J. Mol. Biol. (1997) **273**, 927–948

JMB

Standard Conformations for the Canonical Structures of Immunoglobulins

Bissan Al-Lazikani, Arthur M. Lesk and Cyrus Chothia

MRC Laboratory of Molecular Biology and the Department of Haematology, University of Cambridge Clinical School Hills Road, Cambridge CB2 2QH, England

A comparative analysis of the main-chain conformation of the L1, L2, L3, H1 and H2 hypervariable regions in 17 immunoglobulin structures that have been accurately determined at high resolution is described. This involves 79 hypervariable regions in all. We also analysed a part of the H3 region in 12 of the 15 V_H domains considered here.

On the basis of the residues at key sites the 79 hypervariable regions can be assigned to one of 18 different canonical structures. We show that 71 of these hypervariable regions have a conformation that is very close to what can be defined as a ''standard'' conformation of each canonical structure. These standard conformations are described in detail. The other eight hypervariable regions have small deviations from the standard conformations that, in six cases, involve only the rotation of a single peptide group. Most H3 hypervariable regions have the same conformation in the part that is close to the framework and the details of this conformation are also described here.

Keywords: hypervariable regions; high resolution structures; V_κ; V_λ; V_H

Introduction

The antigen binding site of immunoglobulins is formed by six hypervariable regions: three from the V_L domain and three from the V_H domain (Figure 1; Wu & Kabat, 1970). Although there is great variation in sequence and size of these regions, various studies (Chothia *et al.*, 1986, 1989, 1992; Chothia & Lesk, 1987; Tramontano *et al.*, 1990; Brünger *et al.*, 1991; He *et al.*, 1992; Wu & Cygler, 1993; Tomlinson *et al.*, 1995; Guarné *et al.*, 1996) have shown that five of the six hypervariable regions usually have one of a small number of main chain conformations; these have been called canonical structures. In an antibody the major determinants of the specificity and affinity of these five regions for an antigen are: (i) the canonical structure present in the hypervariable region; (ii) the size, shape and chemical character of their surface residues; and (iii) their positions relative to each other.

The conformation of a particular canonical structure is determined by the length of the loop and the residues present at key sites. These key residues have been discussed in a number of papers: for the most recent work see Tomlinson *et al.* (1995) on the V_κ canonical structures; Chothia & Lesk (1987) and Wu & Cygler (1993) on those in V_λ and Chothia *et al.* (1992) on those in V_H.

To what extent is the local conformation of the individual canonical structures conserved in different immunoglobulins? At the time of the initial work, two high resolution immunoglobulin structures were available and in their L1, L2 and L3 hypervariable regions they had the same canonical structures. Optimal superposition of the three pairs of canonical structures showed that, though they differed somewhat in their positions relative to the framework, they had very similar local conformations: the r.m.s. differences in the position of their main-chain atoms were in the range 0.2 to 0.8 Å (Chothia & Lesk, 1987). Subsequently similar results were obtained for four different canonical structures that were found two or more times in seven high resolution immunoglobulin structures (Bajorath *et al.*, 1995). Recently, 244 hypervariable regions in 49 immunoglobulin Fab or Fv structures determined at resolutions between 1.7 and 3.1 Å were placed in clusters on the basis of their structural features (Martin & Thornton, 1996). Some 85% of these were clustered into groups that were identified as known canonical structures with very similar conformations (the remaining 15% are discussed below).

These studies indicated that the local conformation of canonical structures is well conserved. However, they neither give a complete description

Abbreviations used: CDR, complementarity determining region.

0022–2836/97/440927–22 $25.00/0/mb971354

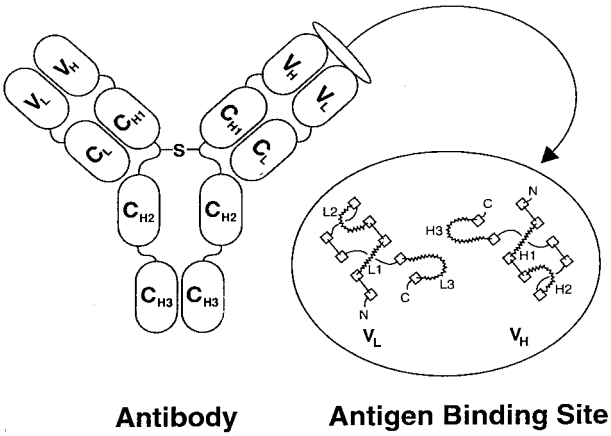


Figure 1. Outline of the structure of an immunoglobulin and its binding site.

of the conformations that are actually present in these structures nor treat any variations that might occur. Here we provide this information by an analysis and comparison of the conformations of the canonical structures in 17 immunoglobulins whose structures have been determined accurately at high resolution. Though the hypervariable regions in these particular immunoglobulins do not cover all the known canonical structures, they do include all those that occur commonly.

The descriptions we give are crucial to a proper understanding of the molecular mechanism of immune recognition. They are also important for those concerned with the design or modelling of antibody structures. Most of the results are summarised in Figures 2 to 20 and these Figures allow those concerned with a new immunoglobulin structure to determine, for most hypervariable regions, whether they contain novel conformations.

Accurate High Resolution Immunoglobulin Structures

As mentioned above, 17 structures determined at high resolution were used in this study. The structures are refined to a resolution of 1.6 to 2.3 Å, with an *R*-factor not greater than 21%. These contained seven V_λ, ten V_κ, and 15 V_H chains. Table 1 gives a list of the structures and summarises some essential information about them. In HIL, 4D5 and 17/9, two molecules are present in one unit cell; in these cases, both molecules were included in the examinations. Ten of the structures comprise the Fab fragment of the antibody, five (SE155-4, POT, H52, 4D5 and D1.3) are Fv fragments and RHE and McPC603 are V_L dimers. SE155-4, CHA255, TE33 and NC6.8 are complexed with antigens. The different structures had been refined using a variety of programs. Most have been refined by one program and a few by several programs. This use of structures refined by different methods allows us to see whether or not the method of refinement has an influence on the fine details of the loop structure. The conformational deviations that are sometimes observed for a few particular canonical structures are found to be independent of the methods used in the refinement of the structure.

Table 1. Accurate high resolution immunoglobulin structures and the canonical structures in their hypervariable regions

V_L	Antibody	PDB File	Resolution (Å)	*R* factor (%)	Canonical structure L1	L2	L3	H1	H2	Ref/Prog	
λ	KOL	2fb4	1.9	19	1	1	2	1	3	a	d
λ	RHE	2rhe	1.6	15	1	1	2	–	–	b	p
λ	NEWM	7fab	2.0	17	2	[a]	1	1	1	c	x
λ	SE155	1mfa	1.7	17	3	1	1	1	2	d	x
λ	CHA255	1ind	2.2	19	3	1	1	1	3	e	xtp
λ	HC19	1gig	2.3	20	3	1	1	1	1	f	px
λ	HIL	8fab	1.8	17	4	1	[b]	1	3	g	x
κ	J539	2fbj	1.95	19	1	1	2	1	3	h	p
κ	POT	1igm	2.3	20	2	1	1	1	3	i	p
κ	H52	1fgv	1.9	18	2	1	1	1	2	j	x
κ	4D5	1fvc	2.2	18	2	1	1	1	2	k	x
κ	D1.3	1vfa	1.8	16	2	1	1	1	1	1	x
κ	17/9	1hil	2.0	20	3	1	1	1	3	m	x
κ	McPC603	2imm	2.0	15	3	1	1	–	–	n	e
κ	4–4–20	1flr	1.85	19	4	1	1	1	4	o	f
κ	TE33	1tet	2.3	15	4	1	1	1	2	p	x
κ	NC6.8	2cgr	2.2	21	4	1	1	1	2	q	x

Structure references: a, Marquart & Huber (1989); Marquart *et al.* (1980); b, Furey *et al.* (1983); c, Saul & Poljak (1992); d, Zdanov *et al.* (1994); e, Love *et al.* (1993); f, Bizebard *et al.* (1994); g, Saul & Poljak (1993); h, Bhat *et al.* (1990); Suh *et al.* (1986); i, Fan *et al.* (1992); j, Eigenbrot *et al.* (1994); k, Eigenbrot *et al.* (1993); l, Bhat *et al.* (1994); m, Rini *et al.* (1992); n, Steipe *et al.* (1992); o, Whitlow *et al.* (1995); p, Shoham (1993); q, Guddat *et al.* (1994).

Prog, refinement programs: x, X-Plor; p, PROLSQ; t, TNT; d, DIAMOND; e, EREF; f, ProFit.

[a] The L2 region of NEWM is deleted.

[b] The L3 of HIL has very high temperature and is not considered here.

Table 2. Kabat and structural numbering of residues in the L1 region of V_κ domains

Canonical structure	Numbering scheme	Residue numbers												
1	Kabat	26	27	29	30	31								32
1	Structure	26	27	28	29	30								32
2	Kabat	26	27	28	29	30							31	32
2	Structure	26	27	28	29	30							31	32
6	Kabat	26	27	27a	28	29	30						31	32
6	Structure	26	27	28	29	30	30a						31	32
5	Kabat	26	27	27a	27b	27c	27d	28			29	30	31	32
5	Structure	26	27	28	29	30	30a	30b			30c	30d	31	32
4	Kabat	26	27	27a	27b	27c	27d	28e	28		29	30	31	32
4	Structure	26	27	28	29	30	30a	30b	30c		30e	30f	31	32
3	Kabat	26	27	27a	27b	27c	27d	27e	27f	28	29	30	31	32
3	Structure	26	27	28	29	30	30a	30b	30c	30d	30e	30f	31	32

Note that residues with the same residue number in different canonical structures do not necessarily have the same conformation: see Table 3.

In some parts of the paper, we use medium resolution structures not included in Table 1. These are described in the appropriate sections.

L1 Hypervariable Regions

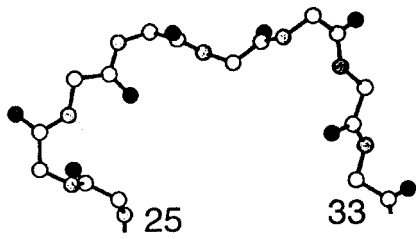
V_κ L1 canonical structures

The L1 region packs across the top of the V_κ domain, bridging the two β-sheets. On the basis of sequence variations, residues 24 to 34 were defined by Kabat *et al*. (1979) as the first CDR region. The number of residues in this region varies and, on the basis of sequence, it has been believed that this variation involves the addition or deletion of residues at the site of residue 27. Later analysis of immunoglobulin structures showed that the region outside the β-sheet framework and therefore potentially able to have different conformations comprises residues 26 to 32. It also showed that the variations in the size of this region involved insertions and deletions after residue 30 (Chothia & Lesk, 1987). In Table 2 we give the Kabat and structural numbering of residues in the V_κ L1 hyper-variable region.

Six canonical structures have been identified for this region (Chothia *et al*., 1989; Rini *et al*., 1993; Haynes *et al*., 1994; Tomlinson *et al*., 1995). These have the serial numbers 1 to 6; the actual number reflecting mainly the historical order in which their

Table 3. Torsion angles of the V_κ L1 canonical structures

Residue	Angle	Canonical structure 1	2	4	3
26	Φ	−71	−80	−87	−75
	Ψ	−16	−20	−22	−15
27	Φ	−158	−132	−134	−136
	Ψ	169	166	166	157
28	Φ	−59	−72	−66	−66
	Ψ	133	128	139	140
29	Φ	−113	−112	−95	−103
	Ψ	151	6	23	9
30	Φ	−76	56	−77	−86
	Ψ	−38	−120	119	130
30a	Φ		−125	−75	−105
	Ψ		29	156	107
30b	Φ			−54	−56
	Ψ			−34	−34
30c	Φ			−72	−59
	Ψ			−4	−54
30d	Φ				1
	Ψ				8
30e	Φ			97	47
	Ψ			−18	45
30f	Φ			−81	−132
	Ψ			156	153
31	Φ			−117	−102
	Ψ			107	128
32	Φ	−157	−88	−89	−95
	Ψ	160	73	77	74

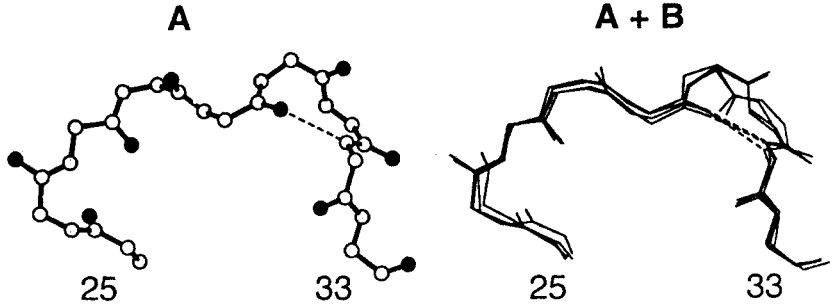


Torsion angles:

Residue	Angle		Angle	
26	Φ	-71	Ψ	-16
27	Φ	-158	Ψ	169
28	Φ	-59	Ψ	133
29	Φ	-113	Ψ	151
30	Φ	-76	Ψ	-38
32	Φ	-157	Ψ	160

Structure	PDB file	Residues
J539	2fbj	SSSVSS

Figure 2. κ L1 canonical structure 1.



Torsion angles:

Residue	Angle	Mean	S.D.	Angle	Mean	S.D.
26	Φ	-80	10	Ψ	-20	33
27	Φ	-132	49	Ψ	166	8
28	Φ	-72	9	Ψ	128	3
29	Φ	-112	8	Ψ	6	6
A:30	Φ	56	16	Ψ	-120	2
B:30	Φ	54	0	Ψ	42	5
A:31	Φ	-126	16	Ψ	21	20
B:31	Φ	67	4	Ψ	6	1
32	Φ	-88	5	Ψ	73	4

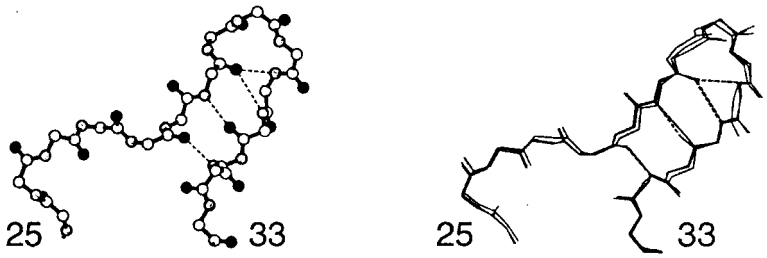
Hydrogen Bonds:

A & B: 29 N - - - O 32
B only : 30 O - - - HO Y 71

R.M.S. deviation for A: 0.4Å
R.M.S. deviation for B: 0.5Å
R.M.S. deviation for all structures: 0.4 - 0.9Å

Structure	PDB file	Residues
A:		
POT	1igm	ASQDISNYL
4D5	1fvc	ASQDVNTAV
B:		
H52	1fgv	ASQDINNYL
D1.3	1vfa	ASGNIHNYL

Figure 3. κ L1 canonical structure 2.



Torsion angles:

Residue	Angle	Mean	S.D.	Angle	Mean	S.D.
26	Φ	-75	4	Ψ	-15	4
27	Φ	-136	1	Ψ	157	3
28	Φ	-66	2	Ψ	140	4
29	Φ	-103	2	Ψ	9	8
30	Φ	-85	4	Ψ	130	3
30a	Φ	-105	10	Ψ	107	3
30b	Φ	-56	6	Ψ	-34	6
30c	Φ	-59	6	Ψ	-54	25
30d	Φ	1	13	Ψ	8	10
30e	Φ	47	8	Ψ	45	0
30f	Φ	-132	21	Ψ	153	7
31	Φ	-102	4	Ψ	128	6
32	Φ	-95	8	Ψ	74	1

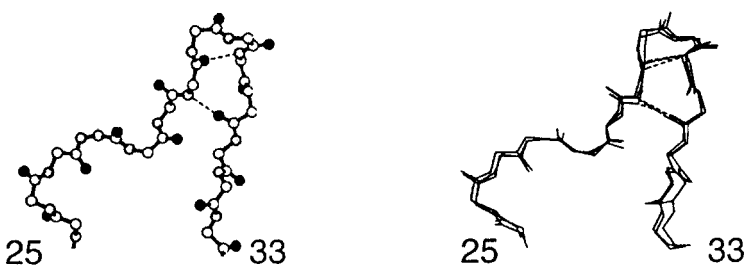
Hydrogen Bonds:

29 O - - - N 32
30a N - - - O 30f
30a O - - - N 30d
30a O - - - N 30e

R.M.S. deviations: 0.3 to 0.5Å

Structure	PDB file	Residues
17/9	1hil	SQSLFNSGKQKNY
MCPC603	2imm	SQSLLNSGNQKNF

Figure 4. κ L1 canonical structure 3.



Torsion angles:

Residue	Angle	Mean	S.D.	Angle	Mean	S.D.
26	Φ	-87	8	Ψ	-22	22
27	Φ	-134	23	Ψ	166	7
28	Φ	-66	4	Ψ	139	3
29	Φ	-95	11	Ψ	13	23
30	Φ	-77	9	Ψ	119	6
30a	Φ	-75	7	Ψ	156	9
30b	Φ	-54	5	Ψ	-34	9
30d	Φ	-72	14	Ψ	-6	8
30e	Φ	97	6	Ψ	-18	7
30f	Φ	-81	8	Ψ	156	4
31	Φ	-117	11	Ψ	107	3
32	Φ	-89	9	Ψ	77	7

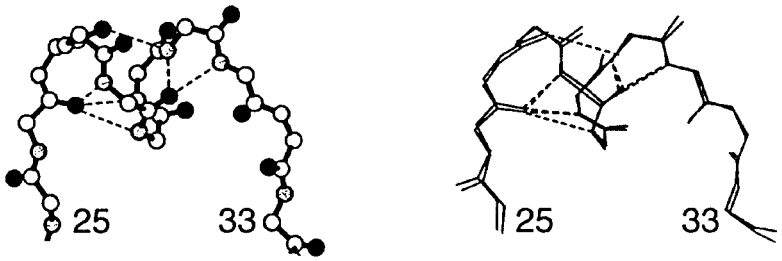
Hydrogen Bonds:

30 N - - - O 31
30 O - - - N 30d

R.M.S. deviations: 0.5 – 0.6Å

Structure	PDB file	Residues
4-4-20	1flr	SQSLVHSNGNTY
NC6.8	2cgr	SQSLVHSNGNTY
TE33	1tet	SQSIVHSSGNTY

Figure 5. κ L1 canonical structure 4.



Torsion angles:

Residue	Angles					
		Mean	S.D.		Mean	S.D.
25	Φ	-132	11	Ψ	-148	8
26	Φ	-134	10	Ψ	178	2
27	Φ	-53	4	Ψ	-35	3
28	Φ	-82	1	Ψ	-18	4
29	Φ	-113	1	Ψ	-90	7
30	Φ	-58	7	Ψ	-40	4
30a	Φ	-60	7	Ψ	-26	9
30b	Φ	-116	6	Ψ	-22	9
31	Φ	-133	1	Ψ	163	7
32	Φ	-75	12	Ψ	147	14
33	Φ	-104	13	Ψ	139	6

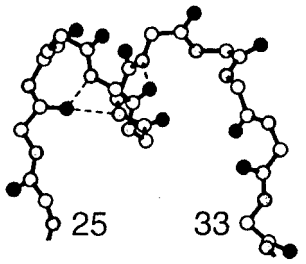
Hydrogen Bonds:

26 O - - - N 29
26 O - - - N 30
26 O - - - N 30a
27 O - - - N 30b
29 O - - - N 30b
29 O - - - N 30

R.M.S. deviation = 0.3Å

Structure	PDB file	Residues
KOL	2fb4	GTSSNIGSSTV
RHE	1rhe	GSATDIGSNSV

Figure 6. λ L1 canonical structure 1.



Torsion angles:

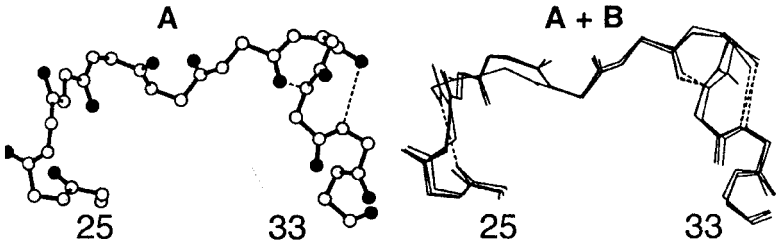
Residue	Angles			
25	Φ	-120	Ψ	-135
26	Φ	-149	Ψ	-180
27	Φ	-55	Ψ	-23
28	Φ	-84	Ψ	-19
29	Φ	-127	Ψ	-88
30	Φ	-61	Ψ	-38
30a	Φ	-74	Ψ	7
30b	Φ	-116	Ψ	26
30c	Φ	70	Ψ	37
31	Φ	-109	Ψ	148
32	Φ	-71	Ψ	157
33	Φ	-119	Ψ	125

Hydrogen Bonds:

26 O - - - N 29
26 O - - - N 30a
29 O - - - N 30b

Structure	PDB file	Residues
NEW	7fab	GSSSNIGAGHNV

Figure 7. λ L1 canonical structure 2.



Torsion angles:

Residue	Angle	Mean	S.D.	Angle	Mean	S.D.
25	Φ	-99	1	Ψ	137	5
26	Φ	-67	1	Ψ	-31	19
27	Φ	-70	6	Ψ	-20	8
A: 28	Φ	161	9	Ψ	-164	0
B: 28	Φ	112	–	Ψ	-29	–
A: 29	Φ	-66	3	Ψ	143	5
B: 29	Φ	-149	–	Ψ	153	–
30	Φ	-89	15	Ψ	136	5
30a	Φ	-124	7	Ψ	173	6
30b	Φ	-60	1	Ψ	-33	5
30c	Φ	-74	14	Ψ	-4	10
31	Φ	-93	6	Ψ	-6	4
32	Φ	43	5	Ψ	52	5
33	Φ	–42	1	Ψ	124	1

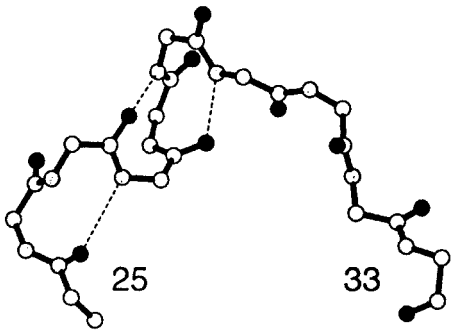
Hydrogen Bonds:

A & B: 30a O - - - N 31
A only: 30b O - - - N 32
B only: 25 O - - - N 28

R.M.S. deviation for A: 0.2Å
R.M.S. deviation all structures: 0.2 – 0.8Å

Structure	PDB file	Residues
A:		
CHA255	1ind	SSTGAVTTSNYA
HC19	1gig	SSTGAVTTSNYA
B:		
SE155	1mfa	SSTGTVTSGNHA

Figure 8. λ L1 canonical structure 3.



Torsion angles:

Residue	Angles			
25	Φ	-172	Ψ	150
26	Φ	-69	Ψ	-25
27	Φ	-80	Ψ	-6
30	Φ	-71	Ψ	-27
30a	Φ	-47	Ψ	-34
30b	Φ	-101	Ψ	10
30c	Φ	-135	Ψ	159
31	Φ	-91	Ψ	152
32	Φ	-105	Ψ	135
33	Φ	-136	Ψ	150

Hydrogen Bonds:

25 O - - - N 28
27 O - - - N 30
28 O - - - N 32

Structure	PDB file	Residues
HIL	8fab	ANALPNQYAY

Figure 9. λ L1 canonical structure 4.

structures were determined. Almost all human germline and expressed V_κ domains have canonical structures 2, 3, 4 or 6 in their L1 regions (Tomlinson *et al.*, 1995). Inspection of the sequences of mouse germline and expressed V_κ domains (Kabat *et al.*, 1991) indicates that almost all their L1 regions have one of the four found in human antibodies or one of the two found in non-human antibody structures: 1 or 5 (our unpublished work).

Examination of the six canonical structures shows that residues 26 to 29 adopt an extended conformation whereas residues between 29 and 32 form short links or hairpin loops. The number of residues between 30 and 32 in the different canonical structures are:

Canonical structure	No. of residues between 30 and 32
1	0
2	1
3	7
4	6
5	5
6	2

Initial studies had suggested that, for all six canonical structures, residues 26 to 29 and 32 have a common conformation that packs against the framework, and the variation just involved the number of residues between 29 and 32. The data for the high resolution structures in Table 3 show that this is true for canonical structures 2 to 4 (and by implication 5 and 6). Canonical structure 1 has residue 31 deleted and the effects of this limit the region of conserved conformation to residues 26 to the C^α of 29 (Table 3).

X-ray structure determinations are available for all six canonical structures; however, accurate high resolution structures are available only for 1, 2, 3 and 4. The conformational details of canonical structures 1, 2, 3 and 4 found in the high resolution structures are given in Figures 2 to 5.

For canonical structure 1, only one high resolution structure is available. The structure of HyHEL-5, determined at a resolution of 2.65 Å (Sheriff *et al.*, 1987; Cohen *et al.*, 1995) has the same canonical structure. In HyHEL-5 the torsion angles in the L1 region are very similar to those in J539 and optimal superposition of their main-chain atoms gives a r.m.s. deviation of atomic positions of 0.3 Å.

Canonical structure 2 has two alternative forms: A and B. These differ in the conformation of the peptide between residues 30 and 31 (Figure 3). The conformation assumed in a particular antibody depends on the key residue occupying the framework position 71, which packs against the loop. In type B, this position is occupied by Tyr and a hydrogen bond is formed between its hydroxyl group and the carboxyl oxygen atom of residue 30. In type A, residue 71 is Phe, which cannot form this hydrogen bond and the peptide rotates 170° relative to its position in B. Nineteen V_κ human germline segments are expected to have canonical structure 2 in L1 region and they all have Phe at position 71 (Tomlinson *et al.*, 1995). Thus, we would expect human antibodies to have form A, except for those cases where a somatic mutation changes residue 71 to Tyr.

Canonical structures 3 and 4 have insertions between residues 30 and 32 of seven and six residues, respectively. These residues form a short β-sheet hairpin with the turns at the top (see Figures 4 and 5). The turn at the top of canonical structure 3 has four residues and that at the top of canonical structure 4 has three residues. In both cases they have the conformations most commonly found for turns of this size (Sibanda *et al.*, 1989; Efimov, 1993).

V_λ L1 canonical structures

As in V_κ, residues 24 to 34 in V_λ domains were defined on the basis of sequence variation as the first CDR region (Kabat *et al.*, 1979). The analysis of the first three V_λ structures indicated that the region outside the β-sheet framework and therefore potentially able to have different conformations comprised residues 26 to 32 (Chothia & Lesk, 1987). Later it was shown that residue 25 is not part of the framework region in certain V_λ structures (Wu & Cygler, 1993).

The conformations of four canonical structures, numbered 1 to 4, have been identified so far (Chothia & Lesk, 1987; Wu & Cygler, 1993; Martin & Thornton, 1996; our unpublished work) and high resolution structures are available for all of them (Table 1). In the region 25 to 32 canonical structure 1 has ten residues; canonical structures 2 and 3 have 11 residues; and canonical structure 4 has nine residues. Inspection of the structures shows that they all have near the centre a hydrophobic residue that packs in a cavity in the framework. We give this residue the number 30 in all four canonical structures and the structural sequence numbering that results from this is given, together with the Kabat numbering, in Table 4.

The conformations of canonical structures 1 and 2 are closely related but 3 and 4 are quite different from 1 and 2 and from each other (Table 5). Canonical structures 1 and 2, shown in Figures 6 and 7, differ by the latter having an additional residue, 30c. This has only a local effect and residues 25 to 30b and 31 to 33 have the same conformation. Residues 27 to 30b of both canonical structures form an irregular helix.

The conformations found for canonical structure 3 are shown in Figure 8. Two different forms exist and differ in the orientation of the peptide between residues 28 and 29. The reason for this difference is not known. In both forms, the loop is extended across the β-sheet core with a helical turn formed by residues 30b to 32.

For canonical structure 4 (called 1b by Martin & Thornton, 1996), residues 26 to 30 form a distorted

Table 4. Kabat and structural numbering of residues in the L1 region of V_λ domains

Canonical structure	Numbering scheme	Residue numbers										
1	Kabat	25	26	27	27a	27b	28	29	30		31	32
1	Structure	25	26	27	28	29	30	30a	30b		31	32
2	Kabat	25	26	27	27a	27b	27c	28	29	30	31	32
2	Structure	25	26	27	28	29	30	30a	30b	30c	31	32
3	Kabat	25	26	27	27a	27b	27c	28	29	30	31	32
3	Structure	25	26	27	28	29	30	30a	30b	30c	31	32
4	Kabat	25	26	27			27a	28	29	30	31	32
4	Structure	25	26	27			30	30a	30b	30c	31	32

Note that residues in different canonical structures with the same residue number do not necessarily have the same conformation: see Table 5.

helix (Figure 9). Only one known immunoglobulin structure, HIL, contains this canonical structure. In the crystal, however, there are two molecules in the asymmetric unit and the r.m.s. difference in the position of the main-chain atoms of 25 to 32 is 0.3 Å.

L2 Hypervariable Region

The L2 hypervariable region is in the hairpin loop linking the C′ and C″ strands (Figure 1). On the basis of sequence variation, residues 50 to 56 were defined by Kabat *et al.* (1979) as the second CDR region. Subsequently it was shown that the region outside the β-sheet framework and therefore potentially able to have different conformations comprises residues 50 to 52 in both V_λ and V_κ domain (Chothia & Lesk, 1987).

The residues in this region form a three-residue hairpin loop, joining framework residues 49 and 53 which are themselves linked by two main-chain hydrogen bonds. The previous work on this region found only one canonical structure in the V_λ and V_κ domains (Chothia & Lesk, 1987; Steipe *et al.*, 1992). All 16 L2 regions in the accurate structures considered here have very similar conformations: superposition of the main-chain atoms gave r.m.s. deviations ranging between 0.1 and 0.5 Å (Figure 10). The three residue L2 loop is a classic γ-turn: the residue at the tip, 51, is in a strained conformation whose angles, +67°, −40°, are close to the average value found in this type of turn: +75°, −60° (Milner-White *et al.*, 1988).

Table 5. Torsion angles of the V_λ L1 canonical structures

Residue	Angle	Canonical structure 1	2	3	4
25	Φ	−132	−120	−99	−172
	Ψ	−148	−135	137	150
26	Φ	−134	−149	−67	−69
	Ψ	178	−180	−31	−25
27	Φ	−53	−55	−70	−80
	Ψ	−35	−23	−20	−6
28	Φ	−82	−84	144	
	Ψ	−18	−19	−119	
29	Φ	−113	−127	−94	
	Ψ	−90	−88	147	
30	Φ	−58	−61	−89	−71
	Ψ	−40	−38	136	−27
30a	Φ	−60	−74	−124	−47
	Ψ	−26	7	173	−34
30b	Φ	−116	−116	−60	−101
	Ψ	−22	26	−33	10
30c	Φ		70	−74	−135
	Ψ		37	−4	159
31	Φ	−133	−109	−92	−91
	Ψ	163	148	−6	135
32	Φ	−75	−71	43	−105
	Ψ	147	157	51	152
33	Φ	−104	−119	−44	−136
	Ψ	139	125	126	150

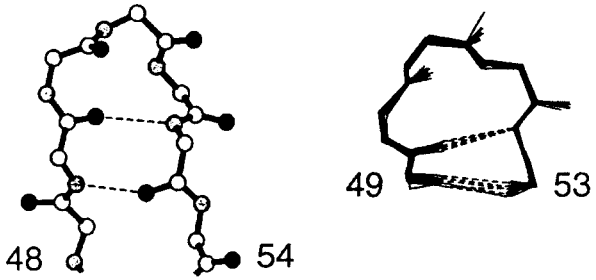
L3 Hypervariable Regions

V_κ L3 canonical structures

The L3 hypervariable region is in the hairpin loop linking the F and G strands (Figure 1). On the basis of sequence variation, residues 89 to 96 were defined by Kabat *et al.* (1979) as the third CDR. Subsequently, it was shown that the region outside the β-sheet framework and therefore potentially able to have different conformations comprises residues 91 to 96 in both V_λ and V_κ domains (Chothia & Lesk, 1987).

So far, six canonical structures have been observed for this region. Five of these are described by Chothia *et al.* (1989), Brüger *et al.* (1991) and He *et al.* (1992). Recently Guarné *et al.* (1996) have described a sixth canonical structure that they found in the accurate high resolution structure of the immunoglobulin CRIS-1. The medium resolution structure of 17E8 (Zhou *et al.*, 1994) has an L3 region the same size as that in CRIS-1 and has a very similar conformation. No atomic co-ordinates are available for CRIS-1 and we therefore describe here just the two for which data are available: canonical structures 1 and 2 (Figures 11 and 12).

Canonical structure 1 is that most commonly observed in the L3 regions of V_κ domains and is found in nine of the structures considered here (Figure 11). Eight are in a Fab or Fv fragment of an immunoglobulin; one (McPC603) is in a dimer of V_κ domains. For the eight canonical structures in natural oligomers, superpositions of their main-chain co-ordinates give r.m.s. differences in position between 0.3 and 0.7 Å.



Torsion angles:

Residue		Angle Mean	S.D.		Mean	S.D.
49	Φ	-128	7	Ψ	158	9
50	Φ	51	8	Ψ	50	9
51	Φ	67	7	Ψ	-40	19
52	Φ	-140	15	Ψ	2	15
53	Φ	-91	10	Ψ	126	11

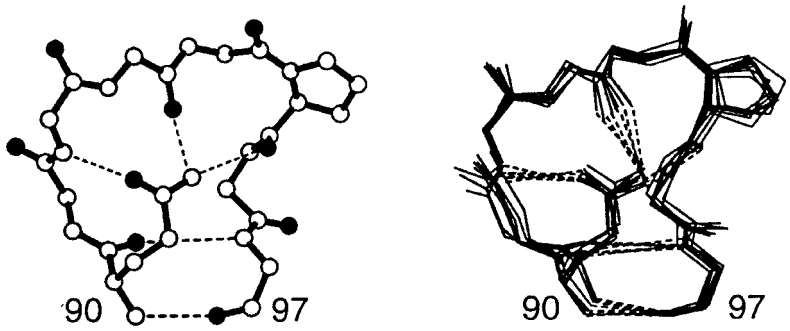
Hydrogen Bonds:

49 N - - - O 53
49 O - - - N 53

R.M.S. deviations : 0.1 - 0.5 Å

Structure	PDB file	Rsidues	Structure	PDB file	Residues
KOL	2fb4	YRDAM	TE33	1tet	YKVSN
J539	2fbj	YEISK	NC6.8	2cgr	YRVSN
POT	1igm	YDASN	MCPC	2imm	YGAST
D1.3	1vfa	YYTTT	RHE	2rhe	YYNDL
H52	1fgv	YYTST	HIL	8fab	YKDTQ
4D5	1fvc	YSASF	SE155	1mfa	GDTNN
17/9	1hil	YWAST	CHA255	1ind	GGTNN
4-4-20	1flr	YKVSN	HC19	1gig	GGTNN

Figure 10. L2 canonical structure 1.



Torsion angles:

Residue	Angle	Mean	s.d.	Angle	Mean	s.d.
90	Φ	-96	19	ψ	131	16
91	Φ	-122	13	ψ	25	12
92	Φ	-91	23	ψ	-39	12
93	Φ	-128	17	ψ	145	22
94	Φ	-86	9	ψ	140	9
95	Φ	-74	6	ψ	144	7
96	Φ	-66	9	ψ	136	8
97	Φ	-130	6	ψ	148	9

95 is a *cis*-Proline

Hydrogen Bonds:

90 N - - - O 97
90 O - - - N 97
90 OE1 - - - N 92
90 NE1 - - - O 93
90 NE - - - O 95

R.M.S. deviations: 0.1 - 0.9Å

Structure	PDB file	Residues
POT	1igm	QYQNLPLT
H52	1fgv	QGNTLPPT
4D5	1fvc	QHYTTPPT
17/9	1hil	NDYSNPLT
4-4-20	1flr	QSTHVPWT
TE33	1tet	QGSHIPFT
NC6.8	2cgr	QGTHVPYT
MCPC603	2imm	NDHSYPLT

Figure 11. κ L3 canonical structure 1.

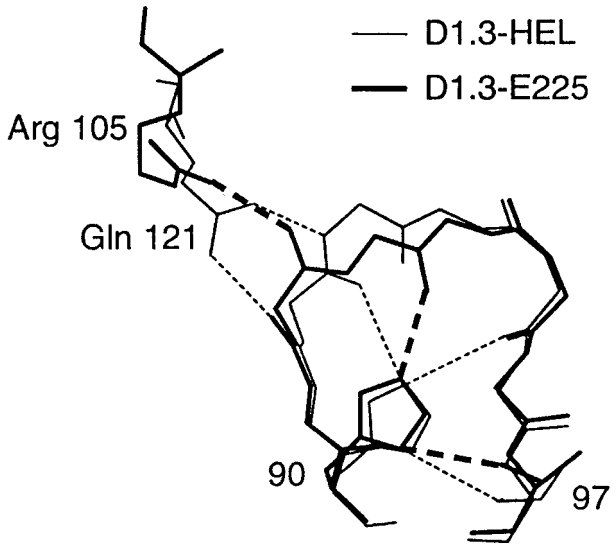
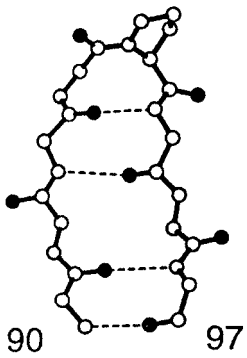


Figure 12. The binding of antibody D1.3 to certain antigens produces a rotation of a peptide in the L3 hypervariable region (see the text). Here we show the κ L3 loop in D1.3 bound to the immunoglobulin E225 and hen egg-white lysozyme (HEL). In the first complex the L3 region hydrogen bonds to Arg105 of E225 with a standard canonical structure 1 conformation. In the second complex the L3 region is able to form hydrogen bonds to Gln121 of HEL by a rotation of the 92–93 peptide group (Bhat *et al.*, 1994; Braden *et al.*, 1996).



Torsion angles:

Residue	Angle			
90	Φ	-106	ψ	161
91	Φ	-133	ψ	131
92	Φ	-153	ψ	100
93	Φ	-45	ψ	126
94	Φ	-101	ψ	31
95	Φ	-121	ψ	147
96	Φ	-126	ψ	131
97	Φ	-134	ψ	141

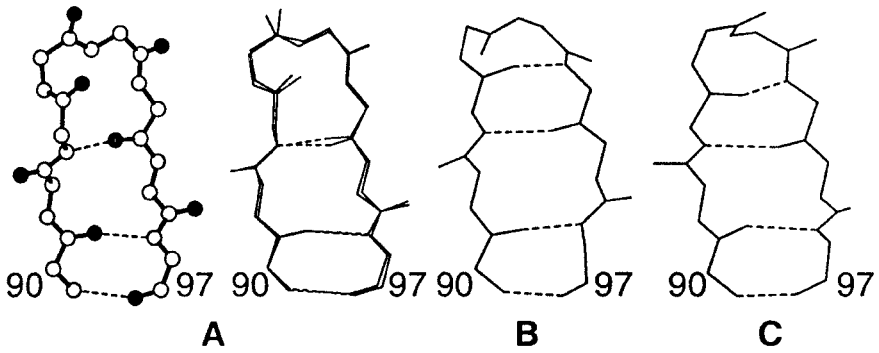
94 is a *cis*-Proline

Hydrogen Bonds:

90 N - - - O 97
90 O - - - N 97
92 N - - - O 95
92 O - - - N 95

Structure	PDB file	Residues
J539	2fbj	QWTYPLIT

Figure 13. κ L3 canonical structure 2.



Torsion angles:

Residue	Angle	A		B	C	Angle	A		B	C
90	Φ	-128	1	-141	-89	ψ	137	11	-160	113
91	Φ	-95	4	-138	-143	ψ	131	2	145	122
92	Φ	-122	4	-134	54	ψ	60	6	94	45
93	Φ	48	14	86	59	ψ	10	64	-124	36
94	Φ	-160	42	-96	-136	ψ	-23	21	2	144
95	Φ	-148	3	-77	-91	ψ	171	10	136	127
96	Φ	-113	9	-93	-119	ψ	143	4	133	129
97	Φ	-132	1	-118	-97	ψ	125	1	127	143

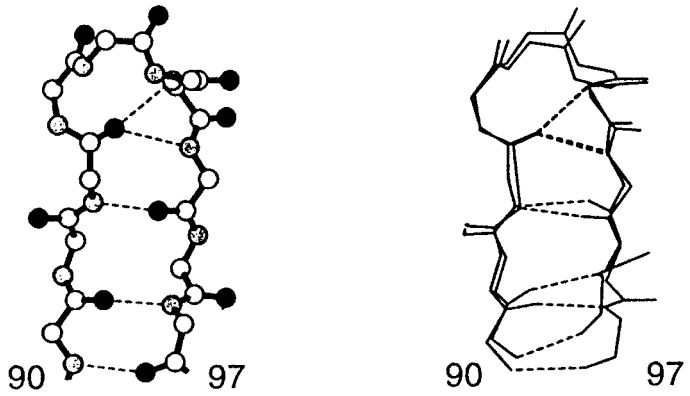
Hydrogen Bonds:

90 N - - - O 97
90 O - - - N 97
92 N - - - O 95
92 O - - - N 95

R.M.S. deviations for A: 0.4Å
R.M.S. deviations for all structures: 0.4 - 1.1Å

Structure	PDB file	Residues
A:		
CHA255	1ind	LWYSNLWV
HC19	1gig	LWYSNHWV
B:		
NEW	7fab	SYDRSLRV
C:		
SE155	1mfa	LWSNNHWI

Figure 14. λ L3 canonical structure 1.



Torsion angles

Residue	Angle	Mean	S.D.	Angle	Mean	S.D.
90	Φ	-164	4	Ψ	150	3
91	Φ	-71	8	Ψ	135	5
92	Φ	-111	11	Ψ	117	5
93	Φ	-66	1	Ψ	-27	3
94	Φ	-79	3	Ψ	-40	6
95	Φ	-84	2	Ψ	-16	7
95a	Φ	54	11	Ψ	47	15
95b	Φ	-159	21	Ψ	166	17
96	Φ	-82	16	Ψ	132	24
97	Φ	-110	4	Ψ	135	17

Hydrogen Bonds:

90 N - - - O 97
90 O - - - N 97
92 N - - - O 95
92 O - - - N 95
92 O - - - N 94

R.M.S. Deviation = 0.8Å

Structure	PDB file	Residues
KOL	2fb4	AWDVSLNAYV
RHE	2rhe	AWNDSLDEPG

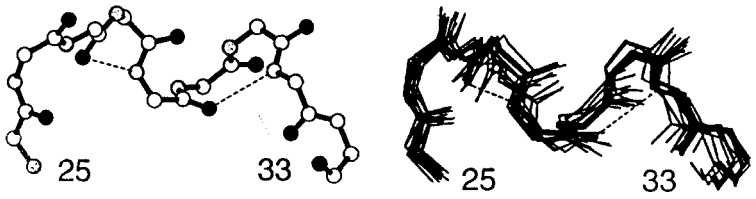
Figure 15. λ L3 canonical structure 2.

Table 6. Kabat and structural numbering of residues in the H1 region of V_H domains

Canonical structure	Numbering scheme	Residue numbers									
1	Kabat	29	30	31			32	33	34	35	36
1	Structure	29	30	31			32	33	34	35	36
2	Kabat	29	30	31	32		33	34	35	35a	36
2	Structure	29	20	31	31a		32	33	34	35	36
3	Kabat	29	30	31	32	33	34	35	35a	35b	36
3	Structure	29	30	31	31a	31b	32	33	34	35	36

Superposition of the McPC603 canonical structure on the other eight gives larger r.m.s. differences in position: 0.6 to 0.9 Å. Viewed from the side the canonical structure is bent and the larger difference between the McPC603 structure and the others is because it is more bent than they are. The structure of the Fab fragment of immunoglobulin McPC603 is known (though only at a resolution of 2.8 Å; Satow *et al.*, 1986). To see whether the conformation found in the V_κ dimer is likely to occur in the V_κ-V_H complex we replaced the V_κ domain in the Fab structure of McPC603 with a V_κ from the dimer. We find that residues Asp97 and His98 in L3 of the dimer V_κ occupy the same space as Trp103 in V_H; a residue that is absolutely conserved in V_H domains. Thus, it seems very unlikely that canonical structure 1 with a conformation the same as that found in the McPC603 V_κ dimer will be found in immunoglobulins.

The L3 region of the antibody D1.3 has canonical structure 1, though the key residue at position 90 is His rather than the more usual Asn or Gln. The complexes formed by D1.3 show cases of a ligand-



Torsion angles:

Residue	Angle	Mean	S.D.	Angle	Mean	S.D.
26	Φ	90	8	Ψ	6	16
27	Φ	-158	8	Ψ	157	42
28	Φ	-88	16	Ψ	111	19
29	Φ	-54	13	Ψ	-40	13
30	Φ	-71	13	Ψ	-9	17
31	Φ	-98	15	Ψ	-6	18
32	Φ	-125	14	Ψ	148	11

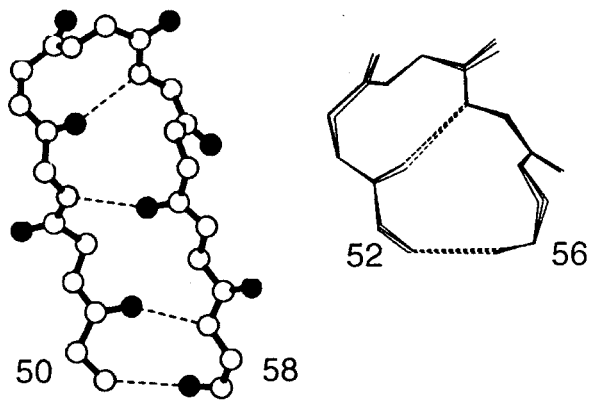
Hydrogen Bonds:

31 N --- O 28
32 N --- O 29

R.M.S. deviations: 0.1 to 1.2Å

Structure	PDB file	Residues	Structure	PDB file	Residues
KOL	2fb4	SGFIFSSYA	TE33	1tet	SGYTFTTYG
J539	2fbj	SGFDFSKYW	SE155	1mfa	SGYTFTNYW
CHA255	1ind	SGFTLSGET	POT	1igm	SGFTFNIFV
HIL	8fab	SGFTFSNYG	NEW	7fab	SGTSFDDYY
17/9 (B)	1hil	SGFSFSSYG	NC6.8	2cgr	TGYTFSEYW
H52	1fgv	SGYTFTEYT	HC19	1gig	SGFLLISNG
4D5	1fvc	SGFNIKDTY	4-4-20	1flr	SGFTFSDYW
D1.3	1vfa	SGFSLTGYG			

Figure 16. H1 canonical structure 1.



Torsion angles

Residue	Angle	Mean	S.D.	Angle	Mean	S.D.
52	Φ	-85	24	Ψ	157	16
53	Φ	-57	7	Ψ	-43	27
54	Φ	-77	8	Ψ	-5	10
55	Φ	103	8	Ψ	-3	12
56	Φ	-65	7	Ψ	142	9

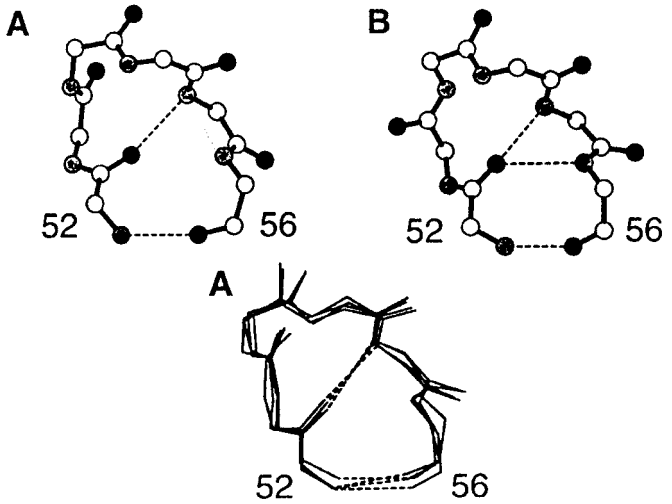
Hydrogen Bonds:

55 N - - - O 52
52 N - - - O 56

R.M.S. deviations: 0.1 to 0.2Å

Structure	PDB file	Residues
NEW	7fab	FYTGT
D1.3	1vfa	WGDGN
HC19	1gig	WAGGN

Figure 17. H2 canonical structure 1.



Torsion angles :

Residue	Angle	A mean	A S.D.	B	Angle	A mean	A S.D.	B
52	Φ	-95	15	-114	Ψ	111	5	106
52a	Φ	-60	16	-67	Ψ	-31	21	125
53	Φ	-61	16	134	Ψ	-49	15	-71
54	Φ	-109	37	-91	Ψ	1	18	-22
55	Φ	70	13	74	Ψ	15	12	24
56	Φ	-69	16	-84	Ψ	128	7	127

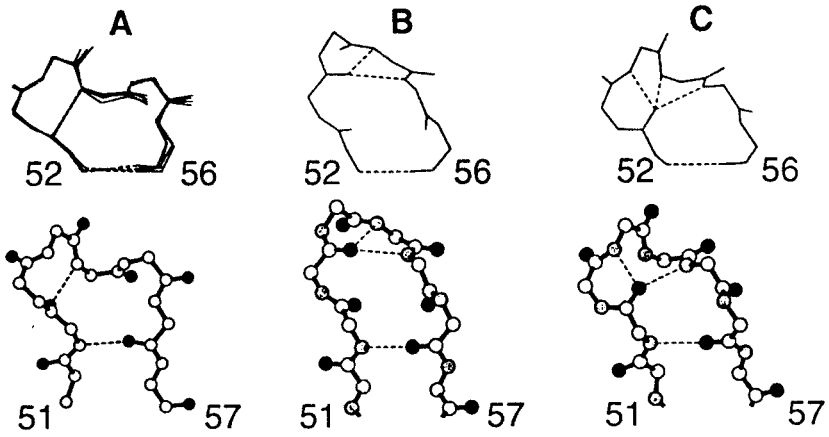
Hydogen bonds:

A & B: 52 N - - - O 56
A & B: 52 O - - - N 55
B only: 52 O - - - N 56

R.M.S. Deviations for A: 0.3 - 0.5Å

Structure	pdb file	Residues
A		
SE155	1mfa	YPGNS
4D5	1fvc	YPTNG
H52	1fgv	NPKNG
TE33	1tet	NTYSG
B		
NC6.8	2cgr	LPGSG

Figure 18. H2 canonical structure 2.



Torsion angles :

		A				
Residue	**Angle**	**mean**	**S.D.**	**Angle**	**Mean**	**S.D.**
52	Φ	-62	7	Ψ	156	10
52a	Φ	-50	9	Ψ	-37	8
53	Φ	-100	12	Ψ	8	21
54	Φ	76	17	Ψ	5	10
55	Φ	-71	3	Ψ	-27	8
56	Φ	-138	22	Ψ	140	17

	B				C			
Residue	**Angles**				**Angles**			
52	Φ	-73	Ψ	155	Φ	-106	Ψ	89
52a	Φ	-71	Ψ	-177	Φ	49	Ψ	-75
53	Φ	80	Ψ	-110	Φ	-72	Ψ	-8
54	Φ	-156	Ψ	32	Φ	75	Ψ	37
55	Φ	-68	Ψ	180	Φ	-83	Ψ	-117
56	Φ	-57	Ψ	160	Φ	-65	Ψ	137

Hydogen bonds:

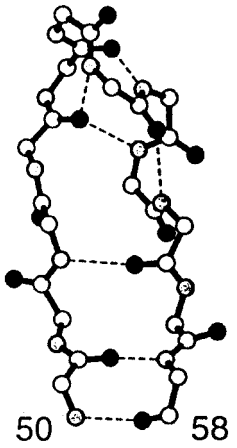
A	B
52 N - - - O 56	52 N - - - O 56
52 O - - - N 54	53 O - - - N 55

C
52 N - - - O 56
52 O - - - N 53
52 O - - - N 55

Structure	**pdb file**	**Residues**
A:		
KOL	2fb4	WDDGSD
HIL	8fab	WYNGSR
J539	2fbj	HPDSGT
17/9	1hil	SNGGGY
B:		
CHA255	1ind	LSGGGF
C:		
POT	1igm	FGSGGN

R.M.S. Deviations for 3A: 0.2 - 0.3Å
R.M.S. Deviations for all structures: 0.2 - 1.5Å

Figure 19. H2 canonical structure 3.



Torsion angles:

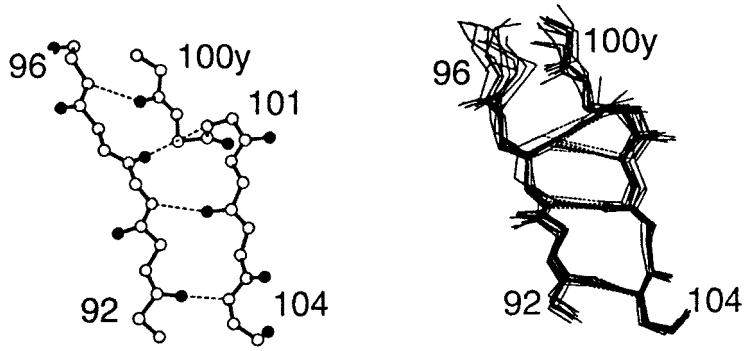
Residue	**Angles**			
52	Φ	-71	Ψ	-172
52a	Φ	-117	Ψ	-174
52b	Φ	-42	Ψ	-65
52c	Φ	-44	Ψ	-30
53	Φ	-112	Ψ	18
54	Φ	49	Ψ	60
55	Φ	24	Ψ	42
56	Φ	-57	Ψ	165

Hydrogen Bonds:

52	N	- - - O	56
53	N	- - - O	52a
54	N	- - - O	52a
55	N	- - - O	52a
54	N	- - - O	52b

Structure	**pdb file**	**Residue**
4-4-20	1flr	RNKPYNYE

Figure 20. H2 canonical structure 4.



Torsion Angles:

Residue	Angle	Mean	S.D.	Angle	Mean	S.D.
92 (C)	ϕ	-101	10	ψ	147	6
93	ϕ	-141	7	ψ	152	13
94	ϕ	-94	15	ψ	138	13
95	ϕ	-102	22	ψ	137	12
96	ϕ	-100	38	ψ	-23	98
100y	ϕ	-78	102	ψ	164	15
100z	ϕ	-105	25	ψ	99	13
101	ϕ	-86	10	ψ	-31	11
102	ϕ	-130	10	ψ	133	7
103	ϕ	-119	8	ψ	149	5
104(G)	ϕ	-79	6	ψ	-176	10

Hydrogen Bonds:

92	O	- - -	N	104
94	O	- - -	N	103
94	N	- - -	O	104
98	O	- - -	N	103

R.M.S. deviation: 0.17-2.24 Å

Structure	PDB file	Structure	PDB file
TE33	1tet	4D5	1fvc
NC6.8	2cgr	HIL (B)	8fab
D1.3	1vfa	HC1	1gig
J539	2fbj	KOL	2fb4
NEW	7fab	POT	1igm
SE155	1mfa	17/9(Complex)	1hil

Figure 21. The common H3 stem conformation (see Shirai *et al.*, 1996 and Morea *et al.*, 1997). Kabat and Wu have placed the insertions in this region after residue 100 and numbered then 100a, 100b etc. Here we use 100y and 100z to represent the last two residues before residue 101, although the numbering in the actual structures will differ. The calculation of the mean ψ value of residue 100y given here does not include the residue from SE155 which has a value of −21°. If D1.3 is excluded from the calculation of the r.m.s. difference in atomic positions in this region of H3, the range of differences is 0.2 to 1.7 Å.

induced conformational change in this canonical structure. Structures of D1.3 are known for a native form and for complexes with hen egg lysozyme (HEL), turkey egg lysozyme (TEL) and the antibody E225 (Braden *et al.*, 1996). In the complexes with TEL and E225, L3 has the conventional canonical structure 1 conformation (Figure 11) with r.m.s. deviations between 0.4 and 0.5 Å when fitted against the high resolution structures listed in Figure 11. In the native form and the complex with HEL the conformation differs with the peptide between residues 92 and 93 rotated by ~120°. For both conformations the peptide orientation is stabilised by hydrogen bonds from the bound antigen or, in the native structure, a neighbouring D1.3 molecule (Figure 12). In this case, the unusual conformation is produced by hydrogen bonds from the antigen that provide energy sufficient to overcome the strain energy that the flipped peptide produces in the conformation of residue 93 where $\phi = +57°$, $\psi = -150°$.

J539 is the only accurately determined structure available for canonical structure 2. This loop forms a regular hairpin with two main-chain hydrogen bonds between residues 92 and 95. Site 94 in the turn is occupied by a *cis*-proline residue (Figure 13). This canonical structure has not been observed so far in any other immunoglobulin structure.

λ L3 canonical structures

Two canonical structures are known for the L3 region of V_λ and accurately determined structures are available for both. Canonical structure 1 is a six

residue hairpin with two residues at the top forming the turn (Figure 14). All four high resolution structures available adopt similar conformations over the majority of the loop but differ in the orientation of the peptide between residues 93 and 94 at the top of the loop. This gives three forms for this canonical structure: A, B and C (see Figure 4). The differences in the orientations of the peptide involve small accommodations in torsion angles of residues on both sides of the peptide. A hydrogen bond between the main-chain oxygen atom of residue 95 and the main-chain nitrogen atom of residue 92 is found in all three forms of the canonical structure. A second hydrogen bond between the carbonyl oxygen atom of residue 92 and the main-chain amino group of residue 95 is present in the B and C forms but not in A. The reasons for the differences in the orientation of the peptide are not known.

The second canonical structure consists of eight residues in a hairpin with four of these residues forming the turn (Figure 15). The two accurately determined examples have very similar conformations that have a r.m.s. difference in the position of their main-chain atoms of 0.8 Å. The four residue turn has the conformation found most commonly in turns of this size (Sibanda *et al.*, 1989; Efimov, 1993).

The H1 Hypervariable Region

The H1 region packs across the top of the V_H domain, bridging the two β-sheets. On the basis of sequence variation, residues 31 to 35 were defined by Kabat *et al.* (1979) as the first CDR. Up to two insertions are found in this region and it has been believed that these involve insertions at a site following residue 35. Structural work has shown that the region outside the β-sheet framework and therefore potentially able to have different conformations comprises residues 26 to 32. It also showed that the variations in the size of this region involved insertions at site 31 (Chothia *et al.*, 1992). In Table 6 we give the Kabat and structural numbering for residues in the V_H first hypervariable region.

Three canonical structures have been observed for this region (Chothia *et al.*, 1992). Of these, canonical structure 1 is the most commonly observed conformation, and the only one for which accurate high resolution structures are available. The other two are seen in only a few structures determined at a medium resolution (see below).

The structures studied here have 15 examples of canonical structure 1 (Table 1). Superpositions of main-chain atoms of residues 26 to 32 give r.m.s. differences in atomic positions between 0.1 and 1.2 Å (see Figure 16). This range in the values of the r.m.s. differences is somewhat larger than that seen for other canonical structures. It is related to the wide range in the size of the hydrophobic residue sites 24, 29 and 34 (Chothia *et al.*, 1992). Earlier work had suggested that two sub-classes may exist for this canonical structure; 1 and 1′ (Chothia *et al.*, 1989). This was because at that time the few known structures suggested that this canonical structure could have small but distinct differences. The 15 high resolution structures studied here have continuous distribution of small changes and the notion of sub-classes is not appropriate.

H2 Hypervariable Region

On the basis of sequence variation, residues 50 to 65 were defined by Kabat *et al.* (1979) as the second CDR region in V_H domains. Subsequently, it was shown that residues 56 to 58 form the short C″ strand and that the region showing variation in conformation is limited to residues 52 to 56. Four canonical structures have been observed for this region.

Canonical structure 1 is the shortest loop, observed in three accurately determined structures. The optimal superposition of the main-chain atoms gives very low r.m.s. values of 0.2 Å (see Figure 17). The two hydrogen bonds shown are seen in all structures. The conformation is that most commonly seen in turns of this size (Sibanda *et al.*, 1989; Efimov, 1993).

Five of the high resolution structures adopt canonical structure 2, which contains four residues. Four of these structures have the same conformation, described as form A (see Figure 18). The fifth structure (NC6.8) has the B conformation: the peptide between 52a and 53 is rotated 160° relative to its position in form A. The reason for this difference is not apparent.

Canonical structure 3 is also a four-residue loop. It, however, assumes a very different conformation from canonical structure 2 (Tramontano *et al.*, 1990). Figure 19 shows the three different forms found for this canonical structure. The most common of these is A, which is found in four accurately determined structures. The B and C conformations each occur in one structure determined at high resolution. In both B and C three of the four residues in the loops are glycine residues. Three glycine residues are also found in the 17/9 loop which has an A form conformation (Figure 19). This suggests that the glycine residues give the loop a conformational freedom that may, or may not, be reflected in the non-standard canonical structure.

In canonical structure 4 (Figure 20), the loop contains six residues. Loops of this length have been observed only in one accurately determined structure (4-4-20). Examination of three other structures that have this canonical structure, BV04-01 (Herron *et al.*, 1991), YST9.1 (Rose *et al.*, 1993) and McPC603 (Satow *et al.*, 1986), showed conformations very close to that in 4-4-20. Atomic superposition gave main-chain r.m.s. deviations ranging from 0.4 Å for BV04-01 to 1.0 Å for YST9.1 and McPC603. Those in YST9.1 (Rose *et al.*, 1993) and McPC603 (Satow *et al.*, 1986) have the same conformation as 4-4-20

except that the peptides between 52a and 52b and between 54 and 55 are rotated by ~120°. This may represent a second form of this canonical structure but it needs to be confirmed by refinement at high resolution.

H3 Hypervariable Region

The H3 hypervariable region is in the hairpin loop linking the F and G strands (Figure 21). On the basis of sequence variation, residues 95 to 102 were defined by Kabat *et al.* (1979) as the third CDR region. The relations between the sequences and structures of the H3 hypervariable region are not as well understood as those for the other hypervariable regions. Though some progress has been made recently (Shirai *et al.*, 1996; Martin & Thornton, 1996; Morea *et al.*, 1997).

Kabat *et al.*, (1979) have placed the insertions in this region after residue 100 and numbered them 100a, 100b, etc. Here we use 100y and 100z to represent the last two residues before residue 101, although the numbering in the actual structures will differ.

In H3 loops that have 11 or more residues between Cys92 and Gly104 it is very common for the residues 92 to 96 and 100y-100z-101 to 104 to form a kinked two-stranded antiparallel β-sheet (Figure 21 and Chothia & Lesk, 1987). This conformation is found in all the structures considered here except CHA255, for which the H3 region is too short; 4-4-20, and the unliganded form of 17/9.

For the 12 high resolution structures that contain this common conformation, torsion angles and hydrogen bonds were compared and rigid-body superposition was performed over five N-terminal residues (92 to 96) and six C-terminal residues (100y to 104) which included the kink. Figure 21 shows that the conformation is well conserved over this region with r.m.s. deviations ranging between 0.4 and 1.7 Å in all cases except for D1.3. For D1.3, large differences in the relative position of residues 96 and 100y at the two ends of the stem push the r.m.s. differences up to 2.2 Å. Note that, although the native form of 17/9 does not have this conformation, it is produced by formation of the complex with the peptide of influenza hemagglutinin (Rini *et al.*, 1992).

Canonical Structures in Medium and Low Resolution Structures

We have discussed here the canonical structures present in immunoglobulin structures determined accurately at high resolution. There are additional canonical structures known only from immunoglobulins, the structures of which have been determined at medium or low resolution and these are listed in Table 7. There are two cases where these canonical structures are known from two or three different immunoglobulins. In each the observed conformations are similar, but not the same to within the range of deviations observed in canonical structures determined at high resolution. The differences involve the orientation of just one peptide in κL1, canonical structure 5 and the conformation of two residues as in H1, canonical structure 3. All these differences involve structures determined at resolutions between 2.7 and 3.0 Å. As discussed below the refinement of such structures at high resolution commonly produces revisions in the conformation of hypervariable regions. This suggests that to be confident that the differences in conformation are true and that the two canonical structures have more than one form requires refinement at high resolution.

The Extent of the Canonical Structure Repertoire in Known Structures

Martin & Thornton (1996) have made a detailed and extensive comparison of the conformations of all the L1, L2, L3, H1 and H2 hypervariable regions in 49 immunoglobulin structures. In all 244 loops were considered. (In one immunoglobulin the L2 region is deleted). They assigned an expected canonical structure class to 217 hypervariable regions. We have examined the 27 regions that were unassigned. Of these, 15 are in V_κ domains, six in V_λ domains and six in V_H domains. On the basis of the key residue descriptions given by Tomlinson *et al.* (1995) and Chothia *et al.* (1992), we can assign an expected canonical structure class to all 15 of the V_κ hypervariable regions and to five of the six V_H hypervariable regions. This means that on the basis of published work we can assign an expected canonical structure of 237 hypervariable regions.

Table 7. Canonical structures known from immunoglobulin structures determined at medium or low resolution

Domain	Hypervariable region	Canonical structure	Immunoglobulins	Resolution (Å)	Reference
V_k	L1	5	40–50 50.1 59.1	2.7 2.8 3.0	a,b,c
	L1	6	1F7	3.0	d
	L3	3	HyHEL5	2.7	e
	L3	4	3D6	2.7	f
	L3	5	ANO2	2.9	g
V_H	H1	2	ANO2	2.9	g
	H1	3	50.1 59.1	2.8 3.0	b,c

a, Jeffrey *et al.* (1993); b, Rini *et al.* (1993); c, Ghiara *et al.* (1994); d, Haynes *et al.* (1994); e, Sheriff *et al.* (1987); f, He *et al.* (1992); g, Brünger *et al.* (1991).

Martin & Thornton (1996) placed the hypervariable regions in clusters on the basis of their structural features. They identified which clusters correspond to the known canonical structures. Their results show that 219 of the 237 hypervariable regions are placed in a cluster that corresponds to their expected canonical structure. The other 18 hypervariable regions are placed in a cluster different from that which was expected. These are listed in Table 8 together with some details of the structures from which they come.

Of these 18 hypervariable regions, 11 come from structures determined at resolutions between 2.7 and 3.1 Å. At this resolution it can be difficult to distinguish carbonyl oxygen atoms from main-chain atoms. This means that when immunoglobulin structures determined in this range of resolution are refined at high resolution it is common for the conformation of some of their hypervariable regions to be revised. For example, the refinement of J539 resulted in the revision of the conformation of H1, H2 and H3 (Suh *et al.*, 1986; Bhat *et al.*, 1990); the refinement of D1.3 resulted in the revision of the conformation of L3 and H1 (Fischmann *et al.*, 1991); the high resolution structure of the V_κ dimer of McPC603 showed that the conformations of L1 and L2 in the Fab structure need to be revised (Steipe *et al.*, 1992). In all these cases, these revisions produced conformations for the hypervariable regions that correspond to those expected from canonical structure predictions. This suggests that conformations that have been determined at resolutions of 2.7 Å or greater and which do not match well the expected canonical structures should be regarded as possible, but uncertain, alternatives that need to be confirmed at high resolution.

This view is supported by the refined structure of immunoglobulin 4-4-20. Martin & Thornton (1996) used in their work the structure determined at 2.7 Å (Herron *et al.*, 1989) and found conformations of L1 and H2 that are not those expected from the predicted canonical structure (Table 8). The structure refined at a resolution of 1.85 Å (Whitlow *et al.*, 1995) revised the conformation of L1 and it now corresponds to the expected conformation (see above).

The H2 conformation in the refined structure of 4-4-20 and the conformations of L3 and H2 in CHA255 were discussed above in the sections on these canonical structures.

The sequences of V_κ domains in BV04-01 and 4-4-20 have only five differences and all are at surface sites. We might expect, therefore, that the conformations of the two domains should be the same. The conformation of L1 in BV04-01 is close to that of the medium resolution structure of 4-4-20 (PDB file 4fab; Herron *et al.*, 1989). But it is different from that in the high resolution structure (PDB file 1flr; Whitlow *et al.*, 1995). Because the structure of BV04-01 was solved by molecular replacement using the medium resolution 4-4-20 structure (BV04-01; Herron *et al.*, 1991), it is probable that further refinement of BV04-01 will also produce a standard L1 conformation.

The L2 region in the structure of 36-71 does not have a standard canonical structure conformation (Table 8). Residues 50 and 51 have partially allowed main-chain torsion angles $(+\phi, +\psi)$ and residues 52 and 53 have non-allowed torsion angles $(+\phi, -\psi)$. Such a large amount of steric strain in a surface loop is implausible and clearly indicates that the structure of 36-71 is in need of further refinement.

This view is supported by the medium resolution structure of the immunoglobulin R19.9 (Lascomb *et al.*, 1990). Except for the H3 region, 36-71 and R19.9 have a 90% residue identity. They also bind the same antigen. Except for a conservative change, Ile to Val at position 34 in V_H, all the sequence differences in 36-71 and R19.9 are at surface sites and we would expect them to have the same canonical structures. The conformations observed for the hypervariable regions in R19.9 differ from those in 36-71 and have the expected canonical structure (Lascombe *et al.*, 1990; Martin & Thornton, 1996).

Table 8. Hypervariable regions in known Fab and Fv structures that do not have the standard conformation of their canonical structure

Immunoglobulin	PDB file	Resolution of structure determination	Regions with unusual conformations	Reference
B72.3	1bbj	3.1	H2	a
1F7	1fig	3.0	L3 H1 H2	b
59.1	1acy	3.0	L1	c
JE142	1jel	2.8	L1	d
R6.5	1rmf	2.8	H2	e
26–10	1igi	2.7	H1	f
4–4–20	4fab	2.7	L1 H2	g
40–50	1ibg	2.7	L1	h
CHA255	1ind	2.2	L3 H2	i
BV04–01	1nbv	2.0	L1 H1 H2	j
36–71	6fab	1.9	L2 H2	k

a, Brady *et al.* (1992); b, Haynes *et al.* (1994); c, Ghiara *et al.* (1994); d, Prasad *et al.* (1993); e, Jedrazejas *et al.* (1995); f, Jeffrey *et al.* (1993); g, Herron *et al.* (1989); h, Jeffrey *et al.* (1995); i, Love *et al.* (1993); j, Herron *et al.* (1991); k, Strong *et al.* (1991).

Conclusions

Here we have made a comparative analysis of the conformations of the L1, L2, L3, H1 and H2 hypervariable regions in 17 immunoglobulin structures: some 79 loops in all. We also analysed part of the H3 region in 12 of the 15 V_H domains considered here.

On the basis of the residues at key sites the 79 hypervariable regions can be assigned to one of 18 different canonical structures. We have shown that 71 of these hypervariable regions have the standard conformation of their canonical structure. Of the remaining eight, six differ in the orientations of a single peptide group, which in some cases can be ascribed to the presence of a particular residue and two in H2 involve non-standard conformations that are allowed by the presence of three glycine residues in a surface loop.

We have drawn attention to the description by Braden *et al*. (1996) of the switch in peptide orientation produced in D1.3 by the antigen. It is also quite possible that new alternative forms may be found for the 18 canonical structures considered here and for those canonical structures for which there is less structural data at present. Nevertheless, our results demonstrate that the large majority of hypervariable regions have a local conformation very close to the standard conformation of one of the canonical structures and most deviations, when they do occur, are small.

Elsewhere, rules have been described for the relations between the sequence and conformation of the stem region of the H3 hypervariable regions (Shirai *et al*., 1996; Morea *et al*., 1997). This work, together with the work described here, suggests that except for the tip of the H3 loop we can predict from sequence with great accuracy the local conformation of the hypervariable regions in most immunoglobulins.

Acknowledgements

We are most grateful for the comments and correction made on an earlier version of this paper by two referees.

References

Bajorath, J., Harris, L. & Novotny, J. (1995). Conformational similarity and systematic displacement of complementarity determining regions loops in high resolution antibody X-ray structures. *J. Biol. Chem.* **270**, 22081–22084.

Bhat, T. N., Padlan, E. A. & Davies, D. R. (1990). Refined crystal structure of the galactan binding immunoglobulin Fab J539 at 1.95 Å resolution. *PDB file:2fbj.*

Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F. P., Mariuzza, R. A. & Poljak, R. J. (1994). Bound water molecule and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl Acad. Sci. USA*, **91**, 1089–1093.

Bizebard, T., Daniels, R., Kahn, R., Golinellipimpaneau, B., Skehel, J. J. & Knossow, M. (1994). Refined 3-dimensional structure of the Fab fragment of a murine IGG1, lambda antibody. *Acta Crystallog. sect. D*, **50**, 768–777.

Braden, B. C., Fields, B. A., Ysern, X., Goldbaum, F. A., Dall'Acqua, W., Schwarz, F. P., Poljak, R. J. & Mariuzza, R. A. (1996). Crystal structure of the complex of the variable domain of antibody D1.3 and turkey egg white lysozyme: A novel conformational change in antibody CDR-L3 selects for antigen. *J. Mol. Biol.* **257**, 889–894.

Brady, R. L., Edwards, D. J., Hubbard, R. E., Jiang, J. S., Lange, G., Roberts, S. M., Todd, R. J., Adair, J. R., Emtage, J. S., King, D. J. & Low, D. C. (1992). Crystal structure of a chimeric Fab' fragment of an antibody binding tumor cells. *J. Mol. Biol.* **227**, 253–264.

Brünger, A. T., Leahy, D. J., Hynes, T. R. & Fox, R. O. (1991). 2.9 Å resolution structure of an anti-dinitrophenyl spin label monoclonal antibody Fab fragment with bound hapten. *J. Mol. Biol.* **221**, 239–256.

Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917.

Chothia, C., Lesk, A. M., Levitt, M., Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. (1986). The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. *Science*, **233**, 755–758.

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

Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). Structural repertoire of the human VH segments. *J. Mol. Biol.* **227**, 799–817.

Cohen, G. H., Sheriff, S. & Davies, D. R. (1995). The refined structure of the monoclonal antibody HY/HEL-5 with its antigen hen egg white lysozyme. *PDB file: 3hfl.*

Efimov, A. V. (1993). Patterns of loop regions in proteins. *Curr. Opin. Struct. Biol.* **3**, 379–384.

Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A. A. (1993). X-ray structure of the antigen-binding domain from three variants of humanised anti-p185^{HER2} antibody 4D5 and comparison with molecular modeling. *J. Mol. Biol.* **229**, 969–995.

Eigenbrot, C., Gonzalez, T., Mayeda, J., Carter, P., Werther, W., Hotaling, T., Fox, J. & Kessler, J. (1994). X-ray structure of fragments from binding and nonbinding versions of a humanised anti-CD18 antibody: structural indications of key role of V_H residues 59 to 65. *Proteins: Struct. Funct. Genet.* **18**, 49–62.

Fan, Z., Shan, L., Guddat, L. W., He, X., Gray., W. R., Raison, R. L. & Edmundson, A. B. (1992). Three-dimensional structure of an Fv from a human IgM immunoglobulin. *J. Mol. Biol.* **228**, 188–207.

Fischmann, T. O., Bentley, G. A., Bhat, T. N., Boulot, G., Mariuzza, R. A., Phillips, S. E. V., Tello, D. & Poljak, R. J. (1991). Crystallographic refinement of the 3-dimensional structure of the Fab D1.3-lysozyme complex at 2.5 Å resolution. *J. Biol. Chem.* **266**, 12915–12920.

Furey, W., Wang, B. C., Yoo, C. S. & Sax, M. (1983). Structure of a novel Bence-Jones protein (RHE) fragment at 1.6 Å resolution. *J. Mol. Biol.* **167**, 661–668.

Ghiara, J. B., Stura, E. A., Stanfield, R. L., Profy, A. T. & Wilson, I. A. (1994). Crystal structure of the principal neutralisation site of HIV-1. *Science,* **264**, 82–85.

Guarné, A., Bravo, J., Calvo, J., Lozano, F., Vives, J. & Fita, I. (1996). Conformation of the hypervariable region L3 without the key proline residue. *Protein Sci.* **5**, 167–169.

Guddat, L. W., Shan, L., Anchin, J. M., Linthicum, D. S. & Edmundson, A. B. (1994). Local and transmitted conformational changes on complexion of an anti-sweetener Fab. *J. Mol. Biol.* **236**, 247–274.

Haynes, M. R., Stura, E. A., Hilvert, D. & Wilson, I. A. (1994). Routes to catalysis: structure of a catalytic antibody and comparison with its natural counterpart. *Science,* **263**, 646–652.

He, X. M., Ruker, F., Casale, E. & Carter, D. C. (1992). Structure of human monoclonal antibody Fab fragment against GP41 of HIV-1. *Proc. Natl Acad. Sci. USA,* **89**, 7154–7158.

Herron, J. N., He, X. M., Mason, M. L., Voss, E. W. & Edmundson, A. B. (1989). Three-dimensional structure of a fluorescein Fab complex crystallised in 2-methyl-2,4-pentanediol. *Proteins: Struct. Funct. Genet.* **5**, 271.

Herron, J. N., He, X. M., Ballard, D. W., Bleir, P. R., Pace, P. E., Bothwell, A. L. M., Voss, E. W. & Edmundson, A. B. (1991). An autoantibody to single stranded DNA- Comparison of the 3-dimensional structures of the unliganded Fab and a deoxynucleotide Fab complex. *Proteins: Struct. Funct. Genet.* **11**, 159–175.

Jedrazejas, M. J., Miglietta, J., Griffin, J. A. & Luo, M. (1995). Structure of a monoclonal anti-ICAM-1 antibody R6.5 Fab fragment at 2.8 Å resolution. *Acta Crystallog. sect. D,* **51**, 380–385.

Jeffrey, P. D., Strong, R. K., Sieker, L. C., Chang, C. Y. Y., Campbell, R. L., Petsko, G. A., Haber, E., Margolies, M. N. & Sheriff, S. (1993). 26-10 Fab-digoxin complex: Affinity and specificity due to surface complementarity. *Proc. Natl Acad. Sci. USA,* **90**, 10310–10314.

Jeffrey, P. D., Schildbach, J. F., Chang, C. Y., Kussie, P. H., Margolies, M. N. & Sheriff, S. (1995). Structure and specificity of the anti-digoxin antibody 40–50. *J. Mol. Biol.* **248**, 344–360.

Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979). *Sequences of Immunoglobulin Chains,* National Institute of Health NIH publication 80–2008.

Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest.* 1, U.S. Department of Health and Human Services.

Lascombe, M. B., Alzari, P. M., Boulot, G., Saludjian, P., Tougard, P., Berek, C., Haba, S., Rosen, E. M., Nisonoff, A. & Poljak, R. J. (1990). 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.

Love, R. A., Villafranca, J. E., Aust, R. M., Nakamura, K. K., Jue, R. A., Major, J. G., Jr, Radhakrishnan, R. & Butler, W. F. (1993). How the anti-(metal chelate) antibody CHA255 is specific for the metal ion of its antigen: X-ray structure for two Fab'/hapten complexes with different metals in the chelate. *Biochemistry,* **32**, 10950–10959.

Marquart, M. & Huber, R. (1989). *PDB file 2fb4.*

Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. (1980). 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.

Martin, A. C. R. & Thornton, J. M. (1996). Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies. *J. Mol. Biol.* **263**, 800–815.

Milner-White, E. J., Ross, B. M., Ismail, R., Balhadjmostefa, K. & Poet, R. (1988). One type of gamma-turn rather than the other gives rise to chain reversal in proteins. *J. Mol. Biol.* **204**, 777–782.

Morea, V., Tramontano, A., Rustici, M., Chothia, C. & Lesk, A. M. (1997). Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J. Mol. Biol.* (in press).

Prasad, L., Sharma, S., Vandonselaar, M., Quail, J. W., Lee, J. S., Waygood, E. B., Wilson, K. S., Dauter, Z. & Delbaere, L. T. J. (1993). Evaluation of mutagenesis for epitope mapping: structure of an antibody-protein antigen complex. *J. Biol. Chem.* **268**, 10705–10708.

Rini, J. M., Schulze-Gahmen, U. & Wilson, I. A. (1992). Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science,* **255**, 959–965.

Rini, J. M., Stanfield, R. L., Stura, E. A., Salinas, P. A., Profy, A. T. & Wilson, I. A. (1993). Crystal structure of a human immunodeficiency virus type-1 neutralising antibody, 50.1, in complex with its V3 loop peptide antigen. *Proc. Natl Acad. Sci. USA,* **90**, 6325–6329.

Rose, D. R., Przbylska, M., To, R. J., Keyden, C. S., Oomen, R. P., Vorberg, E., Young, N. M. & Bundle, D. R. (1993). Crystal structure to 2.45 Å resolution of a monoclonal Fab specific for the brucella-A cell wall polysaccharide. *Protein Sci.* **2**, 1106–1113.

Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). Phosphocholine binding immunoglubulin Fab McPC603: An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593–604.

Saul, F. A. & Poljak, R. J. (1992). Crystal structure of human immunoglobulin fragment Fab NEW refined at 2.0 Å resolution. *Proteins: Struct. Funct. Genet.* **14**, 363–371.

Saul, F. A. & Poljak, R. J. (1993). Structural patterns at residue position 9, position 18, position 86 and position 82 in the V_H framework regions of human and murine immunoglobulins. *J. Mol. Biol.* **230**, 15–20.

Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smithgill, S. J., Finzel, B. C. & Davies, D. R. (1987). 3-dimensional structure of an antibody-antigen complex. *Proc. Natl Acad. Sci. USA,* **84**, 8075–8079.

Shirai, H., Kidera, A. & Nakamura, H. (1996). Structural classification of CDR-H3 in antibodies. *FEBS Letters,* **399**, 1–8.

Shoham, M. (1993). Crystal structure of an anticholera toxin peptide complex at 2.3 Å. *J. Mol. Biol.* **232**, 1169–1175.

Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989). Conformations of β-hairpins in protein structures. *J. Mol. Biol.* **206**, 759–777.

Steipe, B., Plückthun, A. & Huber, R. (1992). Refined crystal structure of a recombinant immunoglobulin domain and a complementarity-determining region 1-grafted mutant. *J. Mol. Biol.* **225**, 739–753.

Strong, R. K., Campbell, R., Rose, D. R., Petsko, G. A., Sharon, J. & Margolies, M. N. (1991). Three-dimensional structure of murine anti-para-azophenylarsonate Fab 36-71.1. X-ray crystallography, site directed mutagenesis, and modelling of the complex with hapten. *Biochemistry,* **30**, 3739–3748.

Suh, S. W., Bhat, T. N., Navia, M. A., Cohen, G. H., Rao, D. N., Rudikoff, S. & Davies, D. R. (1986). The galactan-binding immunoglobulin Fab J539. An X-ray diffraction study at 2.6 Å resolution. *Proteins: Struct. Funct. Genet.* **1**, 74–80.

Tomlinson, I. M., Cox, J. P. L., Gherardi, E., Lesk, A. M. & Chothia, C. (1995). The structural repertoire of the human V kappa domain. *EMBO J.* **14**, 4628–4638.

Tramontano, A., Chothia, C. & Lesk, A. M. (1990). Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J. Mol. Biol.* **215**, 175–182.

Whitlow, M., Howard, A. J., Wood, J. F., Vos, E. W., Jr & Hardman, K. D. (1995). 1.85 Å structure of anti-fluorescein 4-4-20 Fab. *Protein Eng.* **8**, 749–761.

Wu, S. & Cygler, M. (1993). Conformation of complementarity determining region L1 loop in murine IgG λ light chain extended the repertoire of canonical forms. *J. Mol. Biol.* **229**, 597–601.

Wu, T. T. & Kabat, E. A. (1970). 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.

Zdanov, A., Li, Y., Bundle, D. R., Deng, S., MacKenzie, C. R., Narang, S. A., Young, N. M. & Cygler, M. (1994). Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7 Å resolution. *Proc. Natl Acad. Sci. USA,* **91**, 6423–6427.

Zhou, G. W., Guo, J. C., Huang, W., Fletterick, R. J. & Scanlan, T. S. (1994). Crystal structure of a catalytic antibody with a serine protease active site. *Science,* **265**, 1059–1064.

Edited by I. A. Wilson

(Received 24 March 1997; received in revised form 30 June 1997; accepted 6 August 1997)