The EMBO Journal vol.14 no.18 pp.4628–4638, 1995

The structural repertoire of the human V_κ domain

Ian M.Tomlinson[1], Jonathan P.L.Cox[2], Ermanno Gherardi[3], Arthur M.Lesk[4] and Cyrus Chothia[1,5]

[1]MRC Centre for Protein Engineering, [5]MRC Laboratory of Molecular Biology, [3]ICRF Cell Interactions Laboratory, [4]Department of Haematology, University of Cambridge, Hills Road, Cambridge CB2 2QH and [2]School of Chemistry, University of Bath, Claverton Down, Bath, Avon BA2 7AY, UK

In humans, the gene for the V_κ domain is produced by the recombination of one of 40 functional V_κ segments and one of five functional J_κ segments. We have analysed the sequences of these germline segments and of 736 rearranged V_κ genes to determine the repertoire of main chain conformations, or canonical structures, they encode. Over 96% of the sequences correspond to one of four canonical structures for the first antigen binding loop (L1) and one canonical structure for the second antigen binding loop (L2). Junctional diversity produces some variation in the length of the third antigen binding loop (L3) and in the identity of residues at the V_κ–J_κ join. However, this is limited and 70% of the rearranged sequences correspond to one of three known canonical structures for the L3 region. Furthermore, we show that the canonical structures selected during the primary response are conserved during affinity maturation: the key residues that determine the conformations of the antigen binding loops are unmutated or undergo conservative mutation. The implications of these results for immune recognition are discussed.

Keywords: antibodies/canonical structures/human/kappa/V_κ

Introduction

The antigen binding site of an antibody is formed by six loops of polypeptide: three from the light chain variable domain (V_L) and three from the heavy chain variable domain (V_H) (see Figure 1). A diverse repertoire of V genes that encode the V_L and V_H domains is produced by the combinatorial rearrangement of gene segments (Tonegawa, 1983), that are drawn from pools of moderate size. The V_L gene is produced by the recombination of two gene segments, V_L and J_L. The V_L segment codes for residues 1–95 which include the regions that form the first and second antigen binding loops. The third antigen binding loop is formed by the end of the V_L segment and the beginning of the J_L segment. Antibodies encoded by these rearranged genes are believed to be sufficiently diverse to recognize almost all antigens with at least moderate affinity. High-affinity antibodies are produced by somatic mutation of the rearranged genes, followed by selection for improved binding (Berek and Milstein, 1987).

V_L domains are divided into two classes: kappa (κ) and lambda (λ). Work over the past 15 years has resulted in what is probably a complete description of the sequences that encode the human V_κ domain [see Zachau (1993) for a review]. The V_κ segments are arranged in two separate regions on chromosome 2, named after their positions relative to the five J_κ segments. The J_κ-distal region, that contains 36 V_κ segments, is a partial duplicate of the larger J_κ-proximal region, that contains 40 V_κ segments.

Analysis of the structures and sequences of antibodies from humans and mice has shown that the three antigen binding loops of the V_κ domain have a small repertoire of main chain conformations, or canonical structures, that is largely shared by both species (Chothia and Lesk, 1987; Chothia *et al.*, 1989; Brünger *et al.*, 1991; He *et al.*, 1992). The canonical structures are determined by (i) the length of the antigen binding loop and (ii) particular residues, or types of residues, at key sites in the loop itself and in the antibody framework. These sequence features have been described in detail (see the references above) and can be used to predict the canonical structures of the three antigen binding loops of the V_κ domain.

In this paper, we describe the canonical structures implicit in human V_κ sequences. To describe the structural repertoire available to V_κ domains in the primary repertoire, the sequences of the germline V_κ segments are first examined. We go on to look at 736 rearranged V_κ

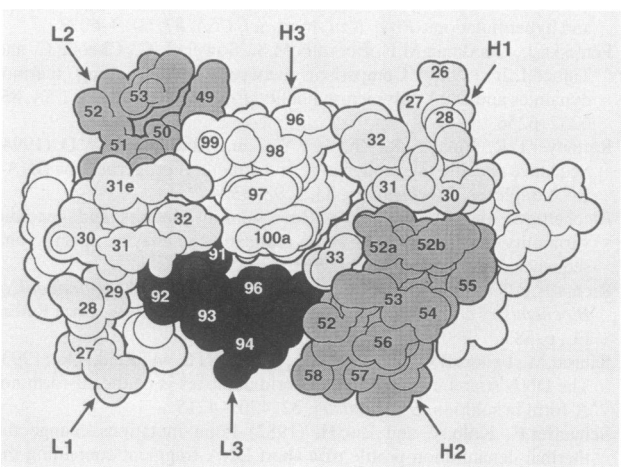

Fig. 1. The antigen binding site of an antibody. A space-filling model of the McPC603 antigen binding site depicting the superposition of five sections cut at 2 Å intervals is shown (Segal *et al.*, 1974). Residues 31a, 31b, 31c, 31d of the V_L domain and residue 52c of the V_H domain are above the plane of the paper. The structural repertoire available to the H1 and H2 regions of the human V_H domain has been previously described (Chothia *et al.*, 1992). Here we discuss the repertoire of main chain conformations available to the L1, L2 and L3 regions of the human V_κ domain.

Table I. Genetic and structural defects in the human germline V_κ segments[a]

V_κ segment	Genetic defects	Structural defects[b]	
O14/O4		88G (C)	Loss of disulphide bridge and creation of an internal cavity
A4	Frame shift in V_κ exon (insertion of one nucleotide at residue 57)		
A21	Frame shift in V_κ exon (insertion of one nucleotide at residue 39)		
A7		23F (C)	Loss of disulphide bridge and small residue replaced by a larger one at an internal site
A29	No start codon at beginning of leader sequence (ATG replaced by ATA)		
A13	No start codon at beginning of leader sequence (ATG replaced by ATA)	37P (L/Q)	Loss of β-sheet hydrogen bonds by P
L22		5I (T) 22I (N/S/T) 24W (R) 56P (S/T) 76I (N/S)	A large number of mutations, at sites on the surface of the V_L domain, replace hydrophilic residues by hydrophobic residues
L10	Defective acceptor splice site in leader intron (AG replaced by GG)		
L13	No start codon at beginning of leader sequence (ATG replaced by GTG)	75F (I)	Increase in size of an absolutely conserved buried residue
		87C (F/Y)	Large reduction in size of a residue buried in V_L–V_H interface
B1	No start codon at beginning of leader sequence (ATG replaced by ATA)		

[a]Genetic and structural defects in the 51 V_κ segments which were described as 'potentially functional' or having 'minor defects' (Schäble and Zachau, 1993). References: O14/O4, Pargent *et al.* (1991); A4, Jaenichen *et al.* (1984); A21, Schäble *et al.* (1994); A7, A29, A13, Lautner-Rieske *et al.* (1992); L22, Pech *et al.* (1985); L10, Pech and Zachau (1994); L13, Huber *et al.* (1993); B1, Lorenz *et al.* (1988). For defects in the 25 'pseudogenes', see Schäble *et al.* (1994).
[b]Normal residues are given in parentheses.

sequences to determine how the assembly of V_κ and J_κ segments influences the canonical structure of the third antigen binding loop. In addition, the extent to which canonical structures are conserved during affinity maturation is determined. The implications of our results for understanding the mechanisms of immune recognition are then discussed.

The repertoire of human germline V_κ segments

For a germline V_κ segment to be functional, it must have (i) intact coding and regulatory sequences, (ii) functional recombination signal sequences (RSS) and (iii) encode a protein that can form a stable three-dimensional structure. On inspection of the sequences of the 76 V_κ segments, Schäble and Zachau (1993) described 25 as pseudogenes (see Schäble *et al.*, 1994).

Genetic and structural defects

The remaining 51 V_κ segments were described as being 'potentially functional' or having 'minor defects' (Schäble and Zachau, 1993). Segments with minor defects are those with one or two point mutations in the coding or regulatory regions that might render the segment non-functional, but which may exist as functional alleles in certain individuals (Schäble and Zachau, 1993). To examine the expression of each V_κ segment and its alleles, we compiled a large database of rearranged V_κ genes, derived from many individuals, containing 736 sequences (I.M.Tomlinson and J.P.L.Cox, in preparation), which extends previous studies of the expressed V_κ repertoire (Klein *et al.*, 1993; Cox *et al.*, 1994).

Eleven of the V_κ segments described as 'potentially functional' or having 'minor defects' have serious genetic and/or structural defects, such as the absence of start codons or mutations which are likely to seriously destabilize the three-dimensional structure of the protein (Table I); these segments are not expressed and are therefore classified as non-functional.

Functional V_κ segments

Thirty-eight of the remaining 40 V_κ segments are seen rearranged *in vivo*. Since the other two segments (A14 and L25) have no obvious defects which would prevent their functional expression, our failure to see them rearranged *in vivo* is probably due to them being located in the rarely used J_κ-distal portion of the V_κ locus (Cox *et al.*, 1994). The amino acid sequences of the 40 functional V_κ segments and their alleles are given in Table II [see Schäble and Zachau (1993) and Cox *et al.* (1994) for nucleotide sequences]. The functional segments are derived from six families: 19 V_κI, nine V_κII, seven V_κIII, one V_κIV, one V_κV and three V_κVI. Fifteen pairs of functional segments differ by <11 amino acids, some even encode identical gene products (Table II). Known alleles of the functional segments can alter the sequence of the V_κ exon, but only by up to three amino acids (see footnote to Table II). The introduction of stop codons in the V_κ exons of L4, L16 and A18, in the leader exon of O12, or the introduction of structural defects in O11, means that there may be as few as 35 functional segments in some individuals. The segments listed in Table II should therefore be regarded as the maximum number of functional V_κ segments in any one individual.

The use of these segments in the expressed repertoire

Table II. Amino acid translations of the human germline V_κ segments

	1 2 3 4 5 6 7 8 9 0 0 0 abcdef 0 0 0 0 0 0	CANONICAL STRUCTURE L1 L2 L3	SEQUENCE REFERENCES	
SUBGROUP I				
O12*/O2	DIQMTQSPSSLSASVGDRVTITCRASQSISS------YLNWYQQKPGKAPKLLIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQSYSTP	2 1 (1)	O12/O2	a/a
O18/O8	DIQMTQSPSSLSASVGDRVTITCQASQDISN------YLNWYQQKPGKAPKLLIYDASNLETGVPSRFSGSGSGTDFTFTISSLQPEDIATYYCQQYDNLP	2 1 (1)	O18/O8	b/b
A20	DIQMTQSPSSLSASVGDRVTITCRASQGISN------YLAWYQQKPGKVPKLLIYAASTLQSGVPSRFSGSGSGTDFTLTISSLQPEDVATYYCQKYNSAP	2 1 (X)	A20	c
A30	DIQMTQSPSSLSASVGDRVTITCRASQGIRN------DLGWYQQKPGKAPKRLIYAASSLQSGVPSRFSGSGSGTEFTLTISSLQPEDFATYYCLQHNSYP	2 1 (1)	A30	d
L14	NIQMTQSPSAMSASVGDRVTITCRARQGISN------YLAWFQQKPGKVPKHLIYAASSLQSGVPSRFSGSGSGTEFTLTISSLQPEDFATYYCLQHNSYP	2 1 (1)	L14	d
L1*	DIQMTQSPSSLSASVGDRVTITCRASQGISN------YLAWFQQKPGKAPKSLIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQYNSYP	2 1 (1)	L1	e
L15*	DIQMTQSPSSLSASVGDRVTITCRASQGISS------WLAWYQQKPEKAPKSLIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQYNSYP	2 1 (1)	L15	e
L4/18a*	AIQLTQSPSSLSASVGDRVTITCRASQGISS------ALAWYQQKPGKAPKLLIYDASSLESGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQFNSYP	2 1 (1)	L4/18a	f
L5/L19	DIQMTQSPSSVSASVGDRVTITCRASQGISS------WLAWYQQKPGKAPKLLIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQANSFP	2 1 (1)	L5/L19	f/g
L8	DIQLTQSPSFLSASVGDRVTITCRASQGISS------YLAWYQQKPGKAPKLLIYAASTLQSGVPSRFSGSGSGTEFTLTISSLQPEDFATYYCQQLNSYP	2 1 (1)	L8	f
L23	AIRMTQSPFSLSASVGDRVTITCWASQGISS------YLAWYQQKPAKAPKLFIYYASSLQSGVPSRFSGSGSGTDYTLTISSLQPEDFATYYCQQYYSTP	2 1 (1)	L23	d
L9	AIRMTQSPSSFSASTGDRVTITCRASQGISS------YLAWYQQKPGKAPKLLIYAASTLQSGVPSRFSGSGSGTDFTLTISCLQSEDFATYYCQQYYSYP	2 1 (1)	L9	f
L24	VIWMTQSPSLLSASTGDRVTISCRMSQGISS------YLAWYQQKPGKAPELLIYAASTLQSGVPSRFSGSGSGTDFTLTISCLQSEDFATYYCQQYYSFP	X 1 (1)	L24	h
L11	AIQMTQSPSSLSASVGDRVTITCRASQGIRN------DLGWYQQKPGKAPKLLIYAASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCLQDYNYP	2 1 (1)	L11	b
L12*	DIQMTQSPSTLSASVGDRVTITCRASQSISS------WLAWYQQKPGKAPKLLIYDASSLESGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCQQYNSYS	2 1 (X)	L12	i
SUBGROUP II				
O11*/O1	DIVMTQTPLSLPVTPGEPASISCRSSQSLLDSDDGNTYLDWYLQKPGQSPQLLIYTLSYRASGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQRIEFP	3 1 (1)	O11/O1	a/a
A17	DVVMTQSPLSLPVTLGQPASISCRSSQSLVYS-DGNTYLNWFQQRPGQSPRRLIYKVSNRDSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGTHWP	4 1 (1)	A17	c
A1	DVVMTQSPLSLPVTLGQPASISCRSSQSLVYS-DGNTYLNWFQQRPGQSPRRLIYKVSNWDSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGTHWP	4 1 (1)	A1	c
A18*	DIVMTQTPLSLSVTPGQPASISCKSSQSLLHS-DGKTYLYWYLQKPGQSPQLLIYEVSSRFSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGIHLP	4 1 (1)	A18	c
A2	DIVMTQTPLSLSVTPGQPASISCKSSQSLLHS-DGKTYLYWYLQKPGQPPQLLIYEVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQSIQLP	4 1 (1)	A2	j
A19/A3	DIVMTQSPLSLPVTPGEPASISCRSSQSLLHS-NGYNYLDWYLQKPGQSPQLLIYLGSNRASGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQALQTP	4 1 (1)	A19/A3	c/k
A23	DIVMTQTPLSSPVTLGQPASISCRSSQSLVHS-DGNTYLSWLQQRPGQPPRLLIYKISNRFSGVPDRFSGSGAGTDFTLKISRVEAEDVGVYYCMQATQFP	4 1 (1)	A23	k
SUBGROUP III				
A27	EIVLTQSPGTLSLSPGERATLSCRASQSVSSS-----YLAWYQQKPGQAPRLLIYGASSRATGIPDRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSP	6 1 (1)	A27	l
A11	EIVLTQSPATLSLSPGERATLSCGASQSVSSS-----YLAWYQQKPGLAPRLLIYDASSRATGIPDRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSP	6 1 (1)	A11	m
L2/L16*	EIVMTQSPATLSVSPGERATLSCRASQSVSS------NLAWYQQKPGQAPRLLIYGASTRATGIPARFSGSGSGTEFTLTISSLQSEDFAVYYCQQYNNWP	2 1 (1)	L2/L16	n/o
L6	EIVLTQSPATLSLSPGERATLSCRASQSVSS------YLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSGTDFTLTISSLEPEDFAVYYCQQRSNWP	2 1 (1)	L6	p
L20	EIVLTQSPATLSLSPGERATLSCRASQGVSS------YLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGPGTDFTLTISSLEPEDFAVYYCQQRSNWH	2 1 (X)	L20	g
L25	EIVMTQSPATLSLSPGERATLSCRASQSVSSS-----YLSWYQQKPGQAPRLLIYGASTRATGIPARFSGSGSGTDFTLTISSLQPEDFAVYYCQQDYNLP	6 1 (1)	L25	d
SUBGROUP IV				
B3	DIVMTQSPDSLAVSLGERATINCKSSQSVLYSSNNKNYLAWYQQKPGQPPKLLIYWASTRESGVPDRFSGSGSGTDFTLTISSLQAEDVAVYYCQQYYSTP	3 1 (1)	B3	q
SUBGROUP V				
B2	ETTLTQSPAFMSATPGDKVNISCKASQDIDD------DMNWYQQKPGEAAIFIIQEATTLVPGIPPRFSGSGYGTDFTLTINNIESEDAAYYFCLQHDNFP	2 1 (1)	B2	r
SUBGROUP VI				
A26/A10	EIVLTQSPDFQSVTPKEKVTITCRASQSIGS------SLHWYQQKPDQSPKLLIKYASQSFSGVPSRFSGSGSGTDFTLTINSLEAEDAATYYCHQSSSLP	2 1 (1)	A26/A10	c/s
A14	DVVMTQSPAFLSVTPGEKVTITCQASEGIGN------YLYWYQQKPDQAPKLLIKYASQSISGVPSRFSGSGSGTDFTFTISSLEAEDAATYYCQQGNKHP	2 1 (1)	A14	s

[a]V_κ segments marked with an asterisk have alleles that alter the amino acid sequence: O11a (ref. a) (35W→C 64G→D); O12a (ref. a) (10S→F 90Q→C 91S→G); for A18, the only known allele has a stop codon at position 88 (ref. c). Two of the rearranged counterparts of A18 have a Cys at position 88. These rearranged genes are probably derived from a functional allele of A18 which has a Cys at position 88; LFVK431 is an allele of L1 (ref. t) (61R→K); L15a (ref. i) (26S→R); the allele L4/18a is functional, but has not been unambiguously mapped to L4 and/or L18 (Schäble and Zachau, 1993). Mapped alleles are L4 (ref. f) (35W→* 93S→N) and L18 (ref. g) (93S→N); L16c (ref. n) (5T→M 94W→*); L12a (ref. d) (50D→K).

First publication of full-length sequence: (a) Pargent *et al.* (1991); (b) Scott *et al.* (1991); (c) Lautner-Rieske *et al.* (1992); (d) Huber *et al.* (1993); (e) Bentley and Rabbitts (1983); (f) Pech *et al.* (1984); (g) Pech *et al.* (1985); (h) Jaenichen *et al.* (1984); (i) Bentley and Rabbitts (1980); (j) Scott *et al.* (1989); (k) Staubinger *et al.* (1988a); (l) Radoux *et al.* (1986); (m) Chen *et al.* (1986); (n) Liu *et al.* (1989); (o) Chen *et al.* (1987); (p) Pech and Zachau (1984); (q) Klobeck *et al.* (1985); (r) Stavnezer *et al.* (1985); (s) Staubinger *et al.* (1988b); (t) Foroni (1990).

is highly biased. Not only is there a strong bias towards use of V_κ segments from the J_κ-proximal region of the locus (97% of the expressed repertoire is derived from V_κ segments in the J_κ-proximal region), but there is also a bias towards the use of particular V_κ segments. For example, the single segment A27 accounts for almost 30% of the expressed repertoire, whereas the 22 least used V_κ segments account for only 6% of the expressed repertoire. Those less frequently used segments include several with defects in the promoter region (L8, L9) or heptamer RSS (A20, A26, A10, A14, L4, L18) which are likely to reduce the efficiency of transcription or rearrangement (Akamatsu *et al.*, 1994; Stiernholm and Berinstein, 1995).

Canonical structures of the first antigen binding loop of the V_κ domain

Sequence variation of residues 24–34 in the V_L domain led Kabat and Wu (1971) to predict that this region is involved in antigen binding and to call it the first complementarity determining region (CDR1). Later, an analysis of the three-dimensional structures showed that residues 24–25 and 33–34 are part of the β-sheet framework: sequence changes at these positions do not alter the main chain conformation (Chothia and Lesk, 1987). The region outside the β-sheet framework, 26–32, is called L1.

Table III.

L1 Canonical structure	Length of L1 region (26–32 inclusive) and the residues between 29 and 32
2	7: 30, 31
6	8: 30, 31, 31a
4	12: 30, 31, 31a, 31c, 31d, 31e, 31f
3	13: 30, 31, 31a, 31b, 31c, 31d, 31e, 31f

Further analysis suggested that nearly all L1 regions in V_κ domains have one of six canonical structures (Chothia and Lesk, 1987; Chothia *et al.*, 1989; unpublished data). All have residues 26, 27, 28, 29 and 32 packed against the framework in the same extended conformation (Figure 2). They differ in the number of residues between positions 29 and 32 (Figure 2). All six are found in murine

Table IV. Residues at key sites for the canonical structures of the V_κ domain

(A) L1 region

Canonical structure	Number of residues in L1 region (26–32 inclusive)	Sequence characteristics of key residues in known structures						Number of germline segments with this canonical structure
		2	25	29		33	71	
L1: 2	7	I: 18	A: 18	I: 16 V: 2		I: 1 L: 15 V: 2	F: 7 Y: 11	26
		2	25	29		33	71	
L1: 6	8	N: 1[a]	A: 1	V: 1		L: 1	Y: 1	3
		2	25	29	31d	33	71	
L1: 4	12	L: 1 V: 9	S: 9 P: 1	I: 3 L: 7	G: 10	F: 1 L: 9	F: 10	7
		2	25	29	31d	33	71	
L1: 3	13	I: 5	S: 5	L: 4 V: 1	E: 1 Q: 3 S: 1	L: 5	F: 5	3

[a]For L1 canonical structure 6, there is only a single crystal structure (the murine antibody 1F7; Haynes *et al.*, 1994). Although the Asn at position 2 in this structure (see Figure 2) is unusual (it is normally a medium-sized hydrophobic residue, either Ile, Leu or Val), one-third of murine V_κVI sequences also contain Asn at this position (Kabat *et al.*, 1991). However, in the 1F7 structure it does not seem to affect the main chain conformation of residues 26–29, which is conserved in all known L1 structures. We would expect Thr 2 in the human V_κ segment B2 to be accomodated in a similar manner.

(B) L2 region

Canonical structure	Number of residues in L2 region (50–52 inclusive)	Sequence characteristics of key residues in known structures		Number of germline segments with this canonical structure
		48	64	
L2: 1	3	I: 39 V: 2	G: 41	40

(C) L3 region

Canonical structure	Number of residues in L3 region (91–96 inclusive)	Sequence characteristics of key residues in known structures								Number of rearranged sequences with this canonical structure
		90	91	92	93	94	95	96	97	
L3: 1	6	H: 2 N: 5 Q: 30	–	–	–	–	P: 37	–	S: 1 T: 36	324
		90	91	92	93	94	95	96	97	
L3: 2	6	Q: 1	–	–	–	P: 1	–	–	T: 1	1
		90	91	92	93	94	95		97	
L3: 3	5	Q: 1	–	–	–	–	–		–	49
		90	91	92	93	94			97	
L3: 4	4	Q: 1	–	–	–	–			S: 1	1
		90	91	92	93	94	95	96 96a	97	
L3: 5	7	Q: 1	–	–	–	–	–	P: 1 –	T: 1	67

antibodies, but only four of the six are present in humans. In humans, the four structures are listed in Table III.

The canonical structure numbers are historical in origin. Mice also encode canonical structure 1, that has one residue between 29 and 32, and canonical structure 5, with six residues between 29 and 32. The non-functional human V_κ segment B1 has the sequence characteristics of canonical structure 5 (Table I).

The key residues that determine the conformation of the known L1 canonical structures for the V_κ domain are at positions 2, 25, 29, 33 and 71, which are involved in the packing of L1 against the framework (Figure 2). In canonical structures 3 and 4, residue 31d has positive ϕ,ψ values which produce some steric strain in residues other than Gly, Asn and Asp. The conformation is facilitated by the presence of these residues, but they are not essential.

There are now several crystal structures available for each of the L1 canonical structures 2, 3 and 4, and a single structure available for canonical structure 6. The residues at the key sites in these structures are given in Table IVA.

The L1 structures implicit in the germline repertoire

Thirty-nine human V_κ segments have L1 regions that correspond in size and key residues to one of four canonical structures (Tables II and IVA). The segment L24 (subgroup I) has all the characteristics of canonical structure 2, except that the key residue at position 25 is Met, rather than Ala or Ser. It is difficult to determine whether the large Met side chain at this site would produce a strained protein with canonical structure 2 or a change in conformation to form a new canonical structure.

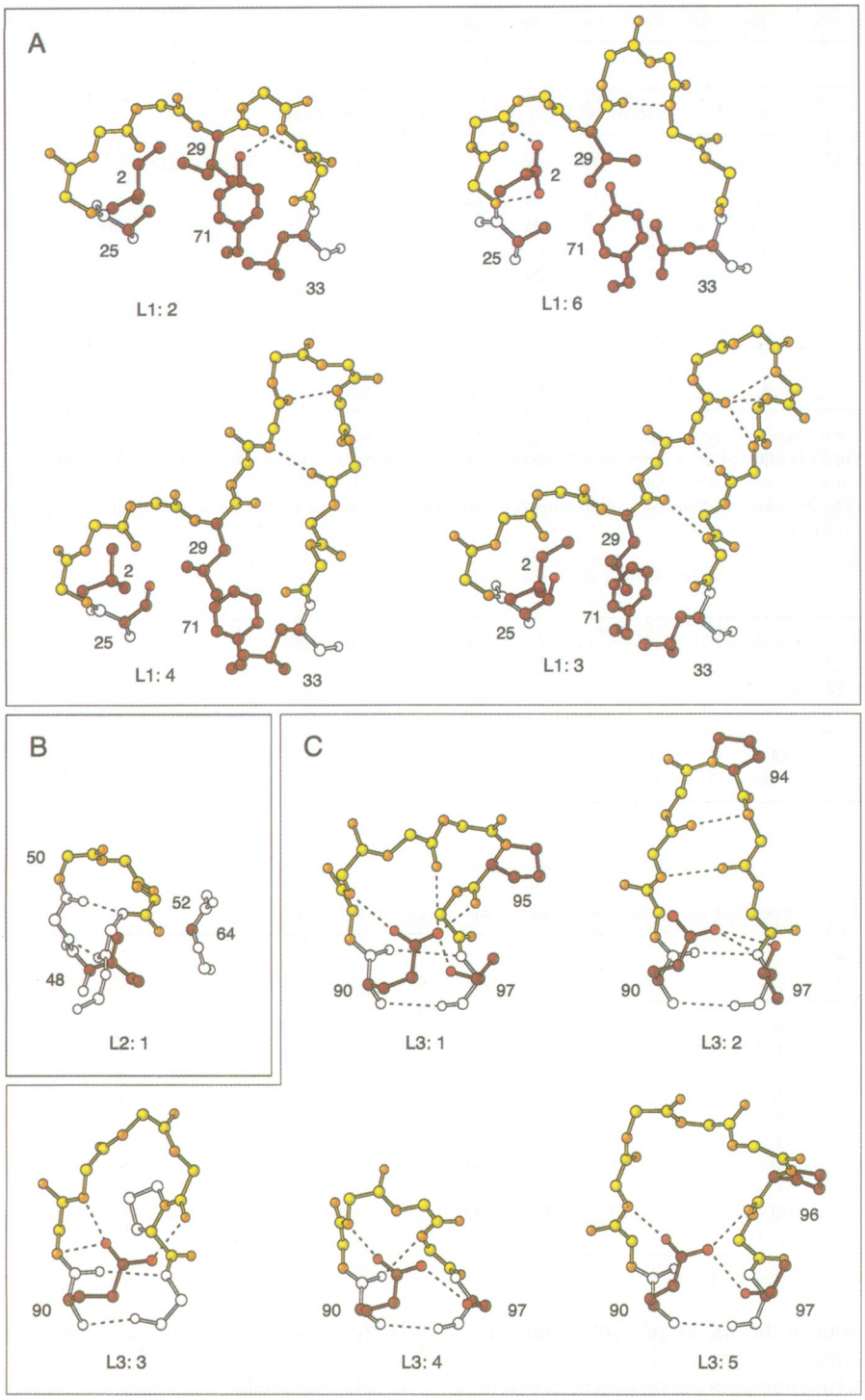

Fig. 2. Canonical structures of the antigen binding loops encoded by human V_κ domains. Each drawing is labelled La:b, where La corresponds to the antigen binding loop, L1 (**A**), L2 (**B**) or L3 (**C**), and b is the number of the canonical structure. In each case, we show the main chain of the antigen binding loops (yellow) and the Cα atom/side chain of the key residues which are the main determinants of its conformation (red). In addition, the conformations of L1: 3 and L1: 4 are facilitated by the presence of Gly, Asn and Asp at 31d which have positive ϕ,ψ values (see the text). The conformations shown here are taken from antibodies whose structures have been determined by X-ray crystallography: L1: 2 from H52 (Eigenbrot *et al.*, 1994); L1: 6 from 1F7 (Haynes *et al.*, 1994); L1: 4 from 4-4-20 (Whitlow, 1995); L1: 3 from 17/9 (Rini *et al.*, 1992); L2: 1 from KOL (Marquart *et al.*, 1980); L3: 1 from REI (Epp *et al.*, 1975); L3: 2 from J539 (Suh *et al.*, 1986); L3: 3 from HyHEL5 (Sheriff *et al.*, 1987); L3: 4 from 3D6 (He *et al.*, 1992); L3: 5 from AN02 (Brünger *et al.*, 1991).

The L1 structures implicit in the expressed repertoire

Our database of rearranged V_κ sequences has 736 entries, although some of them are incomplete. Of these, 401 (55%) are derived from germline V_κ segments which encode canonical structure 2 for the L1 region, 54 (7%) from segments which encode canonical structure 3, 65 (9%) from segments which encode canonical structure 4

Table V. Somatic mutation at the key sites for the L1 and L2 regions in the human V_κ domain

Region	Key residue	Number of sequences containing key residue	Number of sequences with no change in key residue	Conservative changes and their occurrences	Non-conservative changes and their occurrences
L1	2	333	254	I→F: 2 I→L: 67 I→T: 4 I→V: 5 V→I: 1 Total: 79[a]	
	25	715	686	A→G: 1 A→S: 9 A→T: 12 S→A: 1 S→T: 1 Total: 24	A→E: 2 A→V: 2 S→L: 1 Total: 5
	29	714	630	I→L: 1 I→M: 1 I→V: 11 V→I: 44 V→L: 18 V→M: 1 V→T: 1 Total: 77	I→K: 1 V→A: 2 V→F: 3 V→G: 1 Total: 7
	31d	115	108	N→D: 1 Total: 1	G→E: 1 G→R: 2 N→K: 2 N→S: 1 Total: 6
	33	717	701	L→F: 1 L→V: 11 Total: 12	L→S: 4 Total: 4
	71	730	723	F→Y: 3 Total: 3	F→L: 3 Y→C: 1 Total: 4
L2	48	723	709	I→L: 1 I→M: 5 I→V: 4 Total: 10	I→F: 3 I→Y:1 Total: 4
	64	722	713	G→A: 7 Total: 7	G→S: 1 G→V: 1 Total: 2

[a]Most of these changes are probably introduced by PCR primers at the beginning of framework 1.

and 216 (29%) from segments which encode canonical structure 6.

We examined their sequences to determine whether they have insertions, deletions or mutations that produce L1 conformations different to those encoded by the germline segments. None has insertions or deletions, relative to the corresponding germline segments. The number of sequences in which changes have occurred at the key sites, and the nature of these changes, are given in Table V. The changes are classified as conservative or non-conservative. Conservative changes are those that are unlikely to make any large changes in the conformation of the loop, but which may introduce small distortions or slightly shift its mean position and the envelope of alternative low-energy positions available to it. Non-conservative changes are likely to make a major alteration in the conformation or lower its stability. At four sites, 25, 31d, 33 and 71, only a few sequences contain changes. Two other sites, 2 and 29, are more frequently changed. Although the majority of changes at residue 2 are likely to have been introduced by primers based at the beginning of framework 1, those that occur at the other key residues probably represent genuine somatic mutations. The majority of these changes are conservative. Thus, the key sites for L1 are unmutated, or undergo only conservative changes, in >96% of the rearranged sequences and undergo non-conservative changes in <4% of the rearranged sequences.

Canonical structure of the second antigen binding loop of the V_κ domain

Sequence variation of residues 50–56 in V_L domains led Kabat and Wu (1971) to predict that this region is involved in antigen binding and to call it the second complementarity determining region (CDR2). Later, an analysis of the three-dimensional structures showed that residues 53–56 are part of the β-sheet framework structure: sequence changes at these positions do not alter the main chain conformation (Chothia and Lesk, 1987). The region outside the β-sheet framework, 50–52, is called L2.

Further analysis indicated that L2 regions in V_κ domains have a single canonical structure defined by there being a single residue between 50 and 52, Ile or Val at 48 and Gly at 64 (Chothia and Lesk, 1987; Steipe *et al.*, 1992) (Figure 2 and Table IVB).

The L2 structure implicit in the germline and expressed repertoires

All 40 functional germline V_κ segments have residues Ile at site 48, Gly at site 64 and one residue between residues 50 and 52 (Tables II and IVB); this means they all have the features of the single canonical structure known for the L2 region. Consequently, all 736 rearranged sequences are derived from germline V_κ segments with this canonical structure. None of these has insertions or deletions in the L2 region, relative to the corresponding germline segments, and only six have non-conservative changes at L2

Table VI.

J_κ segment	Sequence[a]
1	W T F G Q G T K V E I K R
2	Y T F G Q G T K L E I K R
3	F T F G P G T K V D I K R
4	L T F G G G T K V E I K R
5	I T F G Q G T R L E I K R

[a]First residue is position 96, last residue is position 108.

key sites (Table V). Thus, the key sites for L2 are unmutated or undergo conservative changes in >99% of the rearranged sequences and undergo non-conservative changes in <1% of the rearranged sequences.

Canonical structures of the third antigen binding loop of the V_κ domain

Sequence variation of residues 89–97 in V_L domains led Kabat and Wu (1971) to predict that this region is involved in antigen binding and to call it the third complementarity determining region (CDR3). Later, an analysis of the three-dimensional structures showed that residues 89–90 and 97 are part of the β-sheet framework structure (Chothia and Lesk, 1987). The region outside the β-sheet framework, 91–96, is called L3.

The L3 region is formed by the joining of the V_κ and J_κ segments. The joining process involves trimming and repair of the V and J segments, as well as nucleotide insertions (Milstein *et al.*, 1992). This produces variations in both the length of the L3 region and in the identity of residues at the join. The structure of the L3 region is therefore determined not only by the sequence of the germline V_κ and J_κ segments, but also by the way in which they are assembled. There are five J_κ segments, all functional and with the sequences listed in Table VI (Hieter *et al.*, 1982).

The L3 canonical structures known at present

The analysis of the three-dimensional structures of V_κ L3 regions has so far identified five canonical structures (Chothia and Lesk, 1987; Chothia *et al.*, 1989; Brünger *et al.*, 1991; He *et al.*, 1992). The five structures are illustrated in Figure 2 and in Table IVC we describe their size and sequence characteristics.

Canonical structure 1 has six residues, a Gln, Asn or His at position 90, a *cis*-Pro at 95 and a Ser or Thr at 97. Thirty-seven of the 40 functional V_κ segments have Gln at position 90 and Pro at position 95 (Table II). All five J_κ segments have Thr at 97 (Table VI). Therefore, these 37 V_κ segments can form canonical structure 1 if the joining process juxtaposes the last codon of the V_κ segment with the first codon of the J_κ segment. The two V_κ segments that have Ser and His at position 95 (Table II) can also produce this canonical structure if V_κ–J_κ joining modifies this codon to produce a Pro at position 95.

All the other known canonical structures are produced by modification of the germline segments by the joining process. Canonical structure 2 has six residues, but the Pro is at position 94 rather than 95. Since none of the functional V_κ segments has a Pro at position 94 and V_κ–J_κ joining events that modify this codon are infrequent (I.M.Tomlinson and J.P.L.Cox, in preparation), this conformation is likely to be rare in human V_κ domains. In canonical structure 3, the joining process has deleted one residue, leaving only five (Chothia *et al.*, 1989). In canonical structure 4, the joining process has deleted two residues leaving only four (He *et al.*, 1992). In canonical structure 5, the joining process adds a residue to produce a seven-residue L3 region.

Table VII.

Number of residues in the L3 region, 91–96	Number of rearranged sequences
1	1
2	2
3	2
4	1
5	54
6	427
7	119
8	26
9	1

The L3 structures implicit in the expressed repertoire

Of the 736 rearranged V_κ sequences in our database, 633 contain complete L3 regions without frame shifts. Junctional diversity has produced some variations in the length of L3, but it is limited: in 99% of the rearranged sequences the number of residues in the L3 region, 91–96, is between five and eight (Table VII).

We examined the L3 regions to determine which have the characteristics of one of the known canonical structures (Table IVC). Of those with six residues, three-quarters have the sequence characteristics of canonical structure 1. Most of those with five residues have the characteristics of canonical structure 3 and most of those with seven residues have the characteristics of canonical structure 5. Canonical structures 2 and 4 are rare, each occurring in only one sequence. In total, 70% of the rearranged sequences have the characteristics of one of the five known canonical structures (Table IVC).

The sequences of the remaining 191 rearranged genes do not correspond to a known L3 canonical structure. However, it is likely that most of these will have one of a small number of conformations. For example, we would expect most of the six-residue regions that do not have a Pro at positions 94 or 95 to have a simple hairpin structure with a conventional two-residue turn at its tip. A total of four or five canonical structures (including three for which crystal structures are available) probably describe the conformations of >90% of human L3 regions.

Because it is difficult to distinguish between residue changes introduced during V_κ–J_κ joining and those introduced by somatic mutation at key residues 95 and 97, we cannot exclude the possibility that the canonical structure of L3 changes during affinity maturation due to mutation at these positions. However, of the 729 rearranged sequences that include key residue 90, 90% are unmutated (657 sequences), 5% undergo conservative changes (Gln changes to His in 38 sequences), 4% undergo non-conservative changes (29 sequences with Gln to residues other than Asn or His) and in 1% the nature of the changes

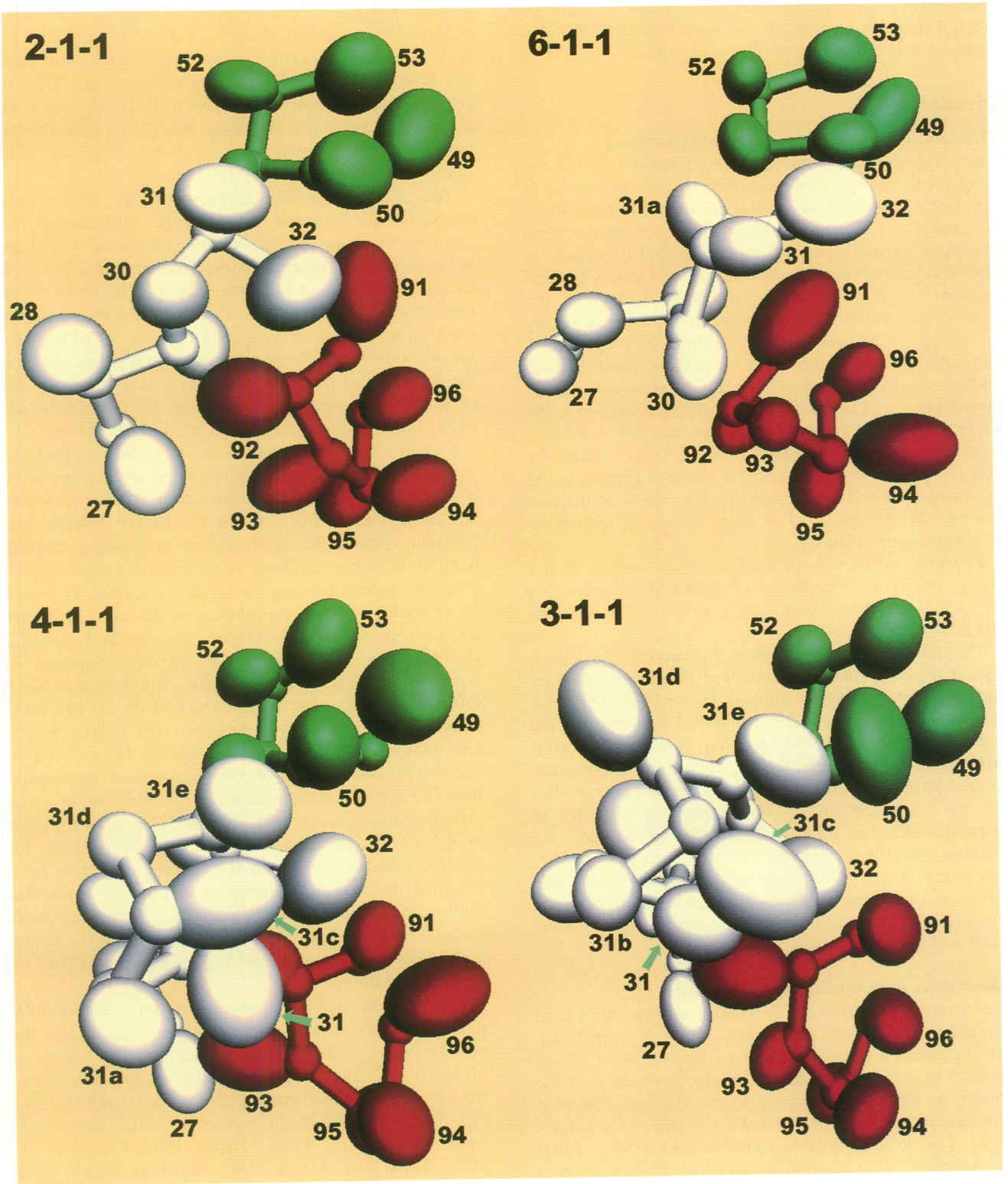

Fig. 3. Structures of the common antigen binding sites in human V_κ domains. Residues are represented by their Cα atoms and ellipsoids for the sidechains. The numbers on the top left of each drawing correspond to the canonical structure of the three antigen binding loops (L1, L2 and L3). More than 96% of the human V_κ domains have canonical structures 2, 3, 4 or 6 for the first antigen binding loop (L1) and canonical structure 1 for the second antigen binding loop (L2). Approximately half of the human V_κ domains have canonical structure 1 for the third antigen binding loop (L3).

is uncertain (five sequences where Lys changes to Gln or Asn). This suggests that L3 canonical structures are also highly conserved during affinity maturation.

Discussion

The conformations of canonical structures

Canonical structures describe the common folds of the antigen binding loops of antibodies. When they occur in different antibodies how similar are their conformations? A number of different antibody structures with the same canonical structures have been determined accurately at high resolution, including several in complex with their antigens. Analysis of these structures shows that the local conformation of the canonical structures is highly conserved (A.M.Lesk and C.Chothia, in preparation). Normally, those that are small or medium sized (e.g. L2 and the smaller L1 regions) have rms differences in atomic position of 0.5 Å or less. For those that are large (e.g. the larger L1 and L3 regions) the values are 0.8 Å or less. Very occasionally, hydrogen bonding from within the variable domain or by the antigen rotates a peptide group into an unusual orientation such that the canonical structure has rms differences of ~1 Å when compared with those of normal conformations.

Sequence differences in the framework regions do not change the local conformation of canonical structures, but can shift their relative positions by up to ~1.3 Å (Chothia and Lesk, 1987).

The interface between V_L and V_H domains has a large central region formed by conserved framework residues and a peripheral region formed by residues from the antigen binding loops. The conserved residues fix the overall packing of the V_L and V_H domains. However, differences in the antigen binding loops, and the contacts formed by antigens that penetrate the contact area, modulate the detailed geometry of the packing such that in extreme cases the relative orientation of V_L and V_H domains can differ by up to ~15° (Colman *et al.*, 1987; Lascombe *et al.*, 1989; Stanfield *et al.*, 1993). These differences in orientation shift the relative positions of the canonical structures in the two domains by 1–2 Å in most cases and by 3 Å in a few.

Mechanisms of immune recognition

The function of antibodies is to bind antigens with high specificity and affinity. In antigen–antibody complexes, binding sites for protein antigens occupy ~800 $Å^2$, which is close to one-third of the total surface presented to the solvent by the antigen binding loops. The sites that bind peptides and haptens are somewhat smaller [see Davies *et al.* (1990) and Wilson and Stanfield (1993, 1994) for reviews]. Examination of accurate high-resolution complexes demonstrates that the antigen–antibody interface is close packed, internal cavities are rare, and hydrogen bonding, van der Waals contacts and solvent molecules combine to make the packing density like that in crystals. These features are the same as those found in the recognition sites of non-immunoglobulin protein–protein complexes (Janin and Chothia, 1990).

The purpose of the genetic mechanisms that produce high-affinity antibodies is therefore to create structures that either have complementary surfaces of sufficient size to bind antigens or can produce such surfaces by low-energy conformational changes. Isolated loop structures, such as those that occur in certain H3 regions with parts free from tertiary constraints, can switch between different conformations (Wilson and Stanfield, 1994) but, in general, the intrinsic flexibility of antigen binding loops is limited with only a small envelope of alternative positions (Bhat *et al.*, 1990; Gerstein *et al.*, 1994).

The precise structural requirements for high-affinity antibodies are achieved in two stages. The assembly of germline segments produces a range of antibodies that are believed to be sufficiently diverse to recognize almost all antigens with at least moderate affinity. The small repertoire of canonical structures (Figure 3) provides a framework on which to display a highly diverse set of residues for selection during the primary immune response. Indeed, the residues that have the highest germline variability are clustered at the centre of the binding site (Chothia *et al.*, 1992; I.M.Tomlinson and J.P.L.Cox, in preparation).

Subsequent somatic mutation and selection leads to an improvement in the affinity of the antigen–antibody interaction. Typically, this maturation process increases affinity constants by a factor of 10^2–10^3 (Berek and Milstein, 1987). We have shown that during affinity maturation, human antibodies conserve the canonical loop structures selected during the primary response. There are no examples of somatic insertions or deletions in the L1 and L2 regions. Together, the key sites for the L1 and L2 regions have undergone non-conservative mutation in <4% of the rearranged sequences. Although it is more difficult to demonstrate conservation of the L3 canonical structure during affinity maturation (see above), in the vast majority of rearranged sequences the key site at position 90 is unmutated or undergoes conservative mutation. Our analysis of the human V_H repertoire demonstrated a similar conservation for the H1 and H2 regions (Chothia *et al.*, 1992). Although there are no data available for somatic mutation in human H3 regions, data have been obtained in mice by sequencing anti-hapten antibodies during an ongoing immune response. Reviewing these studies, Berek and Milstein (1987) noted that the length and certain residues of the H3 region are not modified by somatic mutation.

Conservative mutations engineered into residues that make direct contact with the antigen can change the affinity of the antibody–antigen interaction by factors of up to 10^3, whilst conservative mutations engineered into the key residues that contact the antigen binding loops can change the affinity by factors of 3–10 (Amit *et al.*, 1986; Alzari *et al.*, 1990; Sharon, 1990; Denzin *et al.*, 1991; Foote and Winter, 1992; Hawkins *et al.*, 1992, 1993; Schildbach *et al.*, 1993; Chacko *et al.*, 1995; Chen *et al.*, 1995). Here, we have shown that although non-conservative mutations at key residues that contact the antigen binding loops are rarely observed, >17% of rearranged sequences have undergone conservative mutation at these sites. Whilst these changes are unlikely to disrupt or change the canonical structures of the antigen binding loops (see above), they may well play a role in affinity maturation by producing small shifts in their geometry.

References

Akamatsu,Y., Tsurushita,N., Nagawa,F., Matsuoka,M., Okazaki,K., Imai,M. and Sakano,H. (1994) Essential residues in V(D)J recombination signals. *J. Immunol.*, **153**, 4520–4529.

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.

Alzari,P.M., Spinelli,S., Mariuzza,R.A., Boulot,G., Poljak,R.J., Jarvis,J.M. and Milstein,C. (1990) Three-dimensional structure determination of an anti-2-phenyloxazolone antibody: the role of somatic mutation and heavy/light chain pairing in the maturation of an immune response. *EMBO J.*, **9**, 3807–3814.

Bentley,D.L. and Rabbitts,T.H. (1980) Human immunoglobulin variable region genes—DNA sequences of two V_κ genes and a pseudogene. *Nature*, **288**, 730–733.

Bentley,D.L. and Rabbitts,T.H. (1983) Evolution of immunoglobulin V genes: Evidence indicating that recently duplicated human V_κ sequences have diverged by gene conversion. *Cell*, **32**, 181–189.

Berek,C. and Milstein,C. (1987) Mutation drift and repertoire shift in the maturation of the immune response. *Immunol. Rev.*, **96**, 23–41.

Bhat,T.N., Bentley,G,A., Fischmann, Boulot,G. and Poljak,R.J. (1990) Small rearrangements in structures of Fv and Fab fragments of an antibody D1.3 on antigen binding. *Nature*, **347**, 483–485.

Brünger,A.T., Leahy,D.J., Hynes,T.R. and 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.

Chacko,S., Silverton,E., Kam-Morgan,L., Sandra-Gill,S., Cohen,G. and Davies,D. (1995) Structure of an antibody-lysozyme complex: unexpected effect of a conservative mutation. *J. Mol. Biol.*, **245**, 261–274.

Chen,C., Roberts,V.A., Stevens,S., Brown,M., Stenzel-Poore,M.P. and Rittenberg,M.B. (1995) Enhancement and destruction of antibody function by somatic mutation: unequal occurrence is controlled by V gene combinatorial associations. *EMBO J.*, **14**, 2784–2794.

Chen,P.P., Albrandt,K., Orida,N.K., Radoux,V., Chen,E.Y., Schrantz,R., Liu,F.-T. and Carson,D.A. (1986) Genetic basis for the cross-reactive idiotypes on the light chains of human IgM anti-IgG antibodies. *Proc. Natl Acad. Sci. USA*, **83**, 8318–8322.

Chen,P.P., Robbins,D.L., Jirik,F.R., Kipps,T.J. and Carson,D.A. (1987) Isolation and characterisation of a light chain variable region gene for human rheumatoid factors. *J. Exp. Med.*, **166**, 1900–1905.

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

Chothia,C. *et al.* (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. and Winter,G. (1992) Structural repertoire of the human V_H segments. *J. Mol. Biol.*, **227**, 799–817.

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*, **326**, 358–363.

Cox,J.P.L., Tomlinson,I.M. and Winter,G. (1994) A directory of human germ-line V_κ segments reveals a strong bias in their usage. *Eur. J. Immunol.*, **24**, 827–836.

Davies,D.R., Padlan,E.A. and Sheriff,S. (1990) Antibody-antigen complexes. *Annu. Rev. Biochem.*, **59**, 439–473.

Denzin,L.K., Whitlow,M. and Voss,E.W. (1991) Single-chain site-specific mutations of fluorescein-aminoacid contact residues in high affinity monoclonal antibody 4-4-20. *J. Biol. Chem.*, **266**, 14095–14103.

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

Epp,O., Colman,P., Fehlhammer,H., Bode,W., Schiffer,M. and Huber,R. (1975) Crystal and molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI. *Eur. J. Biochem.*, **45**, 513–524.

Foote,J. and Winter,G. (1992) Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.*, **224**, 487–499.

Foroni,L. (1990) PhD Thesis, University of London, UK.

Gerstein,M., Lesk,A.M. and Chothia,C. (1994) Structural mechanisms for domain movements in proteins. *Biochemistry*, **33**, 6739–6749.

Hawkins,R.E., Russell,S.J. and Winter,G. (1992) Selection of phage antibodies by binding affinity: Mimicking affinity maturation. *J. Mol. Biol.*, **226**, 889–896.

Hawkins,R.E., Russell,S.J., Baier,M. and Winter,G. (1993) The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen: The interaction of mutant D1.3 antibodies with lysozyme. *J. Mol. Biol.*, **234**, 958–964.

Haynes,M.R., Stura, E.A., Hilvert,D. and 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. and Carter,D.C. (1992) Structure of a human monoclonal antibody Fab fragment against gp41 of human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA*, **89**, 7154–7158.

Hieter,P.A., Maizel,J.V. and Leder,P. (1982) Evolution of human immunoglobulin κ J region genes. *J. Biol. Chem.*, **257**, 1516–1522.

Huber,C., Schäble,K.F., Huber,E., Klein,R., Meindl,A., Thiebe,R., Lamm,R. and Zachau,H.G. (1993) The V_κ genes of the L regions and the repertoire of V_κ gene sequences in the human germ line. *Eur. J. Immunol.*, **23**, 2868–2875.

Jaenichen,H.-R., Pech,M., Lindenmaier,W., Wildgruber,N. and Zachau,H.G. (1984) Composite human V_κ genes and a model of their evolution. *Nucleic Acids Res.*, **12**, 5249–5263.

Janin,J. and Chothia,C. (1990) The structure of protein–protein recognition sites. *J. Biol. Chem.*, **265**, 16027–16030.

Kabat,E.A. and Wu,T.T. (1971) Attempts to locate complementarity determining residues in the variable positions of light and heavy chains. *Ann. NY Acad. Sci.*, **190**, 382–393.

Kabat,E.A., Wu,T.T., Perry,H.M., Gottesman,K.S. and Foeller,C. (1991) *Sequences of Proteins of Immunological Interest*, 5th edn. Public Health Service, NIH, Washington, DC.

Klein,R., Jaenichen,R. and Zachau,H.G. (1993) Expressed human immunoglobulin kappa genes and thier hypermutation. *Eur. J. Immunol.*, **23**, 3248–3271.

Klobeck,H.-G., Bornkamm,G.W., Combriato,G., Mocikat,R., Pohlenz,H.-D. and Zachau,H.G. (1985) Subgroup IV of human immunoglobulin K light chains is encoded by a single germline gene. *Nucleic Acids Res.*, **13**, 6515–6529.

Lascombe,M.-B. *et al.* (1989) Three dimensional structure of Fab R19.9: a monoclonal murine antibody specific for the p-azobenzene-arsonate group. *Proc. Natl Acad. Sci. USA*, **86**, 607–611.

Lautner-Rieske,A., Huber,C., Meindl,A., Pargent,W., Schäble,K.F., Thiebe,R., Zocher,I. and Zachau,H.G. (1992) The human immunoglobulin κ locus. Characterization of the duplicated A regions. *Eur. J. Immunol.*, **22**, 1023–1029.

Liu,M.-F., Robbins,D.L., Crowley,J.J., Sinha,S., Kozin,F., Kipps,T.J., Carson,D.A. and Chen,P.P. (1989) Characterization of four homologous L chain variable region genes that are related to 6B6.6 idiotype positive human rheumatoid factor L chains. *J. Immunol.*, **142**, 688–694.

Lorenz,W., Schäble,K.F., Thiebe,R., Stavnezer,J. and Zachau,H.G. (1988) The J_κ proximal region of the human K locus contains three uncommon Vκ genes which are arranged in opposite transcriptional polarities. *Mol. Immunol.*, **25**, 479–484.

Marquart,M., Deisenhofer,J., Huber,R. and 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.

Milstein,C., Even,J., Jarvis,J.M., Gonzalez-Fernandez,A. and Gherardi,E. (1992) Non-random features of the repertoire expressed by the members of one V_κ gene family and of the V-J recombination. *Eur. J. Immunol.*, **22**, 1627–1634.

Pargent,W., Meindl,A., Thiebe,R., Mitzel,S. and Zachau,H.G. (1991) The human immunoglobulin κ locus. Characterization of the duplicated O regions. *Eur. J. Immunol.*, **21**, 1821–1827.

Pech,M. and Zachau,H.G. (1984) Immunoglobulin genes of different subgroups are interdigitated within the V_κ locus. *Nucleic Acids Res.*, **12**, 9229–9236.

Pech,M., Jaenichen,H.-R., Pohlenz,H.-D., Neumaier,P.S., Klobeck,H.-G. and Zachau,H.G. (1984) Organization and evolution of a gene cluster for human immunoglobulin variable regions of the kappa type. *J. Mol. Biol.*, **176**, 189–204.

Pech,M., Smola,H., Pohlenz,H.-D., Staubinger,B., Gerl,R. and Zachau,H.G. (1985) A large section of the gene locus encoding human immunoglobulin variable regions of the kappa type is duplicated. *J. Mol. Biol.*, **183**, 291–299.

Radoux,V., Chen,P.P., Sorge,J.A. and Carson,D.A. (1986) A conserved

human germline V_κ gene directly encodes rheumatoid factor light chains. *J. Exp. Med.*, **164**, 2119–2124.
Rini,J.M., Schulze-Gahmen,U. and Wilson,I. (1992) Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science*, **255**, 959–965.
Schäble,K.F. and Zachau,H.G. (1993) The variable genes of the human immunoglobulin κ locus. *Biol. Chem. Hoppe-Seyler*, **374**, 1001–1022.
Schäble,H.F., Thiebe,R., Flügel,A., Meindl,A. and Zachau,H.G. (1994) The human immunoglobulin κ locus: Pseudogenes, unique and repetitive sequences. *Biol. Chem. Hoppe-Seyler*, **375**, 189–199.
Schildbach,J.F., Near,R.I., Bruccoleri,R.E., Haber,E., Jeffrey,P.D., Novotny,J., Sheriff,S. and Margolies,M.N. (1993) Modulation of antibody affinity by a non-contact residue. *Protein Sci.*, **2**, 206–214.
Scott,M.G., Crimmins,D.L., McCourt,D.W., Zocher,I., Thiebe,R., Zachau,H.G. and Nahm,M.H. (1989) Clonal characterization of the human IgG antibody repertoire to *Haemophilus influenzae* type b polysaccharide. III. A single V_κII gene and one of several J_κ genes are joined by an invariant arginine to form the most common L chain V region. *J. Immunol.*, **143**, 4110–4116.
Scott,M.G., Crimmins,D.L., McCourt,D.W., Chung,G., Schäble,K.F., Thiebe,R., Quenzel,E.-M., Zachau,H.G. and Nahm,M.N. (1991) Clonal characterization of the human IgG antibody repertoire to *Haemophilus influenzae* type b polysaccharide. IV. The less frequently expressed VL are heterogenous. *J. Immunol.*, **147**, 4007–4013.
Segal,D., Padlan,E., Cohen,G., Rudikoff,S., Potter,M. and Davies,D. (1974) The three dimensional structure of a phosphorylcholine binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Natl Acad. Sci. USA*, **71**, 4298–4302.
Sharon,J. (1990) Structural correlates of high antibody affinity: three engineered amino acid substitutions can increase the affinity of an anti-p-azophenylarsonate antibody 200 fold. *Proc. Natl Acad. Sci. USA*, **87**, 4814–4817.
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.
Stanfield,R.L., Takimoto-Kamimura,M., Rini,J.M., Profy,A.T. and Wilson,I.A. (1993) Major antigen induced domain rearrangements in an antibody. *Structure*, **1**, 83–93.
Staubinger,B., Huber,E., Lorenz,W., Osterholzer,E., Pargent,W., Pech,M., Pohlenz,H.-D., Zimmer,F.-J. and Zachau,H.G. (1988a) The human V_κ locus. Characterization of a duplicated region encoding 28 different immunoglobulin genes. *J. Mol. Biol.*, **199**, 23–34.
Staubinger,B., Thiebe,R., Huber,C., Osterholzer,E. and Zachau,H.G. (1988b) Two unusual human immunoglobulin V_κ genes. *Biol. Chem. Hoppe-Seyler*, **369**, 601–607.
Stavnezer,J., Kekish,O., Batter,D., Grenier,J., Balazs,I., Henderson,E. and Zegers,B.J.M. (1985) Aberrant recombination events in B cell lines derived from a κ-deficient human. *Nucleic Acids Res.*, **13**, 3495–3514.
Steipe,B., Plückthun,A. and 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.
Stiernholm,N.B.J. and Berinstein,N.L. (1995) A mutated promoter of a human Ig Vλ gene segment is associated with reduced germ-line transcription and a low frequency of rearrangement. *J. Immunol.*, **154**, 1748–1761.
Suh,S.W., Bhat,T.N., Navia,M.A., Cohen,G.H., Rao,D.N., Rudikoff,S. and Davies,D.R. (1986) The galactan-binding immunoglobulin Fab J539: an x-ray diffraction study at 2.6 Å resolution. *Proteins*, **1**, 74–80.
Tonegawa,S. (1983) Somatic generation of antibody diversity. *Nature*, **302**, 575–581.
Whitlow,M. (1995) Brookhaven Protein Databank (PDB) entry 1FLR.
Wilson,I.A. and Stanfield,R.L. (1993) Antibody-antigen interactions. *Curr. Opin. Struct. Biol.*, **3**, 113–118.
Wilson,I.A. and Stanfield,R.L. (1994) Antibody-antigen interactions: new structures and new conformational changes. *Curr. Opin. Struct. Biol.*, **4**, 857–867.
Zachau,H.G. (1993) The immunoglobulin κ locus—or—what has been learned from looking closely at one-tenth of a percent of the human genome. *Gene*, **135**, 167–173.

Received on April 10, 1995; revised on June 30, 1995