PROTEINS: Structure, Function, and Genetics 6:382–394 (1989)

Structural Determinants of the Conformations of Medium-Sized Loops in Proteins

Anna Tramontano,[1] Cyrus Chothia,[2,3] and Arthur M. Lesk[1,2]

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 1022.09, 6900 Heidelberg, Federal Republic of Germany, [2]MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, England, and [3]Christopher Ingold Laboratory, University College London, 20 Gordon Street, London WC1H 0AJ, England

ABSTRACT **Loops are integral components of protein structures, providing links between elements of secondary structure, and in many cases contributing to catalytic and binding sites.**

The conformations of short loops are now understood to depend primarily on their amino acid sequences. In contrast, the structural determinants of longer loops involve hydrogen-bonding and packing interactions within the loop and with other parts of the protein. By searching solved protein structures for regions similar in main chain conformation to the antigen-binding loops in immunoglobulins, we identified medium-sized loops of similar structure in unrelated proteins, and compared the determinants of their conformations.

For loops that form compact substructures the major determinant of the conformation is the formation of hydrogen bonds to inward-pointing main chain atoms. For loops that have more extended conformations, the major determinant of their structure is the packing of a particular residue or residues against the rest of the protein.

The following picture emerges: Medium-sized loops of similar conformation are stabilized by similar interactions. The groups that interact with the loop have very similar spatial dispositions with respect to the loop. However, the residues that provide these interactions may arise from dissimilar parts of the protein: The conformation of the loop requires certain interactions that the protein may provide in a variety of ways.

Key words: immunoglobulins, hydrogen bonding, hairpin loops

INTRODUCTION

Surface loops generally comprise about a third of the polypeptide chain in proteins. They have a structural role in connecting elements of secondary structure—helices and sheets—and often have functional importance, forming parts of active sites. An understanding of the determinants of conformations of loops is of intrinsic interest, and also of practical significance for engineering of new properties of proteins.

Hairpin loops have been the subjects of numerous investigations (a hairpin is a loop connecting two successive strands of antiparallel β-sheet). From the work of Venkatachalam,[1] Thornton and her colleagues,[2–4] Efimov,[5] and others,[6–9] the determinants of the conformations of short hairpins, three or four residues in length, are understood at least in outline. For a short peptide to reverse direction, a residue in a conformation in the nonallowed region of the Ramachandran diagram is generally required. Therefore the conformations of short loops depend primarily on the position within the loop of special residues—usually Gly, Asn, or Pro—that allow the chain to take up an unusual conformation.

Not all loops in proteins are short, however. Leszczynski and Rose have shown that proteins often contain loops 6–10 residues in length; some loops are much longer.[7] The factors that determine the conformation of these longer loops have not previously been understood.

Our interest in the structures of loops derives from our attempts to understand how the sequences of immunoglobulins determine the conformation of their antigen-binding sites. In immunoglobulins, the antigen-binding site contains six loops linked to strands in a conserved β-sheet framework. In some cases these are short hairpins, but most are medium to large in size: 6–15 residues.

In earlier work, we attempted to identify conformational determinants of most of the longer antigen-binding loops, from an analysis of known immunoglobulin structures.[8] We found the loops to be stabilized by hydrogen bonding and packing interactions. For at least five of the six loops there is a discrete repertoire of conformations, called "canonical structures," stabilized by a small proportion of the residues. (Other residues are relatively free to vary, to modulate the surface topography and

Received May 22, 1989; revision accepted July 21, 1989.

Address reprint requests to Arthur M. Lesk, European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 1022.09, 6900 Heidelberg, Federal Republic of Germany.

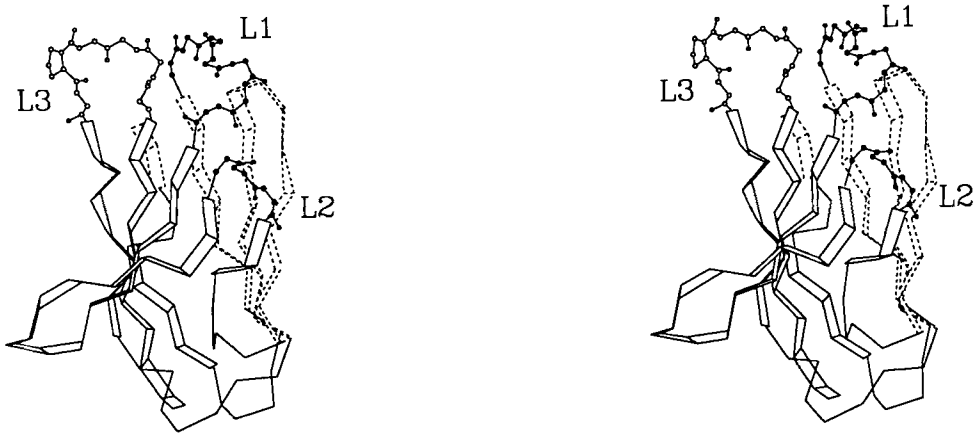

Fig. 1. Folding pattern of the immunoglobulin domain, showing antigen-binding loops. This figure shows the V_L domain of REI[35] and its three antigen-binding loops. The framework is shown as ribbons (the front sheet is drawn in solid lines and the back sheet in broken lines), and the main chain atoms of the loops are shown in a ball-and-stick representation: L1 and L2 shaded, and L3 open circles. The V_H domain is qualitatively similar.

charge distribution of the antigen-binding site.) These results have allowed us to make predictions of the structures of antigen-binding regions of immunoglobulins of unknown structure.[10]

To extend this analysis, we searched known protein structures for loops similar in main chain conformation to antigen-binding loops in immunoglobulins. Jones and Thirup have shown that loops of similar conformation can occur in unrelated proteins.[11] Here we describe several examples of regions with similar conformations in immunoglobulins and other proteins, and compare the interactions that stabilize their conformations. The structural similarities are not the result of evolutionary relationships. Nevertheless, we find that similar conformations are stabilized by similar interactions. However, the surroundings of the loops, that provide these interactions, may be constructed in different ways in different proteins.

The Loops That Form the Antigen-Binding Sites of Immunoglobulins

In this section we describe the basic features of the structures of immunoglobulins and their antigen-binding loops.

Immunoglobulins are built of domains with a common fold: two β-sheets packed face to face (Fig. 1). The antigen-binding site is formed by six loops, three from the V_L domain and three from the V_H domain. These are denoted L1, L2, and L3 and H1, H2, and H3, respectively. L2, L3, H2, and H3 are hairpins: they connect adjacent antiparallel strands in the same β-sheet. L1 and H1 connect strands of different sheets within V_L and V_H domains.

Variations in the lengths and amino acid sequences of these loops generate differences in specificity and affinity of immunoglobulins. In the immunoglobulins of known structure, L1 varies in length between 6 and 13 residues, L2 has 3 residues, L3 has 5–8 residues, H1 has 7 residues, H2 has 3–6 residues, and H3 has 3–15 residues. Two different classes of light chains—V_λ and V_κ—show systematic differences in the conformation of L1 and L3.[8,13]

Some of the loops are short hairpins, having conformations that follow principles described previously.[1–9] These include the L2 region, the L3 region of V_λ domains, and in most cases the H2 region.We shall not discuss them further.

It is the conformational determinants of the longer loops that we analyze here. These include L1 of V_λ and V_κ light chains, L3 of V_κ light chains, H1, and H3.

Identification of Regions With Conformations Similar to Antigen-Binding Loops of Immunoglobulins

Using methods similar to those of Jones and Thirup[11] we searched the protein data bank[12] for regions of main chain conformation similar to antigen-binding loops of immunoglobulins of known structure (Table I). For each loop, the program reported regions with rms deviation (△) of the N, C_α, and C atoms ≤ 1.0 Å.

RESULTS

In this section we describe the structures and conformational determinants of individual antigen-binding loops, and compare them with regions of similar conformation in other proteins. In Figures 2a, 3a, 6a–e, 8, 9, and 11 we illustrate these comparisons. In each of these figures the surface of the protein is nearest the viewer and other parts of the protein are below.

L1

The L1 region packs across the top of the V_L domain, connecting strands in different β-sheets. V_λ

TABLE I. Immunoglobulin Variable Domains of Known Atomic Structure

Molecule	Chain type L	Chain type H	Protein data bank designation	Reference
Fab' NEWM	λI	γII	3FAB	31
Fab KOL	λI	γIII	1FB4	32
V_L RHE	λI		2RHE	33
Fab McPC603	κ	γI	1MCP	34
Fab J539	κ	γIII	1FBJ	35
Fab HyHEL-5	κ	γII	2HFL	36
V_L REI	κ		1REI	37

and V_κ domains have very different conformations of L1. In V_λ L1 is 7–8 residues long and includes one turn of helix. It has the unusual feature that a residue from the loop is very deeply inserted into the framework, between the β-sheets.[13] No other structure in the protein data bank contains a loop of similar conformation.

In the known V_κ structures, L1 contains 6 residues (J539 and HyHel-5), 7 residues (REI), or 13 residues (McPC603). Despite the variation in length, they have extended conformations that are closely related in that the determinants of their structures show a common feature: the packing of a large hydrophobic residue at the fourth position of the loop (29_L Val, Ile, or Leu) into a cavity between the β-sheets (Figs. 2b and 3b). (Residues in immunoglobulins are numbered as in Kabat et al.[14] with a subscript L or H for light or heavy chain.)

The stabilization of an extended loop by packing a residue into the protein is one of the two main types of interactions that determine the conformations of medium-sized loops. It is illustrated schematically in Figure 4.

The data base search identified two proteins containing regions of conformation similar to the 6-residue L1 loop of J539 and HyHEL-5, and one containing a region similar to the 7-residue L1 loop of REI (Figs. 2a and 3a).

V_κ domain	L1 length	Other protein	Residues	Δ/Å	Reference
HyHEL-5	6	γ-Crystallin, 1GCR	19–24	0.4	15, 16
HyHEL-5	6	Lobster GPD, 2GPD	117–122	0.4	17
REI	7	T_4 Lysozyme, 2LZM	51–57	0.7	18

No region was found similar in structure to the 13-residue McPC603 L1.

Although the regions identified are similar in conformation to the L1 regions in immunoglobulins, they appear in different structural contexts. In immunoglobulins, L1 joins strands of two β-sheets. In γ-crystallin the region joins a helix and a strand. In lobster glyceraldehyde-3-phosphate dehydrogenase (GPD), it connects two sheets, within a domain of entirely different topology from that of the immunoglobulins. In T_4 lysozyme it links two helices.

Nevertheless, there are notable similarities in the interactions stabilizing these loops. In the regions of γ-crystallin, lobster GPD and T_4 lysozyme having conformations similar to V_κ L1 loops, the packing of the side chain of the fourth residue is similar to that in immunoglobulins. Although the search was conducted using main chain atoms only, the position of the side chain of the fourth residue of the loop is very similar in all cases (Figs. 2 and 3). This residue is smaller than in immunoglobulins (Cys in γ-crystallin and T_4 lysozyme, and Ser in lobster GPD). However, as in immunoglobulins, the side chain of the fourth residue packs into a cavity formed partly by residues in the elements of secondary structure to which the loop is attached, and partly by residues distant in the sequence (Figs. 2c, d and 3c).

In addition to the packing interactions, in all cases the first residue, and except in 2GPD, the N of the fourth residue, form hydrogen bonds (Table II).

Additional, different interactions further stabilize the buried side chain. In γ-crystallin the fourth residue is a Cys that forms a hydrogen bond with the carbonyl of Asn-24, or, in the oxidized form, can make an S–S bridge to Cys-18 with its side chain in a conformation similar to that of the corresponding residue of L1.[15,16] This side chain orientation occurs also in GPD, where the hydroxyl of Ser-120 forms a hydrogen bond to the $O\gamma_1$ of Thr-99.

A common feature of the determinants of the conformation of these loops is that the residue at the fourth position appears especially sensitive, with packing and hydrogen bonding interactions to fix its position. We are led from the general question of what interactions involve the loop, to its restatement: What interactions does a region require to take up a particular conformation, and how may the protein—perhaps in a variety of ways—provide them?

L3

L3 loops of V_κ domains illustrate the second major type of interaction that determines the conformations of medium-sized loops: hydrogen bonding of inward-pointing main chain polar atoms. This is illustrated schematically in Figure 5.

In V_κ domains the most common form of L3 contains six residues—91_L–96_L—in the conformation illustrated in Figure 6a. REI and McPC603 both have L3 loops of this form, very similar in structure. The conformation is determined mainly by the following features: the *cis*-peptide of Pro-95_L, and hydrogen bonds formed by the sidechain of Gln- or Asn-90_L to the main chain of residues within the loop. Kabat et al. observed the conservation in V_κ sequences of Gln-90_L and Pro-95_L, and suggested

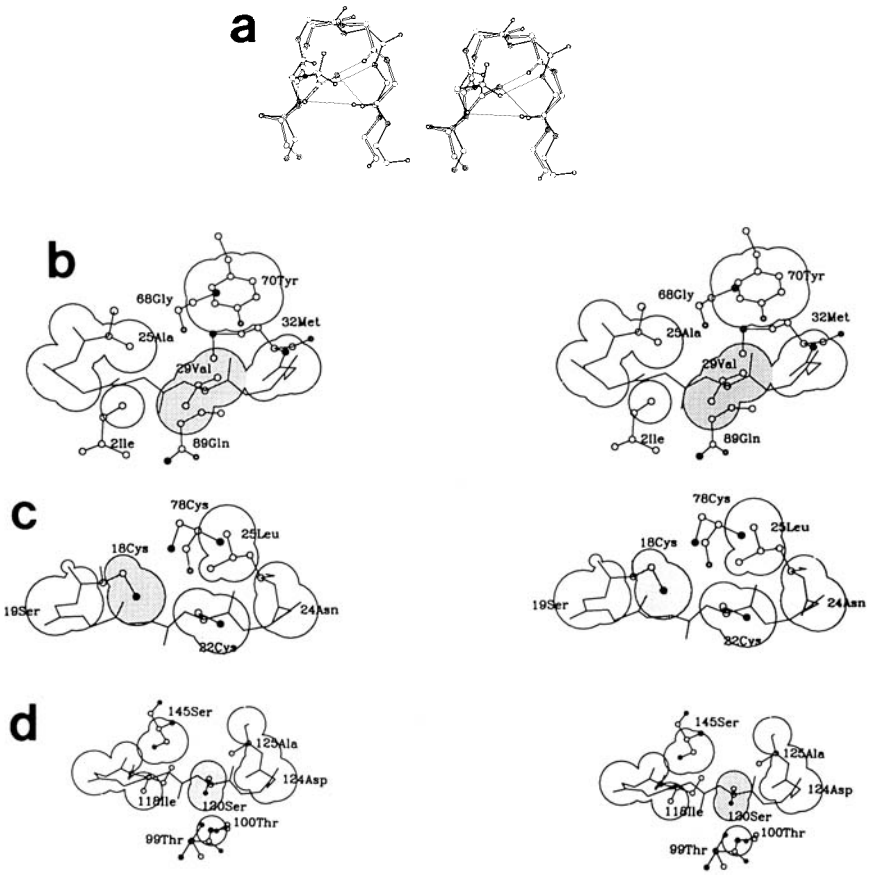

Fig. 2. **(a)** Superposition of L1 from Fab HyHEL-5 and regions of similar conformation from γ-crystallin and lobster GPD. **(b)** L1 of Fab HyHEL-5, showing the packing of Val-29_L. **(c)** Region of similar conformation in γ-crystallin. **(d)** Region of similar conformation in lobster GPD. In **(b–d),** van der Waals contours around the fourth residue are shaded.

that they have structural roles.[19] Of 207 human and mouse V_κ sequences[14] with six-residue L3 regions, 160 contain Gln- or Asn-90_L and Pro-95_L.

A loop of very similar conformation occurs in tomato bushy stunt virus, 2TBV,[20] residues 355–360, which fit L3 of McPC603 with $\triangle$ = 0.7 Å and L3 of REI with $\triangle$ = 0.9 Å. (Fig. 6b,c).

These loops have different structural contexts (Fig. 7). In immunoglobulins, L3 is a hairpin. In the virus the loop connects strands from different sheets (a structural role similar to L1 and H1 in immunoglobulins).

Nevertheless, the structural determinants of the virus loop are very similar to those of the immunoglobulin loop. There is a *cis*-proline at the equivalent position (359 in 2TBV). Hydrogen bonds to the main chain of the loop, similar to those made by Asn-90_L in McPC603, are made in the virus by the amino nitrogen of Ala-331 (Table IIIa), which occupies the same position in space relative to the loop that the side chain of residue 90_L does in V_κ chains (Fig. 6c).

A region with a conformation related to the L3 region of V_κ McPC603 and REI occurs in residues 12–17 of cytochrome c_3 from *Desulfovibrio vulgaris,* 2CDV,[21] $\triangle$ = 0.7 Å and 0.8 Å, respectively (Fig. 6d,e). This region in 2CDV does not have a *cis*-proline. The corresponding residue is Gln, and the peptide is *trans*. However, this causes only a local distortion of the loop. Table IV shows fits that isolate the peptide different in conformation. These regions have similar structural contexts: both are hairpins.

In this case also there is a similarity in the stabilizing interactions. In cytochrome c_3, hydrogen bonds are formed by one of the propionyl groups of a heme (Table IIIb). This carboxyl group occupies the same region of space relative to the loop as the amide group of the Asn of McPC603 (Fig. 6e).

As in the case of L1, the conformation of the loop makes certain demands on the rest of the protein for specific interactions, which can be supplied in a variety of ways.

H1

This loop connects strands in the two β-sheets of V_H domains. Its structural context within its domain is similar to that of L1. In all known structures it is seven residues long: 26_H to 32_H. Typically the

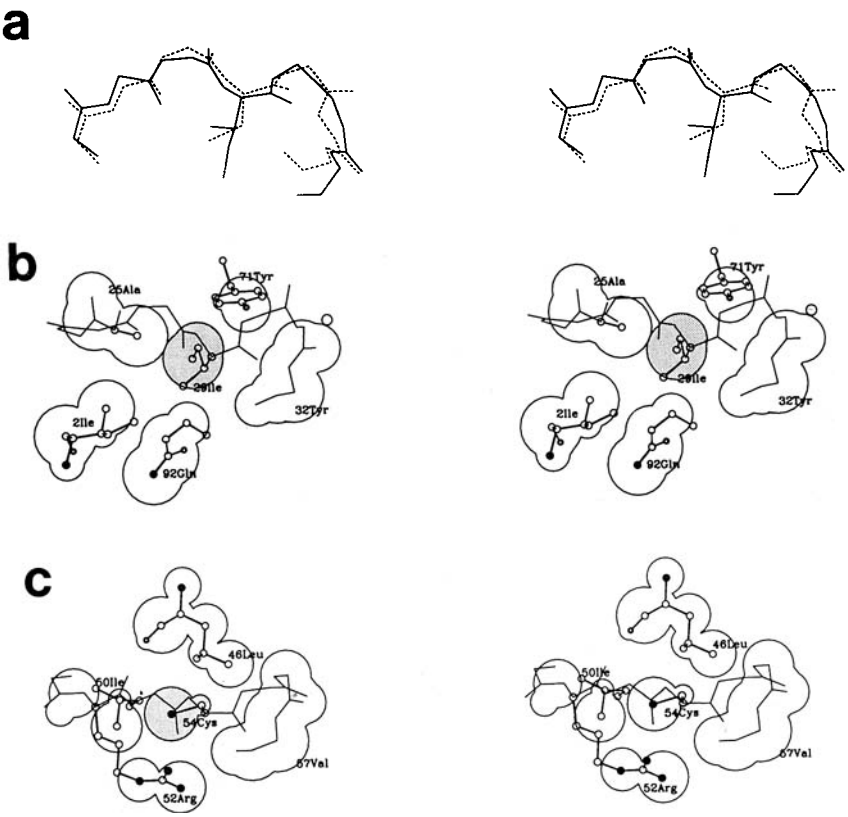

Fig. 3. **(a)** Superposition of L1 from V_κ REI (solid) and the region of similar conformation from T_4 lysozyme (broken line). **(b)** REI L1 and **(c)** a region of similar conformation in T_4 lysozyme, showing similar packing interactions. Corresponding to Ile-2_L and Gln-92_L in REI on one side of the loop is Arg-52 in 2LZM. On the other side, in REI, Tyr-71_L interacts with the fourth residue. At approximately the same relative position Leu-46 packs against Cys-54 in 2LZM. In **(b–c)**, van der Waals contours around the fourth residue are shaded.

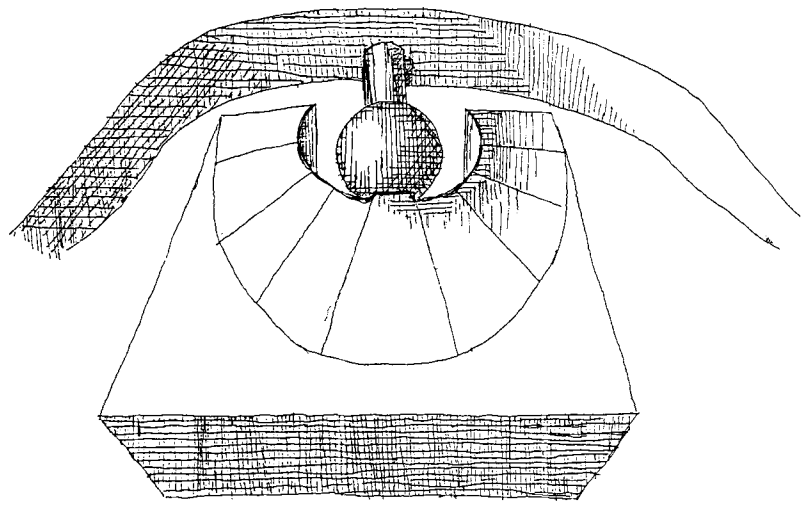

Fig. 4. An extended loop is generally stabilized by the packing of a residue into the protein. Loops L1 and H1 of immunoglobulins, and regions similar in conformation in other proteins, are of this type.

chain is coiled, with residues 28_H–32_H forming a distorted helix: the carbonyl oxygens of residues 28_H and 29_H form hydrogen bonds to the main chain nitrogen of 31_H and 32_H, respectively. A hydrophobic side chain at the fourth position, 29_H, is buried in a cavity between the β-sheets.

TABLE II. Hydrogen Bonds of the Main Chain Atoms of the V_κ L1 Loops of Immunoglobulins and in Loops of Similar Conformation*

Residue	Atom	HY5	J539	γ-Crystallin	2GPD	REI	2LZM
1	*N*	*O* 3_L *Val*	*O* 3_L *Val*	*O* *178 Hoh*	*O* *93 Glu*	*O* 3_L *Gln*	*O* *47 Asp*
	O						
2	N						O 47 Asp
	O				Oγ 113 Ser		
3	N						
	O			Oγ 20 Ser	Oγ1 99 Thr		Nh1 52 Arg
							Nh1 52 Arg
4	*N*	*O* 68_L *Gly*	*O* *837 Hoh*	*O* *78 Cys*		*Oδ2* 28_L *Asp*	*Oδ1* *47 Asp*
						O 68_L Gly	
	O	OH 70_L Tyr		N 78 Cys		N 32_L Tyr	
5	N		Oγ1 92_L Thr			Oδ2 28_L Asp	O 170 Hoh
	O			N 49 Asn		N 68_L Gly	
6	O		N 91_L Trp				N 17 Ile
7	N					O 29_L Ile	
	O					O 91_L Tyr	

*The first column indicates the position of the residue in the loop, and the second column the atom of the loop. Subsequent columns show the atoms to which the loop atoms form hydrogen bonds in the different structures discussed. Lines printed in italics indicate that the nitrogen of the first residue and, except in 2GPD, the nitrogen of the fourth residue form hydrogen bonds in all the structures.

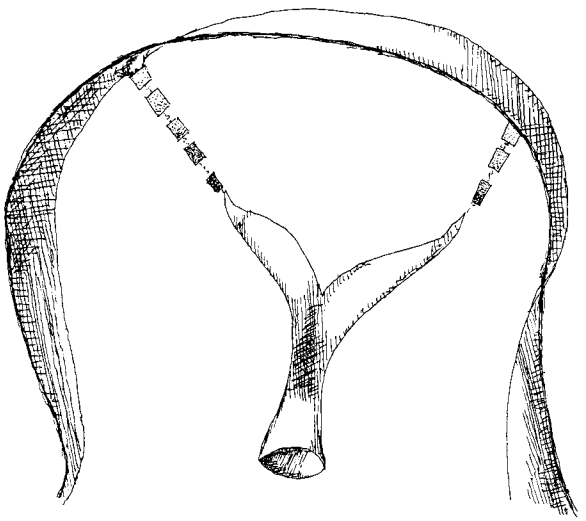

Fig. 5. A compact hairpin loop of medium size is generally stabilized by hydrogen bonding to inward-pointing polar atoms. L3 of immunoglobulins, and regions similar in conformation in other proteins, are of this type.

Residues 111–117 of the insect globin, *Chironomus* erythrocruorin,[22] have the same conformation as the H1 regions of immunoglobulins (Fig. 8a). (Superposition of this loop with the H1 region of KOL gives $\triangle$ = 0.7 Å for *all* main chain atoms.) These loops have a different structural context. In the globin the loop links two helices (G and H) that pack together, with axes approximately antiparallel.

In the globin loop residues 114–116 form a distorted helix with the carbonyl oxygen of residue 114 forming a hydrogen bond to the main chain nitrogen of residue 117. As in the immunoglobulin H1 loop the residue in the fourth position, 114 in erythrocruorin, is a Phe buried in a cavity.

How does nature form cavities to pack the same or similar side chains out of very different structural elements: sheets in the immunoglobulins and helices in erythrocruorin? In KOL the side chain of Phe-29_H is packed between four sets of residues: these include the two strands linked by the loop and additional residues in the region 71_H–78_H (Fig. 8b). In erythrocruorin, Phe-114 also has four sets of neighbors, arising from ridges formed by side chains adjacent on the helix surface (Fig. 8c). In both cases the elements of secondary structure linked by the loop contribute to the packing of the buried residue; in KOL and other immunoglobulins these are supplemented by other residues.

Although in both structures the buried residue is a Phe, mutations in the residues lining the cavity could allow a quite different residue to occupy this site. Indeed, the residues that form the cavities in the immunoglobulins and in the globin are quite different.

Comparison of L1 and H1: Homologous Loops With the Same Size, Related Interactions, but Different Conformations

The L1 region of V_L REI and the H1 regions of V_H domains each has seven residues, link strands in two different β-sheets, and pack the side chain of the fourth residue between the β-sheets. However, their conformations are quite different (Fig. 9), primarily because of a difference in the spatial relationship between the loop and the framework.

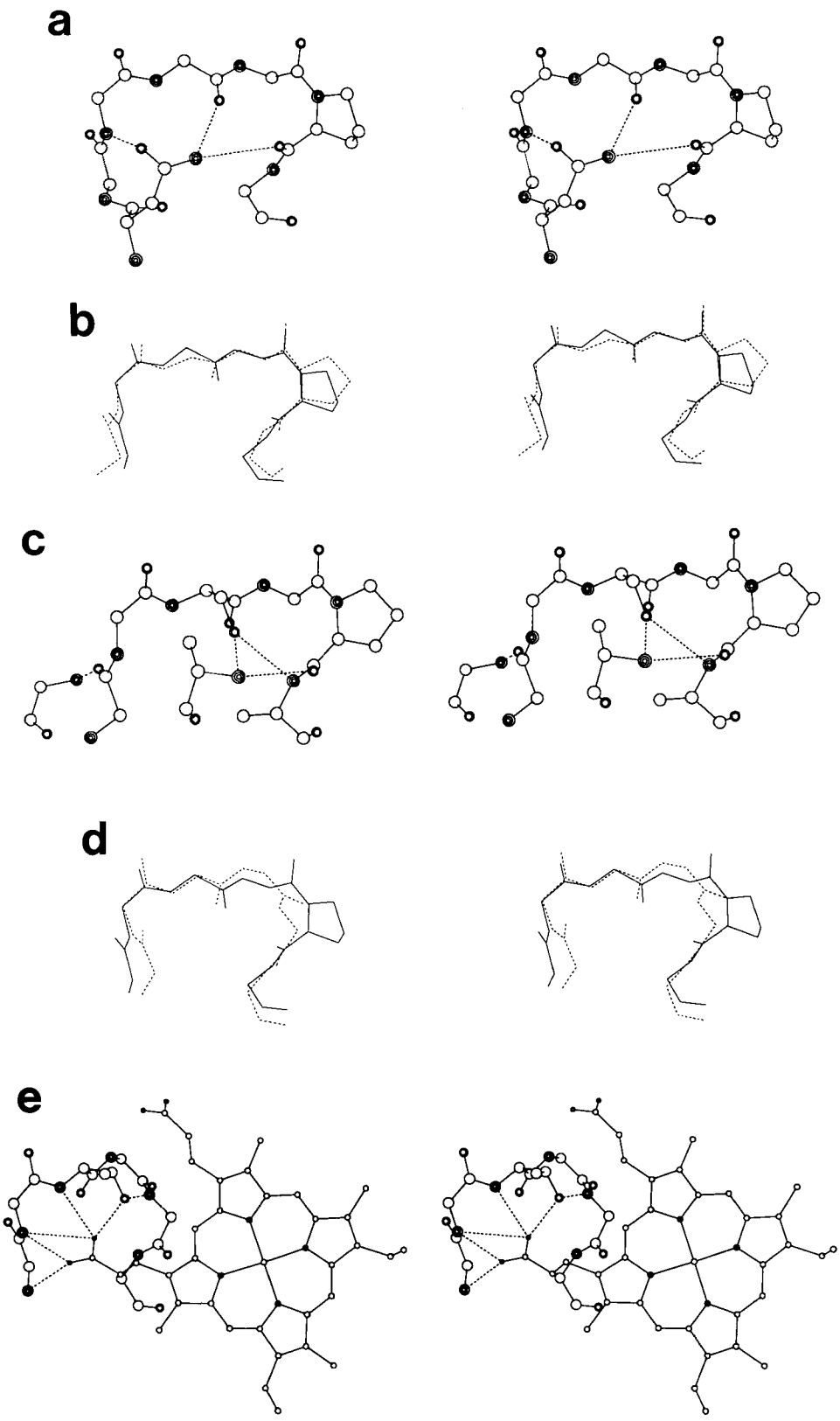

Fig. 6. **(a)** Typical V_κ L3, from McPC603, showing *cis*-proline at position 95_L and hydrogen bonds from Asn-90_L. **(b)** Superposition of L3 from McPC603 (solid) and a region of similar conformation in tomato bushy stunt virus, 2TBV (broken lines). **(c)** Region in tomato bushy stunt virus of conformation similar to L3 of McPC603 and REI, showing *cis*-proline and hydrogen bonding from Ala-331. **(d)** Superposition of L3 from McPC603 (solid) and a region of similar conformation in cytochrome c_3 (2CDV) (broken lines). **(e)** Region in cytochrome c_3 of conformation similar to L3 of McPC603 and REI; in this case there is no *cis*-peptide, and hydrogen bonds are made to a propionyl group of one of the heme groups. In all three cases hydrogen bonding groups occupy the same region of space relative to the loop.

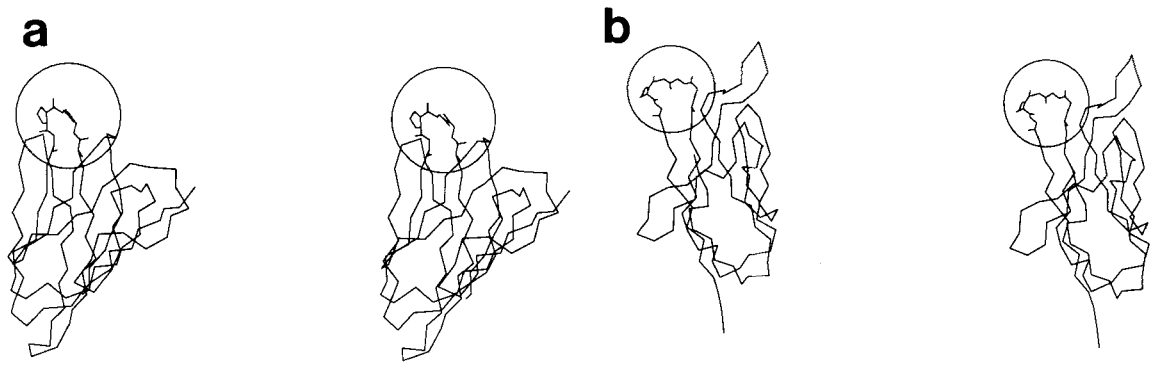

Fig. 7. Difference in structural context between **(a)** McPC603 L3, a hairpin, and **(b)** the region of similar conformation in tomato bushy stunt virus, which links strands between two β-sheets.

TABLE IIIa. Hydrogen Bonds to Corresponding Residues in Regions of Similar Conformation in McPC603 L3 and Tomato Bushy Stunt Virus*

McPC603							2TBV						
91_L	Asp	Oδ1	—	Nδ2	95_H	Asn	355	Val	N	—	O	210	Gly
91_L	Asp	Oδ2	—	Oδ1	95_H	Asn	355	Val	O	—	N	210	Gly
92_L	His	Nδ1	—	Oγ	93_L	Ser							
93_L	*Ser*	*O*	—	*Oδ1*	*90_L*	*Asn*	*357*	*Ser*	*Oγ*	—	*N*	*331*	*Ala*
							357	*Ser*	*Oγ*	—	*N*	*360*	*Ala*
94_L	Tyr	OH	—	Oε2	35_H	Glu							
							359	Pro	O	—	N	331	Ala

TABLE IIIb. Hydrogen Bonds to Corresponding Residues in Regions of Similar Conformation in REI L3 and Cytochrome c_3*

REI							2CDV						
91_L	Tyr	N	—	O	32_L	Tyr	12	Asp	N	—	Oδ1	108	Hem
92_L	Gln	N	—	Oε1	90_L	Gln	13	Lys	N	—	Oδ1	108	Hem
							13	Lys	N	—	Oδ2	108	Hem
93_L	*Ser*	*O*	—	*Nε2*	*90_L*	*Gln*	*14*	*Thr*	*N*	—	*Oδ2*	*108*	*Hem*
							14	*Thr*	*Oγ1*	—	*Oδ2*	*108*	*Hem*
							14	*Thr*	*Oγ1*	—	*N*	*16*	*Gln*
95_L	Pro	O	—	Nε2	90_L	Gln	16	Gln	N	—	Oγ1	14	Thr

*Note that in all four cases the third residue (in italics) is hydrogen bonded, but with residues coming from different parts of the protein, or with a ligand.

TABLE IV. Fits of the N, Cα and C-atoms of selected portions of the L3 loop of REI to the corresponding atoms of residues 12–17 of cytochrome c_3 (2CDV). Arrows indicate the regions used in the fitting; the numbers report the RMS deviation.

REI	2CDV		△/Å		
91_L Tyr	12 Asp				
92_L Gln	13 Lys				
93_L Ser	14 Thr		0.4		
		N			
94_L Leu	15 Lys	Cα			0.7
		C			
				0.4	
95_L Pro	16 Gln				
96_L Tyr	17 Pro				

A comparison of the hydrogen bonding in V_L and V_H domains shows that each of the strands linked by H1 is a residue longer than the strands linked by L1 (Fig. 3 of ref. 7). Therefore the ends of the loops are at different positions in space with respect to the domain: this is shown in Figure 9 in which V_L REI and V_H KOL are superposed on the framework, and the α-carbons of the terminal residues of L1 and H1—26_L and 32_L in REI and 26_H and 32_H—are shaded.

For the 29_H Phe in KOL to pack into the framework in a position similar to that of 29_L Ile of REI, it must descend further into the framework, relative to the ends of the loop, than the corresponding residue in REI. This also necessitates a displacement out of the plane of the loop (Fig. 9b), and is associated with a change in the shape and position of the pocket.

H3

The H3 regions are hairpins. Of all the six antigen-binding loops, they vary most widely in size,

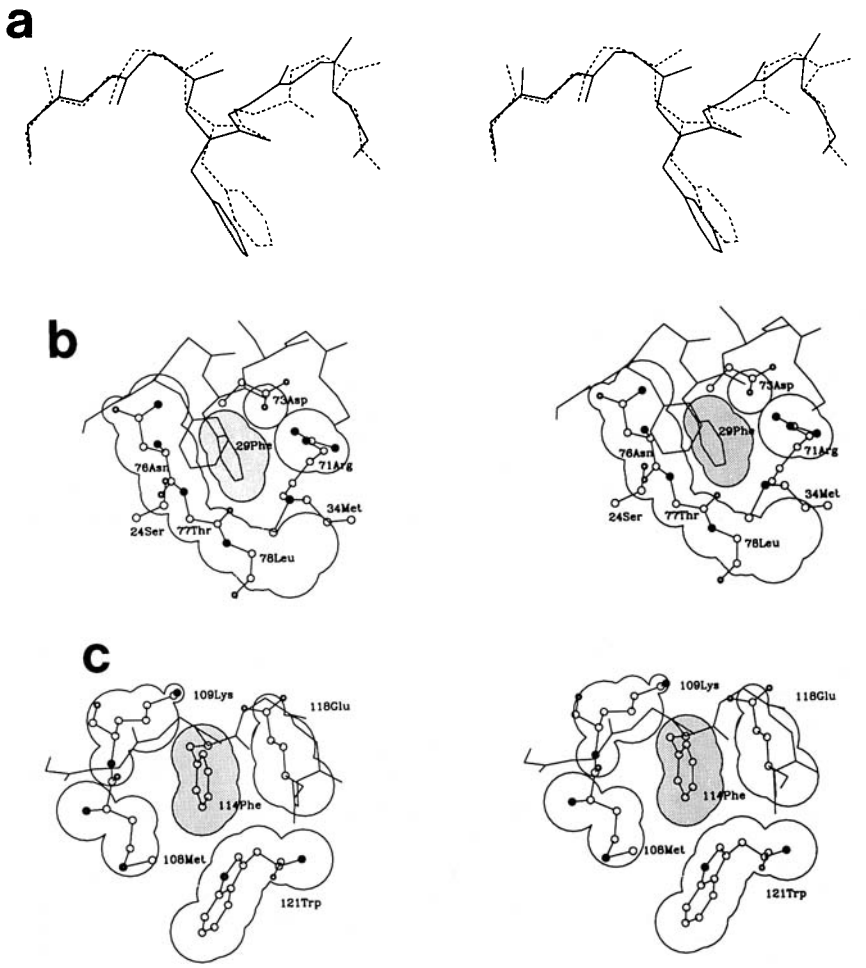

Fig. 8. **(a)** Superposition of H1 from McPC603 (solid) and residues 111–117 of *Chironomus* erythrocruorin (broken lines). **(b)** Packing of residues around Phe-29_H in KOL. **(c)** Packing of residues around Phe-114 in *Chironomus* erythrocruorin. In **(b)** and **(c)** van der Waals contours around the fourth residue are shaded.

sequence, and conformation. No protein in the data bank was found to contain a loop with a conformation as similar to any of the H3 regions in known structures as the loop in tomato bushy stunt virus is to V_κ L3 or the loop in erythrocruorin is to H1. However, the conformation of the loop formed by residues 340–345 in α_1-antitrypsin[23] is closely related to the H3 region of HyHEL-5. Both are hairpins.

In HyHEL-5, the last β-sheet hydrogen bonds occur between residues 94_H and 102_H (Fig. 10). If β-sheet hydrogen bonds occurred also between backbone atoms of residues 96_H and 100_H, the structure would be a two-residue hairpin (see Fig. 10). In α_1-antitrypsin there is one hydrogen bond between the corresponding residues: Asp-341 N · · · O Gly-344. (The loop in α_1-antitrypsin may be thought of as a distorted 2-residue hairpin of the XXXG type.[8]) But in both structures the chain is twisted so that the backbone atoms of the C-terminal part of the loop make hydrogen bonds to a *side chain* in the N-terminal part (Fig. 11). In HyHEL-5, Asn-96_H $N\delta_2$ forms hydrogen bonds to the carbonyl oxygens of Asp-99_H and Phe-100_H. In α_1-antitrypsin the corresponding residue is Asp-341, and with the change in chemical character the hydrogen bond partners change: Asp-341 $O\delta_1$ forms a hydrogen bond to Gly-344 N and Gly-344 O forms a hydrogen bond with Asp-341 N. (Figs. 10 and 11). In α_1-antitrypsin the carbonyl oxygen of Lys-343 points outside the loop, and forms a hydrogen bond to Trp-194. In these structures, different residues occupy nonallowed regions in the Ramachandran diagram, to turn the chain. In α_1-antitrypsin, 344G is in a $\phi > 0$, $\psi < 0$ conformation; in HyHEL-5 Phe-100_H is in a $\phi > 0$, $\psi > 0$ conformation.

Although one face of the loop in α_1-antitrypsin is on the surface, there are very extensive hydrogen-bonding interactions with the rest of the protein. This is true of both the loop in α_1-antitrypsin and that in HyHEL-5. In α_1-antitrypsin, the carbonyl O of every residue between 339 and 346 forms a hy-

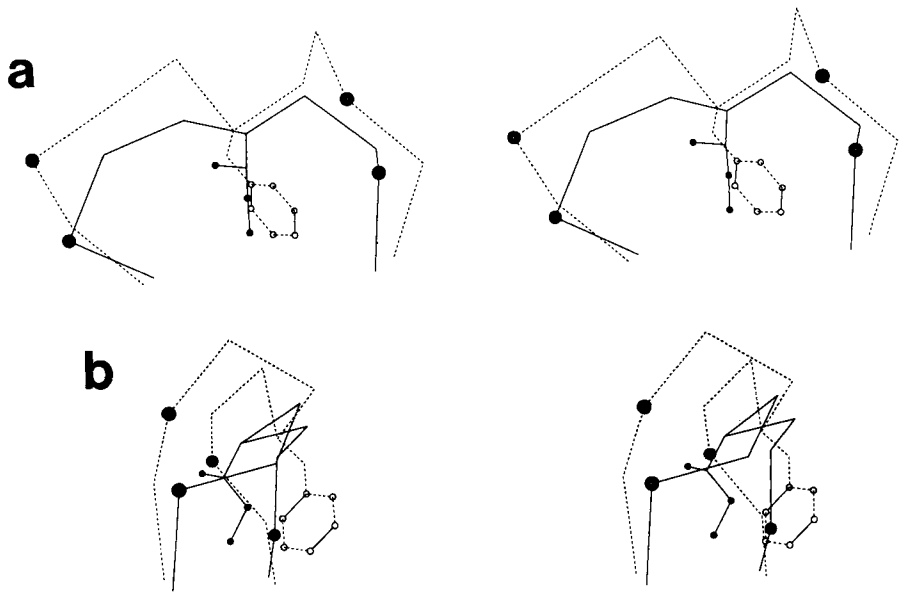

Fig. 9. Comparison of L1 from REI (solid) and H1 from KOL (broken lines), two homologous loops of the same length but different conformations. **(a)** View perpendicular to the plane of the loop. **(b)** View in the plane of the loop.

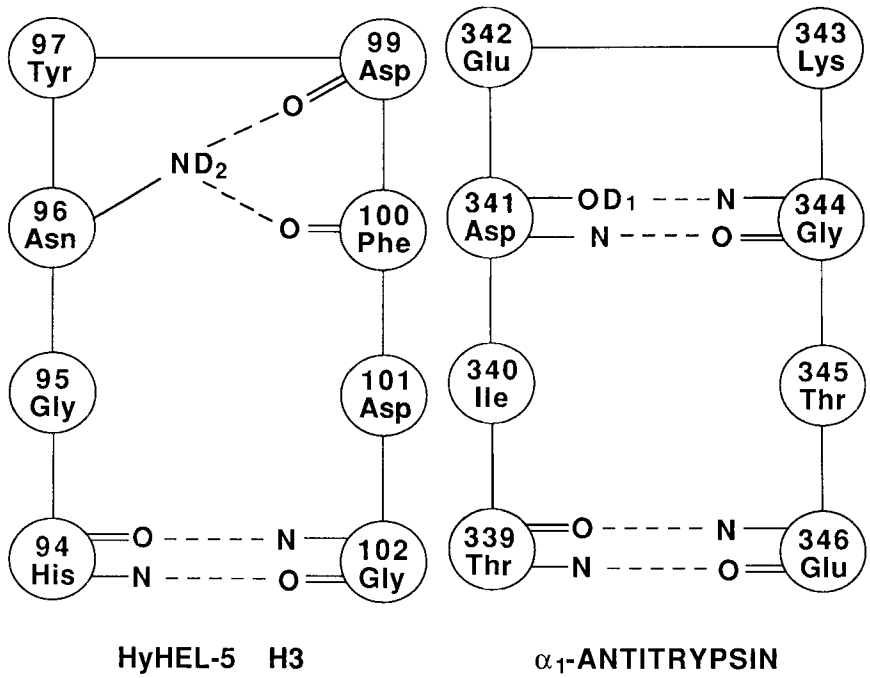

Fig. 10. Hydrogen-bonding in **(a)** H3 from HyHEL-5 and **(b)** a region of related conformation from α_1-antitrypsin, 6API.

drogen bond with another residue, as do the sidechains of Thr-339, Asp-341, Glu-342, Thr-345, and Glu-346. Only the sidechain of Lys-343 does not. In the loop in HyHEL-5, 5 of the 8 carbonyl groups form hydrogen bonds, as do all but one of the charged or polar side chains: His-94_H, Asn-96_H,Tyr-97_H and Asp-101_H. The side chain of Asp-99_H in HyHEL-5, like that of the corresponding residue Lys-343 in α_1-antitrypsin, is not hydrogen bonded to another protein atom.

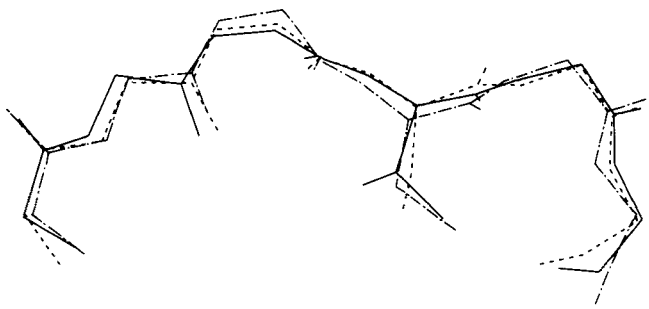

Fig. 11. Superposition of H3 from HyHel-5 and a region of related conformation from α_1-antitrypsin, 6API.

CONCLUSIONS

The results presented here, taken together with previous work,[8] imply certain general rules governing the conformations of medium-sized loops. In discussing these rules it is convenient to distinguish compact loops, with ends close together, from those with ends widely separated.

For loops of more than a few residues, there will be several combinations of allowed values of ϕ and ψ that place their ends in the positions required by the secondary structures that they link.[24–28] The main chain conformation taken up in a native protein structure is determined by the environment of the loop; that is, by hydrogen bonding and packing interactions.

For medium-sized loops that form compact structures, the major conformational determinants are hydrogen bonds to the inward-pointing main chain polar atoms of the loop. These hydrogen bonds can be formed from the loop atoms to main chain or side chain atoms or even ligands, and the atoms that make these hydrogen bonds can be from residues within or adjacent to the loop or can belong to other parts of the protein. Thus, in immunoglobulins, the principal conformation of V_κ L3 is stabilized by hydrogen bonding from a side chain adjacent to the loop. In regions of similar conformation in tomato bushy stunt virus and in cytochrome c_3 from *Desulfovibrio vulgaris,* a related pattern of hydrogen bonds arises from a residue distant in the sequence (in the virus) or from the propionyl group of a heme ligand (in cytochrome c_3). The group that in each case provides these hydrogen bonds occupies the same region in space relative to the loop.

For loops which link more distant secondary structures, the major determinant of the observed conformation is the packing of a particular residue against the rest of the protein. In the cases described here a loop residue is buried in a cavity between major elements of secondary structure. In immunoglobulin L1 and H1 loops the double-β-sheet framework provides the cleft for a side chain. In *Chironomus* erythrocruorin, a corresponding residue packs between two helices. From the point of view of the loop this interaction appears similar, but the protein creates the cavity out of different material.

In some cases, the interactions that create similar loop conformations in unrelated proteins involve analogous residues from the loops, in very similar ways; the similarity in the packing of a Phe, in the H1 loops of immunoglobulins and the G–H loop of erythrocruorin, is a good example of this. Such loops appear to contain specific sites or "nodal points" that require correct positioning and that are the focus of interactions; the interactions of these specific residues appear necessary to fix the conformation of the rest of the region.

On the basis of these results, the nature of the similarity in the interactions made by regions of similar conformation can be stated more precisely: *In many cases the conformation of the loop dictates the stabilizing interactions required, but different proteins may use a variety of topological arrangements in the surroundings of the loops to achieve the correct relative positioning of residues to provide them.*

Exceptions to these statements about the conformational determinants of medium-sized loops may occur in special cases in which the loop is formed by special repetitive sequences; for example, the SPXX sequences found in gene regulatory proteins.[29] However, the generality of our conclusions is supported by an examination of the structures of other loops in other proteins, not discussed here, and by the pro-

tein engineering experiments of Hynes et al.[30] who transferred a five-residue loop from concanavalin A to staphylococcal nuclease. The transferred loop is compact, with hydrogen bonds to inward-pointing main chain atoms formed by residues within the loop itself. Transfer of the loop to the new protein context did not change its conformation significantly.[30]

If, in sets of homologous proteins, the residues responsible for a particular loop conformation are conserved, the loop conformation will be conserved. This provides the basis for predictions of antigen-binding sites, for immunoglobulin sequences can be examined to see whether they contain the set of residues responsible for an observed loop structure. However, it would not be possible to identify medium-sized loops of similar conformation in unrelated proteins from amino acid sequences, because, as we have shown, quite different sets of residues can generate regions of similar conformation. This implies that it is unlikely that rules relating amino acid sequence and conformation of short hairpins can be extended to medium-sized loops.

ACKNOWLEDGMENTS

We thank Dr. C. Slingsby and co-workers at Birkbeck College, S. Sheriff and L. Banaszak for refined coordinate sets, Mr. V. I. Lesk for Figures 4 and 5, our colleagues for discussion, and the Royal Society for support.

REFERENCES

1. Venkatachalam, C. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 6:1425–1436, 1968.
2. Sibanda, B.L., Thornton, J.M. β-Hairpin families in globular proteins. Nature (London) 316:170–174, 1985.
3. Wilmot, C.M., Thornton, J.M. Analysis and prediction of the different types of β-turns in proteins. J. Mol. Biol. 203: 221–232, 1988.
4. Sibanda, B.L., Blundell, T.L., Thornton, J.M. 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, 1989.
5. Efimov, A.V. Standard conformations of polypeptide chains in irregular regions of proteins. Mol. Biol. (USSR) 20:208–216, 1986.
6. Rose, G.D., Gierasch, L.M., Smith, J.A. Turns in peptides and proteins. Adv. Protein Chem. 37:1–109, 1985.
7. Leszczynski, J.F., Rose, G.D. Loops in globular proteins: A novel category of secondary structure. Science 234:849–855, 1986.
8. Chothia, C., Lesk, A.M. Canonical structures for the hypervariable regions of immunoglobulins. J. Mol. Biol. 196: 901–918, 1987.
9. Milner-White, E.J., Ross, B.M., Ismail, R., Belhadj-Mostefa, K., Poet, R. One type of gamma-turn, rather than the other gives rise to chain-reversal in proteins. J. Mol. Biol. 204:777–782, 1988.
10. Chothia, C., Lesk, A.M., Levitt, M., Amit, A.G., Mariuzza, R.A., Phillips, S.E.V., Poljak, R. The predicted structure of immunoglobulin D1.3 and its comparison with the crystal structure. Science 233:755–758, 1986.
11. Jones, T.A., Thirup, S. Using known substructures in protein model building and crystallography. EMBO J. 5:819–822, 1986.
12. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein databank: A computer-based archival file for macromolecular structure. J. Mol. Biol. 112:535–542, 1977.
13. Lesk, A.M., Chothia, C. The evolution of proteins formed by β-sheets. II. The core of the immunoglobulin domains. J. Mol. Biol. 160:325–342, 1982.
14. Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M., Gottesman, K.S. "Sequences of Proteins of Immunological Interest, 4th. ed. Washington D.C.: Public Health Service, N.I.H., 1987.
15. Summers, L., Wistow, G., Narebor, M., Moss, D., Lindley, P., Slingsby, C., Blundell, T., Bartunik, H., Bartels, K. X-ray studies of the lens specific proteins: The crystallins. In: "Peptide and Protein Reviews," Vol. 3. Hearn, M., ed. New York and Basel: Dekker, 1984, 147.
16. Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P., Blundell, T. X-ray analysis of the eye lens protein γ-II crystallin at 1.9 Å *resolution*. J. Mol. Biol. 170:175–202, 1983.
17. Murthy, M.R.N., Garavito, R.M., Johnson, J.E., Rossmann, M.G. The structure of lobster apo-D-*glyceraldehyde*-3-phosphate dehydrogenase at 3.0 Å resolution. J. Mol. Biol. 138:859–872, 1980.
18. Weaver, L.H., Matthews, B.W. Structure of bacteriophage T_4 lysozyme refined at 1.7 Å resolution. J. Mol. Biol. 193: 189–199, 1987.
19. Kabat, E.A., Wu, T.T., Bilofsky, H. Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. J. Biol. Chem. 252:6609–6616, 1977.
20. Olson, A.J., Bricogne, G., Harrison, S.C. Structure of tomato bushy stunt virus. IV. The virus particle at 2.9 Å resolution. J. Mol. Biol. 171:61–93, 1983.
21. Higuchi, Y., Kusunoki, M., Matsuura, Y., Yasuoka, N., Kakudo, M. Refined structure of cytochrome c_3 at 1.8 Å resolution. J. Mol. Biol. 172:109–139, 1984.
22. Steigemann, W., Weber, E. Structure of erythrocruorin in different ligand states refined at 1.4 Å resolution. J. Mol. Biol. 127:309–338, 1979.
23. Loebermann, H., Tokuoka, R., Deisenhofer, J., Huber, R., Human α_1-proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. J. Mol. Biol. 177:531–556, 1984.
24. Gō, N., Scheraga, H.A. Ring closure and local conformational deformation of chain molecules. Macromolecules 3: 178–187, 1980.
25. Moult, J., James, M.N.G. An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. Proteins: Structure, Function Genetics 1:146–163, 1986.
26. Fine, R.M., Wang, H., Shenkin, P.S., Yarmush, D.L., Levinthal, C. Predicting antibody hypervariable loop conformations. II. Minimizing and molecular dynamics studies of McPC603 from many randomly generated loop conformations. Proteins: Structure, Function, Genetics 1: 342–362, 1986.
27. Bruccoleri, R.E., Karplus, M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. Biopolymers 26:137–168, 1987.
28. Bruccoleri, R.E., Haber, E., Novotny, J. Structure of antibody hypervariable loops reproduced by a conformational search algorithm. Nature (London) 335:564–568, 1988.
29. Suzuki, M. SPXX, a frequent sequence motif in gene regulatory proteins. J. Mol. Biol. 207:61–84, 1989.
30. Hynes, T.R., Kautz, R.A., Goodman, M.A., Gill, J.F., Fox, R.O. Transfer of a β-turn structure to a new protein context. Nature (London) 339:73–76, 1989.
31. Saul, F.A., Amzel, L.M., Poljak, R.J. Preliminary refinement and structural analysis of the Fab fragment from the human immunoglobulin New at 2.0 Å. J. Biol. Chem. 253: 585–597, 1978.
32. Marquart, M., Deisenhofer, J., Huber, R., Palm, W. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. J. Mol. Biol. 141: 369–391, 1980.

33. Furey, W., Jr., Wang, B.C., Yoo, C.S. Sax, M. Structure of a novel Bence-Jones protein (Rhe) fragment at 1.6 Å resolution. J. Mol. Biol. 167:661–692, 1983.
34. Satow, Y., Cohen, G.H., Padlan, E.A., Davies, D.R. Phosphocholine binding to immunoglobulin Fab McPC603. An x-ray diffraction study at 2.7 Å. J. Mol. Biol. 190:593–604, 1986.
35. Suh, S.W., Bhat, T.N., Navia, M.A., Cohen, G.H., Rao, D.N., Rudikov, S., Davies, D.R. The galactan-binding immunoglobulin FabJ539: An x-ray diffraction study at 2.6-Å resolution. Proteins: Structure, Function, Genetics 1: 74–79, 1986.
36. Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C., Davies, D.R. Three-dimensional structure of an antibody-antigen complex. Proc. Natl. Acad. Sci. USA 84:8075–8079, 1987.
37. Epp, O., Lattman, E.E., Schiffer, M., Huber, R., Palm, W. The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0 Å resolution. Biochemistry 14:4943–4952, 1975.