities. Q. Rev. Biophys. 10:35–65, 1977.
39. Amzel, L.M., Poljak, R.J. Three-dimensional structure of immunoglobulins. Annu. Rev. Biochem. 48:961–997, 1979.
40. Davies, D.R., Metzger, H. Structural basis of antibody function. Annu. Rev. Immunol. 1:87–117, 1983.
41. Chothia, C., Lesk, A.M. Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. 196: 901–917, 1987.
42. Chothia, C., Novotny, J., Bruccoleri, R., Karplus, M. Domain association in immunoglobulin molecules. The packing of variable domains. J. Mol. Biol. 186:651–663, 1985.
43. Navia, M.A., Segal, D.M., Padlan, E.A., Davies, D.R., Rao, N., Rudikoff, S., Potter, M. Crystal structure of galactan-binding mouse immunoglobulin J539 Fab at 4.5-Å resolution. Proc. Natl. Acad. Sci. USA 76:4071–4074, 1979.
44. Suh, S.W., Bhat, T.N., Navia, M.A., Cohen, G.H., Rao, D.N., Rudikoff, S., Davies, D.R. The galactan-binding immunoglobulin Fab J539: An X-ray diffraction study at 2.6-Å resolution. Proteins 1:74–80, 1986.
45. Epp, O., Lattman, E.E., Schiffer, M., Huber, R., Palm, W. The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0 Å resolution. Biochemistry 14:4943–4952, 1975.
46. Padlan, E.A., Segal, D.M., Spande, T.F., Davies, D.R., Rudikoff, S., Potter, M. Structure at 4.5 Å resolution of a phosphorylcholine-binding Fab. Nature [New Biol.] 245: 165–167, 1973.
47. Amzel, L.M., Poljak, R.J., Saul, F., Varga, J.M., Richards, F.F. The three dimensional structure of a combining region–ligand complex of immunoglobulin NEW at 3.5-Å resolution. Proc. Natl. Acad. Sci. USA 71:1427–1430, 1974.
48. Lascombe, M.-B., Alzari, P.M., Boulot, G., Saludjian, P., Tougard, P., Berek, C., Haba, S., Rosen, E.M., Nisonoff, A., Poljak, R.J. Three-dimensional structure of Fab R19.9, a monoclonal murine antibody specific for the p-azobenzenearsonate group. Proc. Natl. Acad. Sci. USA 86:607–611, 1989.
49. Kendrew, J.C. Myoglobin and the structure of proteins. Science 139:1259–1266, 1963.
50. Kraut, J. Serine proteases: Structure and mechanism of catalysis. Annu. Rev. Biochem. 46:331–358, 1977.
51. Salemme, F.R. Structure and formation of cytochrome c. Annu. Rev. Biochem. 46:299–329, 1977.
52. Poljak, R.J., Amzel, L.M., Chen, B.L., Phizackerley, R.P., Saul, F. The three-dimensional structure of the Fab' fragment of a human myeloma immunoglobulin at 2.0-Å resolution. Proc. Natl. Acad. Sci. USA 71:3440–3444, 1974.
53. Novotny, J., Haber, E. Structural invariants of antigen binding: Comparison of immunoglobulin VL-VH and VL-VL domain dimers. Proc. Natl. Acad. Sci. USA 82:4592–4596, 1985.
54. Bell, J.I., Denny, D.W., Jr., McDevitt, H.O. Structure and

polymorphism of murine and human class II major histocompatibility antigens. Immunol. Rev. 84:51–71, 1985.
55. Lopez de Castro, J.A., Barbosa, J.A., Kragel, M.S., Biro, P.A., Strominger, J.L. Structural analysis of functional sites of class 1 HLA antigens. Immunol. Rev. 85:149–168, 1985.
56. Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C. Structure of the human class I histocompatibility antigen, HLA-A2. Nature 329:506–512, 1987.

APPENDIX A: MOUSE SEQUENCES

The mouse V(lambda) sequences that were included in this survey were from the proteins MOPC104E (sequence No. 1); Y5444, Y5606, S176, H2020, RPC20, and IG 303 LAMBDA'CL (Nos. 14–19); S43'CL (No. 21); RZ5-8'CL (No. 23); MOPC315, TEPC952, MA8–13, 6-2, 5-7, and MOPC315-26'CL' (Nos. 25–30); and MOPC315-37'CL (No. 32) on pp. 129–130 of Kabat et al.[11] In all of these sequences, residues 1–107 (numbering of Kabat et al.[11] have been unequivocably identified.

The mouse V(kappa) sequences that were included in this survey were from the following proteins (page numbers refer to Kabat et al.[11]). On page 78, S107 (sequence No. 1). On pages 85–88, TEPC(CAL20)105, TEPC(CAL20)-119, TEPC817, TEPC821, and H51-5.2 (Nos. 1–5); F17.170.2'CL, PC2205(NZB), 26-10, FLOPC1, PC2567(NZB), G8CA1.7, L XIX27'CL, G5BB2.2, H56.406.48, G7AB2.9, and G6BD2.6 (Nos. 7–17); TEPC-602 (No. 27); G8AD3.8 (No. 30); BALB/C 1210.7 and 36-60 CRI^- (Nos. 52 and 53); 17S29.1 (No. 65); 2S1.3 (No. 67); 7S34.1 (No. 72); MOPC167'CL and C57BL 2857 (Nos. 75 and 76); and MOPC167 and MOPC511 (Nos. 78 and 79). On pages 95–98, ABPC17, PC2413(NZB), MOPC321, PC7043(NZB), and PC7183(NZB) (Nos. 10–14), PC7769(NZB), and PC6684(NZB) (Nos. 16 and 17); PC7175(NZB), PC2485(NZB), and PC4039(NZB) (Nos. 19–21); 50S10.1 (No. 27); ABPC22, PC9245(NZB), and PC4050(NZB) (Nos. 55–57), and PC2154(NZB) (No. 65). On pages 105–109, R16.7 CRI^+ and 93G7 CRI^+ (Nos. 1 and 2); 10K26-12A1 (No. 4); 124E1 CRI^+ and 10K44-7A1 (Nos. 6 and 7); 91A3 CRI^- (No. 23); 123E6 CRI^+ (No. 26); MOPC41 (No. 39); T1'CL (No. 48); UPC61 and J606 (Nos. 51 and 52); W3082 and EPC109 (Nos. 54 and 55); MOPC149 (No. 79); MPC11 and CEA66-E3'CL (Nos. 93 and 94); VTNP'CL (No. 96); C.C58M75'CL (No. 100); and MOPC21 (No. 105). On pages 117–118, TEPC191, TEPC601, HYGAL1, HYGAL2, HYGAL4, HYGAL11, HYGAL12, HYGAL3, HYGAL10, SAPC10, XRPC24, and XRPC44 (Nos. 1–12); HYGAL7, HYGAL9, and J539 (Nos. 15–17); HYGAL6 (No. 20); 45.21.1'CL (No. 24); 14.6b.1'CL and 26.4.1'CL (Nos. 29 and 30); NQ11 7.12'CL (No. 32); and 70Z/3'CL (No. 43). In all of these sequences, residues 1–107 (numbering of Kabat et al.[11]) have been unequivocably identified.

The mouse VH sequences that were included in this survey were from the following proteins (page numbers refer to Kabat et al.[11]). On pages 176–177, 36-60 CRI^- (sequence No. 1); BALB/C 1210.7, M460'CL, and D35'CL (Nos. 3–5); LB8'CL (No. 18); and GAM3-2'CL (No. 28). On page 182, MC101'CL (No. 24); VH101'CL and A8.1'CL (Nos. 26 and 27); and MOPC141'CL (No. 29). On pages 186–188, MOPC104E, J558, HDEX2, HDEX3, HDEX6, HDEX7, HDEX25, HDEX1, HDEX4, HDEX5, HDEX11, HDEX24, HDEX37, HDEX31, HDEX12, HDEX10, HDEX14, and HDEX9 (Nos. 1–18); AC38 205.12 (No. 24); 45.21.1'CL and HDEX8 (Nos. 27 and 28); A6/24'CL (No. 32); MU-3.2'CL (No. 40); 14.6b.1'CL, 26.4.1'CL, and BCL1'CL (Nos. 47–49). On pages 192–193, B1-8'CL and B1.8.DELTA1V3 (Nos. 1 and 2); B1-48'CL (No. 4); S2D8'CL (No. 7); N-HYB'CL (No. 9); B1-8.V1/V2 and B1-8.V1'CL (Nos. 13 and 14); S43'CL (No. 16); F17.170.2 (No. 20); and TEPC1017'CL (No. 23). On page 200, G5BB2.2(AB1); G8CA1.7; GAT50'CL; and G8AD3.8 (Nos. 1–4); L4.13.2'CL; 17.2.25'CL; and L3.10.5'CL (Nos. 12–14); and P10.15.1'CL (No. 19). On pages 204–208, TEPC15 (No. 1); HPCM1, HPCM2, HPCM3, S63, and Y5236 (Nos. 3–7); HPCM6 (No. 10); NQ10.3.8'CL (No. 14); HPCG8 (No. 16); HPCG14 and MCPC603 (Nos. 19 and 20); 36-7'CL, 36-5'CL, and HPCG13 (Nos. 49–51); MOPC167'CL (No. 62); 36-17'CL (No. 65); CBBPC-3 (No. 71); C57BL 2857 (No. 74); CBA/J 7C6 (No. 77); and 6G6'CL (No. 88). On pages 214–215, XRPC44 (No. 1); HYGAL1, HYGAL2, HYGAL3, HYGAL4, HYGAL8, TEPC601, UPC10'CL, HYGAL11, and HYGAL12 (Nos. 3–11); HYGAL10 and HYGAL6 (Nos. 13 and 14); J539 and XRPC24 (Nos. 17 and 18); MOPC173 (No. 22); and HPCA97'CL (No. 25). On page 219, HPC76'CL (No. 6). On pages 226–227, 93G7CRI^+'CL (No. 1); 36-65 CRI^+ (No. 3); R16.7 CRI^+'CL and R16.7 CRI^+ (Nos. 5 and 6); IF6 CRI^+ ARS^- (No. 16); 91A3 CRI^- (No. 18); and MPC11 (No. 29). On pages 235–237, MOPC21, MOPC21'CL, X63.2A RI-16, and NL-1'CL (Nos. 1–4); 36-1'CL (No. 7); CEA66-E3'CL (No. 10); U9A5 CRI^+ and U22B5 CRI^+ (Nos. 44 and 45). In all of these sequences, residues 1–113 (numbering of Kabat et al.[11]) have been unequivocably identified.

The mouse constant domain sequences that were included in this survey were as follows (page numbers refer to Kabat et al.[11]). On page 282, the C(kappa) from protein C.C58 M75'CL (Sequence No. 19). On page 288, the C(lambda) from proteins MOUSE L2'CL, MOUSE L4'CL, MOUSE L1'CL and MOUSE L3'CL (Nos. 26–29). On pages 295–297, the CH1 from proteins IGM'CL (No. 58); MOPC104E (No. 60); MOUSE IGG3'CL (No. 68); IGG1'CL (No. 70); MOPC21 (No. 72); IGG2B(A)'CL, IGG2B(A)'CL', and IGG2B(B)'CL (Nos. 75–77); IGG2A(A)'CL, IGG2A(B)'CL, and IGG2A(B)'CL' (Nos. 81–83); MOPC173 (No. 85); IGA'CL, IGA'CL', and MOPC47A (Nos. 87–89); and IGE'CL (No. 92).

On pages 309–311, the CH2 from proteins IGM′CL (No. 58); MOUSE IGG3′CL (No. 68); IGG1′CL (No. 70); IGG2B(A)′CL (No. 75); IGG2B(B)′CL (No. 77); IGG2A(A)′CL and IGG2A(B)′CL (Nos. 81 and 82); IGA′CL, IGA′CL′, and MOPC47A (Nos. 87–79); and MOPC511, IGE′CL, and IGE′CL′ (Nos. 91–93). On pages 316–318, the CH3 from proteins IGM′CL, IGM′CL′, and MOPC104E (Nos. 58–60); IGD′CL (No. 64); MOUSE IGG3′CL (No. 68); IGG1′CL (No. 68); MOPC21 (No. 72); IGG2B(A)′CL (No. 75); IGG2B(B)′CL (No. 77); IGG2A(A)′CL, IGG2A (B)′CL, and IGG2A(B)′CL′ (Nos. 81–83); CBPC101 and IGA′CL (Nos. 86 and 87); and IGE′CL and IGE′CL (Nos. 92 and 93). On pages 323–324, the CH4 from proteins IGM′CL (No. 58); MOPC104E (No. 60); and IGE′CL and IGE′CL′ (Nos. 92 and 93).