Identification of specificity-determining residues in antibodies

EDUARDO A. PADLAN[1], CHANTAL ABERGEL AND JENNIFER P. TIPPER

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0560, USA

ABSTRACT **The successful identification of the residues that contact ligand has important implications, especially in view of the increasing use of antibodies in various medical and industrial applications. Analysis of the crystallographically derived, three-dimensional structures of five antibody-antigen complexes and of the available amino acid sequence data on antibody variable regions reveals that the residues that contact antigen are in the main also the most variable. It is proposed that a good first guess of the identity of the specificity-determining residues can be made from an examination of the variability values at sequence positions. New boundaries for the complementarity-determining regions are proposed.—Padlan, E. A., Abergel, C., Tipper, J. P. Identification of specificity-determining residues in antibodies. *FASEB J.* 9, 133–139 (1995)**

Key Words: *complementarity-determining regions* · *humanization*

THE INCREASING USE OF ANTIBODIES IN A variety of medical and industrial applications emphasizes the need to know which residues in these molecules are important for their varied reactivities. The identification of the critical residues may be achieved by visualization of the three-dimensional structure of the antibody, or of the relevant antibody-ligand complex, by X-ray crystallography or other technique, but this is not always easy.

It may also be possible to identify the residues that are most likely to be involved in a particular reactivity by correlating the available sequence data and three-dimensional information, or by modeling studies, with subsequent verification by protein engineering experiments. The first procedure was used in the attempt to identify the residues that are important in C1q binding (1) and in the attempt to localize the Fcγ receptor binding site (2). In addition, there have been numerous attempts to model antibody combining sites and antibody-ligand interactions (see, for example, the collection of papers in Ref. 3).

Here, we attempt to identify the residues that are directly involved in the interaction with antigen, i.e., the specificity-determining residues (SDRs)[2], by analyzing the large amount of sequence data that has been accumulated on antibody variable domains (4) and the three-dimensional structures already available for several antibody combining sites (reviewed in Ref. 5). The ability to identify the SDRs has important implications. For example, it can facilitate the design of protein engineering studies that attempt to alter or improve the ligand-binding properties of an antibody. It can also assist in the design of humanization protocols.

In view of their exquisite specificity, antibodies are excellent agents for molecular and cellular targeting and, as such, are being used in diagnostic imaging, treatment of tumors, etc. (see, for example, Refs. 6–8). However, antibodies of pre-specified reactivities are more easily obtained from nonhuman sources and their immunogenicity has to be minimized for protracted use in human therapy, i.e., they first have to be humanized.

Ideally, humanization would preserve antibody reactivity while simultaneously eliminating immunogenicity. In order to preserve antigen-binding specificity, the structure of the combining site must be faithfully reproduced in the humanized molecule. This means that all the residues that determine the three-dimensional structure of the combining site should be retained. X-ray crystallographic studies have shown that antibody combining sites are formed primarily by residues from the hypervariable or complementarity-determining regions (CDRs) (9), with occasional contribution from nonhypervariable, or framework, residues. The procedures that have been proposed to reduce the immunogenicity of xenogeneic antibodies (10–13) advocate the retention of all the CDR residues, since it is not usually known a priori which residues are involved in the interaction with the antigen. Since only about a fifth to a third of the CDR residues participate in the binding to antigen (see below), residual immunogenicity may persist because of the presence of nonessential, nonhuman residues in the humanized molecule. The retention of only the SDRs will reduce immunogenicity to a minimum.

By definition (4, 9), the CDRs of the light chain are bounded by the residues at positions 24 and 34 (CDR1-L), 50 and 56 (CDR2-L), and 89 and 97 (CDR3-L), and those of the heavy chain by residues 31 and 35b (CDR1-H), 50 and 65 (CDR2-H), and 95 and 102 (CDR3-H; numbering convention of Kabat et al. (4)). Some CDRs display wide variations in length. CDR1-L, for example, ranges in length from 10 to 17 amino acids. CDR3-H displays the widest length variation, with some having as few as four residues and others having as many as 26 (14). On the other hand, except in a very few cases, CDR2-L always has seven amino acid residues (4).

Within each CDR, there are residue positions that exhibit greater variability than others. The results of very early surveys (15, 16) have suggested that these more variable posi-

[1]To whom correspondence should be addressed, at Laboratory of Molecular Biology, NIDDK, Building 5, Room 303, National Institutes of Health, Bethesda MD 20892-0560, USA.

[2]Abbreviations: SDR, specificity-determining residue; CDR, complementarity-determining region; HPr, histidine-containing phosphocarrier protein; ASD, average structural dissimilarity.

tions are the ones involved in the interaction with ligand and are therefore specificity-determining, whereas the more conserved residues in the CDRs play an ancillary role and mainly serve to stabilize the combining site structure. Much more three-dimensional and sequence data have since become available and the validity of that suggestion is tested in this study. Indeed, we find that combining sites are in the main constructed from the amino acids in the most variable CDR positions. A preliminary account of our findings has been presented elsewhere (17).

MATERIALS AND METHODS

Analysis of the three-dimensional data

The antibody-antigen complexes for which three-dimensional structures have been elucidated are those involving the murine antibodies D1.3 (18–20), HyHEL-5 (21) and HyHEL-10 (22) and chicken lysozyme, NC41 and influenza virus neuraminidase (23, 24), and Jel42 and the histidine-containing phosphocarrier protein (HPr) from *E. coli* (25). Atomic coordinates for these complexes are available from the Protein Data Bank (26, 27) as entries: 1FDL (D1.3), 2HFL (HyHEL-5), 3HFM (HyHEL-10), 1NCA (NC41), and 1JEL (Jel42), respectively.

The antibody residues that contribute to the interaction with antigen were identified by computation of interatomic contacts. Two atoms are designated as being in contact if the distance between them is at most the sum of their van der Waals' radii (28) plus 0.5 Å. Molecular surfaces were computed with program MS of M. L. Connolly (29) using a probe radius of 1.7 Å.

Analysis of the sequence data

The sequences surveyed were those of the human and murine heavy chain and light chain variable regions listed in the tabulation of Kabat et al. (4). Only those CDRs for which all positions had been unambiguously and completely determined were included in the study. Duplicate sequences were excluded. We limited our survey to those cases where at least four unique sequences are available for comparison.

The sequence variability at a given residue position is customarily computed using the Wu-Kabat formulation (9), which defines variability as the number of different amino acids occurring at that position divided by the frequency of the most commonly occurring residue. This method, however, does not differentiate between conservative substitutions, like leucine for isoleucine, and the more drastic ones, like tryptophan for glycine. We prefer to compute variability by a method that takes into account the physicochemical properties of the different amino acids.

Here, we evaluate variability on the basis of the average structural dissimilarity (ASD) of the various amino acids occurring at each position, using the formula:

$$ASD = 100 \times \frac{1}{N} \sum \frac{D_{ij}}{\langle D \rangle_j}$$

where D_{ij} is the structural dissimilarity between residue i and the most commonly occurring residue j and the summation is over the N amino acids occurring at the given position; $\langle D \rangle_j$ is the mean dissimilarity of the 20 naturally occurring residue types relative to amino acid type j (16). The structural dissimilarities among the various amino acids are taken from the dissimilarity matrix of Grantham (30).

The Wu-Kabat variability and the ASD values have been shown to be correlated (31), although ASD values are not as sensitive to the number of se-

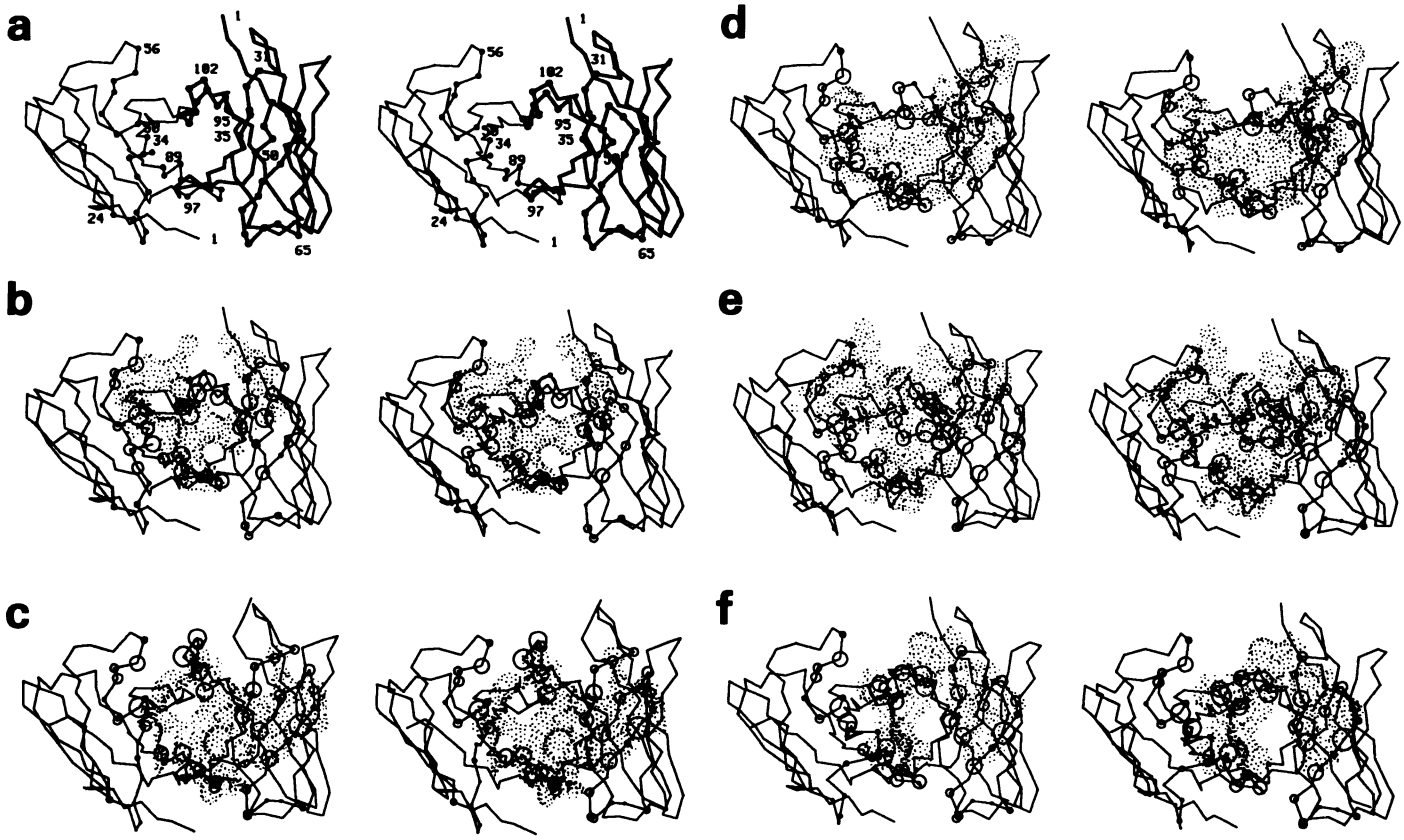

Figure 1. Stereodrawings of the Fv of antibodies D1.3, HyHEL-5, HyHEL-10, NC41, and Jel42 showing the combining site surfaces and the sequence variability in the CDRs. (a) Cα trace of the Fv of D1.3 as seen end on. V_L is at the left (thinner lines) and V_H is at the right (thicker lines). The CDR residues are indicated by filled circles and the CDR boundaries are labeled according to the numbering convention of Kabat et al. (4). The NH_2-termini of the two domains are also labeled. The combining sites of D1.3 (b), HyHEL-5 (c), HyHEL-10 (d), NC41 (e), and Jel42 (e) are represented by molecular surface dots. The structural variability (ASD) values for CDRs of the same length are displayed as circles on the Cα trace of the Fvs, with the radii proportional to the ASD values. The variability values for position 49 of the light chain and position 30 of the heavy chain are also displayed.

TABLE 1. *CDR residues in contact with the ligand in antibody-ligand complexes of known three-dimensional structure*[a]

	CDR1-L	CDR2-L	CDR3-L	CDR1-H	CDR2-H	CDR3-H	PDB Code
	24 27abcdef 30	50 56	90 95a	31 35ab	52abc 60	95 100abc	
D1.3	....------....Y..	Y..T...	..FWS..-..	GY...--	..W---GD.....	..DYR.---..	1FDL
HyHEL-5	....-------..NY.Y	D.....	..WGR.--P.	..W.E--	E.L.--.SGSTN.....	GNY..----..	2HFL
HyHEL-10	....------..GNN..	Y..Q...	..SN.W.-Y.	SDY..--	Y.S---YS.S.Y.....	W..------..	3HFM
NC41	....------.....	W.ST.HI	..HYSP.-W.	NY...--	..N.--N.....	.EDNF.SL...	1NCA
Jel42	.......HG-N...Y..	.....	.....-Y.	TYA..--	..SP--SS.Y.....	.MGE.Y.--..	1JEL
D11.15	....------..S....	.....	...NEY.-W.	S.W..--	..Y.--.D.Y.....	D.NY...--..	1JHL
B13I2	.........-D.D.Y..	.....	..G..V.-P.	R.A..--	.ISS--G.SY.F.....	Y...PF..-..	2IGF
17/9	............Y..	.....	..DYSN.-L.	.....--	T.SN--G.GY.Y.....	RER..E.G...	1IFH
26/9	............F..	.....	..DYSH.-L.	.....--	T.SN--GGGY.Y.....	R.R..E.G...	1FRG
50.1	.......---.....H	.....I.	Q.S..D.-L.	.....	H.F---WD.D.R.....	EGY------I.	1GGI
59.1	......DS--Y...F..	.....	..NNED.-..	..N.C..	R.C---YE.S.....	..HM..T....	1ACY
TE33	.......H. SGN.Y..	K.....	..GS...-F.	TYG..--	W.NT--Y.....	RSW..----..	1TET
Se155-4	.....---....H..	.....	..W..N.-W.	..W.H--	A...--.....F.....	.GHG...--..	1MFA
BV04-01	.......HS-....Y.H	.....	..STHV.-..	TNA..--	R.RS..N.....	DQTGTAW.-..	1CBV
DB3	.......H.-.....	.....	..S.HV.-P.	..G.N--	W...--.....	GDYV.W.F-..	1DBB
26-10	.........-.....	.....	..T....-P.	..Y.N--	Y...--.....	S....WAM-..	1IGJ
4-4-20	.......H.-....Y.R	.....	..S....-W.	..W..--	......Y.....	SYY..----..	4FAB
McPC603	.....	.....	..DH.Y.-L.	..Y..--	..R.....	N.....W....	2MCP
AN02	....-------...Y.Y	D.....	Q.W.....I.	.....-	...---.....	.WP.-----..	1BAF
CHA255	.....---....Y..	.....	..W....-W.	..T..--	T.L.--...F.F.....	HR.------..	1IND
1F7	....------.....	.....	.....Y.-..	.HN.N--	N...--.....	R.D..Y.F-..	1FIG
NC6.8	.......H.-N...Y.H	.....	..G....-Y.	..W.E--	E...--...R.N.....	.YSSM----..	1CGR

[a]Only the residues which contact ligand are shown; the others are indicated simply by dots. Hyphens signify gaps. PDB = Protein Data Bank. The results for antibodies 50.1 (PDB Entry 1GGI) and 26-10 (PDB Entry 1IGJ) are for the first complex in the entry. For antibodies D1.3, HyHEL-5, HyHEL-10, NC41, Jel42, and D11.15, the ligand is an intact (protein) antigen; for the rest, the ligand is a peptide (B1312, 17/9, 26/9, 50.1, 59.1, and TE33), a trisaccharide (Se155-4), a trinucleotide (BV04-01), or a small molecule (DB3, 26-10, 4-4-20, McPC603, AN02, CHA255, 1F7, and NC6.8). Antibodies Se155-4 and CHA255 have lambda light chains; all the others have kappa chains.

TABLE 2. *Sequence variability in murine light-chain CDRs*[a]

Kappa chains

CDR1-L:

Position:	24	25	26	27	a	b	c	d	e	f	28	29	30	31	32	33	34	
(10 residues)	28.6	11.7	6.9	0.0	-	-	-	-	-	-	-	8.8	6.0	19.0	7.3	4.8	48.9	(29 sequences)
(11 residues)	23.9	5.5	2.2	13.6	-	-	-	-	-	-	24.3	13.2	53.1	55.4	74.7	10.2	60.1	(84 sequences)
(12 residues)	31.0	9.0	16.7	10.0	4.0	-	-	-	-	-	12.6	8.0	6.6	41.7	38.2	4.0	34.6	(22 sequences)
(15 residues)	2.2	1.7	4.9	23.6	5.3	7.0	37.6	75.2	-	-	80.3	23.0	78.3	18.1	20.1	3.2	23.9	(41 sequences)
(16 residues)	5.3	0.0	7.0	19.3	13.2	3.3	21.7	31.7	18.4	-	19.3	0.9	42.1	1.8	0.0	0.8	86.7	(59 sequences)
(17 residues)	10.9	3.2	4.8	2.4	7.5	2.3	8.7	42.6	5.5	53.7	26.6	17.5	14.6	16.4	20.6	0.0	20.1	(31 sequences)

CDR2-L (including residue 49):

Position:	49	50	51	52	53	54	55	56	
(7 residues)	23.8	91.3	38.8	14.3	48.3	37.8	72.0	27.2	(161 sequences)

CDR3-L:

Position:	89	90	91	92	93	94	95	a	96	97	
(8 residues)	66.9	9.5	77.1	77.0	56.3	49.4	-	-	57.0	0.0	(19 sequences)
(9 residues)	51.4	6.1	81.6	70.0	58.7	64.8	5.2	-	45.0	0.3	(228 sequences)
(10 residues)	29.7	9.4	65.9	60.4	45.8	78.0	3.4	12.6	34.1	0.0	(13 sequences)

Lambda chains

CDR1-L:

Position:	24	25	26	27	a	b	c	28	29	30	31	32	33	34	
(14 residues)	0.0	0.0	24.4	17.5	0.0	0.0	8.5	0.0	13.5	17.6	0.0	0.0	0.0	17.0	(9 sequences)

CDR2-L (including residue 49):

Position:	49	50	51	52	53	54	55	56	
(7 residues)	0.0	9.0	7.9	23.9	49.1	0.0	25.8	0.0	(10 sequences)

CDR3-L:

Position	89	90	91	92	93	94	95	96	97	
(9 residues)	0.0	0.0	0.0	30.2	13.8	16.7	4.0	13.3	12.9	(8 sequences)

[a]ASD values that exceed the structural variability value for nonconservative substitutions (ASD = 47.3) are underlined.

quences compared as are the Wu-Kabat variability values. Indeed, when comparing N sequences, the maximum Wu-Kabat value would be N × N, for N up to 20, and 400, for N greater than or equal to 20. In contrast, the ASD values, in theory, are independent of the number of sequences compared. Furthermore, a measure of conservative and nonconservative substitutions can be estimated when using structural dissimilarities. Using the Grantham dissimilarity matrix, an ASD value of 18.5 or lower would signify a conservative replacement, while a value of 47.3 or higher would be nonconservative (16).

To eliminate the uncertainties that can result from the presence of gaps, only sequences with the same number of residues were compared with each other. The variability at two framework positions, 49 in the light chain and 30 in the heavy chain, were also computed, in view of the involvement of these positions in some of the antibody-antigen complexes (e.g., in the complexes of D1.3 and HyHEL-10 with lysozyme, of NC41 with neuraminidase, and of Jel42 with *E. coli* HPr). Although the residue at the heavy chain position 47 was found to contribute to the interaction with lysozyme in HyHEL-5, this position in murine and human heavy chains is almost always occupied by tryptophan so that the variability at this position is low.

RESULTS AND DISCUSSION

The CDRs form a continuous surface approximately 2800 Å^2 in area at the NH_2-terminal portion of the Fab, and X-ray crystallographic analysis has shown that antigen binding primarily involves this surface. The six CDRs are disposed (**Fig. 1**) such that the NH_2-terminal half of CDR1-L and the COOH-terminal parts of CDR2-L and CDR2-H are farther from the center of the CDR surface, while CDR1-H, CDR3-H, CDR3-L, the COOH-terminal half of CDR1-L, and the NH_2-terminal parts of CDR2-L and CDR2-H are closer to the center.

Only a fraction of the CDR surface is found to constitute the combining site. In the antibodies D1.3, HyHEL-5, HyHEL-10, NC41, and Jel42, the surface that is used in the

TABLE 3. *Sequence variability in murine heavy-chain CDRs*[a]

CDR1-H (including residue 30):

Position:	30	31	32	33	34	35	a	
(5 residues)	38.9	46.2	34.0	**76.0**	9.2	**66.9**	-	(223 sequences)
(6 residues)	13.3	20.7	44.9	0.0	**64.8**	3.9	27.5	(15 sequences)

CDR2-H:

Position:	50	51	52	a	b	c	53	54	55	56	57	58	59	60	61	62	63	64	65	
(16 residues)	38.1	1.0	**84.7**	-	-	-	**63.6**	37.9	6.7	40.6	17.7	**64.3**	8.9	9.7	35.5	36.3	6.0	28.5	11.0	(79 sequences)
(17 residues)	**81.6**	4.1	**52.1**	22.5	-	-	**51.4**	43.0	25.4	**106.2**	26.1	**75.1**	4.2	37.9	34.4	37.8	11.7	10.5	26.5	(309 sequences)
(19 residues)	**74.2**	**58.3**	1.8	41.1	5.6	37.9	40.8	**56.8**	12.6	30.4	9.4	28.4	0.0	31.5	39.7	0.0	1.0	1.1	3.5	(51 sequences)

CDR3-H:

Position:	95	96	97	98	99	100	a	b	c	d	e	f	g	h	i	101	102	
(3 residues)	**75.5**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**47.5**	0.0	(5 sequences)
(4 residues)	**76.8**	**72.0**	-	-	-	-	-	-	-	-	-	-	-	-	-	**71.2**	18.1	(23 sequences)
(5 residues)	**61.1**	**93.9**	**61.5**	-	-	-	-	-	-	-	-	-	-	-	-	**54.6**	14.7	(44 sequences)
(6 residues)	**61.9**	**96.2**	**86.8**	33.9	-	-	-	-	-	-	-	-	-	-	-	46.3	0.0	(23 sequences)
(7 residues)	**83.3**	**76.6**	**95.2**	**84.9**	**47.4**	-	-	-	-	-	-	-	-	-	-	**49.5**	15.1	(70 sequences)
(8 residues)	**71.7**	**82.9**	**83.2**	**85.5**	**76.2**	32.7	-	-	-	-	-	-	-	-	-	24.6	12.7	(91 sequences)
(9 residues)	**89.8**	**91.8**	**74.3**	**77.1**	**95.1**	**74.4**	28.3	-	-	-	-	-	-	-	-	29.8	11.6	(116 sequences)
(10 residues)	**104.9**	**91.3**	**74.9**	**62.3**	**74.9**	**91.8**	**72.1**	14.1	-	-	-	-	-	-	-	32.0	16.8	(97 sequences)
(11 residues)	**85.9**	**76.5**	**65.7**	**76.9**	**96.5**	**103.5**	**69.2**	**72.8**	19.0	-	-	-	-	-	-	19.7	14.6	(84 sequences)
(12 residues)	**53.5**	**73.2**	34.9	**64.7**	**53.6**	**56.0**	**79.9**	44.0	**60.6**	9.5	-	-	-	-	-	6.9	24.0	(86 sequences)
(13 residues)	**56.6**	**79.4**	**82.6**	**53.2**	**68.0**	**95.9**	**84.4**	**73.9**	**81.4**	**80.4**	14.2	-	-	-	-	20.9	22.9	(31 sequences)
(14 residues)	**64.7**	**47.9**	**87.5**	36.8	**103.4**	**58.6**	**59.3**	**94.7**	**71.6**	**78.9**	**74.5**	13.5	-	-	-	0.0	14.6	(19 sequences)
(15 residues)	**61.5**	**68.2**	**71.9**	**62.3**	**63.9**	**59.3**	**71.3**	**67.2**	**93.3**	**89.6**	**61.9**	**57.0**	10.6	-	-	0.0	5.6	(11 sequences)
(16 residues)	**58.2**	**67.1**	**87.6**	**49.0**	38.2	**85.4**	**62.3**	**55.3**	**81.7**	**72.2**	**67.8**	**65.0**	**56.3**	22.4	-	0.0	28.4	(10 sequences)
(17 residues)	**70.5**	**50.0**	**56.1**	15.4	**85.1**	**70.5**	**68.6**	**62.7**	**50.0**	28.6	**68.6**	**61.6**	44.8	**67.4**	13.0	28.5	44.8	(4 sequences)

[a]ASD values that exceed the structural variability value for nonconservative substitutions (ASD = 47.3) are underlined.

TABLE 4. *Sequence variability in human light-chain CDRs*[a]

Kappa chains																		
CDR1-L:																		
Position:	24	25	26	27	a	b	c	d	e	f	28	29	30	31	32	33	34	
(11 residues)	13.5	4.7	7.3	4.3	-	-	-	-	-	-	35.4	10.6	38.6	54.5	65.1	4.5	51.0	(52 sequences)
(12 residues)	8.2	0.0	8.6	8.2	14.9	-	-	-	-	-	30.8	19.1	24.6	25.8	11.2	2.0	6.9	(18 sequences)
(16 residues)	7.7	0.0	0.0	0.0	0.0	0.0	15.2	51.9	0.0	-	10.0	0.0	83.2	59.7	0.0	0.0	35.6	(8 sequences)
(17 residues)	7.2	0.0	0.0	0.0	0.0	18.2	0.0	44.8	0.0	14.0	10.4	32.4	26.2	16.7	0.0	0.0	29.4	(4 sequences)
CDR2-L (including residue 49):																		
Position:	49	50	51	52	53	54	55	56										
(7 residues)	10.8	70.8	21.6	6.4	40.2	44.8	56.4	42.7	(55 sequences)									
CDR3-L:																		
Position:	89	90	91	92	93	94	95	a	96	97								
(8 residues)	0.0	0.0	0.0	37.7	32.4	88.4	-	-	83.5	0.0	(5 sequences)							
(9 residues)	23.2	2.2	42.2	68.5	36.3	79.8	15.6	-	59.3	9.8	(57 sequences)							
(10 residues)	3.9	8.6	20.1	48.2	50.3	72.1	0.0	39.2	27.3	0.0	(8 sequences)							
Lambda chains																		
CDR1-L:																		
Position:	24	25	26	27	a	b	c	28	29	30	31	32	33	34				
(11 residues)	21.0	5.8	10.4	67.1	-	-	-	1.5	26.9	45.5	44.4	62.0	36.8	94.9	(39 sequences)			
(13 residues)	26.7	17.2	15.8	18.1	29.3	21.3	-	0.3	16.5	39.1	30.0	42.6	0.0	60.7	(21 sequences)			
(14 residues)	35.0	11.0	37.4	36.4	15.9	22.3	17.6	24.9	49.3	86.4	39.5	59.7	8.8	26.8	(17 sequences)			
CDR2-L (including residue 49):																		
Position:	49	50	51	52	53	54	55	56										
(7 residues)	5.9	44.6	49.7	35.1	62.0	4.0	4.3	7.6	(68 sequences)									
CDR3-L:																		
Position:	89	90	91	92	93	94	95	a	b	96	97							
(9 residues)	31.7	43.1	18.6	32.9	39.2	50.4	65.2	-	-	71.8	14.6				(17 sequences)			
(10 residues)	46.6	8.3	12.8	63.1	33.0	35.9	56.1	52.1	-	49.7	9.2				(15 sequences)			
(11 residues)	58.6	57.8	18.2	11.6	44.5	21.5	71.0	46.7	68.1	52.6	7.6				(33 sequences)			

[a]ASD values that exceed the structural variability value for nonconservative substitutions (ASD = 47.3) are underlined.

binding to antigen represents only 21%, 27%, 28%, 32%, and 22%, respectively, of the total surface formed by the CDRs; the CDR residues that contact the antigen represent only 25%, 37%, 36%, 33%, and 27%, respectively, of the total number of CDR residues.

The CDR residues found to be involved in the contact with antigen in the D1.3, HyHEL-5, HyHEL-10, NC41, and Jel42 complexes are presented in **Table 1.** It is seen that there is wide variation in the utilization of the various CDR segments in antigen binding. In D1.3, HyHEL-5 and HyHEL-10, all six CDRs participate in the binding; in NC41 and Jel42, only five of the six CDRs contribute. In NC41, there is no participation of CDR1-L, while five of the seven residues in CDR2-L are involved; in Jel42, the opposite is found and CDR2-L is seen not to participate at all, whereas CDR1-L contributes four residues to the interaction.

The specificity-determining residues or SDRs are seen to originate mostly from the COOH-terminal part of CDR1-L, from the first and sometimes also the middle positions in CDR2-L, from the middle portion of CDR3-L, from most of CDR1-H, from the NH_2-terminal part and the middle portion of CDR2-H, and from most of CDR3-H except for the terminal residues. These regions constitute the central region of the CDR surface. For completeness, we have included in Table 1 the CDR residues from complexes involving smaller ligands. The residues that have been observed to interact with ligand in these latter complexes are also seen to come mainly from the above regions.

The results of analysis of the sequences are presented in **Table 2** for the CDRs of the murine light chains and in **Table 3** for the murine heavy chains. Those for the human sequences are presented in **Table 4** for the light chains and in **Table 5** for the heavy chains. The results are different for the different analyses, due in part from sampling differences. For example, there is little variation seen among the murine lambda chains, but these chains are known to be very similar to each other (4). Some trends are apparent, however. High variability values are observed for the COOH-terminal positions in CDR1-L, for the first and some of the middle positions in CDR2-L, for most of CDR3-L and CDR1-H, for the NH_2-terminal portion of CDR2-H, and for most of CDR3-H. In other words, the hypervariable positions are lo-

TABLE 5. *Sequence variability in human heavy-chain CDRs*[a]

CDR1-H (including residue 30):

Position:	30	31	32	33	34	35	a	b	
(5 residues)	22.3	44.5	54.1	77.0	17.9	52.9	-	-	(75 sequences)
(6 residues)	0.0	0.0	26.9	42.7	16.0	0.0	26.9	-	(4 sequences)
(7 residues)	18.0	35.6	57.8	35.1	34.9	99.4	33.8	38.3	(18 sequences)

CDR2-H:

Position:	50	51	52	a	b	c	53	54	55	56	57	58	59	60	61	62	63	64	65	
(16 residues)	87.2	12.7	68.7	-	-	-	61.8	32.9	12.9	31.6	28.9	71.3	13.0	36.5	27.1	2.6	6.3	14.9	18.2	(28 sequences)
(17 residues)	75.6	16.6	70.9	62.0	-	-	65.3	54.6	35.8	53.9	43.3	76.9	9.7	22.3	42.4	47.7	33.1	39.5	11.5	(56 sequences)
(18 residues)	0.0	0.0	0.0	0.0	7.7	-	0.0	0.0	0.0	40.0	16.7	5.2	40.3	0.0	36.0	0.0	0.0	0.0	0.0	(4 sequences)
(19 residues)	10.9	10.1	18.3	30.8	0.0	41.0	43.4	40.6	32.9	18.4	12.1	57.7	2.5	39.6	0.0	14.8	0.0	0.0	0.0	(10 sequences)

CDR3-H:

Position:	95	96	97	98	99	100	a	b	c	d	e	f	g	h	i	j	k	101	102	
(8 residues)	51.1	82.3	62.2	73.7	52.6	19.8	-	-	-	-	-	-	-	-	-	-	-	9.7	32.8	(9 sequences)
(9 residues)	68.0	95.3	67.0	67.5	64.9	100.7	9.0	-	-	-	-	-	-	-	-	-	-	17.4	51.9	(5 sequences)
(10 residues)	73.2	73.7	77.5	79.7	35.1	84.1	68.7	47.2	-	-	-	-	-	-	-	-	-	2.3	38.3	(9 sequences)
(11 residues)	77.6	85.1	55.0	89.2	62.6	80.6	48.9	61.1	3.8	-	-	-	-	-	-	-	-	14.5	92.5	(6 sequences)
(12 residues)	44.0	79.7	81.3	61.0	66.3	69.6	64.5	64.6	58.6	20.0	-	-	-	-	-	-	-	14.3	10.8	(8 sequences)
(13 residues)	50.6	83.8	84.0	87.6	72.7	69.5	51.9	76.0	82.5	65.1	37.6	-	-	-	-	-	-	49.1	43.6	(15 sequences)
(14 residues)	50.7	74.3	83.9	88.4	59.0	77.8	88.4	86.6	57.9	66.4	70.1	20.2	-	-	-	-	-	14.2	39.7	(6 sequences)
(15 residues)	63.0	55.8	41.7	78.9	48.8	67.0	44.4	39.1	88.6	93.5	77.2	79.0	5.8	-	-	-	-	11.8	63.5	(5 sequences)
(16 residues)	46.5	52.7	59.0	72.9	99.1	81.4	73.3	44.6	90.6	99.9	46.5	40.3	75.9	34.3	-	-	-	12.2	32.5	(7 sequences)
(17 residues)	79.6	70.3	64.1	83.8	80.7	50.2	82.3	58.1	90.8	62.6	59.2	88.5	61.2	66.2	22.2	-	-	0.0	14.9	(8 sequences)
(18 residues)	67.3	72.8	47.9	79.3	65.9	48.8	47.3	43.2	67.5	54.5	87.8	25.7	46.4	81.1	62.5	12.7	-	0.0	25.0	(4 sequences)
(19 residues)	50.8	44.4	31.4	92.7	42.8	49.9	71.2	104.3	47.7	57.2	58.5	67.7	54.4	37.6	93.3	51.0	9.2	0.0	13.1	(5 sequences)

[a]ASD values that exceed the structural variability value for nonconservative substitutions (ASD = 47.3) are underlined.

cated mostly at the center of the CDR surface, just like the SDRs. This coincidence of high structural variability and involvement in antigen binding is demonstrated in Fig. 1 where ASD values are displayed together with the combining-site surfaces of the D1.3, HyHEL-5, HyHEL-10, NC41, and Jel42 antibodies.

The SDRs are probably unique to each antibody and their identification may only be possible with the determination of the three-dimensional structure of the complex with antigen. Nevertheless, our findings suggest that some generalizations can be made that may allow the identification of those positions that are likely to be involved in the interaction with specific ligand.

We propose that the SDRs can be found in the set of the highly variable residues listed in Tables 2 through 5 (or similar listings for CDR lengths not included in this survey). As a first guess, antigen binding probably will not involve the NH_2-terminal residues in CDR1-L or the COOH-terminal residues in CDR2-H. Residues from several positions would also seem unlikely to be specificity-determining, including positions 56 and 97 in the light chain. Identification of the SDRs will become more precise as more sequence and three-dimensional information become available.

In order to reproduce the antigen-binding properties of an antibody during humanization, we suggest that only the potential SDRs, i.e., the residues at the positions that display high variability, be retained. If whole segments are transplanted, abbreviated versions of the CDRs should suffice; these, we propose, should include only the residues bounded by positions 27d and 34, 50 and 55, and 89 and 96 in the light chain; and 31 and 35b, 50 and 58, and 95 and 101 in the heavy chain (numbering convention of Kabat et al. (4)). Of course, as before, it may be necessary to retain a few non-hypervariable residues also, to insure the preservation of the combining-site structure. FJ

REFERENCES

1. Duncan, A. R., and Winter, G. (1988) The binding site for C1q on IgG. *Nature (London)* **332,** 738–740
2. Duncan, A. R., Woof, J. M., Partridge, L. J., Burton, D. R., and Winter, G. (1988) Localization of the binding site for the human high-affinity Fc receptor for IgG. *Nature (London)* **332,** 563–564
3. Padlan, E. A., ed. (1992) Strategies for Modeling Antibody Combining Site Structures, ImmunoMethods Vol. 1, No. 2, Academic, San Diego
4. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S., and Foeller,

C. (1991) Sequences of Proteins of Immunological Interest., 5th Edition, US Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda (NIH Publication No 91-3242)
5. Padlan, E. A. (1994) Antibody-Antigen Complexes, R. G. Landes, Austin. (in press)
6. Pastan, I., and FitzGerald, D. (1991) Recombinant toxins for cancer treatment. *Science* **254,** 1173-1177
7. Schlom, J. (1991) Antibodies in cancer therapy: Basic principles and applications. In Biologic Therapy of Cancer: Principles and Practice (DeVita, V. T., Jr., Hellman, S., and Rosenberg, S. A., eds) pp. 464-481, J. B. Lippincott, Philadelphia
8. Waldmann, T. A. (1991) Monoclonal antibodies in diagnosis and therapy. *Science* **252,** 1657-1662
9. Wu, T. T., and Kabat, E. A. (1970) An analysis of the sequences of the variable regions of Bence-Jones proteins and myeloma light chains and their implication for antibody complementarity. *J. Exp. Med.* **132,** 211-249
10. Boulianne, G. L., Hozumi, N., and Shulman, M. J. (1984) Production of functional chimaeric mouse/human antibody. *Nature (London)* **312,** 643-646
11. Morrison, S. L., Johnson, M. J., Herzenberg, L. A., and Oi, V. T. (1984) Chimeric human antibody molecules: Mouse antigen-binding domains with human constant region domains. *Proc. Natl. Acad. Sci. USA* **81,** 6851-6855
12. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S., and Winter, G. (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature (London)* **321,** 522-525
13. Padlan, E. A. (1991) A possible procedure for reducing the immunogenicity of antibody variable domains while preserving their ligand-binding properties. *Mol. Immunol.* **28,** 489-498
14. Wu, T. T., Johnson, G., and Kabat, E. A. (1993) Length distribution of CDRH3 in antibodies. *Proteins: Struc. Funct. Genet.* **16,** 1-7
15. Kabat, E. A., Wu, T. T., and Bilofsky, H. (1977) 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
16. Padlan, E. A. (1977) Structural implications of sequence variability in immunoglobulins. *Proc. Natl. Acad. Sci. USA* **74,** 2551-2555
17. Abergel, C., Tipper, J. P., and Padlan, E. A. (1994) Structural significance of sequence variability in antibody complementarity-determining regions. *Res. Immunol.* **145,** 49-53
18. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V., and Poljak, R. J. (1986) Three-dimensional structure of an antigen-antibody complex at 2.8 Å resolution. *Science* **233,** 747-753
19. Fischmann, T. O., Bentley, G. A., Bhat, T. N., Boulot, G., Mariuzza, R. A., Phillips, S. E. V., Tello, D., and Poljak, R. J. (1991) Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-Å resolution. *J. Biol. Chem.* **266,** 12915-12920
20. Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F. P., Mariuzza, R. A., and Poljak, R. J. (1994) Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* **91,** 1089-1093
21. Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C., and Davies, D. R. (1987) Three-dimensional structure of an antibody-antigen complex. *Proc. Natl. Acad. Sci. USA* **84,** 8075-8079
22. Padlan, E. A., Silverton, E. W., Sheriff, S., Cohen, G. H., Smith-Gill, S. J., and Davies, D. R. (1989) Structure of an antibody-antigen complex: Crystal structure of the HyHEL-10 Fab-lysozyme complex. *Proc. Natl. Acad. Sci. USA* **86,** 5938-5942
23. Colman, P. M., Laver, W. G., Varghese, J. N., Baker, A. T., Tulloch, P. A., Air, G. M., and Webster, R. G. (1987) Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. *Nature (London)* **326,** 358-363
24. Tulip, W. R., Varghese, J. N., Laver, W. G., Webster, R. G., and Colman, P. M. (1992) Refined crystal structure of the influenza virus N9 neuraminidase-NC41 Fab complex. *J. Mol. Biol.* **227,** 122-148
25. Prasad, L., Vandonselaar, M., Lee, J. S., and Delbaere, L. T. J. (1988) Structure determination of a monoclonal Fab fragment specific for histidine-containing protein of the phosphoenolpyruvate:sugar phosphotransferase system of *Escherichia coli*. *J. Biol. Chem.* **263,** 2571-2574
26. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112,** 535-542
27. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., and Weng, J. (1987) Protein Data Bank. In Crystallographic Databases—Information Content, Software Systems, Scientific Applications (Allen, F. C., Bergerhoff, G., and Sievers, R., eds) pp. 107-132, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester
28. Gelin, B. R., and Karplus, M. (1979) Side-chain torsional potentials: effect of dipeptide, protein, and solvent environment. *Biochemistry* **18,** 1256-1268
29. Connolly, M. L. (1983) Analytical molecular surface calculation. *J. Appl. Crystallogr.* **16,** 548-558
30. Grantham, R. (1974) Amino acid difference formula to help explain protein evolution. *Science* **185,** 862-864
31. Kabat, E. A. (1978) The structural basis of antibody complementarity. *Adv. Prot. Chem.* **32,** 1-75

Received for publication September 1, 1994.
Accepted for publication October 10, 1994.