J. Mol. Biol. (1998) **275**, 269–294

JMB

Conformations of the Third Hypervariable Region in the VH Domain of Immunoglobulins

Veronica Morea[1], Anna Tramontano[1], Mauro Rustici[2], Cyrus Chothia[3] and Arthur M. Lesk[4]*

[1]*Istituto di Ricerche di Biologia Molecolare P. Angeletti Via Pontina Km 30.600 00045 Pomezia (Roma), Italy*

[2]*Università degli Studi di Sassari, Dipartimento di Chimica, Via Vienna 2 I-07100 Sassari, Italy*

[3]*MRC Laboratory of Molecular Biology, MRC Centre Hills Road, Cambridge CB2 2QH, UK*

[4]*Department of Haematology University of Cambridge Clinical School, MRC Centre Hills Road, Cambridge CB2 2QH, UK*

Antigen-combining sites of antibodies are constructed from six loops from VL and VH domains. The third hypervariable region of the heavy chain is far more variable than the others in length, sequence and structure, and was not included in the canonical-structure description of the conformational repertoire of the three hypervariable regions of Vκ chains and the first two of VH chains. Here we present an analysis of the conformations of the third hypervariable region of VH domains (the H3 regions) in antibodies of known structure.

We define the H3 region as comprising the residues between 92Cys and 104Gly. We divide it into a torso comprising residues proximal to the framework, four residues from the N terminus and six residues from the C terminus, and a head. There are two major classes of H3 structures that have more than ten residues between 92Cys and 104Gly: (1) the conformation of the torso has a β-bulge at residue 101, and (2) the torso does not contain a bulge, but continues the regular hydrogen-bonding pattern of the β-sheet hairpin. The choice of bulged *versus* non-bulged torso conformation is dictated primarily by the sequence, through the formation of a salt bridge between the side-chains of an Arg or Lys at position 94 and an Asp at position 101. Thus the torso region appears to have a limited repertoire of conformations, as in the canonical structure model of other antigen-binding loops.

The heads or apices of the loops have a very wide variety of conformations. In shorter H3 regions, and in those containing the non-bulged torso conformation, the heads follow the rules relating sequence to structure in short hairpins. We surveyed the heads of longer H3 regions, finding that those with bulged torsos present many very different conformations of the head.

We recognize that H3, unlike the other five antigen-binding loops, has a conformation that depends strongly on the environment, and we have analysed the interactions of H3 with residues elsewhere in the VH domain, in the VL domain, and with ligands, and their effects on the conformation of H3. We tested these results by attempts to predict the conformations of H3 regions in antibody structures solved after the results were derived.

The general conclusion of this work is that the conformation of H3 shows some regularities, from which rules relating sequence to conformation can be stated, but to a less complete degree than for the other five antigen-binding loops. Accurate prediction of the torso conformation is possible in most cases; predictions of the conformation of the head is possible in some cases. However, our understanding of the sequence–structure relationships has reduced the uncertainty to no more than a few residues at the apex of the H3 region.

Keywords: antibody structure; loop conformation; canonical structures; antigen-binding; ligand-induced conformational change

**Corresponding author*

0022–2836/98/020269–26 $25.00/0/mb971442

Introduction

The antigen-binding sites of antibodies are constructed from six loops between strands of the β-sheets in the variable domains of the light and heavy chains (Figure 1). The light-chain loops are called L1, L2, L3 and the heavy-chain loops are H1, H2, and H3. Five of the six loops have been shown to have a small discrete repertoire of main-chain conformations determined by only a few key residues in each loop and in the framework, called canonical structures (Chothia & Lesk, 1987; Chothia *et al.*, 1989; Tramontano *et al.*, 1990; Brünger *et al.*, 1991; Wu & Cygler, 1993; Martin & Thornton, 1996; Al-Lazikani *et al.*, 1997). Other residues in these loops are free to vary, to modulate the topography and charge distribution of the antigen-binding site. The key residues form signature patterns from which the main chain conformations of the loops can be recognized in the sequences of antibody structures in ''blind tests'' (Chothia *et al.*, 1986, 1989). The canonical structure model has been established for the three hypervariable regions of VL chains (L1, L2 and L3), and the first two of VH chains (H1 and H2).

The third hypervariable region of VH chains, or H3, includes the region between the conserved residues Cys92 and Gly104 (Figure 1(b)).† In previous work we defined the H3 loop as residues 96 to 101 of the VH domain, as these were the residues outside the common β-sheet hydrogen bonds (Chothia & Lesk, 1987). However, as the determinants of the conformation of the H3 loop include interactions with residues outside positions 96 to 101, and the conserved residues Cys92 and Gly104 provide useful landmarks to identify this region in a new sequence, we treat the extended H3 region, residues 92 to 104, here.

The H3 region is much more variable in length and sequence than the other antigen-binding loops (Kabat *et al.*, 1991; Rock *et al.*, 1994). Because it falls in the region of the V-D-J join in the assembly of the immunoglobulin heavy chain gene (Figure 1(c)), several mechanisms contribute to generation of its diversity, including selection of VH, D and J gene segments, and alternative splicing patterns. In expressed antibodies, H3 appears prominently at the centre of the binding site, and interacts with the VL domain as well as with other parts of the VH domain (Figure 1(a)).

† Residue numbering is identical to that of Kabat *et al.* (1991) except for residues denoted 100a, 100b and 100c, following Chothia & Lesk (1987); in the tabulation of Kabat *et al.* (1991) these residues are labelled 100I, 100J, and 100K, respectively. The nomenclature used here is consistent with the recent literature. Sometimes we shall refer to a position relative to the conserved residues Cys92 and Gly104, as C + 4 or G − 5, to indicate the position 4 after Cys92 or 5 before Gly104. When it is necessary to distinguish light and heavy chains the subscript L or H will be used; residue numbers without subscripts refer to the heavy chain.

Given the central position of the H3 region in the antigen-binding site, there are significant interactions with other loops, as well as with the framework, with the light-chain partner, and with ligands, that potentially influence the conformation of H3. However, we show here that in cases where the antibody changes conformation in response to ligation, the structure of the H3 region in the bound form is regular. This has implications for testing predicted models of antigen-binding sites against crystal structures of unligated antibodies. Pei *et al.* (1997) have solved a structure containing the VH domain of antibody B1-8 combined with two different VL domains, showing two very different conformations of H3. This important observation implies that general rules governing the conformation of H3 must involve interactions outside the local region.

Nevertheless, some of the H3 regions are short hairpins, and their structures should follow the rules governing sequence–conformation correlations in short hairpins (Venkatachalam, 1968; Sibanda & Thornton, 1985; Rose *et al.*, 1985; Chothia & Lesk, 1987; Wilmot & Thornton, 1988; Sibanda *et al.*, 1989). The conformations of short hairpins depend primarily on the sequence of the residues within the loop, notably on the position of a Gly, Asn or Asp, which are more frequently observed to have conformations with positive values of ϕ and ψ, or a Pro which can more easily accommodate a *cis* peptide. Such loops have conformations that are determined locally. However, many of the H3 regions are longer than the range covered by these rules.

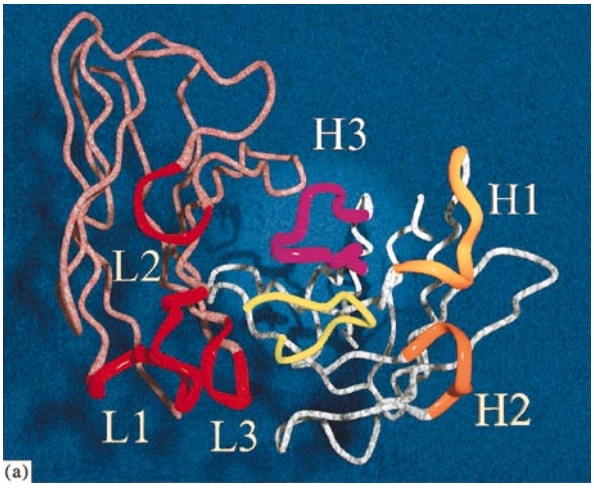
Previous work on the conformation of H3

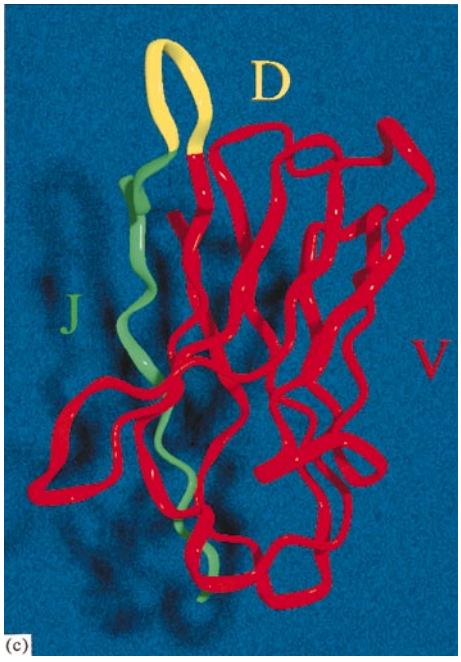
Chothia & Lesk (1987) described determinants of H3 conformation in the context of the analysis of canonical structures for the other five antigen-binding loops. That paper focused attention on two sets of residues: 94Arg and 101Asp, which form a salt bridge, and residues 100b and 100c which pack against the VL domain; the significance of the 94Arg-101Asp hydrogen bond was noted, and was included in a set of conditions on the sequence for predicting that the H3 conformation would resemble that of McPC603, which contains the bulged torso region. This set the stage for subsequent work, which developed as more structures became available.

Novotny *et al.* (1990) and Searle *et al.* (1994) also noted the 94Arg-101Asp hydrogen bond. Mas *et al.* (1992) described two conformations of the proximal residues which they called ''kinked'' and ''extended,'' but did not infer the choice of structure from the sequence. Rees *et al.* (1994) and Searle *et al.* (1994) proposed an empirical classification of seven types of H3 conformations, but this cannot survive the appearance of the structures with the variety of conformations now available. Martin & Thornton (1996) proposed conformation-determining residues for 12-residue H3 loops

(called seven-residue loops in their notation), suggesting unspecified roles for residues at positions 99_H and 101_H, but their method does not identify the important 94Arg–101Asp interaction.

In work most closely related in approach and results to our own, Shirai *et al*. (1996) have analysed the conformations of 55 H3 regions. They distinguished bulged from non-bulged torso regions (which they call kinked and extended), and proposed rules based on sequence for predicting which will appear. Although they did not precisely define the extent of the region of conserved structure, these rules are equivalent to ours. They noted the second bulge, appearing in e.g. Se155-4. They correctly observe that some of the tips of the loops are hairpins of standard types, and present a classification of the more regular H3 regions based on established rules for relating hairpin conformation and sequence; but they only mention the extent and variety of irregular structures. In particular they do not observe the distinction between two types of non-bulged torso regions (see below).





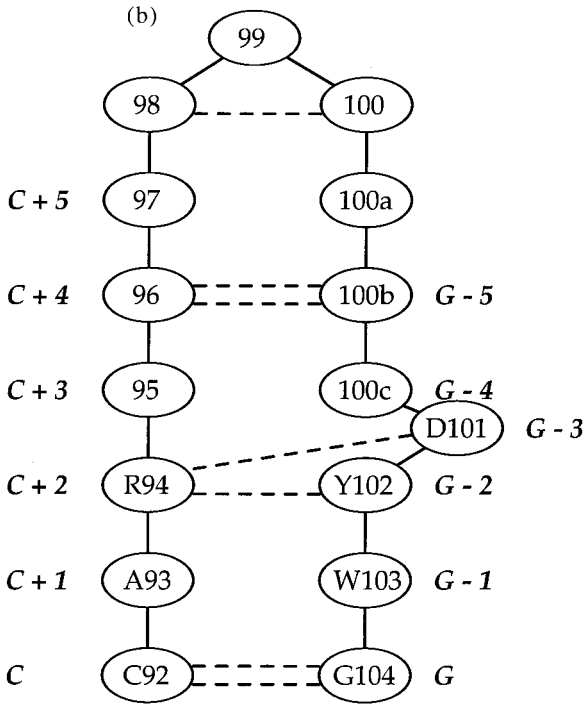


Figure 1. The conformation and structural context of the H3 region. (a) General view of the antigen-binding site of Fab TE33 (1tet), with fragment of cholera toxin. Light chain hypervariable regions L1, L2 and L3 in red; H1 and H2 in orange; H3 ''spotlit'' in magenta. Antigen in yellow. (b) Schematic diagram of the H3 region illustrating the residue nomenclature and numbering. This diagram shows the hydrogen bonding in the H3 region of McPC603; other antibodies have different hydrogen-bonding patterns. The residue numbering agrees with that of Kabat *et al*. (1991) except for residues 100a, 100b and 100c which are labelled 100 I, 100 J, and 100 K, respectively in Kabat *et al*. (1991). (c) H3 appears in the region of the V-D-J join in the assembly of the heavy chain gene. This picture shows the VH domain of McPC603 (1mcp), with the residues corresponding to the V segment in red, the D segment in yellow and the J segment in green. The region encoded by the D segment coincides almost exactly with the H3 region as defined here.

They also examined the differences in conformation between ligated and unligated structures. We note that where our investigations overlap our results are in complete agreement.

Modelling, engineering and database searching

Discussions of H3 conformation have also appeared in the context of the problem of modelling, engineering, and design of antibody combining sites; in many cases, the goal has not been to derive empirical sequence–structure correlations, but to model the loops by *ab initio* or database searching methods (Bruccoleri *et al*., 1988; Martin *et al*., 1989, 1991; Rees *et al*., 1994). The *ab initio* methods have the power of generality, and have often been combined with careful observation of the known structures. Moreover, rules that can produce models for the conformation of the proximal residues of the loop greatly simplify the problem of completion of the loop by *ab initio* methods. From our point of view, however, the importance of such investigations is: (a) the features of the structures that are identified as essential for H3 conformation, and (b) the sets of assumptions that are required for correct prediction of loop conformation.

With this motivation we have also carried out database searches for antigen-binding loops (Tramontano & Lesk, 1992; Rustici & Lesk, 1994). We report here database searches for H3 loops in other known protein structures, in order to determine the extent to which the conformations of H3 regions are determined by their sequences, or by the panoply of interactions. We searched a data bank of well-determined protein structures for regions that matched, in conformation, the framework residues flanking the H3 region. In these calculations we used the residues outside the loop to identify stretches of polypeptide chain in well-refined protein structures that bridge the selected endpoints, and checked for the presence in the sequence of special residues that can take on unusual main-chain conformations and thereby often determine loop conformation. The pattern of appearance of these residues in the sequence of the loop is necessary to predict that a loop identified from database searching will have the same conformation as the target loop.

Because no information about the interaction of the loops with their surroundings is used, successful identification of the conformations of H3 regions by this method is evidence that the conformation is determined by residues in the region itself, and suggestive of the existence of rules that relate sequence and structure for these cases, even if these rules cannot yet be succinctly enunciated. (As an alternative to database searching, other investigators have generated conformations that bridge fixed given endpoints *ab initio*, without reference to known protein structures: see, e.g., Bruccoleri *et al*. (1988)).

Summary of results

Here we present detailed descriptions of the structures of H3 regions in known antibody structures, and analysis of the determinants of their conformations. Certain features of H3 conformation have become clear; in general, the closer to the framework, the clearer. Scanning up the C-terminal residues of the region (see Figure 1(b)) the first major choice is at residue 101, which can either form a β-bulge or continue the regular pattern of β-sheet hydrogen bonding; this distinguishes bulged from non-bulged torso regions. In some cases a second bulge is formed by residue G−5. Residue 100b, the side-chain of which usually packs against the V_L domain, is the site of considerable action: If it is a Gly, as in KOL (2fb4), the packing against V_L is affected. In some cases, the side-chain of residue 100b points into its own domain instead of across the interdomain interface, affecting the size and shape of the antigen-binding pocket.

Our main results are:

(1) Clarification of the distinction between the torso region (which appears to have a limited repertoire of conformations as in the canonical structure model of other antigen-binding loops) and the head or apex of the loop; and statement of sequence-based rules for deciding which conformation is adopted (similar to those of Shirai *et al*. (1996)).

(2) Analysis of patterns of interactions between H3 and other VH residues, regions of the VL domain, and ligands; and statement of the possible varieties of H3 conformation that can arise.

(3) A survey of the conformations of the heads or apices of the loops in well-determined structures, in relation to the structure of the torso.

(4) Database searches for loops, in order to clarify the roles of local interactions and to provide a possible basis for modelling of the H3 region.

(5) Tests of the conclusions derived by evaluating their efficacy in predicting the conformations of H3 loops in antibody structures solved after the results were derived.

These conclusions, based on the many new high resolution structures now available, extend the results of our earlier paper (Chothia & Lesk, 1987). A preliminary description of some of our results has appeared in Morea *et al*. (1997).

Coordinates and Calculations

We assembled a non-redundant set of antibody structures from the Protein Data Bank (release no. 72, April 1995; Bernstein *et al*., 1977), including only structures of resolution $\leqslant 2.8$ Å which are complete in the region of interest, i.e. no missing residues, and the deposited structure is not limited to C^{α} atoms (see Padlan, 1996). In some cases two or more independent structures are available, from

molecules in the same unit cell, or from different structure determinations of the same protein. If the root mean square (rms) deviation of the backbone (N, C^{α}, C, O) atoms of two or more such regions is $\leqslant$1.0 Å, we retained only one, chosen from a structure determined at the highest resolution.

The H3 regions of the selected structures had lengths between 10 and 22 residues (Table 1A). Thirty-two of the 41 structures used had lengths of the H3 region $\leqslant$16. Calculations were carried out using InsightII (Biosym Corp.) and by programs written by one of us (Lesk, 1986, and references contained therein).

Classes of H3 Conformations

The proximal and distal regions: the torso and the head

Analysis of common structural patterns and the interactions responsible for them suggests considering the region proximal to the framework separately from the tip of the loop. These regions can be thought of as the torso and head of the structure. Many regions of different lengths share a similar torso structure although they differ at their tips.

H3 conformations can be divided into three classes (cf. Chothia & Lesk, 1987):

(1) Most H3 regions have a torso with a β-bulge at residue 101; this torso conformation can be regarded as standard.
(2) In some H3 regions, the torso does not have a β-bulge, but a regular β-sheet hairpin structure.
(3) Some very short H3 regions form a separate class, of short hairpins.

To make precise the distinction between torso and head, and to identify appropriate classes of H3 conformations, we superposed the backbone atoms of different selected subsets of the H3 region, and used the matrices of pairwise rms deviations as the distance measure in a clustering algorithm (a modified unweighted pair group method). Using four residues from the N terminus of the region and six from the C terminus, we obtained a single major cluster of structures with a common conformational pattern, for which all pairwise rms deviations for main-chain atoms are $\leqslant$1.2 Å (Table 2). We define these ten residues as the torso of the region. The main cluster of torso structures has a β-bulge at residue 101; the other structures have a regular β-sheet structure. We will refer to these as the ''bulged'' and ''non-bulged'' torso conformations.

The Bulged Torso Conformation

In this section we discuss the sequence motifs associated with the bulged torso conformation, describe the interactions that stabilize it, and analyse how variations in these interactions cause variation in the structure. Identification of the interactions that stabilize the bulged conformation reveals rules relating torso conformation to sequence. Figure 2(a) shows the bulge and two interactions that commonly stabilize it: a salt bridge between Arg94 and Asp101, and a hydrogen bond between the side-chain of Trp103 and the carbonyl oxygen of the residue at the third position before it (residue 100c = G−4).

Arg94 and Asp101 are strongly but not absolutely conserved, and when they are present a bulged conformation is observed. (Antibody 17E8 (leap), solved after this work was carried out, contains an unusual residue a position 93, K. The sequence of its H3 region is CKRSYYGSSYVDYWG. In this structure a salt bridge is formed between 93Lys and 101Asp, which permits the torso to adopt a non-bulged conformation; this exception is discussed below.) Lys94 replaces Arg94 in POT (1igm) and 3D6 (1dfb). Regions lacking Arg94 but containing Asp101, e.g. JEL103 (1mrf), show a non-bulged torso structure (Figure 2(b)). The Asp, deprived of its preferred partner, forms a hydrogen bond to the side-chain of Trp103, precluding formation of the bulged structure. In contrast, loss of the Asp is consistent with the bulge, which can be stabilized adequately by the Trp103 hydrogen bond to the residue in position G−4, as in J539 (2fbj) (Figure 2)(c).

In the assembly of the heavy-chain gene, position 94 arises from the V segment and 101 from the J segment. Of the 51 human VH germline genes (Cook & Tomlinson, 1995), all but five contain R or K at position 94. Of the six human JH germline genes (Tonegawa, 1983), all but one contains D at position 101.

Interactions within H3

Shirai *et al.* (1996) and Morea *et al.* (1997) have stated rules for predicting torso conformation from the sequence of the H3 region. When Asp101 is present without a positively charged side-chain at position 94 to form a salt bridge, the regions have a non-bulged conformation; in other cases, they have a bulged structure:

		Arg / Lys94	
		present	absent
Asp101	present	bulged	non-bulged
	absent	bulged	no examples of length > 10

During the course of this work an exception to these rules appeared: 17E8 (1eap), in which an alternative salt-bridge, between the side-chains of 93Lys and 101Asp, is formed. Shirai *et al.* (1996) extended the rules to include this case, stating that if Asp101 is present and the residues at positions

Table 1.

A. *Antibody structures used in this work*

Length	Ig	PDB Code	Resol. (Å)	*R*-factor %	Sequence	State of ligation	Torso structure	Reference
10	Cha255	1ind	2.2	10	CASH----------RFVHWG	L	TB	a
10	NQ10/12.5		2.8	18	CARD----------AGAYWG	L	TB	b
10	50.1	1ggim1	2.8	19	CVQE----------GYIYWG	L	TB	c
10	50.1	1ggb	2.8	20	CVQE----------GYIYWG		I	c
12	4-4-20	1flr	1.85	19	CTGSYY----------GMDYWG		TNB-1	d
12	JEL103	1mrf	2.4	18	CANLRG----------YFDYWG		TNB-2	e
12	HyHEL-5	2hfl	2.6	24	CLHGNY----------DFDGWG	L	TNB-2	f
12	TE33	1tet	2.3	15	CARRSW----------YFDVWG	L	TB	g
12	NC6.8	2cgr	2.2	21	CTRGYS----------SMDYWG	L	TB	h
13	D1.3	1vfa	1.8	16	CARERD---------yRLDYWG		TB	i
13	Yst9.1	1mam	2.45	22	CTRDPY---------gPAAYWG		I	j
14	8F5	1bbd	2.8	19	CDGYYS--------YYDMDYWG		TNB-1	k
14	D11.15	1jhl	2.4	21	CTRDDN--------YGAMDYWG	L	TB	l
14	J539	2fbj	1.9	19	CARLHY--------YGYNAYWG		TB	m
14	NEWM	7fab	2.0	17	CARNLI--------AGGIDVWG		TB	n
14	SE155-4	1mfb	2.1	16	CTRGGH--------GYYGDYWG	L	TB	o
14	Je142	1jel	2.8	19	CARVMG--------EQYFDVWG	L	TB	p
15	R6.5	1rmf	2.8	19	CARGGWL-------LLSFDYWG		TB	q
15	17-Ia	1for	2.75	17	CARSGNY-------PYAMDYWG		TB	r
15	26-10	1igjm1	2.5	18	CAGSSGN-------KWAMDYWG	L	TB	s
15	BV04-01	1nbv	2.0	25	CVRDQTG-------TAWFAYWG		TB	t
15	BV04-01	1cbv	2.66	19	CVRDQTG-------TAWFAYWG	L	TB	t
15	DB3	1dbb	2.7	21	CTRGDYV-------NWYFDVWG	L	TB	u
15	B13I2	1igfm1	2.8	18	CTRYSSD-------PFYFDYWG	L	TB	v
16	4D5	1fvcm1	2.2	18	CSRWGGD------GFYAMDYWG		TB	w
16	4D5	1fvdm1	2.5	18	CSRWGGD------GFYAMDVWG		TB	w
16	4D5	1fvdm2	2.5	18	CSRWGGD------GFYAMDVWG		TB	w
16	McPC603	1mcp	2.7	22	CARNYYG------STWYFDVWG		TB	x
16	NC41	1nca	2.5	19	CARGEDN------FGSLSDYWG	L	TB	y
16	26/9	1frg	2.8	19	CARRERY------DEKGFAYWG	L	TB	z
16	17/9	1himm1	2.9	20	CARRERY------DENGFAYWG	L	TB	aa
16	17/9	1hilm1	2.0	20	CARRERY------DENGFAYWG		I	aa
17	POT	1igm	2.3	20	CAKHRVS-----YVLTGFDSWG		TB	bb
17	HIL	8fabm1	1.8	17	CARDPDI-----LTAFSFDYWG		TB	cc
17	HIL	8fabm2	1.8	17	CARDPDI-----LTAFSFDYWG		TB	cc
17	36-71	6fab	1.9	20	CARSEYY-----GGSYKFDYWG		TB	dd
19	HC19	1gig	2.3	20	CARDFYDY---DVFYYAMDYWG		TB	ee
20	R19.9	1fai	2.7	19	CARSFYGG--SDLAVYYFDSWG		TB	ff
22	KOL	2fb4	1.9	19	CARDGGHGFCSSASCFGPDYWG		TB	gg
22	3d6	1dfb	2.7	18	CVKGRDYYDSGGYFTVAFDIWG		TB	hh
22	R45-45-11	1ikf	2.5	16	CTRHTLYDTLYGNYPVWFADWG	L	TB	ii

[a]Love *et al.* (1993); [b]Alzari *et al.* (1990); [c]Rini *et al.* (1993); [d]Whitlow *et al.* (1995); [e]Pokkuluri *et al.* (1994); [f]Sheriff *et al.* (1987); [g]Shoham (1993); [h]Guddat *et al.* (1994); [i]Bhat *et al.* (1994); [j]Rose *et al.* (1993); [k]Tormo *et al.* (1992); [l]Chitarra *et al.* (1993); [m]Suh *et al.* (1986); [n]Saul & Poljak (1992); [o]Zdanov *et al.* (1994); [p]Prasad *et al,* (1993); [q]Jedrzejas *et al.* (1995); [r]Liu *et al.* (1994); [s]Jeffrey *et al.* (1993); [t]Herron *et al.* (1991); [u]Arevalo *et al.* (1993); [v]Stanfield *et al.* (1990); [w]Eigenbrot *et al.* (1993); [x]Satow *et al.* (1986); [y]Tulip *et al.* (1992); [z]Churchill *et al.* (1994); [aa]Rini *et al.* (1992); [bb]Fan *et al.* (1992); [cc]Saul & Poljak (1993); [dd]Strong *et al.* (1991); [ee]Bizebard *et al.* (1994); [ff]Lascombe *et al.* (1992); [gg]Marquardt *et al.* (1980); [hh]He *et al.* (1992); [ii]Altschuh *et al.* (1992).

B. *Antibody structures solved after anaysis of those in Table 1A was completed*

Length	Ig	PDB Code	Resol. Å	*R*-factor %	Sequence	State of Ligation	Torso	Reference
12	D44.1	1mlb	2.1	18	CARGD---------GNYGYWG		TB	a
13	PLG	1plg	2.8	16	CARGGK--------FAMDYWG		TB	b
15	17E8	1eap	2.5	19	CKRSYYG------SSYVDYWG	L	Irregular	c
16	MOPC21	1igc	2.6	17	CARWGNYP-----YYAMDYWG	L	TB	d
18	40-50	1ibg	2.7	21	CARFRFASY---YDYAVDYWG	L	TB	e
18	L5MK16	1lmkm1	2.6	20	CARGEDYYA---YWYVLDYWG		TB	f
18	NC10	1nmb	2.5	21	CARSGGSYR---YDGGFDYWG	L	TB	g
21	OPG2	1opg	2.0	16	CTRHPFYRYDGGNYYAMDHWG		TB	h

[a]Braden *et al.* (1994); [b]Evans *et al.* (1995); [c]Zhou *et al.* (1994); [d]Derrick & Wigley (1994); [e]Jeffrey *et al.* (1995); [f]Perisic *et al.* (1994); [g]Malby *et al.* (1994); [h]Kodandapani *et al.* (1995).

A PDB code followed by m1 or m2 refers to one of several crystallographically non-equivalent molecules in the same crystal, e.g. 1fvcm1 refers to molecule 1 of the coordinate set 1fvc. L = ligated. TB = Bulged torso conformation. TNB-2 and TNB-2 = non-bulged torso conformations, I = irregular conformation (see text)

Table 2. rms deviations of backbone atoms of four N-terminal and six C-terminal residues from H3 regions: 92CXXX...XXXXXG104

			1vfa 1for 1dfb 1mcp 2cgr 1himm1 1jhl 1fai 8fabm2 1nbv 1ikf	1gig 1fvdm1 1fvcm1 7fab 1dbb 1igjm1 2fbj 2fb4 1igfm1 1cbv 1ind	6fab 1fvdm2 1tet 1rmf 1frg 1je1 1igm 1fabm1 1nca 1mfb NQ10/12.5	1ggim1	1hilm1	1ggb	1mrf	1mam	2hfl	1fir	1bbd
1vfa 1for 1dfb 1mcp 2cgr 1himm1 1jhl 1fai 8fabm2 1nbv 1ikf	1gig 1fvdm1 1fvcm1 7fab 1dbb 1igjm1 2fbj 2fb4 1igfm1 1cbv 1ind	6fab 1fvdm2 1tet 1rmf 1frg 1je1 1igm 1fabm1 1nca 1mfb NQ10/12.5	<1.20									
		1ggim1 1hilm1 1ggb 1mrf 1mam 2hfl 1fir 1bbd	1.08–2.63			0.00 1.66 1.40 2.32 2.17 2.61 1.95 2.26	0.00 1.73 2.74 2.65 3.02 2.09 2.54	0.00 2.93 2.72 3.28 2.53 3.05	0.00 0.79 1.23 1.43 1.34	0.00 1.52 1.42 1.43	0.00 1.99 1.52	0.00 1.13	0.00

93 and 94 are basic, the structure is non-bulged. Although only one example is known so far, this rule is likely to be true, because the residue at position 93_H is generally buried, and would be unlikely to contain an uncompensated charge.

Other than 17E8 (1eap), among the 37 structures with H3 regions of length >10 studied there are only three other exceptions: 26-10 (1igi, 1igj molecule 1, and 1igj molecule 2), Yst9.1 (1mam), and one of the structures of 17/9 (1hil). The first two were also noted by Shirai *et al.* (1996). Antibody 26-10 has a bulged torso region although it contains Asp101 and a Gly at position 94. In molecule 1 of 26-10, coordinate set 1igj, the aspartate side-chain forms a hydrogen bond to the OH of Tyr27 of H1, which, in the absence of Arg94 replaces the expected interaction of Asp101 with Trp103. In the H1 of 26-10, Phe29, generally pointing into the domain and buried, points out into the solvent; and indeed the conformation of H1 of this antibody is peculiar and cannot be related to the structure of any other known H1 loop. In this structure the H1 region is involved in crystal packing contacts, which may be distorting the molecule from its solution structure.

Yst9.1 (1mam) and 17/9 (1hil) contain Arg94 and lack Asp101; however each has an unusual structure, the torso regions of which are related to neither the bulged conformation nor the non-bulged one. (The overall conformations of the H3 regions are unrelated to each other; they are of different length.) The presence in Yst9.1 (1mam) of a Pro in position 100b (G−5), and many contacts with a second antibody molecule in the crystal structure, might be partly responsible for the H3 conformation.

The case of 17/9 is interesting because the structure of H3 depends on the state of ligation. With Arg94 present we expect the bulged torso conformation, and the bulged conformation is observed when antigen is bound (in coordinate sets 1ifh, 1him, 1hin). However, in the free antibody (1hil) the torso conformation is irregular. In the unligated state, the H3 region occupies the region of space in which the antigen binds, and the conformational change moves it out of the way upon antigen binding (Figure 3). The occurrence of a regular conformation in the presence of antigen but not in its absence is also seen in antibody 50-1. Antibodies differ in the magnitude of their conformational changes in response to changes in state of ligation (Table 3).

The importance of $Arg94_H$ in determining the conformation of H3 is in agreement with properties of mutants. Panka *et al.* (1988) and Tempest *et al.* (1991) proposed that the reduced or lost affinity of reshaped antibodies arises from a mutation of $Arg94_H$, leading to the introduction (Ser →Arg) or deletion (Arg→Ser) of a salt-bridge with Asp101. In both cases, the conversion of this residue back

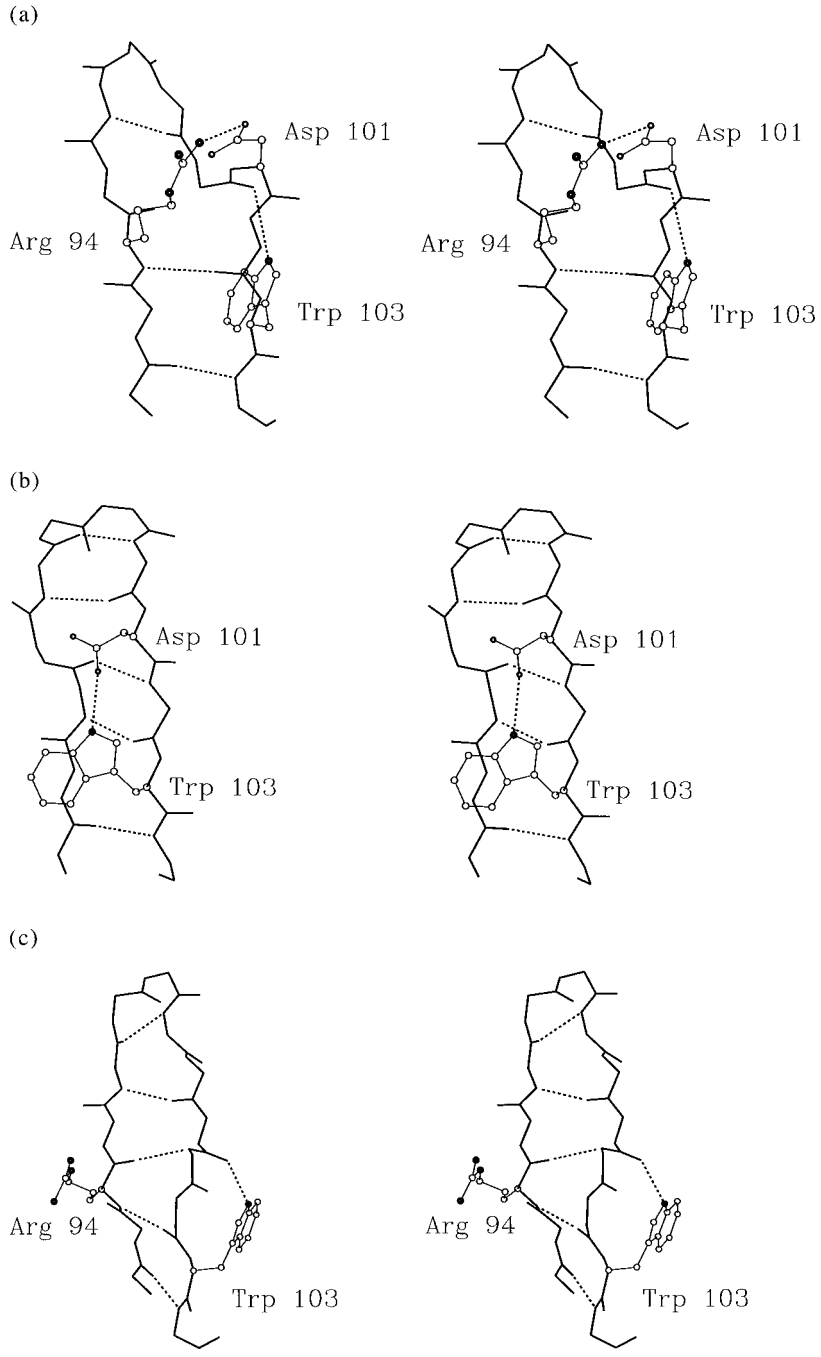


Figure 2. (a) The typical bulged torso conformation, from antibody TE33 (1tet), stabilized by the salt bridge between Arg94 and Asp101, and the hydrogen bond between the side-chain of Trp103 and the carbonyl oxygen of Phe100 K, three residues before it. (b) Non-bulged torso region (1mrf). Note the interaction between D101 and Trp103. (c) Stabilization of the bulged conformation in J539 (2fbj), which contains Arg94 but not Asp101. The Trp103 hydrogen bond appears, and Arg94 forms hydrogen bonds to the side-chain of His96 (not shown).

to its original amino acid significantly restores affinity.

Interactions between H3 and the H1 loop

Residues 27 and 32 of H1 are aromatic residues in most of the structures studied. These residues interact with hydrophobic residues of the H3 region. When H3 is bulged, in all cases of H3 regions longer than 10 excepting only 26-10 (1igj), an Arg is present in position 94 and the long hydrophobic proximal portion of its side-chain packs against the aromatics at positions 27 and 32, in H1. In the non-bulged H3 regions, the aromatics of H1 pack with an aromatic residue in position 96 or 99 of H3. This alternative packing produces two different conformations of these loops, discussed below.

Interactions between H3 and the light chain

Residues from the C-terminal part of the H3 region interact with the light chain (Chothia *et al.*, 1985). Consistent with the observation that different heavy and light chains can dimerize promiscuously, some of these interactions recur in many

Figure 3. Ligand-induced conformational change in Fab 17/9. Superposition of H3 from Fab 17/9, in ligated (1himm1, red) and unligated (1hilm1, green) forms. The H3 region in the ligated form has the expected bulged torso conformation, but the unligated form does not. The position of the antigen (black) in the ligated form would clash with the position that H3 occupies in the unligated form, and the residues in the region that would clash with the antigen change conformation by the largest amount.

structures. The VL-VH interface involves a sheet-to-sheet packing including residues from both the framework (about 3/4 of the interface) and the hypervariable regions (about 1/4 of the interface).

The structure of TE33 (1tet) illustrates a number of the common interactions between H3 and the light chain (Figure 4(a)). Residues near the C terminus of the H3 region of TE33 (1tet) pack against residues adjacent to L2 and L3 of its light chain. $100b_H$Y packs against 49_LY of the light chain, just N-terminal to L2, $100c_H$F packs into a cavity in contact with 97_LY and 99_LF, just C-terminal to L3, 101_HD packs into a cavity near 55_LF, just C-terminal to L2, and 103_HW is in contact with 99_LF.

The main chain N of residue $100c_H$ (G−4) often forms a hydrogen bond to the OH of the side-chain of residue 36_LY, in the framework strand C-terminal to L1, and just C-terminal to the conserved Trp. Residue 36_L is Y in all structures studied except NQ10/12.5 and 17-Ia (1for) in which it is F and CHA255 (1ind), Se155-4 (1mfb) and HC19 (1gig) in which it is V (it is also V in HyHEL-10 (3hfm)). Two structures make alternative hydrogen bonds: In HC19 (1gig) the side-chain of residue 34_LN forms a hydrogen bond to the main chain of residue G−6 in as well as to G−4, and in Se155-4 (1mfb) the side-chain of residue 34_LN forms hydrogen bonds to the main chain of residues G−5 and G−7.

The 36_LY OH-N $100c_H$ hydrogen bond appears in all cases of H3 regions with bulged torso conformations paired with light chains in which residue 36_L is a Y, except for BV04-01 (1nbv and 1cbv). In BV04-01 (1nbv and 1cbv) the carbonyl of residue $100b_H$ points opposite to its usual direction because of a change in peptide orientation, and alternative hydrogen bonds from 36_LY OH to the carbonyl O of $100b_H$ (1nbv) or to the carbonyl O of $100c_H$ (1cbv). This is probably related to the change in orientation of the side-chain of residue $100b_H$, which is discussed below.

Structural Effects of Sequence Variation in VL-VH Contact Residues

Some of the contacts are conserved and we cannot therefore decide whether they are essential for the structure. In some cases the contact pattern

Table 3. rms deviations of backbone atoms of H3 regions between ligated and unligated states of the same antibody

Antibody	Structure		rms deviation of backbone atoms	Length of H3
	Unligated	Ligated		
50.1	1ggb	1ggi m1	1.40	10
JEL103	1mrc	1mrf	0.35	12
D44.1	1mlb	1mlc	0.38	12
NC6.8	1cgs	2cgr	0.77	12
D1.3	1vfa	1fdl	0.34	13
DB3	1dba	1dbb	0.35	15
26-10	1igi	1igj m1	0.55	15
B13I2	1igf	2igf	0.41	15
BV04−01	1nbv	1cbv	1.12	15
McPC603	1mcp	2mcp	0.18	16
17/9	1hil m1	1him m1	1.67	16

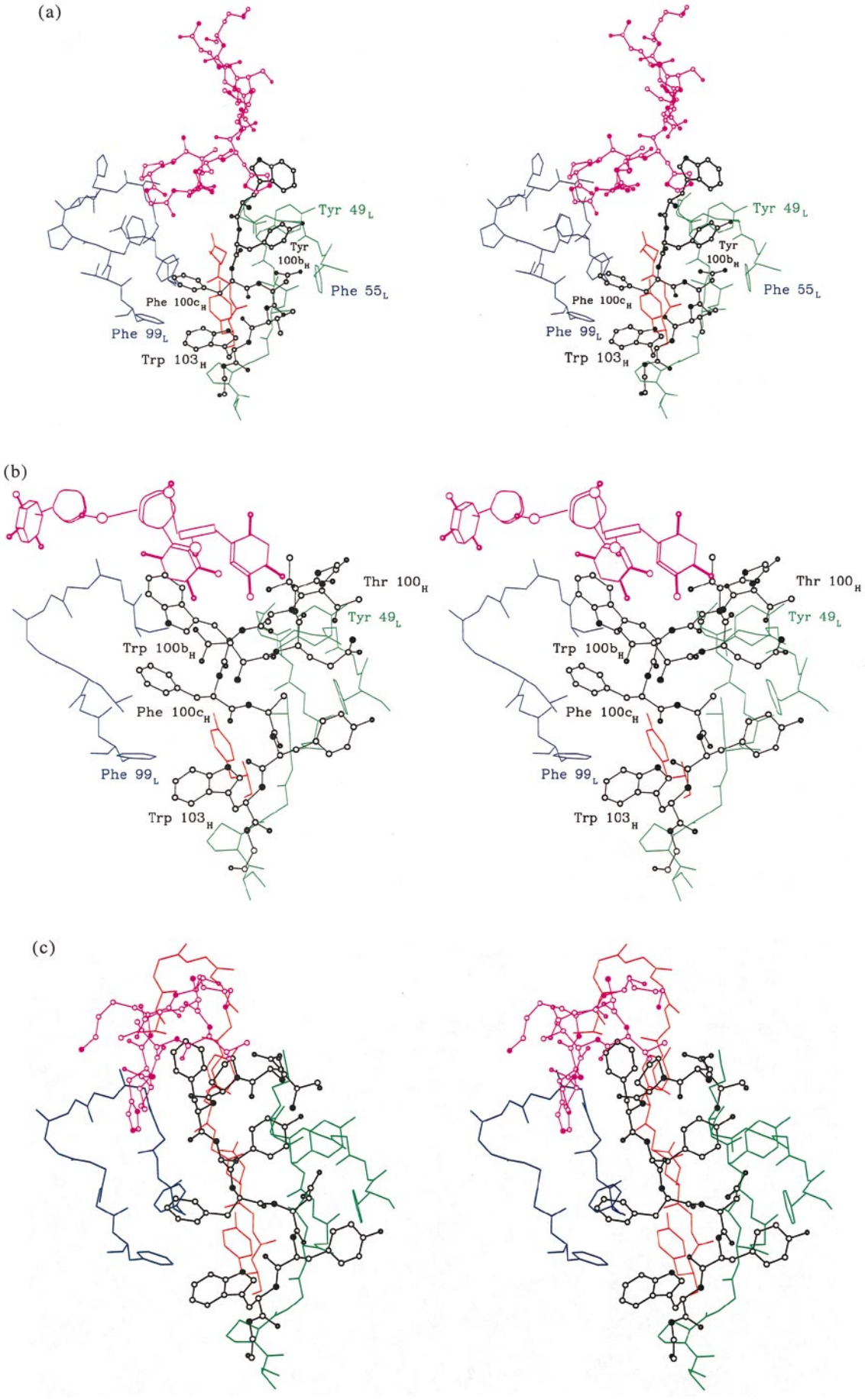


Figure 4. Interactions of H3 regions. Colour assignments in these Figures: black, C-terminal region of H3; red, residues near L1; green, L2 and residues N-terminal to it; blue, L3; magenta, antigen. In (a) and (c) the antigen is polypeptide and is represented as a ball-and-stick model; in (b) the antigen is oligonucleotide and is represented schematically. (a) Fab TE33 (1tet) shows many of the common contacts between H3 and VL. (b) BV04-01 (1cbv): Packing of residues of C-terminal region of H3 against VL. (c) B13I2 (1igfm1): Packing of residues of C-terminal region of H3 against VL.

changes but the H3 conformation does not, suggesting that these interactions are non-essential. In the case of the interactions of residue 100b changes in interactions are correlated with changes in H3 conformation.

Effects of sequence variation in VL-VH contact residues: residues on the VL side

Some of the VL-VH contacts appear in many antibody structures, and the residues that form them tend to be conserved. In the case of the 49_L-$100b_H$ contact: residue 49_L is Y in all Fab fragments of known structure except for the anti-cyclosporin antibody R45-45-11 (1ikf), in which it is F, HyHEL-10 (3hfm) in which it is K, and Cha255 (1ind), HC19 (1gig) and Se155-4 (1mfb) in which it is G. Of these, HyHEL-10 (3hfm) and Se155-4 (1fb) have glycines at position $100b_H$, the H3 region of Cha255 (1ind) is too short to permit a normal 49_L-$100b_H$ packing, and HC19 has a normal $100b_H$ conformation. In the case of the contacts with residues 55_L and 89_L: residue 55_L is F in approximately a third of the known structures. Residue 89_L is often but not always hydrophobic but is deleted in certain antibodies. No correlation between the conformation of H3 and the residues at positions 55_L and 89_L is observed. Residue 99_L is F in all known Fab structures; and the contacting pair 99_LF-103_HW is absolutely conserved in the known Fab structures.

Effects of sequence variation in VL-VH contact residues: residues on the VH side

An important factor in determining the conformation of H3 and the shape of the antigen-binding pocket is the conformation of residue $100b_H$ (G−5), the residue just N-terminal to the XDXWG motif at the C terminus of the H3 region. In TE33 (1tet), residue $100b_H$ is a Y, the side-chain of which points to the right (in Figure 4(a)) and packs against 49_LY of the light chain, adjacent to the L2 region. In certain other structures, the side-chain of this residue points in the opposite direction and packs against L3; for instance BV04-01 (1cbv, 1nbv) (Figure 4(b)), which contains the light chain residue 49_LY but in which the side-chain of the heavy chain residue $100b_H$W packs to the left (in Figure 4(b)) against the L3 region, and residue 100_HT packs against 49_LY. These structural features are reminiscent of the side-chain of Phe B25 of insulin, which in molecule 1 packs into its own monomer but in molecule 2 points out across the dimer boundary, in both 2Zn and 4Zn forms (Chothia *et al.*, 1983).

Residue $100b_H$ is often a large hydrophobic, or Y. In KOL (2fb4) it is G, but in this case the bulge is regular with residue Phe $100a_H$ instead of $100b_H$ packing against 49_LY (Chothia & Lesk, 1987). This conformation is also observed in C3 (1fpt), 26/9 (1frg), 17/9 (1hil) and POT (1igm). In Se155-4 (1mfb) residue G−4 is a glycine, and in this case a second bulge is formed at residue G−5.

The position of the side-chain of residue $100b_H$ fixes the backbone conformation of this residue, and constrains the direction the chain takes in forming the apex of the H3 region. Thus the conformation of residue $100b_H$ affects the size and shape of the antigen-binding pocket. In some antibodies, part of an antigen packs between L3 and H3. If this region is closed off, by the packing of $100a_H$ or $100b_H$ against L3, the antigen tends to lie on an almost-flat surface above this region (see, e.g. BV04-01 (1cbv) in Figure 4(b)). Conversely, if the conformation of residues $100a_H$ and $100b_H$ is such as to leave this space open, an antigen can bury itself quite deeply, as in B1312 (1igfm1) (Figure 4(c)).

The conformation of residue $100b_H$ is not in general an effect of ligation: in two available examples, the conformation of $100b_H$ does not change with state of ligation. For BV04-01 (1cbv and 1nbv) and DB3 (1dba and 1dbb) structures of free and ligated forms are known. In both BV04-01 structures, free and ligated, the side-chain of residue $100b_H$W points towards L3; in both DB3 structures, free and ligated, the side-chain of residue $100b_H$Y points towards L2. Although the conformation of H3 and the VL-VH packing in BV04-01 do change with state of ligation, these changes do not affect the conformation of residue $100b_H$.

For the side-chain of residue $100b_H$ to point towards L3, it must occupy a region of space near (and often crowded by) L3 or by a ligand. In contrast, for this side-chain to point towards L2, there is generally room for it in the region near residue 49_L (usually a Y). Several possible effects may deny access to the region near L3 to residue $100b_H$. In TE33 (1tet) and Se155-4 (1mfb) a residue at the C terminus of L3 (104_LF in TE33 and 101_LW in Se155-4) occupies the space that residue $100b_H$ would need if it were to point towards L3. In DB3 (1dfa) and McPC603 (1mcp), the residue from the N terminus of L3 (97_LS in DB3 and 97_LD in McPC603) occupies this space. D1.3 (1vfa) and 59.1 (1acy) are other examples. In R6.5 (1rmf) residue $100b_H$ is small (Ser) and is not sterically hindered despite the occupancy of the space adjacent to H3 by residues from L3.

It appears from these and other examples that, when residue $100b_H$ is large and there is room for residue $100b_H$ to point towards L3 it does so; when there is no room for it to do so it (necessarily) points towards L2. That is, we have seen no case in which there is room for a large residue $100b_H$ to point towards L3 but it does not do so.

Non-Bulged Torso Regions

We distinguish two structures of non-bulged torso regions: 4-4-20 (1flr) and 8F5 (1bbd) have one structure, and HyHEL-5 (2hfl) and JEL103 (1mrf)

have the other. In all non-bulged structures an aromatic residue from H3 packs against residues 27 and/or 32 of H1. In the first class of structures (4-4-20 and 8F5) the aromatic residue is in position 96, whereas in the second (JEL103 and HyHEL-5) it occupies position 99. These are referred to in Table 1 as TNB-1 and TNB-2. Table 4 shows the rms deviations of the backbone atoms of the five N-terminal and five C-terminal residues of the H3 regions of non-bulged structures. Figure 5 shows the H3 regions from JEL103 (1mrf) and 4-4-20 (1flr), superposed on the residues nearest the framework. The H3 region of 8F5 (1bbd) is similar to that of 4-4-20 (1flr): 4-4-20 has length 12 and 8F5 has length 14, but the rms deviation of the backbone atoms from all but the central two residues of 4-4-20 from the corresponding atoms in 8F5 is 0.9 Å.

Table 4. rms deviations of backbone atoms of five N-terminal and five C-terminal residues from H3 regions with non-bulged torsos

	1flr	1bbd	1mrf	2hfl
1flr	0.00	0.86	1.31	1.62
1bbd	0.86	0.00	1.46	1.50
1mrf	1.31	1.46	0.00	0.88
2hfl	1.62	1.50	0.88	0.00

Conformation of the Head of the H3 Loop in Relation to Torso Conformation

H3 regions with bulged torsos

12-Residue H3 regions

There are five 12-residue regions. Clustering according to the rms deviations of their backbone atoms gives four classes (Table 5). TE33 (1tet) and NC6.8 (2cgr) have bulged torsos and a common overall conformation. Figure 6 shows the superposition of the H3 regions of TE33 (1tet) and NC6.8 (2cgr).

13-Residue regions

There are two 13-residue H3 regions: in D1.3 (1vfa) and Yst9.1 (1mam). The H3 region of D1.3 (1vfa) has a regular bulge and a four-residue turn at the apex that has the sequence and one of the structures expected for the sequence XXXX (Chothia & Lesk, 1987). In specifying sequence motifs we use X to represent a ''wild card'' – a residue that can be any amino acid except proline. Figure 7(a) and (b) shows the structure of the H3 region of 1vfa, and the structure and main-chain conformational angles of the apex. (As in Al-Lazikani *et al.* (1997), we illustrate the structures determined at a resolution of 2.3 Å or better.)

The H3 region of Yst9.1 (1mam) is very different from that of D1.3 (1vfa), with a bifurcated apex. Its sequence contains two proline residues and one glycine. Figure 7(c) shows the structure.

14-Residue regions

The six structures with 14-residue H3 regions include five with bulged torso regions and one,

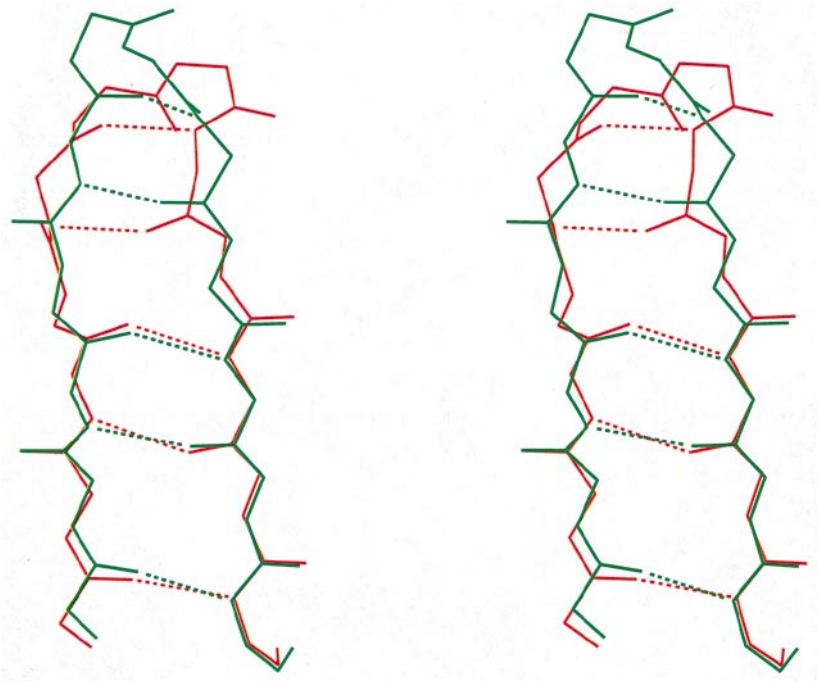


Figure 5. Two non-bulged torso conformations: superposition of H3 regions from JEL103 (1mrf; red) and 4-4-20 (1flr; green).

Table 5. rms deviations of backbone atoms of 12-residue H3 regions

	1tet	2cgr	1flr	1mrd	2hfl
1tet	0.00	0.44	1.77	2.16	2.53
2cgr	0.44	0.00	1.90	2.26	2.62
1flr	1.77	1.90	0.00	1.69	2.12
1mrd	2.16	2.26	1.69	0.00	1.45
2hfl	2.53	2.62	2.12	1.45	0.00

8F5 (1bbd), without the bulge. The five bulged regions have similar torso conformations (the rms deviation of the backbone atoms is $\leqslant$0.78 Å), except for Se155-4 (1mfb) which deviates in residue G−4 (Table 6A).

Two unusual residues, one in H3 and one in VL, may be responsible for the different conformation of Se155-4 (1mfb). Residue G−4 is a glycine in Se155-4 (1mfb), the only occurrence of this residue at this position in an immunoglobulin of known structure. In all of the other bulged loops, the side-chain of the G−4 residue packs with the conserved Trp103; in 1mfb, the side-chain of Tyr in position G−5 replaces this contact (see Figure 8).

In most light chains, the residue C-terminal to the conserved Trp35_L is a Tyr, the hydroxyl of which often forms a hydrogen bond to the main-chain N of residue 100c (G−4) (see above). Among the 14-residue H3 regions with bulged torsos this hydrogen bond is observed in J539 (2fbj), D11.15 (1jhl), NEWM (7fab) and Je142 (1jel). In contrast, in Se155-4 (1mfb) the residue at position 36_L is V; the main-chain N of residue 100c forms an alternative hydrogen bond to the side-chain of light chain residue 34_LAsn, which also forms a hydrogen bond to the main chain of residue $100a_H$Gly (see Figure 8):

$$34_L\text{Asn ND2} \cdots \text{N } 100c_H\text{Tyr}$$
$$34_L\text{Asn OD1} \cdots \text{O } 100a_H\text{Gly}$$

Note that in J539 (2fbj) but not in all others there is an alternative hydrogen bond between residues $100a_H$ and 34_L (see Figure 8).

The apices of these regions differ in conformation (Table 6B). The apex of Je142 (1jel) differs from that of 1mfb only by a change in orientation of one peptide.

15-Residue regions

There are seven structures with 15-residue H3 regions. All have the bulged conformation of the torso region (rms deviation of the backbone atoms $\leqslant$1.0 Å; see Table 7A). Superpositions of the backbone atoms of the five residues at the tip of the loop show that all are fairly similar (up to orientations of individual peptide planes) except B13I2 (1igfm1; Table 7B). Figure 9 shows the structure and conformational angles of the apex of unligated BV04-01 (1nbv), the exemplar determined at the highest resolution (2.0 Å).

16-Residue regions

All the regions of this length in known structures have the bulged conformation of their torso regions,

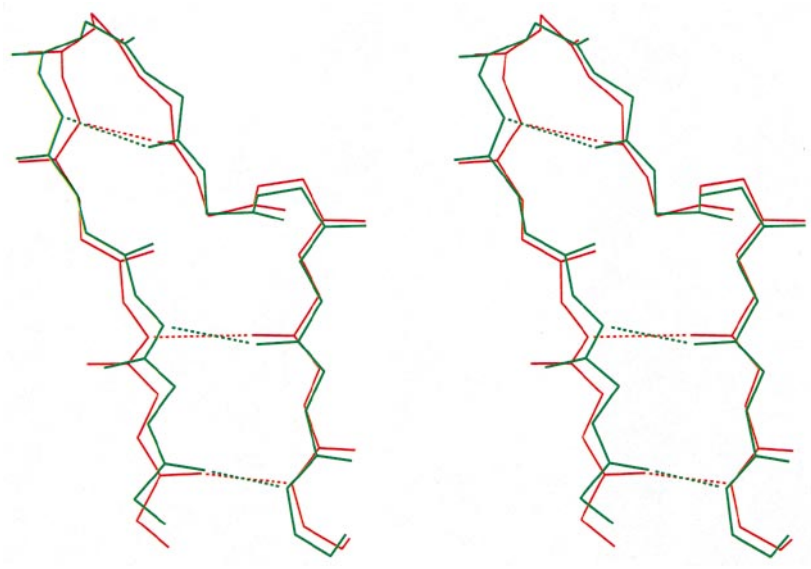


Figure 6. Twelve-residue H3 regions. Superposition of the H3 regions of TE33 (1tet; red) and NC6.8 (2cgr; green). The conformational angles of the five residues at the apices of these turns (residues 96 to 100b, or C+3 to G−5) are:

residue	1		2		3		4		5	
	φ	ψ	φ	ψ	φ	ψ	φ	ψ	φ	ψ
TE33(1tet)	−131	131	−117	−90	−102	1	−164	169	− 93	91
NC6.8(2cgr)	−149	144	−140	−74	−129	2	−146	168	−114	86

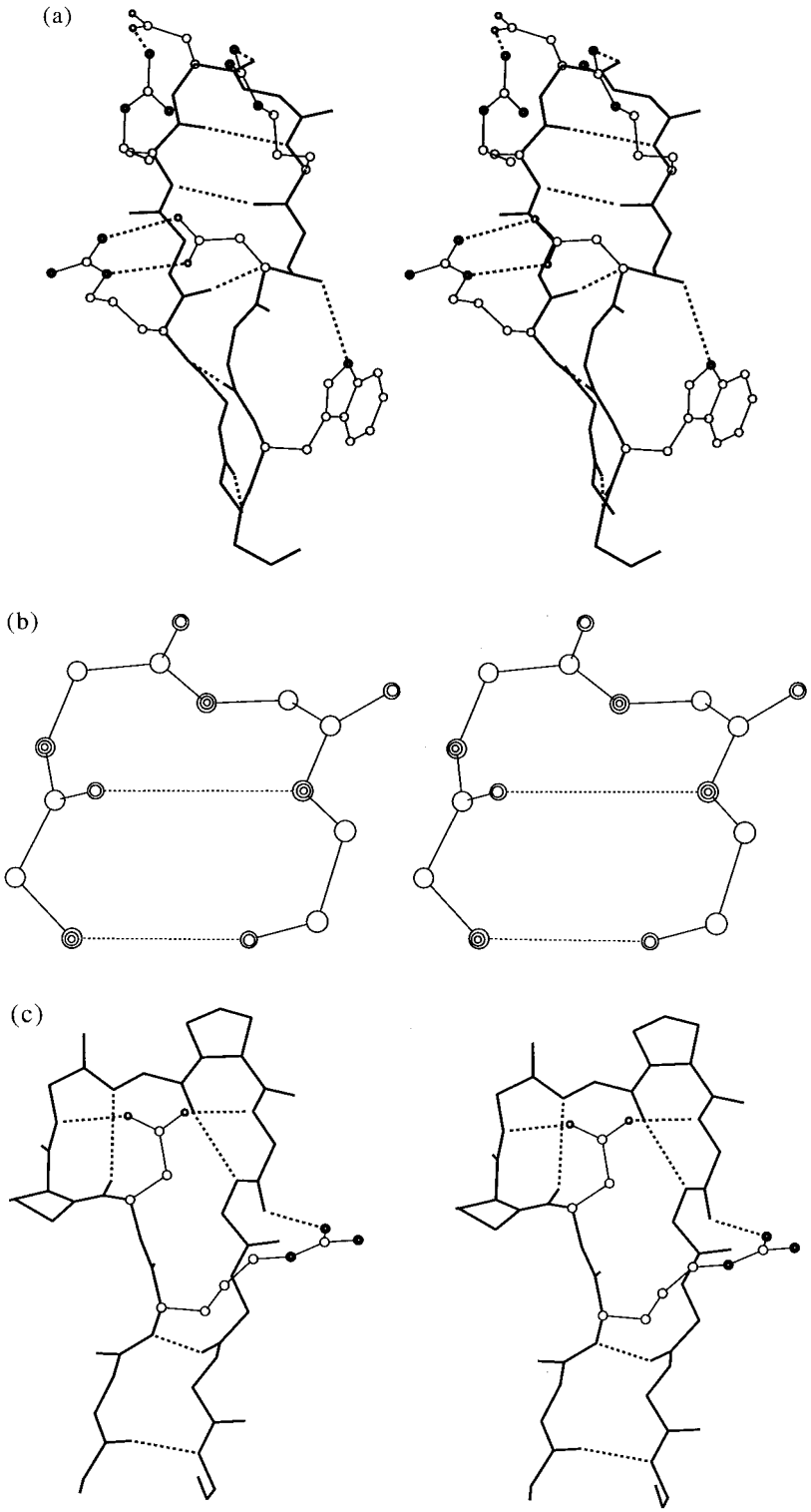


Figure 7. Thirteen-residue H3 regions. (a) H3 from D1.3 (1vfa). (b) The apex of H3 from D1.3 (1vfa). (Residues C+4 to G−5). The conformational angles are:

Arg		Asp		Tyr		Arg	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−151	135	40	63	60	42	−159	154

(c) The H3 region in Yst9.1 (1mam). This ''hammerhead'' structure is quite unusual.

Table 6.

A. *rms deviation of backbone atoms of torso residues C to C+3 and G−5 to G from 14-residue H3 regions with bulged torso conformations*

	2fbj	1jhl	7fab	1jel	1mfb
2fbj	0.00	0.33	0.55	0.63	0.99
1jhl	0.33	0.00	0.66	0.70	0.97
7fab	0.55	0.66	0.00	0.78	1.28
1jel	0.63	0.70	0.78	0.00	0.85
1mfb	0.99	0.97	1.28	0.85	0.00

B. *rms deviation of backbone atoms of apex residues C+3 to G−5 from 14-residue H3 regions*

	2fbj	1jhl	1jel	1mfb	7fab
2fbj	–				
1jhl	1.27	–			
1jel	1.53	1.65	–		
1mfb	1.51	1.95	1.18	–	
7fab	1.95	1.70	1.65	2.04	–

and the pairwise rms deviations of the backbone atoms are $\leqslant$0.9 Å except for one of the structures of 17/9 (1hil, molecule 1), because of the different conformation of residue G−3 (Ala 101b), which is bulged in the other regions. (Table 8A and Figure 10(a) to (c)).

The apices of these loops have different conformations. (Table 8b and Figure 10(d–f)).

17-Residue regions

There are four 17-residue H3 regions of known structure (Figure 11(a) to (d).) The conformation of the bulged torso of all these regions is similar; the maximum rms deviation of the backbone atoms is 0.69 Å (Table 9A). However, the apices are different (Table 9B and Figure 11(e) to (h).)

Regions longer than 17 residues

The 19-, 20- and 22-residue H3 regions have the same conformation of the bulged torso (rms deviation of backbone atoms $\leqslant$0.9 Å).

Heads associated with non-bulged torso regions

Fabs 4-4-20 (1flr), JEL103 (1mrf), HyHEL-5 (2hfl) have 12-residue H3 regions and that of 8F5 (1bbd) has 14 residues. Their torsos form

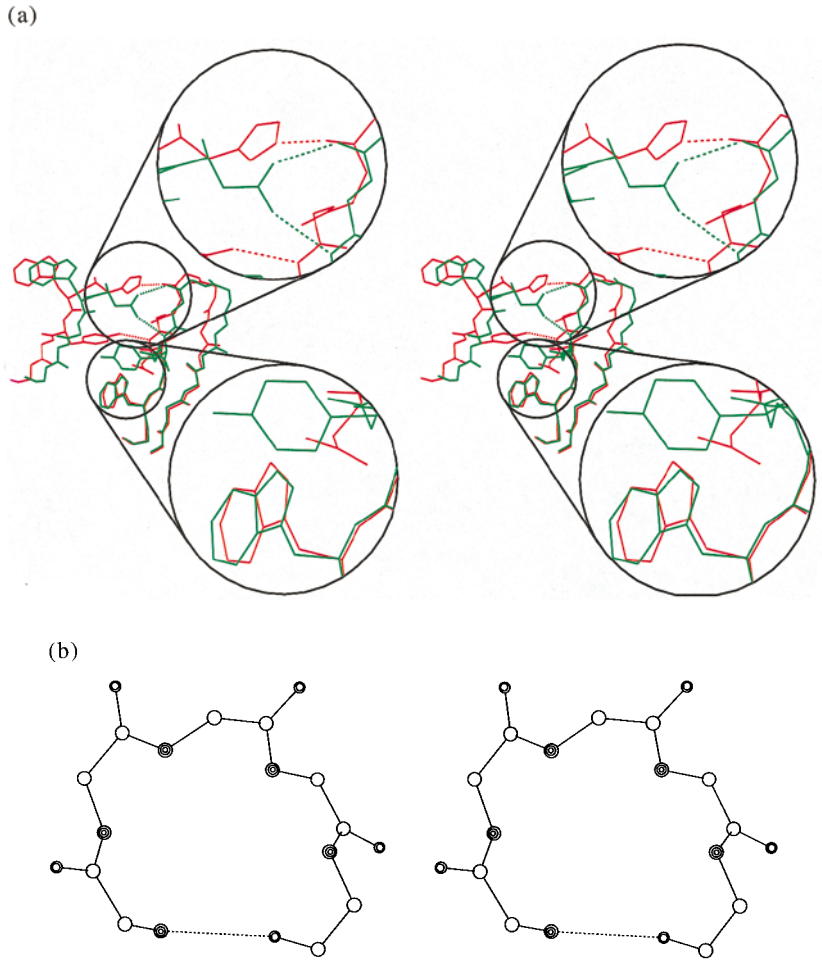


Figure 8(a–b) (*legend on page* 284)

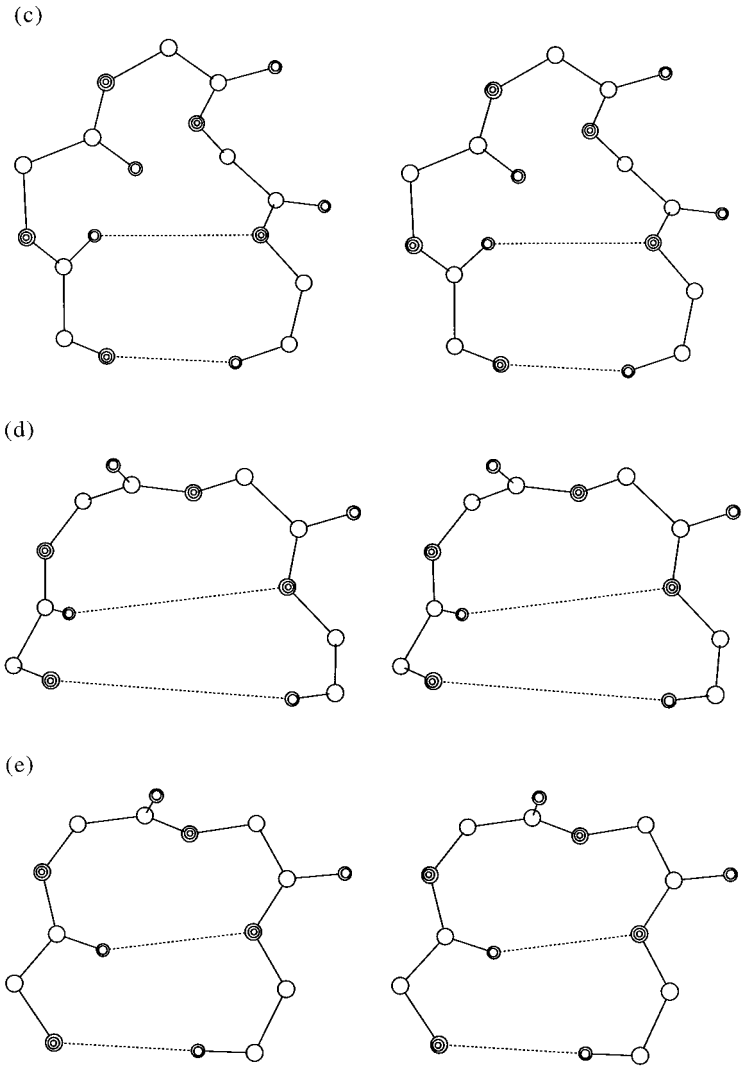


Figure 8. Fourteen-residue H3 regions. (a) Comparison of conformations, and hydrogen boding interactions with VL, of the H3 regions of J539 (2fbj) (red) and SE155-4 (1mfb) (green). The Figure shows the H3 region and the residues from VL on the strand C-terminal to L1. The large circles at the right ''blow up'' specific regions of interest. The lower circle shows the packing of the side-chain residue G − 4 (Asn) against the Trp at G − 1 in J539 (2fbj) (red) and the packing of G−5 (Tyr) against the Trp in SE155-4 (1mfb) (green). The upper circle shows the 34_L-$100a_H$ hydrogen bond in J539 (2fbj) (red) and the 36_L-$100a_H$ and 36_L-$100c_H$ hydrogen bonds in SE155-4 (1mfb) (green). (b to e) Apices of H3 regions of

(b) D11.15 (1jhl). The conformational angles are:

Asp		Asn		Tyr		Gly		Ala	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−63	−6	101	−13	−98	−30	123	−10	−91	169

(c) J539 (2fbj). The conformational angles are:

His		Tyr		Tyr		Gly		Tyr	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−117	−179	−58	145	74	19	61	44	−95	134

(d) NEWM (7fab). The conformational angles are:

Ile		Ala		Gly		Gly	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
73	−64	−118	120	134	−13	94	−175

(e) SE155-4 (1mfb). The conformational angles are:

Gly		His		Gly		Tyr	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−154	−155	−74	115	88	5	−150	158

Table 7. rms deviation of backbone atoms

A. *rms deviation of backbone atoms of the torso region of 15-residue H3 regions*

	1rmf	1for	1igjm1	1nbv	1cbv	1dbb	1igfm1
1rmf	–						
1for	0.53	–					
1igjm1	0.73	0.58	–				
1nbv	0.92	0.83	0.78	–			
1cbv	1.00	0.76	0.86	0.74	–		
1dbb	0.53	0.61	1.04	0.88	1.00	–	
1igfm1	0.87	0.71	0.73	0.86	0.93	0.81	–

B. *rms deviation of backbone atoms of the apices of 15-residue H3 regions (residues C+4 to G−6)*

	rmf	1for	1igjm1	1nbv	1cbv	1dbb	1igfm1
1rmf	–						
1for	1.00	–					
1igjm1	1.26	1.12	–				
1nbv	1.38	1.26	1.38	–			
1cbv	1.62	1.42	1.66	1.10	–		
1dbb	1.19	1.05	1.25	1.16	1.16	–	
1igfm1	2.47	2.08	2.00	1.92	2.20	2.01	0.00

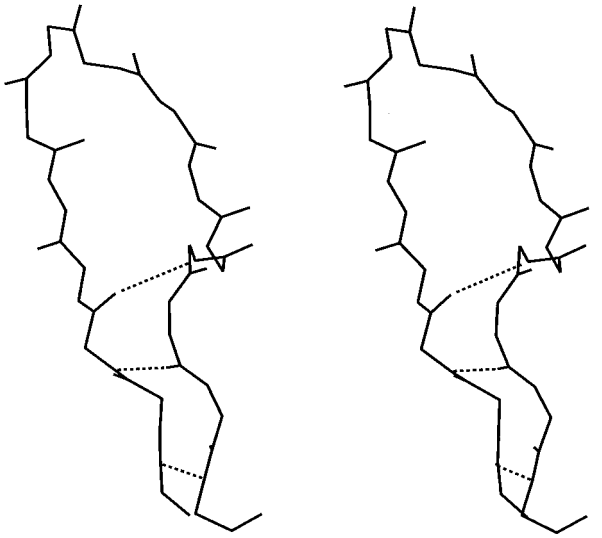


Figure 9. The apex of the 15-residue H3 region of the unligated form of BV04-01 (1nbv).

regular antiparallel strands without a bulge. The four residues that form the apices of 12-residue H3 regions follow the sequence–conformation correlations expected for four-residue regions (Chothia & Lesk, 1987). For instance, the apex of 4-4-20 (1flr) is a regular hairpin with sequence XXGX, and the angles expected for this sequence; it is a type I′ turn (Figure 12). Similarly, the apex of JEL103 (1mrf) is regular hairpin with sequence XGXX and the angles expected for this sequence.

HyHEL-5 (2hfl) also has a four-residue loop at the apex, but does not assume the expected conformation with standard angles; unusual hydrogen bonds from the side-chain amide group of $Asn100_H$ to main-chain carbonyls of $100b_H$ and $100c_H$ produces a distortion of the β-sheet likely to be responsible for the non-standard conformation in HyHEL-5.

The apex of 8F5 (1bbd), two residues longer than the others, forms a regular hairpin with the conformation expected for a six-residue loop of sequence XXXX(N or G or D or X)X, the actual sequence being YSYYDM.

Table 8.

A. *rms deviation of backbone atoms of the torso region of 16-residue H3 regions*

	1fvcm1	1fvdm1	1fvdm2	1mcp	1nca	1himm1	
1fvcm1	–						
1fvdm1	0.43	–					
1fvdm2	0.50	0.30	–				
1mcp	0.60	0.43	0.38	–			
1nca	0.83	0.80	0.81	0.79	–		
1himm1	0.59	0.66	0.64	0.69	0.88	–	
1hilm1	1.22	1.13	1.15	1.25	0.95	1.25	–

B. *rms deviation of backbone atoms of the apices of 16-residue H3 regions (residues C + 4 to G−6)*

	1fvcm1	1fvdm1	1fvdm2	1mcp	1nca	1himm1	1hil
1fvcm1	–						
1fvdm1	1.08	–					
1fvdm2	2.27	2.25	–				
1mcp	1.77	1.74	2.36	–			
1nca	1.83	2.08	2.13	1.94	–		
1himm1	2.06	2.07	2.32	1.88	1.65	–	
1hil	1.67	1.52	1.44	1.70	1.78	1.76	–

Short H3 Regions

Regions of length 10 are too short to cluster with either the bulged or non-bulged torso regions of longer H3 regions. However, typically the first three N-terminal residues and the last five C-terminal residues are similar in conformation to the corresponding residues of H3 regions with bulged torsos. Three of the structures with 10-residue H3 regions (NQ10/12.5, ligated 50.1 (1ggim1) and CHA255 (1ind)) contain the Trp103 hydrogen bond, and have similar main-chain conformations: the rms deviations of the backbone atoms in pairwise fits of these regions are $\leqslant 0.74$ Å (Table 10).

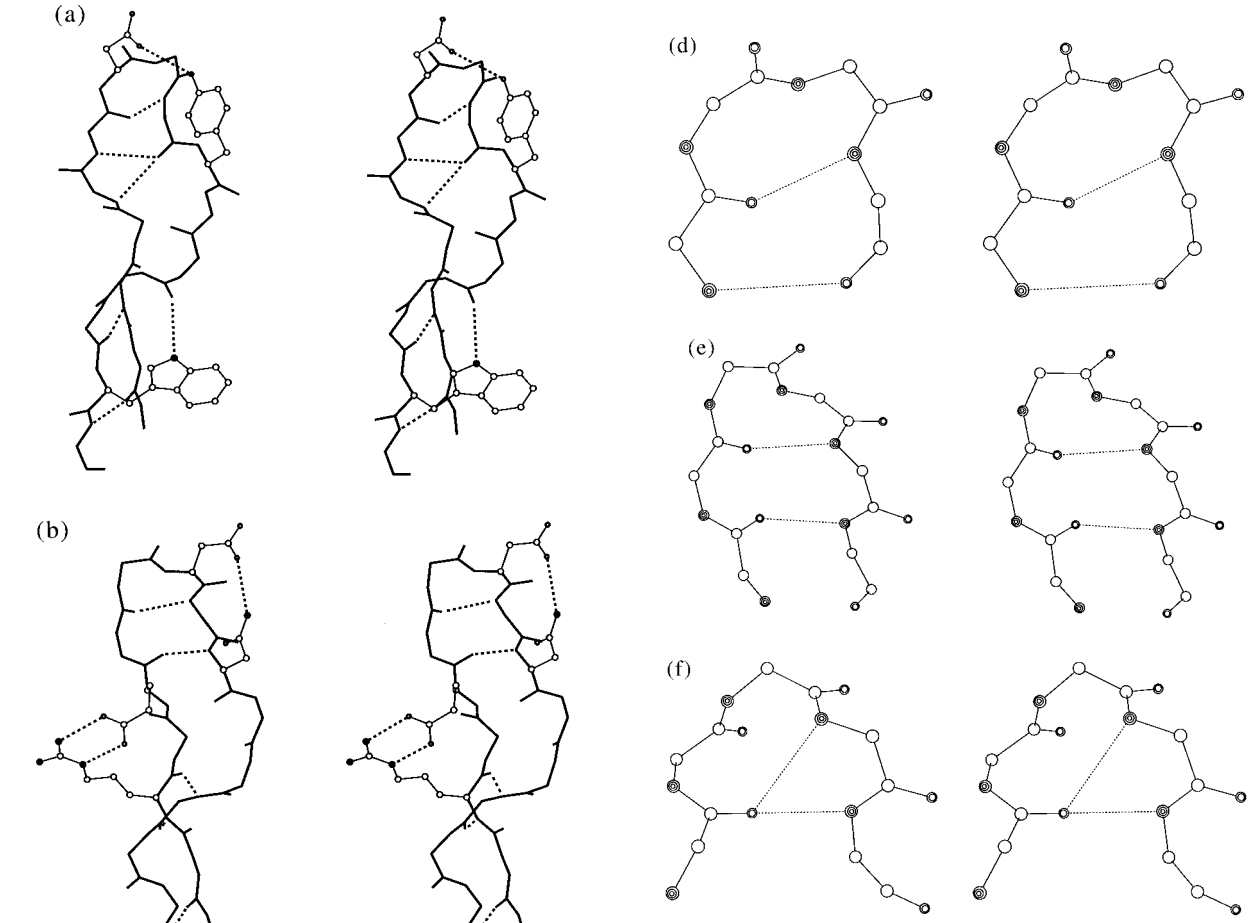


Figure 10. Sixteen-residue H3 regions of (a) the unligated form of 4D5 (1fvcm1), (b) the unligated form of 17/9 (1hilm1) and (c) the ligated form of 17/9 (1himm1). Apices of:
(d). The unligated form of 4D5 (1fvcm1). The conformational angles are:

Gly		Gly		Asp		Gly		Phe	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−50	−44	166	−153	−83	120	83	7	−120	84

(e) The unligated form of 17/9 (1hilm1). The conformational angles are:

Glu		Arg		Tyr		Asp		Glu		Asn	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−76	153	−79	−178	62	29	63	20	−75	−41	57	−127

(f) The ligated form of 17/9 (1himm1). The conformational angles are:

Glu		Arg		Tyr		Asp		Glu	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−86	81	−50	−47	−78	−55	−94	−5	45	59

Figure 13 shows the similarity in conformation of these proximal residues of the 10-residue region of CHA255 (1ind) and the longer, bulged torso conformation of D1.3 (1vfa). The unligated form of 50.1 (1ggb) has a different structure. As in the case of 17/9, the ligated state of 50.1 (1ggi, molecule 1) has a regular conformation and the unligated state an irregular one.

Effect of state of ligation

In most of the antibody structures crystallized with and without antigen the H3 region has the same conformation (below 0.8 Å rms deviation for all backbone atoms; Table 3). In two cases, 50.1 and 17/9, the state of ligation has a large effect on the structure. In both instances, the conformation of the H3 torso region is regular (bulged) in the presence of antigen and irregular in the unligated state. This throws new light on the role of induced fit in antigen recognition: the antibody selected because of its capacity to interact with the antigen does undergo somatic mutations that, although not preventing the selected conformation to occur in presence of the antigen, could destabilize this conformation in the unligated antibody, in favour of the different one that we observe.

In both crystal structures of BV04-01: 1cbv (ligated) and 1nbv (unligated) the 15-residue H3 region has a regular bulged torso, but a different tip conformation. Shirai *et al.* (1996) describe the hydrogen bond pattern of the apex of the ligated form of BV04-01 (1cbv) as regular and that of the unligated form (1nbv) as less regular; this is consistent with our observation that the ligated conformations are more regular, representing a case in which the effect is observed in the head rather than in the torso region.

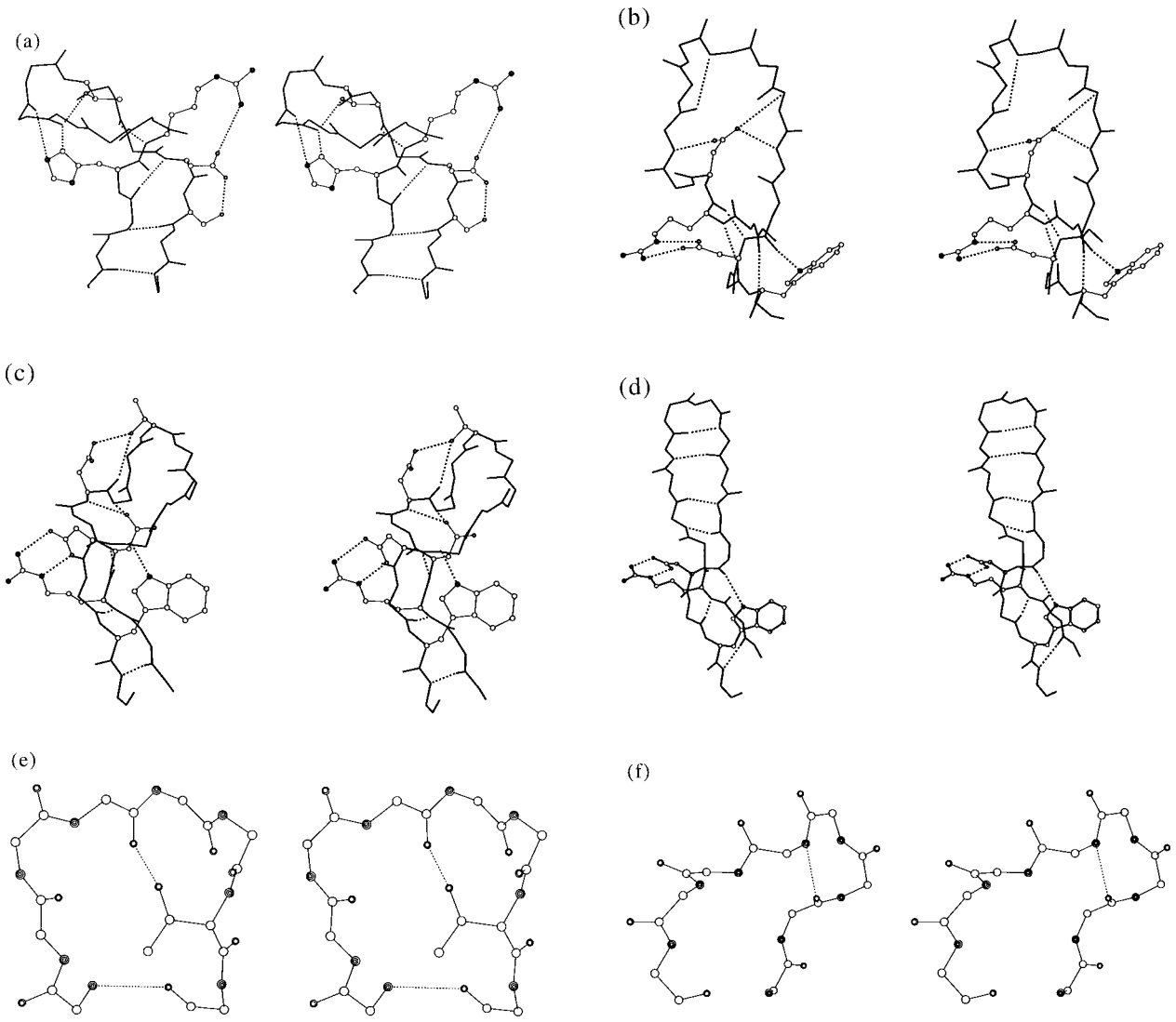


Figure 11(a–f) (*legend on page* 288)

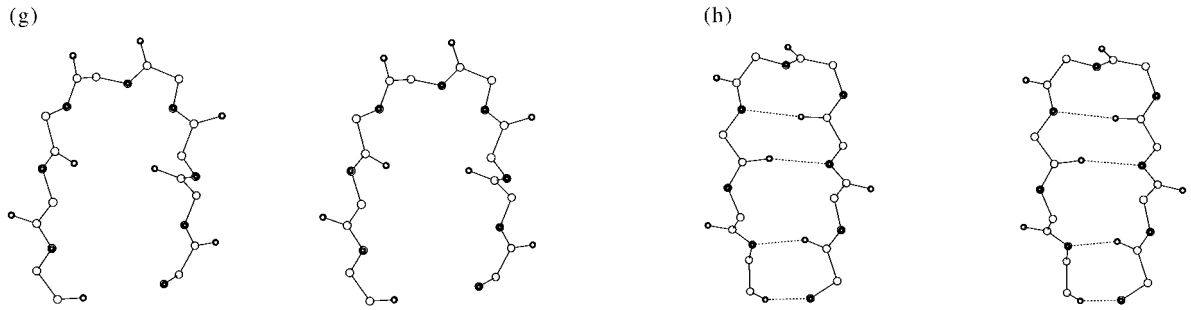


Figure 11. Seventeen-residue H3 regions (a) POT (1igm), (b) HIL (8fabm1), (c) HIL (8fabm2), (d) 36-71 (6fab). Apices of 17-residue H3 regions (residues C+4 to G−5):

(e) POT (1igm)

Arg		Val		Ser		Tyr		Val		Leu		Thr		Gly	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−55	−43	−121	179	−90	79	−173	155	−64	54	−76	80	−150	152	−72	180

(f) HIL (8fabm1)

Pro		Asp		Ile		Leu		Thr		Ala		Phe		Ser	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−65	−39	−96	127	−55	−34	−82	7	68	103	−80	−20	57	49	−138	−180

(g) HIL (8fabm2)

Pro		Asp		Ile		Leu		Thr		Ala		Phe		Ser	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−65	−30	−173	96	−82	101	−152	−57	−68	−86	−155	172	−79	149	−158	168

(h) 36-71 (6fab)

Glu		Tyr		Tyr		Gly		Gly		Ser		Thr		Lys	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−133	135	−75	136	−143	120	14	70	133	−48	−147	−167	−116	142	−110	146

Database Searching

Standard structures exist that encompass all residues of H3 regions of up to 14 residues long and the torsos of longer regions. Three of the four non-bulged H3 regions are hairpins with the dihedral angles expected for their sequence patterns. For the apices of bulged regions, the situation is rather different in that the pattern of hydrogen bonds in the torso is not that of regular hairpins, because of the bulge, and the same rules can therefore not always be applied.

One way to address the question of the extent to which the conformation of a region is determined locally, rather than varying with the molecular context, is to search the database for other loops that fit onto the same flanking residues, or ''stems'' (Jones & Thirup, 1986; Sutcliffe *et al.*, 1987; Tramontano & Lesk, 1992; Rustici & Lesk, 1994).

Table 9.

A. *rms deviation of backbone atoms of the torso regions of 17-residue H3 regions*

	8fabm1	8fabm2	1igm	6fab
8fabm1	–			
8fabm2	0.15	–		
1igm	0.51	0.52	–	
6fab	0.69	0.61	0.58	–

B. *rms deviation of backbone atoms of the apices of 17-residue H3 regions (residues C+4 to G−5)*

	8fabm1	8fabm2	1igm	6fab
8fabm1	–			
8fabm2	1.84	–		
1igm	1.83	2.33	–	
6fab	2.82	2.34	2.78	–

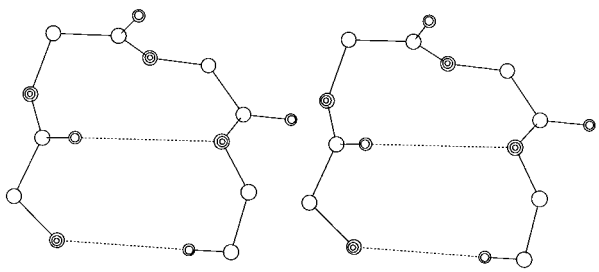


Figure 12. The apex of the H3 region in 4-4-20 (1flr) (residues C+4 to G−4). The conformational angles are:

Tyr		Tyr		Gly		Met	
ϕ	ψ	ϕ	ψ	ϕ	ψ	ϕ	ψ
−163	115	49	62	75	−19	−87	163

Table 10. rms deviation of backbone atoms of ten-residue H3 regions

	1ind	NQ10/12.5	1ggi	1ggb
1ind	0.00	0.44	0.74	1.09
NQ10/12.5	0.44	0.00	0.65	1.12
1ggi	0.74	0.65	0.00	1.40
1ggb	1.09	1.12	1.40	0.00

For ten-residue H3 regions, we searched a database of high-resolution protein structures (<2.0 Å resolution) for fragments of structures that match the stem of the loop: C+1 to C+3 and G−1 to G−3 (for ten-residue regions there are two residues at the apex between these two stems). For Cha255 (1ind) the best fit to the stem occurs in penicillopepsin (3app) and the fit of the backbone atoms of eight residues comprising the six used for the search and the two central ones gave an rms deviation of 0.4 Å. The highest deviations are at the ends of the region: eliminating the C-terminal residue reduced the rms deviation to 0.27 Å, an extremely small value. Moreover, both 1ind and 3app show values of ϕ, ψ angles at the tip of the loop similar to those expected for XXXX turns (that is, characteristic of turns which contain no residues from the set Gly, Asn, Asp, Pro that tend to produce unusual main-chain conformations (Chothia & Lesk, 1987)).

Does a good fit of the stem imply a good fit of the whole region? Figure 14 shows the correlation for CHA255 (1ind), a ten-residue region, suggesting that we can trust the loop we identify by data base searching only if the stem fit is very good. These considerations suggest that for ten-residue regions it would be possible to predict the conformation of the apex of the loop from the structure of the proximal residues and the sequence of the loop.

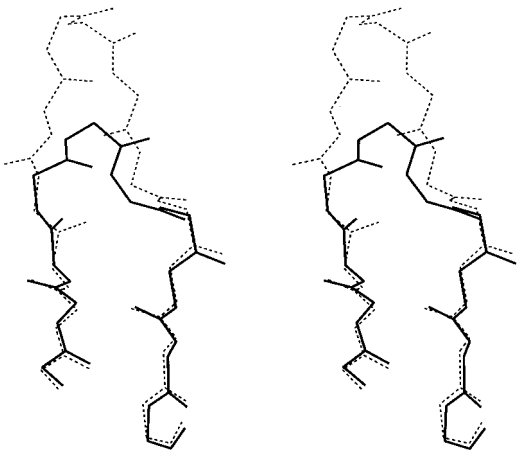


Figure 13. Similarity in structure of the proximal residues of the ten-residue H3 regions of CHA255 (1ind) (continuous lines) and the normal bulged torso conformation of D13 (1vfa) (broken lines).

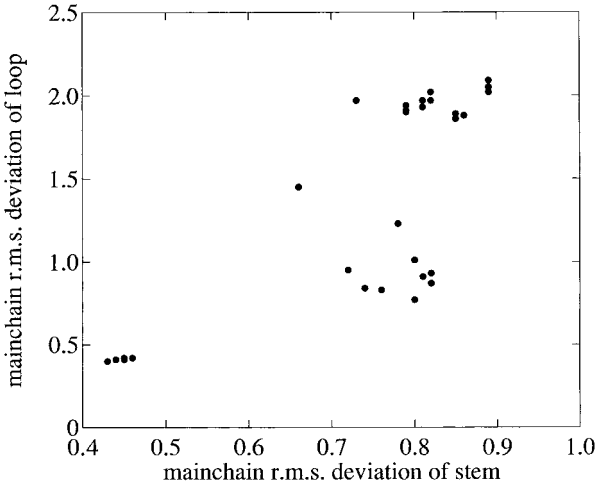


Figure 14. The correlation between fit of stem as used in database searching with the fit of the entire region found, for CHA255 (1ind).

To investigate the heads of the longer bulged H3 regions we searched the database of known protein structures for loops of the appropriate length with flanking residues that match in structure the corresponding regions of the torso of the H3 region.

We searched all protein structures determined at better than 2.0 Å resolution, for the two residues preceding and the three following the apices of the bulged H3 regions (in order to take into account the unusual H-bond pattern) of length 14 residues or longer. We imposed a threshold on backbone rms deviation for these five ''stem'' residues of 0.6 Å. We subsequently selected among the loops found in our database search those for which the pattern in the sequence of residues that can take on unusual backbone conformations matches that of the apex of the target H3 region, and calculated the rms deviation of the backbone atoms N, C^{α} and C between the loop from the database and the target H3 region.

In order for such a procedure to provide a useful method for predicting loop conformation, the database search should extract loops with a reasonable match of the flanking regions, and the appropriate sequence pattern. Failure to find the searched-for sequence pattern would result in an inability to predict the loop with the current database but would not produce incorrect predictions. In contrast, a high rms deviation between the target loop and a loop with the expected sequence pattern would result in incorrect predictions.

Regions of length 14

For D11-15 (1jhl), J539 (2fbj) and NEWM (7fab), regions with stems of similar conformation and the same sequence pattern XXXG of the apex can be

found and, in all cases, the tip of the target loop and the database hits are very similar (in some cases the main-chain conformations differ primarily in peptide orientation; this is reflected in the column in the following Table labelled -O, reporting the fit of N, C^{α} and C atoms only):

		rms deviation of backbone atoms			
Target Ig	Model	stem	apex	apex(-O)	stem+apex
1jhl	1coy	0.42	1.02	0.55	0.92
2fbj	2cga	0.33	1.01	0.60	1.02

No loop with the same XGXX sequence pattern of the apex as that of Je142 (1jel) was found in the database when searching for its stems.

In Se155-4 (1mfb), a glycine at position G−5 creates a deviation from the normal bulged conformation. The resulting conformation is fairly uncommon and in the database search we found very few hits, none of which had the same sequence pattern. Using the current database, we would be unable to predict the conformation of the apex of the H3 loop of Se155-4 (1mfb) by the methods described here.

Regions longer than 14 residues

For regions longer than 14 residues, the apex is longer than the hairpins for which correlations between sequence patterns and conformations have been established. We applied the database search procedure to regions of length 15 and 16.

For the regions of length 15, a reasonably similar loop in the database was identified for two out of the eight cases:

		rms deviation of backbone atoms			
Target Ig	Model	stem	apex	apex(-O)	stem+apex
26−10 (1igjm1)	2bbk	0.59	1.12	0.97	1.05

In the other six cases, no structure was identified with the proper apex sequence pattern.

For regions of length 16, database searching would provide a model for McPC603 (1mcp) but no hit could be found with the same sequence pattern of the other 16-residue regions.

		rms deviation of backbone atoms			
Target Ig	Model	stem	apex	apex(-O)	stem+apex
McPC603(1mcp)	1pha	0.57	0.87	0.67	1.12

The conclusions are that the correct loop can often be found by searching the database of known protein structures for regions having similar stem conformations and the same sequence pattern at the apex. Although in some cases no similar stem could be found, in no case did the search procedure give an incorrect prediction. This suggests that the conformations of these regions are most likely determined predominantly by local interactions, in particular by the position of a glycine.

Tests and extensions using newly determined immunoglobulin structures

Although the results described above are based on a relatively large number of structures, they leave open two important questions, particularly in view of the very great variety of H3 conformations that we know must exist. To what extent will the rules presented here: (1) apply to structures determined after this work was carried out? (2) provide a method of predicting H3 conformations in ''blind'' tests?

The high-resolution antibody structures solved and deposited in the PDB since this analysis was carried out appear in Table 1B.

In all cases, a bulged torso region would be predicted, and in all cases but one a regular bulged torso region is observed. The exception is 17E8 (1eap), which contains the sequence CKRS at the N terminus of the H3 region. Lys93 does not appear in this position of any human VH germline gene; usually the residue at this position is Ala (Cook & Tomlinson, 1995). At least one functional mouse germline gene (the source organism of 17E8) has K at position 93 (Rathbun *et al.*, 1988). In 17E8 (1eap) a salt bridge is formed between Lys93 and Asp101 rather than Arg94 and Asp101, and the β-bulge does not appear (Figure 15). The Asp101-Trp103 hydrogen bond is present. As noted above, this requires a generalization of the rules for predicting the presence of a bulge.

For the apices of the bulged H3 regions, neither the rules of sequence−structure correlation in hairpins nor database searching produced predictions of the conformation.

Conclusions

The H3 region of immunoglobulins is an important source of antibody diversity, showing much greater variability in length, sequence and conformation than the other five antigen-binding loops.

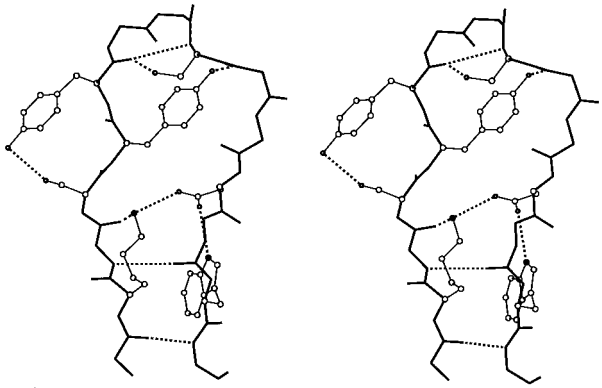


Figure 15. The H3 region of 17E8 (1eap).

Analysis of H3 conformations shows that the region proximal to the framework has a restricted set of conformations, most commonly a bulged structure stabilized by a salt bridge between Arg/Lys94 and Asp101, and a hydrogen bond between the carbonyl O of the residue at position 100c (G−4) and the side-chain of Trp103. In the absence of Arg or Lys at position 94 (or with a Lys at position 93 that can form an alternative salt bridge) the strand does not form a bulge. The conformation of the torso region, the four N-terminal and six C-terminal residues of the H3 region, can in almost all cases be predicted from the sequence of the region. These results are in agreement with those of Shirai *et al*. (1996).

The conformation of the apex of the loop is determined by many interactions: within the H3 region itself, between H3 and other residues of the heavy chain, between H3 and the light chain, and between the antibody and antigen. We have classified the structures that appear, and find that in many cases we can rationalize the conformation of the tip of the H3 region in the known structures. Prediction of H3 conformation by rules derived here from the analysis of known structures is often successful. The rules for sequence−structure correlation in hairpins do not generally apply to the apices of bulged structures, but database searching can be useful in modelling loops provided that the stems can be matched well in a structure with the correct sequence pattern in the head.

For most H3 regions of common lengths, our understanding of the sequence−structure relationships has reduced the uncertainty in the predictable conformation to no more than a few residues at the tip of region.

Acknowledgements

We thank Drs M. Whitlow and K. D. Hardman for the coordinates of 4-4-20 (1flr) prior to their deposition in the PDB, R. J. Poljak and C. Milstein for the coordinates of NQ10/12.5, and Dr E. Gherardi for discussion of the manuscript. A.M.L. thanks Dr R. Cortese for his hospitality at the Instituto di Ricerche di Biologia Molecolare, and EMBO for Short-term Fellowship ASTF 8064.

References

Al-Lazikani, B., Lesk, A. M. & Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273**, 927–948.

Altschuh, D., Vix, O., Rees, B. & Thierry, J.-C. (1992). A conformation of cyclosporin A in aqueous environment revealed by the X-ray structure of a cyclosporin-Fab complex. *Science,* **256**, 92–94.

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

Arevalo, J. H., Stura, E. A., Taussig, M. J. & Wilson, I. A. (1993). Three-dimensional structure of an anti-steroid Fab′ and progesterone-Fab′ complex. *J. Mol. Biol.* **231**, 103–118.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Bhat, T. N., Bentley, G. A., Boulot, G., Green, M. I., Tello, D., Dall'acqua, W., Souchon, H., Schwarz, F. P., Mariuzza, R. A. & Poljak, R. J. (1994). Bound water molecules 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.

Bruccoleri, R. E., Haber, E. & Novotny, J. (1988). Structure of antibody hypervariable loops reproduced by a conformational search algorithm. *Nature,* **335**, 564–568.

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

Braden, B. C., Souchon, H., Eisele, J.-L., Bentley, G. A., Bhat, T. N., Navaza, J. & Poljak, R. J. (1994). Three-dimensional structures of the free and the antigen-complexed Fab from monoclonal anti-lysozyme antibody D44.1. *J. Mol. Biol.* **243**, 767–781.

Chitarra, V., Alzari, P. M., Bentley, G. A., Bhat, T. N., Eisele, J.-L., Houdusse, A., Lescar, J., Souchon, H. & Poljak, R. J. (1993). Three-dimensional structure of a heteroclitic antigen-antibody cross-reaction complex. *Proc. Natl Acad. Sci. USA,* **90**, 7711–7715.

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., Dodson, G. G. & Hodgkin, D. C. (1983). Transmission of conformational change in insulin. *Nature,* **302**, 500–505.

Chothia, C., Lesk, A. M., Levitt, M., Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. L. (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. R., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. L. (1989). Conformations of immunoglobulin hypervariable regions. *Nature,* **342**, 877–883.

Chothia, C., Novotny, J., Bruccoleri, R. E. & Karplus, M. (1985). Domain association in immunoglobulin molecules: The packing of variable domains. *J. Mol. Biol.* **186**, 651–663.

Churchill, M. E. A., Stura, E. A., Pinilla, C., Appel, J. R., Houghten, R. A., Kono, D. H., Balderas, R. S., Fieser, G. G., Schulze-Gahmen, U. & Wilson, I. A. (1994). Crystal structure of a peptide complex of anti-influenza peptide antibody Fab 26/9. Compari-

son of two different antibodies bound to the same peptide antigen. *J. Mol. Biol.* **241**, 534–556.

Cook, G. P. & Tomlinson, I. M. (1995). The human immunoglobulin VH repertoire. *Immunology Today,* **16**, 237–242.

Derrick, J. P. & Wigley, D. B. (1994). The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab. *J. Mol. Biol.* **243**, 906–918.

Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A. A. (193). 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.

Evans, S. V., Sigurskjold, B. W., Jennings, H. J., Brisson, J.-R., To, R., Tse, W. C., Altman, E., Frosch, M., Weisgerber, C., Kratzin, H., Klebert, S., Vaesen, M., Bittersuermann, D., Rose, D. R., Young, N. M. & Bundle, D. R. (1995). Evidence for the extended helical nature of polysaccharide epitopes. The 2.8 Å resolution structure and thermodynamics of ligand binding of an antigen binding fragment specific for alpha-(2→8)-polysialic acid. *Biochemistry,* **34**, 6737–6744.

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

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

He, X. M., Ruker, F., Casale, E. & Carter, D. C. (1992). Rotes to catalysis – Structure of a catalytic antibody and comparison with its natural counterpart. *Science,* **263**, 646–652.

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

Jedrzejas, M. J., Migliette, 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., 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., Schidbach, 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.

Jones, T. A. & Thirup, S. (1986). Using known substructures in protein model building and crystallography. *EMBO J.* **5**, 819–822.

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

Kodandapani, R., Veerapandian, B., Kunicki, T. J. & Ely, K. R. (1995). Crystal structure of the OPG2 Fab: an antireceptor antibody that mimics an RGD cell adhesion site. *J. Biol. Chem.* **270**, 2268–2273.

Lascombe, M.-B., Alzari, P. M., Poljak, R. J. & Nisonoff, A. (1992). Three-dimensional structure of two crystal forms of Fab R19.9, from a monoclonal anti-arsonate antibody. *Proc. Natl Acad. Sci. USA,* **89**, 9429–9433.

Lesk, A. M. (1986). Integrated access to sequence and structural data. In *Biosequences: Perspectives and User Services in Europe* (Saccone, C., ed.), pp. 23–28, EEC, Bruxelles.

Liu, H., Smith, T. J., Lee, W.-M., Mosser, A. G., Rueckert, R. R., Olson, N. H., Cheng, R. H. & Baker, T. S. (1994). Structure determination of an Fab fragment that neutralizes human rhinovirus 14 and analysis of the Fab-virus complex. *J. Mol. Biol.* **240**, 127–137.

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

Malby, R. L., Tulip, W. R., Harley, V. R., McKimm-Breschkin, J. L., Laver, W. G., Webster, R. G. & Colman, P. M. (1994). The structure of a complex between the NC10 antibody and influenza virus neuraminidase and comparison with the overlapping binding site of the NC41 antibody. *Structure,* **2**, 733–746.

Marquardt, 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.

Martin, A. C. R., Cheetham, J. C. & Rees, A. R. (1989). Modelling antibody hypervariable loops: a combined algorithm. *Proc. Natl Acad. Sci. USA,* **86**, 9268–9272.

Martin, A. C. R., Cheetham, J. C. & Rees, A. R. (1991). Molecular modelling of antibody combining sites. *Methods Enzymol.* **203**, 121–153.

Mas, M. T., Smith, K. C., Yarmush, D. L., Aisaka, K. & Fine, R. M. (1992). Modeling the anti-CEA antibody combining site by homology and conformational search. *Proteins: Struct. Funct. Genet.* **14**, 483–498.

Morea, V., Tramontano, A., Rustici, M., Chothia, C. & Lesk, A. M. (1997). Antibody structure, prediction and redesign. *Biophys. Chem.* **68**(1–3), 9–16.

Novotny, J., Bruccoleri, R. E. & Haber, E. (1990). Computer analysis of mutations that affect antibody specificity. *Proteins: Struct. Funct. Genet.* **7**, 93–98.

Padlan, E. (1996). X-ray crystallography of antibodies. *Advan. Protein Chem.* **49**, 57–133.

Panka, D. J., Mudgett-Hunter, M., Parks, D. R., Peterson, L. L., Herzenberg, L. A., Haber, E. & Margolies, M. N. (1988). Variable region framework differences result in decreased or increased affinity of variant anti-digoxin antibodies. *Proc. Natl Acad. Sci. USA,* **85**, 3080–3084.

Pei, X. Y., Holliger, P., Murzin, A. G. & Williams, R. L. (1997). The 2.0-Å resolution crystal structure of a trimeric antibody fragment with noncognate V_H-V_L domain pairs shows a rearrangement of V_H CDR3. *Proc. Natl Acad. Sci. USA,* **94**, 9637–9642.

Pokkuluri, P. R., Bouthillier, F., Li, Y., Kuderova, A., Lee, J. & Cygler, M. (1994). Preparation, characterization and crystallization of an antibody Fab fragment that recognizes RNA. Crystal structures of native Fab and three Fab-mononucleotide complexes. *J. Mol. Biol.* **243**, 283–297.

Perisic, O., Webb, P. A., Holliger, P., Winter, G. & Williams, R. (1994). The structure of a bivalent diabody. *Structure,* **2**, 1217–1226.

Prasad, L., Sharma, S., Vandonsellar, 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.

Rathbun, G. A., Otani, F., Milner, E. C. B., Capra, J. D. & Tucker, P. W. (1988). Molecular characterization of the A/J J588 family of heavy chain variable region gene segments. *J. Mol. Biol.* **202**, 383–395.

Rees, A. R., Staunton, D., Webster, D. M., Searle, S. J., Henry, A. H. & Pedersen, J. T. (1994). Antibody design: beyond the natural limits. *Tibtech,* **12**, 199–211.

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 an HIV-1 neutralizing antibody 50.1 in complex with its V3 loop peptide antigen. *Proc. Natl Acad. Sci. USA,* **90**, 6325–6329.

Rock, E. P., Sibbald, P. R., Davis, M. M. & Chien, Y.-H. (1994). CDR3 length in antigen-specific immune receptors. *J. Exp. Med.* **179**, 323–328.

Rose, D. R., Przybylska, R. J., To, C. S., Kayden, R., Oomen, R., 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.

Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985). Turns in peptides and proteins. *Advan. Protein Chem.* **37**, 1–109.

Rustici, M. & Lesk, A. M. (1994). Three-dimensional searching for recurrent structural motifs in databases of protein structures. *J. Comp. Biol.* **1**, 121–132.

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**, 353–371.

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.

Searle, S. J., Petersen, J. T., Henry, A. H., Webster, D. M. & Rees, A. R. (1994). Antibody structure and function. In *Antibody Engineering* (Borreback, C. A. K., ed.), pp. 3–51, Oxford University Press, Oxford.

Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. & Davies, D. R. (1987). Three-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 2.4 Å. *J. Mol. Biol.* **232**, 1169–1175.

Sibanda, B. L. & Thornton, J. M. (1985). β-hairpin families in globular proteins. *Nature,* **316**, 170–174.

Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989). Conformation of β-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* **206**, 759–777.

Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1989). The phosphocholine binding immunoglobulin Fab McPC603: an X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593–604.

Strong, R. K., Campbell, R., Rose, D. R., Petsko, G. A., Sharon, J. & Margolies, M. N. (1991). Three-dimensional structure of murine anti-*p*-azophenylarsonate Fab 36-71. 1. X-ray crystallography, site-directed mutagenesis, and modeling of the complex with hapten. *Biochemistry,* **30**, 3739–3748.

Stanfield, R. L., Fieser, T. M., Lerner, R. A. & Wilson, I. A. (1990). Crystal structures of an antibody to a peptide and its complex with peptide antigen at 2.8 Å. *Science,* **248**, 712–719.

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 A resolution. *Protein Eng.* **1**, 74–80.

Sutcliffe, M. J., Haneef, I., Carney, D. & Blundell, T. L. (1987). Knowledge based modelling of homologous proteins, part I: three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng.* **1**, 377–384.

Tempest, P. R., Bremner, P., Lambert, M., Taylor, G., Furze, J. M., Carr, F. J. & Harris, W. J. (1991). Reshaping a human monoclonal antibody to inhibit human respiratory syncytial virus infection *in vivo*. *Bio/Technology,* **9**, 266–271.

Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature,* **302**, 575–581.

Tormo, J., Stadler, E., Skern, T., Auer, H., Kanzler, O., Betzel, C., Blaas, D. & Fita, I. (1992). Three dimensional structure of the Fab fragment of a neutralizing antibody to human rhinovirus serotype 2. *Protein Sci.* **1**, 1154–1161.

Tulip, W. R., Varghese, J. N., Laver, W. G., Webster, R. G. & Colman, P. M. (1992). Refined crystal structure of the influenza virus N9 neuraminidase-NC41 Fab complex. *J. Mol. Biol.* **227**, 122–148.

Tramontano, A. & Lesk, A. M. (1992). Common features of the conformations of antigen-binding loops in immunoglobulins and application to modelling of loop conformations by database screening. *Proteins: Struct. Funct. Genet.* **13**, 231–245.

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 V_H domains of immunoglobulins. *J. Mol. Biol.* **215**, 175–182.

Venkatachalam, C. (1968). Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers,* **6**, 1425–1436.

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

Wilmot, C. M. & Thornton, J. M. (1988). Analysis and prediction of the different types of β-turns in proteins. *J. Mol. Biol.* **203**, 221–232.

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

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

Edited by I. A. Wilson

(Received 22 May 1997; received in revised form 26 September 1997; accepted 2 October 1997)