Protein Engineering vol.10 no.10 pp.1131–1141, 1997

β-Hairpins in proteins revisited: lessons for *de novo* design

K.Gunasekaran, C.Ramakrishnan and P.Balaram[1]

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

[1]To whom correspondence should be addressed

β-Hairpins with short connecting loops (1–5 residues) have been identified from a data set of 250 non-homologous, high resolution (≤2.0 Å) protein crystal structures. The conformational preferences of the loop segments have been analyzed with the specific aim of identifying frequently occurring motifs. Type I′ and II′ β-turns were found to have a high propensity for occurrence in two residue loops. For three and four residue loops, the major conformational motif in the linking segments is α_R-α_R-α_L (type I β-turn followed by a residue in a left-handed helical conformation) and α_R-α_R-α_R-α_L (a π-turn motif), respectively. The present larger data set confirms the high occurrences of these motifs which have been identified in earlier analyses. In addition to type I′ and type II′ β-turns, several examples of type I β-turn nucleated two residue loop hairpins, in spite of having an opposing sense of twist to that of type I′ β-turn, have also been observed. Examination of these frequently occurring motifs (flanked by extended conformation [β]) in the data set reveals that the motifs β-α_R-α_R-α_L-β and β-type I′-β have equal propensity and type II′ indeed having highest propensity to nucleate β-hairpins. The larger number of examples in this study allows the estimation of the specific amino acid preferences for loop positions in two, three and four residue loops. Small polar residues Asn, Asp, Ser, Thr, Gly and Pro in general have a high propensity for the loop positions but they reveal specific positional preferences in these frequently occurring motifs. There are no strong compositional preferences in the strand segments. Amino acid pair correlations across strands also do not show any significant pattern, with the exception of Cys–Cys pairs. Several Cys–Cys pairs have been identified at the non-hydrogen bonded positions of β-hairpins; as many as six are disulfide bonded pairs. An examination of longer loop length hairpins reveals that the distortions of hairpins nucleated by tight turns (two residues) are much less frequently observed. The results presented in this study provide inputs for the *de novo* design of consensus loop segments in synthetic hairpins.

Keywords: β-hairpins/hairpin design/protein data analysis/short loop motifs/turns in proteins

Introduction

De novo protein design approaches attempt to construct novel polypeptide sequences that fold into well defined secondary and tertiary structures resembling native globular proteins (DeGrado, 1988; Richardson *et al.*, 1992; Betz *et al.*, 1993; Kametkar, *et al.*, 1993) The success of these studies relies heavily on the ability to design relatively short stretches of polypeptides that can adopt stable secondary structures. Helices have been the most widely studied class of secondary structures (Barlow and Thornton, 1988; Presta and Rose, 1988; Richardson and Richardson, 1988; Nagarajaram, *et al.* 1993; Aurora *et al.*, 1994; Seale *et al.* 1994) and the factors that stabilize helical folding patterns have been extensively investigated experimentally (Lyu *et al.*, 1990; O'Neil and DeGrado, 1990; Bruch *et al.*, 1991; Lyu *et al.*, 1992; Scholtz and Baldwin, 1992; Padmanabhan and Baldwin, 1994; Baldwin, 1995; Chakrabartty and Baldwin, 1995; Doig and Baldwin, 1995; Munoz and Serrano, 1995). β-Hairpins, on the other hand, have only been the focus of several recent synthetic studies (Blanco *et al.*, 1993, 1994; Haque *et al.*, 1994; Alba *et al.*, 1995; Awasthi *et al.*, 1995; Constantine *et al.*, 1995; Karle *et al.*, 1996b; Ramirez-Alvarado, *et al.*, 1996; Sieber and Moe, 1996; Struthers *et al.*, 1996), most of which are based on early analyses of β-hairpin conformations in protein structures (Lifson and Sander, 1979, 1980; Sibanda and Thornton, 1985, 1991, 1993; Sibanda *et al.*, 1989). The major feature to emerge from the classical analysis carried out by Sibanda and Thornton (1985) was that β-hairpins in proteins are frequently nucleated by type I′ or type II′ β-turns (Venkatachalam, 1968). In both of these tight turns the dihedral angle ϕ_{i+1} is positive, thus greatly restricting the choice of amino acids that may be placed at this position in β-hairpin design, since L-amino acids do not have an appreciable preference for positive φ values in Ramachandran space (Ramachandran *et al.* 1963; Ramakrishnan and Ramachandran, 1965). Indeed, as a consequence of the analysis, D-proline for which φ is ideally restricted to positive values (60 ± 20°), has been successfully employed for hairpin nucleation (Richardson *et al.*, 1992). The current availability of a much larger database of high resolution protein crystal structures, as compared with earlier studies (Sibanda and Thornton, 1989), prompted a re-examination of β-hairpin conformations in proteins, with a view towards expanding the scope and nature of connecting loops that may be used in synthetic hairpin design. The results presented in this paper suggest that the rational design of loops with more than two residues may indeed be possible. This analysis also highlights several important features of nucleating loop conformations and inter-strand residue recognition.

Methods

A data set of 250, largely non-homologous, high-resolution (≤2.0 Å) protein structures from the Brookhaven Protein Data Bank (PDB) (Bernstein *et al.*, 1977) was examined. The data set consisted of the PDB entries given in Table I (polypeptide chain identifiers are indicated wherever homologous multiple chains are present).

The strands that form anti-parallel β-sheets are picked up by an algorithm that uses virtual bond angles, virtual torsion angles and the end-to-end distance of the C^α positions of a tripeptide (C^α_i to C^α_{i+3}) segment (Ramakrishnan and Soman, 1982; Soman and Ramakrishnan, 1986). The selected anti-

Table I. The data set examined

1AAN	1AAZ A	1ABE	1ABK	1ACF	1ACX	1AFG A
1AHC	1AK3 A	1ALC	1ALD	1ALK A	1AMP	1ANK A
1AOZ A	1APM E	1ARB	1ARP	1ARS	1AST	1BBH A
1BBP A	1BGC	1BGH	1BMD A	1BRS D	1BSA A	1BYB
1CBN	1CCR	1CEW I	1CGT	1CHM A	1CMB A	1COT
1CPC A	1CPC B	1CPN	1CSE E	1CSE I	1CSH	1CTF
1CUS	1DDT	1DFN A	1DMB	1DRI	1DSB A	1ECA
1ESL	1EZM	1FAS	1FDN	1FGV H	1FIA A	1FKF
1FLP	1FLV	1FNA	1FRR A	1FUS	1FX1	1FXD
1GD1 O	1GIA	1GKY	1GLQ A	1GLT	1GOG	1GOX
1GP1 A	1GPR	1HEL	1HIP	1HLE A	1HLE B	1HOE
1HPI	1HSB A	1HSB B	1HSL A	1HUW	1HVK A	1HYP
1IAG	1IFB	1ISA A	1ISU A	1LCF	1LEC	1LIB
1LIS	1LLD A	1LTS A	1LTS C	1LTS D	1MBA	1MBD
1MDC	1MJC	1MOL A	1MPP	1NAR	1NBA A	1NLK R
1NPC	1NSC A	1OLB A	1ONC	1OPA A	1OVA A	1PDA
1PGB	1PHC	1PHP	1PII	1PK4	1PMY	1POC
1POH	1PPA	1PPB H	1PPB L	1PPF E	1PPT	1PRN
1PTF	1PTS A	1R69	1RBP	1RDG	1REC	1RIS
1RNH	1ROP A	1SAC A	1SBP	1SGT	1SHA A	1SHF A
1SHG	1SIM	1SLT A	1SMR A	1SRD A	1STN	1TCA
1TEN	1TFG	1TGN	1TGS I	1TGX A	1THB A	1TML
1TON	1TRB	1TRK A	1UBQ	1UTG	1WHT A	1WHT B
1XIB	1YPI A	256B A	2ACQ	2ACT	2ALP	2APR
2BBK H	2BBK L	2BMH A	2CAB	2CCY A	2CDV	2CHS A
2CI2 I	2CMD	2CPL	2CTV A	2CY3	2CYP	2END
2FCR	2GBP	2GST A	2HAD	2HBG	2HMQ A	2LH7
2LHB	2LTN A	2LTN B	2LZM	2MCM	2MLT A	2MNR
2MSB A	2OHX A	2OVO	2PAB A	2PIA	2PLT	2POR
2PRK	2RHE	2RSP A	2SAR A	2SCP A	2SGA	2SN3
2SPC A	2TRX A	2TSC A	2WRP R	2ZTA A	351C	3APP
3B5C	3BCL	3BLM	3C2C	3CHY	3CLA	3COX
3DFR	3DNI	3DRC A	3EBX	3EST	3GRS	3IL8
3MDS A	3PSG	3RP2 A	3RUB L	3RUB S	3SDH A	3TGL
4AZU A	4BP2	4CPV	4ENL	4FXN	4GCR	4I1B
4ICB	4INS C	4INS D	4MT2	4TNC	5CHA A	5CPA
5FD1	5P21	5PTI	5RUB A	6LDH	7ACN	7RSA
8DFR	8FAB A	8FAB B	9WGA A			

parallel strand segments were further examined for backbone dihedral angles; a minimum length of four residues in the extended conformation (ϕ = –180 to –30° and ψ = 60 to 180° and –180 to –150°) must be present in each strand (Sowdhamini *et al.*, 1992). The inter-strand registering of residues was identified by examining the hydrogen bonding pattern (using the criterion that the N–O distance lies between 2.5 and 3.5 Å) (Baker and Hubbard, 1984). The number of residues in the loop region was counted as the number of intervening residues connecting the last hydrogen bonding pair (Milner-White and Poet 1986; Pavone, 1988). For example, for the definition of a two residue loop, a hydrogen bond must be present between the N–H of the amino acid residue B + 1 and the C–O of the amino acid residue B – 1 (4 → 1 or N–O type) or between the C–O of the amino acid residue B + 1 and the N–H of the amino acid residue B – 1 (1 → 4 or O–N type) or both (4 → 1 and 1 → 4). There should be both N–O and O–N type hydrogen bonds observed between the residues B – 3 and B + 3. The definition of loop residues for the two residue loop class is shown in Figure 1a. Figure 1b shows a flow chart of the scheme used in the selection of β-hairpins. In an earlier analysis (Sibanda *et al.*, 1989), a distinction was made based on the hydrogen bonding pattern that occurs between the B – 1 and B + 1 residues (if 4 → 1 and 1 → 4 was present then the defined two residue loop class is called 2:2 class and if only 1 → 4 was present then it is called 2:4 class. However, if only a 4 → 1 hydrogen bond was present then it will fall in neither of the two classes). We have not made such a distinction as frequently occurring motifs in the loops of β-hairpins mostly fall into one group. The length of strands is also determined by the successive hydrogen bonds down the β-ladder.

Results and discussion

β-Hairpins were identified from a data set of 250, largely non-homologous, high resolution (≤2.0 Å) protein structures. Figure 2 shows the histogram representing the distribution of hairpins classified on the basis of the length of the connecting loops. Approximately 60–70% of the examples in the data set correspond to short loops (≤5 residues). Consequently, the subsequent analysis has been restricted to loop lengths ≤5 residues. Loop segments were examined for the occurrence of classical reverse turn conformations. The results are summarized in Table II, which also provides a comparison with the earlier analysis of Sibanda *et al.* (1989). The data set used in the present study affords a much larger number of examples and provides new insights into the stereochemistry of loop segments in β-hairpins as detailed below.

Loop stereochemistry

One residue loops. One residue loops were identified on the basis that either 3 → 1 or 1 → 3 hydrogen bonds are present at the turn segment. Of the seven examples identified, only one example (2SGA 90–105) corresponds to a classical γ-turn

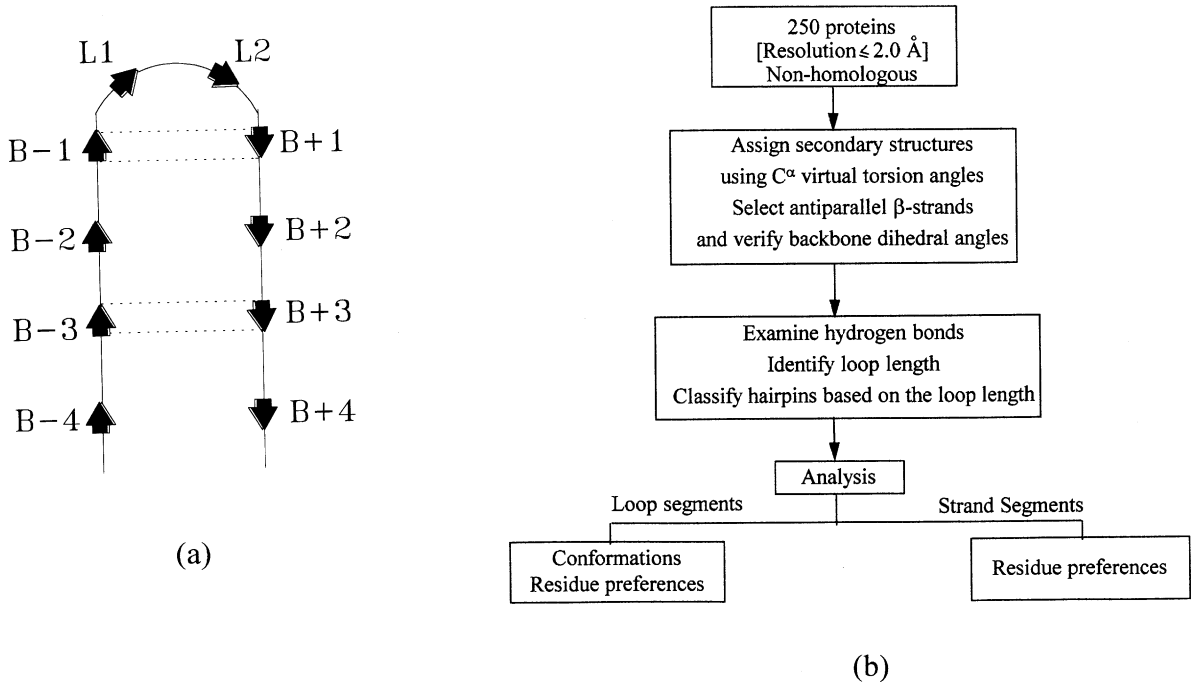

Fig. 1. (a) Definition of a two residue loop β-hairpin (with a strand length of four residues) used in the selection procedure. The strand and loop positions are designated as B ± *x* and L**x**, respectively. A hydrogen bond of the type N–O or O–N or both must be present between the residues B + 1 and B – 1 and there should be both N–O and O– N type hydrogen bonds observed between the residues B + 3 and B – 3 for selection (see Methods). (**b**) Flow chart of the scheme used in the identification of hairpins in a data set consisting 250 protein crystal structures (see Methods for the data set).

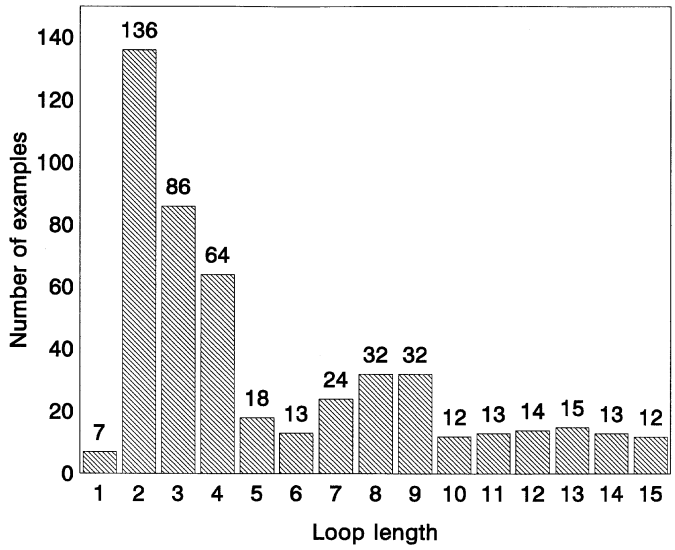

Fig. 2. Distribution of hairpins selected using the scheme shown in Figure 1b and classified on the basis of number of residues in the loop region. (see Figure 1a for the definition of the loop region in the case of two residue loops).

conformation (Mathews, 1972; Milner-White *et al.*, 1988; Milner-White, 1990). In all the other cases, analysis of the conformational angles of residues *i* – 1, *i* and *i* + 1 did not reveal any preference for a specific motif. It is likely that many of the examples in this category may be more approximately classified under five residue loops, because of the distortions near the turning segments. γ-Turns with both 3 → 1 and 1 → 3 hydrogen bonds do not appear to be important elements in β-hairpins.

Two residue loops. By far the most abundant connecting elements in β-hairpins are two residue loops, the overwhelming majority of which adopt classical β-turn conformations. The pioneering analysis of Sibanda and Thornton (1985) leads to the conclusion that type I′ and II′ β-turn conformations are very strongly preferred in two residue loops. The results in Table II confirm these observations. In addition, it is observed that hairpins incorporating type I β-turns are also fairly widespread with as many as 29 examples being identified in the present data set. Further, four examples of hairpins nucleated by type II β-turns are also observed. Figure 3 illustrates hairpins formed with I and I′ β-turns. Sibanda and Thornton (1985) concluded that the preference for the type I′ β-turns was a consequence of compatibility of the twist in the strand segment with the twist of the turns ('... type I′ and II′ turns give acceptable hairpins, whilst the strands rapidly diverge when a type I or II turn is included'). The twist of the turns can be estimated by examining the virtual torsion angle (θ) defined by C^{α}_{B-1}, C^{α}_{L1}, C^{α}_{L2} and C^{α}_{B+1} for the two residue loops (θ is –50° for type I′ and +50° for type I ideal β-turns). The distribution of virtual torsion angles (θ) for the two residue loop class is shown in Figure 4. It is evident that, although the distribution shows a larger number of examples occurring in the interval –60 to –20°, several examples are found to have positive values for θ in the two residue loop class. An estimate of average value of θ for the frequently occurring motifs in the two residue loop class is given in column 4 of Table III and, as expected, it is positive for type I and negative for type I′ β-turns. The observation of a large number of type I β-turn containing hairpins prompted us to examine the distribution of strand length. The results are summarized in Figure 5. Interestingly, there are only two examples (out of 29) of hairpins with a strand length of four residues which are nucleated by type I β-turn compared with 23 examples (out

Table II. Conformational preferences of loop residues in β-hairpins

Loop size[a]	Turn type[b]	Present study	Sibanda *et al.* (1989)	Preferred motif[c]	Hydrogen bonding pattern[g]
One residue	γ	1	0	–	3→1 & 1→3 (2)
	Total	7	0		3→1 (3)
					1→3 (2)
					4→1 & 1→4 (98)
Two residues	β_I	29	7	α_L-α_L (53)	
	β_{II}	4	0	E'-α_R (37)	
	$\beta_{I'}$	53	18	α_R-α_R (30)	4→1 (2)
	$\beta_{II'}$	34	11		
	Others[d]	16	0		1→4 (36)
	Total	136	36		
Three residues	β_I-*x*[e]	54	10	α_R-α_R-α_L (52)	5→1 & 1→5 (10)
	x-β_I	10	–		
	β_{II}-*x*	2	–		5→1 (0)
	x-β_{II}	4	–		
	$\beta_{I'}$-*x*	5	–		1→5 (76)
	x-$\beta_{I'}$	2	–		
	Others[d]	9	7		
	Total	86	17		
Four residues	β_I-*x*-*x*	42	5	α_R-α_R-α_R-α_L (36)	6→1 & 1→6 (31)
	Various turns[f]	10	–		
	Others[d]	12	2		6→1 (0)
	Total	64	7		1→6 (33)
Five residues	Various turns[f]	11	–	–	7→1 & 1→7 (3)
	Others[d]	7	5		7→1 (0)
	Total	18	5		1→7 (15)

[a]See Figure 1a for the definition of loop residues.
[b]β-Turns were identified based on the backbone dihedral angles (φ, ψ) at the central positions ($i + 1$ and $i + 2$) of the turn. Note that the angles can vary up to ±30° from the standard values suggested by Venkatachalam (1968). However, a deviation up to ±45° in only one of the four torsion angles is allowed. For the present study, type I and III β-turns are grouped together as type I; similarly, types I′ and III′ are grouped as type I′.
[c]Numbers in parentheses are the number of examples in the present data set. [α_R (φ = –140° to –30° and ψ = –90° to 45°), α_L (φ = 20° to 125° and ψ = –45° to 90°) and E' (φ = 30° to 180° and ψ = –180° to –60°)].
[d]Loop residues are not in any β-turn conformations.
[e]Various conformations.
[f]Different types of β-turns were found to occur in various positions of the loop segment.
[g]Hydrogen bonding pattern near the loop segment; number of examples having hydrogen bonds between the residues B – 1 and B + 1 are given in parentheses. The hydrogen bond occurs between the N–H of residue B – 1 and C–O of residue B 1 (4→1, in the case of two residue loops) or between the C–O of residue B – 1 and N–H of residue B + 1 (1→4, in the case of two residue loops) or both (4→1 and 1→4) types.

of 53) of type I′ β-turn. Therefore, hairpins nucleated by type I β-turns have a significantly greater strand length than those formed by type I′ β-turns. This observation suggests that inter-strand hydrogen bonding may indeed compensate for any unfavorable interactions involved in adjusting the strand stereochemistry.

Three residue loops. Most three residue loops incorporated type I β-turns. As many as 52 examples had the conformational motif α_R-α_R-α_L, where residues 1 and 2 in the loop formed type I β-turns, with residue 3 lying in the left-handed α-helical region. This motif can also be termed a G1 bulge associated with a type I β-turn (Richardson, 1981). There are very few examples of type I′ and II′ β-turns in three residue loops. In contrast to the two residue loop class, where the type I′ and type II′ are preferred over the type I and type II β-turns in the ratio of 1:3, the three residue loop class prefers type I and type II over the type I′ and type II′ β-turns in the ratio of 10:1. Hence, conformational requirements for three residue loops appear to be more sharply defined than two residue loops. Figure 6 illustrates a typical example of a hairpin nucleated by a three residue loop. The Type I-α_L motif also has left-handed twist, as can be seen by defining the virtual torsion angle ($\theta = C^{\alpha}_{B-1}$–$C^{\alpha}_{L1}$–$C^{\alpha}_{L3}$–$C^{\alpha}_{B+1}$). The average virtual torsion angle for this motif, calculated using the 52 examples of type I-α_L, is –31° and the distribution for the entire three residue loop class is also shown in Figure 4. It is interesting that this motif also has a twist comparable to the twist of type I′ turns which also have a high occurrence. These observations suggest that the type I-α_L motif may also be a strong β-hairpin nucleator. Recently, the amino acid residues in the loop region of a β-hairpin forming 16 residue peptide derived from the N-terminal sequence of ubiquitin were modified in order to maximize the probability of forming a β-turn (Searle *et al.*, 1995). The three residue loop (Leu-Thr-Gly) was mutated to a two residue loop (Pro-Asp). Surprisingly, the type I-α_L conformation which was present in the native sequence was re-established by adjusting the strand registering with a three residue loop (Pro-Asp-Gly) in the analogous peptide. This experimental evidence also supports that the type I-α_L motif may be a stronger β-hairpin nucleating motif.

In order to compare the propensity for nucleation of β-hairpins of type I-α_L motif with other frequently occurring motifs, their occurrence with β (extended conformation) as the flanking conformation in protein crystal structures was

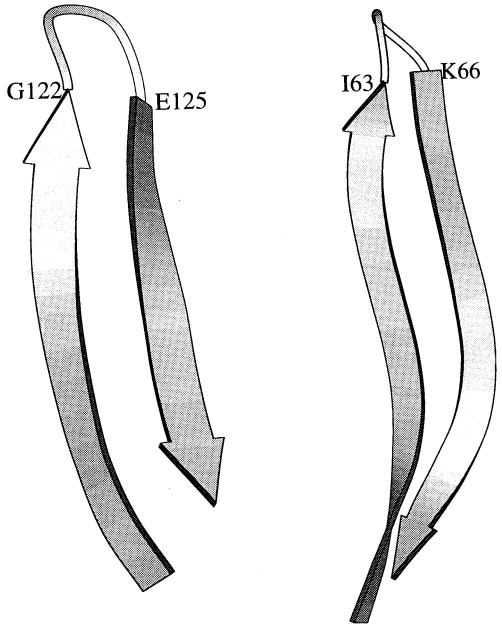

Fig. 3. Examples of two residue loop hairpins nucleated by (**a**) type I and (**b**) type I′ β-turn conformations. (**a**) Residues 115–131 of neuraminidase (1NSC A). (**b**) Residues 53–75 of bilin binding protein (1BBP A). The ribbon diagrams were prepared using the program MOLSCRIPT (Kraulis, 1991).

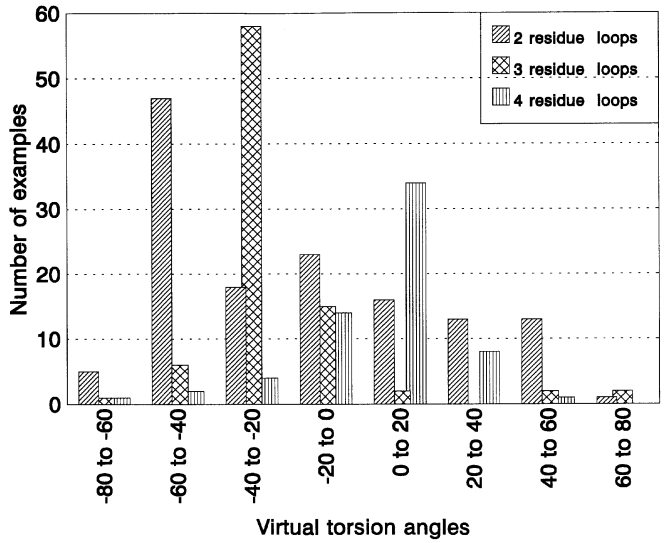

Fig. 4. Histogram showing the distribution of virtual torsion angles in the two, three and four residue loop classes. The virtual torsion angle is defined by the four C^{α} positions, for two residue loops: C^{α}_{B-1}–C^{α}_{L1}–C^{α}_{L2}–C^{α}_{B+1}; for three residue loops: C^{α}_{B-1}–C^{α}_{L1}–C^{α}_{L3}–C^{α}_{B+1}, and for four residue loops: C^{α}_{B-1}–C^{α}_{L1}–C^{α}_{L4}–C^{α}_{B+1}.

examined. The propensity (*P*) of the preferred motifs (m) which occur frequently in the β-hairpins {for two residue loops, type I, type I′ and type II′ β-turns; for three residue loops, type I -α_L; for four residue loops, α_R-α_R-α_R-α_L [see the next section (four residue loops)]} were calculated using the equation

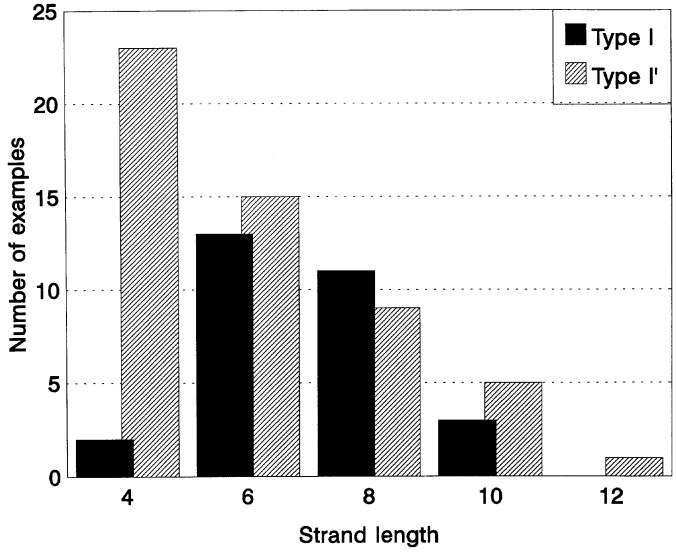

Fig. 5. Histogram showing the strand length distribution for the two residue loop β-hairpins. The distribution is shown separately for the loop region containing type I and type I′ β-turn conformations.

$$P_m = \frac{H_m}{\sum_{m=1}^{5} H_m} \Big/ \frac{D_m}{\sum_{m=1}^{5} D_m}$$

where H_m is the number of times a motif occurs in the loop segments of β-hairpins and D_m is the number of times the motif occurs in the data set with extended conformation in the flanking positions (extended conformation was identified using the limit: ϕ = –30 to –180° and ψ = 60 to 180° and –180 to –150°). There were 138 examples of the β-type I-α_L-β motif found in the data set, of which 52 nucleated β-hairpins, and there were 135 examples of β-type I′-β motif, of which 53 nucleated β-hairpins. Consequently, the propensities for these two motifs to nucleate β-hairpins were comparable (2.377 for β-type I-α_L-β and 2.476 for β-type I′-β) and β-type II′-β had the highest propensity (3.021).

Four residue loops. In contrast to the earlier analysis, the present study reveals a larger number of examples which could be classified as β-hairpins with four residue loops. A very large number of these examples contain overlapping type I/ type III β-turns corresponding to a single turn of a 3_{10}-/α-helical segment. Interestingly, the most widespread conformational motif in the four residue loop segment was of the type α_R-α_R-α_R-α_L (36 examples). This motif has been identified at the C-terminus end of helices in proteins, with the α_L position invariably being occupied by amino acids Gly or Asn (Schellman, 1980; Nagarajaram *et al.*, 1993). Further, a 6 → 1 hydrogen bond between the N–H of the amino acid residue following the α_L residue (B + 1) and the C–O of the amino acid preceding the first α_R residue (B – 1) is observed (the motif containing 6 → 1 and 5 → 2 hydrogen bonds is also termed a Schellman motif; if the hydrogen bond occurs between the C–O of B + 1 and the N–H of B – 1 then it is termed a 1 → 6 type hydrogen bond). These features may also be termed π turns because of the formation of a 16-membered (C_{16}:6 → 1) hydrogen bond (Nagarajaram, 1995; Rajashankar and Ramakumar, 1996). In the case of four residue loops in β-hairpins, of the 36 examples of this conformational motif, 23 have 6 → 1 hydrogen bonds (N–O distance ⩽3.5 Å), while

Table III. Stereochemistry of the preferred motifs which occur frequently in β-hairpins[a]

Loop motif (turn)	Mean backbone dihedral angles (°)	Distance (C^{α}_{B-1}–C^{α}_{B+1}) (Å)	Virtual torsion angle[b] (°)	Hydrogen bonding patterns near the loop	Preferred amino acids[c]
α_R-α_R (β_I)	L1: –63(12), –32(13) L2: –104(20), –15(22)	5.38(0.25)	29(23)	4→1 & 1→4: 9 4→1: 1 1→4: 20	L1: Pro, Asp, Trp, Asn L2: SeR, Phe, Lys
α_L-α_L ($\beta_{I'}$)	L1: 53(10), 43(10) L2: 77(11), 1(17)	5.33(0.17)	–51(7)	4→1 & 1→4: 50 4→1: 0 1→4: 3	L1: Asn, His, Gly, Asp L2: Gly
E'-α_R ($\beta_{II'}$)	L1: 63(30), –115(15) L2: –94(15), –0(15)	5.28(0.25)	–11(15)	4→1 & 1→4: 34 4→1: 0 1→4: 3	L1: Gly L2: Asn, Asp, Ser
α_R-α_R-α_L (β_I-α_L)	L1: –62(7), –25(12) L2: –86(13), 5(11) L3: 85(12), 7(19)	5.39(0.23)	–31(7)	5→1 & 1→5: 0 5→1: 0 1→5: 52	L1: Pro, Ser L2: Asp, Asn L3: Gly
α_R-α_R-α_R-α_L	L1: –68(9), –30(15) L2: –73(12), –39(12) L3: –100(12), –9(11) L4: 64(13), 33(20)	5.37(0.17)	9(11)	6→1 & 1→6: 23 6→1: 0 1→6: 13	L1: Pro, Arg L2: Ser, Lys L3: Ser, Thr, Asn L4: Gly, Asn, Lys, Asp

[a]Numbers in parentheses indicate standard deviations from the mean values.
[b]Virtual torsion angle is defined by the four C^{α} atoms, for two residue loops C^{α}_{B-1}–C^{α}_{L1}–C^{α}_{L2}–C^{α}_{B+1}, for three residue loops C^{α}_{B-1}–C^{α}_{L1}–C^{α}_{L3}–C^{α}_{B+1} and for four residue loops C^{α}_{B-1}–C^{α}_{L1}–C^{α}_{L4}–C^{α}_{B+1}.
[c]Only those amino acids which have a propensity value more than 2.0 are listed in the descending order for each loop position.

Fig. 6. Example of a β-hairpin nucleated by a three residue loop observed in galactose oxidase (1GOG 496–512). The ribbon diagram was prepared using the program MOLSCRIPT (Kraulis, 1991).

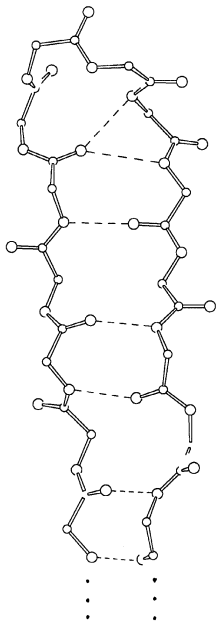

Fig. 7. ORTEP diagram representing a π-turn nucleated β-hairpin observed in bilin binding protein (1BBP A 118–131). The hairpin segment is from residue 113 to 136; only a part of the segment is shown. For clarity, only backbone atoms of the segment are shown. The diagram was prepared using a modified version of ORTEP run on an IBM PC.

10 examples have N–O distances between 3.5 and 4.0 Å and only three examples lie between 4 and 4.4 Å. Since connecting loops are frequently solvent exposed (Rose *et al.*, 1985; Leszczynski and Rose, 1986; Ring *et al.*, 1992; Martin *et al.*, 1995), hydration of the π-turn motif can result in a larger N–O distance. Indeed, in high resolution crystal structures of short helical peptides terminated by α_L conformations, solvent insertion into the 6 → 1 hydrogen bond has been observed (Karle *et al.*, 1996a). Figure 7 illustrates a π-turn nucleated β-hairpin observed in bilin binding protein (1BBP A 112–127). This motif, as can be seen from Table III and Figure 4, is a planar unit with virtual torsion angle (θ) centered around 9° (θ = C^{α}_{B-1}–C^{α}_{L1}–C^{α}_{L4}–C^{α}_{B+1}). Also, the motif β-α_R-α_R-α_R-α_L-β, like β-type II′-β, is fairly common in protein structures and has a propensity value of 2.294 for the nucleation of β-hairpins.

There appears to be greater propensity for the occurrence of *cis* peptide conformations in four and five residue loops; *cis* peptides were observed in as many as seven examples in

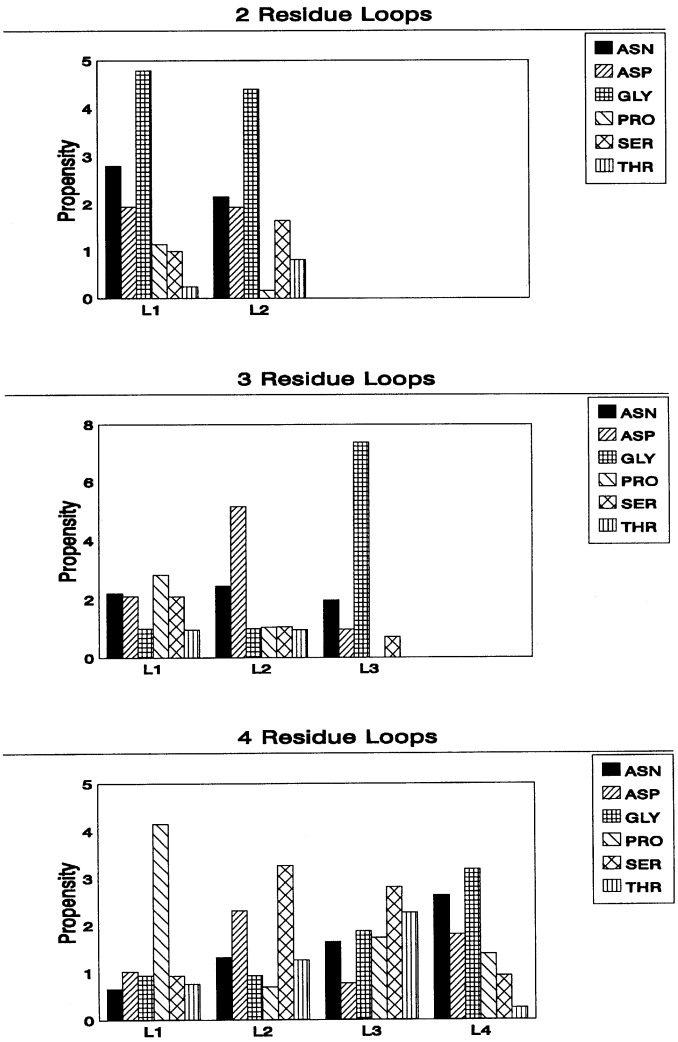

Fig. 8. Amino acid distributions in the loop segments for the two, three and four residue loop classes. Only those residues having propensity values (see text for definition) ⩾2.0 in at least one position in the loop region of any class are indicated.

four residue loops and in three examples in five residue loops, in contrast to one example in two residue loops and three examples in three residue loops. The occurrence of *cis* peptide units (Stewart *et al.*, 1990) in the total data set is 152 out of 52 497 residues, of which 135 examples are in X-Pro segments (Nataraj, 1996). Closer examination reveals that all 10 cases in hairpin loops involved X-Pro bonds. Of seven examples in four residue loops, five could be classified as being part of type VI β-turns with Pro at the $i + 2$ position [ideal type VI β-turns have a *cis* peptide unit between $i + 1$ and $i + 2$ (Richardson, 1981)].

Five residue loops. The data set contained relatively few examples (18) of five residue loops with the majority (11 examples) containing either type I or type II β-turns in the loop segment (Table II). No specific conformational motifs are immediately apparent.

Longer loops. Inspection of Figure 2 reveals that the distribution of longer loop lengths appears to have peaks associated with eight and nine residue loops. It should be noted that the longer loops may in fact be related to the loops which are four residues shorter. This is because if the inter-strand hydrogen bonding near the turning segment is distorted, the present procedure would incorporate the non-hydrogen-bonded residues into the loops, by definition. For example, distortions of two residue loops should transform into six residue loops, while three residue loops are related to seven residue loops. It is noteworthy that there are very few examples of six residue loops, suggesting that the distortions of hairpins nucleated by tight turns (two residues) are much less frequent.

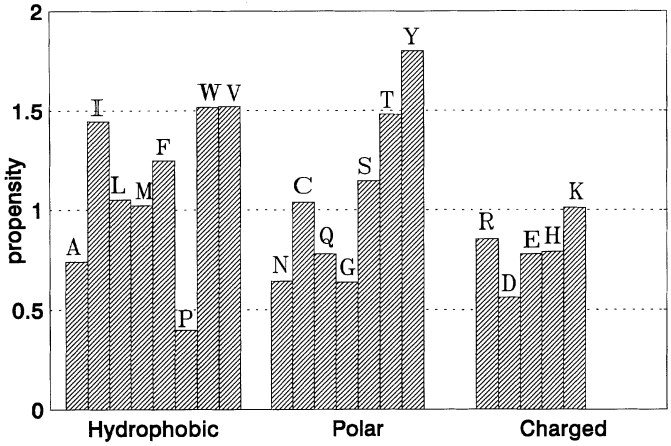

Fig. 9. Amino acid distribution in the strand segments of β-hairpins (see text for the definition of propensity value).

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	9																			
R	5	2																		
N	3	3	5																	
D	7	10	10	2																
C	1	2	0	3	9															
Q	7	7	3	8	1	1														
E	6	13	9	5	1	5	2													
G	20	5	6	8	4	5	6	11												
H	1	10	3	1	0	1	1	3	0											
I	17	4	9	5	7	9	15	13	5	14										
L	27	15	5	7	2	12	11	16	5	34	8									
K	14	5	11	18	6	6	24	7	4	15	18	6								
M	4	2	0	1	2	3	4	1	0	6	10	6	2							
F	15	4	5	2	3	6	5	14	1	17	21	8	3	9						
P	5	2	4	4	2	1	2	2	3	2	4	0	3	3	2					
S	25	7	12	14	3	6	18	14	4	10	18	18	4	9	10	20				
T	27	15	13	11	4	8	19	20	10	18	16	27	9	13	1	35	29			
W	3	3	2	2	2	2	2	4	3	11	7	8	3	5	3	5	2	2		
Y	23	9	4	8	6	7	10	12	4	25	31	12	1	14	8	20	18	8	4	
V	27	9	7	5	5	10	18	24	5	45	52	19	8	28	7	24	32	6	32	26
Total	**247**	**133**	**114**	**131**	**63**	**108**	**176**	**196**	**64**	**282**	**319**	**232**	**72**	**185**	**68**	**278**	**329**	**85**	**256**	**390**

Fig. 10. Pair occurrence of facing residues in the strand segments of β-hairpins (values ⩾20 are highlighted).

Amino acid compositional preferences in loops

The propensities of all the 20 amino acids occurring at any of the loop positions were calculated using the equation

$$P_{ij} = \frac{F_{ij}}{\sum_{i=1}^{20} F_{ij}} \Bigg/ \frac{D_i}{\sum_{i=1}^{20} D_i}$$

where F_{ij} is the number of times residue i occurs in a loop position j and D_i is the number of times residue i occurs in the data set. A value of $P_{ij} > 1$ indicates preference and a value $P_{ij} < 1$ indicates disfavor. Figure 8 shows the distribution of residues in the loop segments with only those residues having $P_{ij} > 2$ in at least one position in any class (two, three and four residue loops) being indicated. It is clear that Gly, Pro and small polar residues Asn, Asp, Ser and Thr are most frequently found in the connecting loops. This is not surprising since these residues have a high tendency to occur in surface-exposed loops in proteins (Rose *et al.*, 1985; Leszczynski and Rose, 1986; Srinivasan, *et al.*, 1991; Ring *et al.*, 1992; Martin *et al.*, 1995) and also have a low preference for occurring in secondary structures (Chou and Fasman, 1974, 1978). Striking

Table IV. Preferred facing pairs in the strands of β-hairpins

Reference/data set	Amino acid pairs	Pair occurrence	g_{ij}[b]
Present study[a]	Cys-Met–Cys-Met	13	2.85
Data set consists of 250 proteins	Arg-Lys-His–Asp-Glu	67	2.10
Total number of residues in the	Asn-Gln–Asp-Glu	32	1.88
strands of β-hairpins = 3728	Leu–Val	52	1.63
Confidence level[e] = 99.5%	Ile–Val	45	1.60
	Arg-Lys-His–Thr	52	1.51
Lifson and Sander (1980)	Met-Cys–Met-Cys	8	2.6
Date set consists of 30 proteins	Asp-Asn-His–Asp-Asn-His	20	2.0
Total number of	Tyr–Tyr	12	2.0
residues in antiparallel	Ile–Ala-Pro	20	1.9
β-sheets = 1576	Ser–Thr	21	1.9
Confidence level = 98.9%	Lys-Arg–Glu-Gln	16	1.8
	Leu–Met-Cys	11	1.8
	Ile–Val	23	1.7
	Val–Leu	25	1.5
	Thr–Thr	18	1.5
	Ser–Ser	14	1.5
Wouters and Curmi (1995)	H-bonded[c]:		
Data set consists of 253 proteins	Cys–Cys	7	4.9
Total number of	Glu–Lys	40	3.4
residues in antiparallel	Glu–Arg	32	3.4
β-sheets = 7231	Gln–Arg	16	2.5
Confidence level = 99%	Phe–Phe	24	2.4
	Ser–Ser	15	2.2
	Asp–Lys	15	2.1
	Gln–Lys	17	2.1
	Thr–Asn	19	2.0
	Non-H-bonded[d]:		
	Cys–Cys	20	9.9
	Glu–Lys	29	3.2
	Asp–His	7	3.0
	Ser–Asn	15	2.1
	Thr–Thr	34	2.0

[a]Minimal groupings of amino acids were made for the purpose of calculating pair correlation. Note that the grouping is slightly different from that used by Lifson and Sander (1980).
[b]Pair correlation g_{ij} value; for equation see text.
[c]Hydrogen-bonded site: facing pairs, in the strands of β-sheets, whose backbone atoms are hydrogen bonded.
[d]Non-hydrogen-bonded site: facing pairs whose backbone atoms are not hydrogen bonded.
[e]Confidence level at which the random pairing hypothesis was rejected, decided by statistical significance (χ^2) test.

Table V. Cys–Cys facing pairs in the present data set

Protein Code	Hairpin segment	Loop segment	Cys–Cys pair	Position[a]	Loop size (residues)	Strand length
1ACX	34–43	38–39	34–43	B ± 4	2	4
1DDT	459–473	465–467	461–471	B ± 4	3	6
1ESL	131–144	137–138	133–142[b]	B ± 4	2	6
1NSC A	114–132	122–123	121–126	B ± 2	2	8
	273–292	281–283	278–288	B ± 4	3	8
			276–290	B ± 6	3	8
2MCM	36–46	40–42	36–46	B ± 4	3	4
2SN3	37–50	43–44	41–46[b]	B ± 2	2	6
3EBX	24–41	32–33	24–41[b]	B ± 8	2	8

[a]Position of the Cys–Cys pair in the strands, counted from the first hydrogen-bonded pair near the loop (see Figure 1a).
[b]Not a disulfide-bonded pair.

positional preferences are observed. Gly has the highest preference for both L1 and L2 positions in two residue loops, a feature consistent with the high β-turn propensity of Gly (Wilmot and Thornton, 1988; Hutchinson and Thornton, 1994; Nataraj, 1996) and a preponderance of type I′ and II′ β-turns in two residue loops, which require residues having a strong tendency to adopt positive φ values (Richardson, 1981). The amino acids having the highest propensity for the L2 and L3 positions in three residue loops are Asp and Gly, respectively. The preference of Gly for the L3 position is undoubtedly a consequence of its ability to adopt α_L conformations, which is an almost essential prerequisite at this position. Four residue loops are characterized by an overwhelming preference for Pro at the L1 position. Serine has a high occurrence at the L2

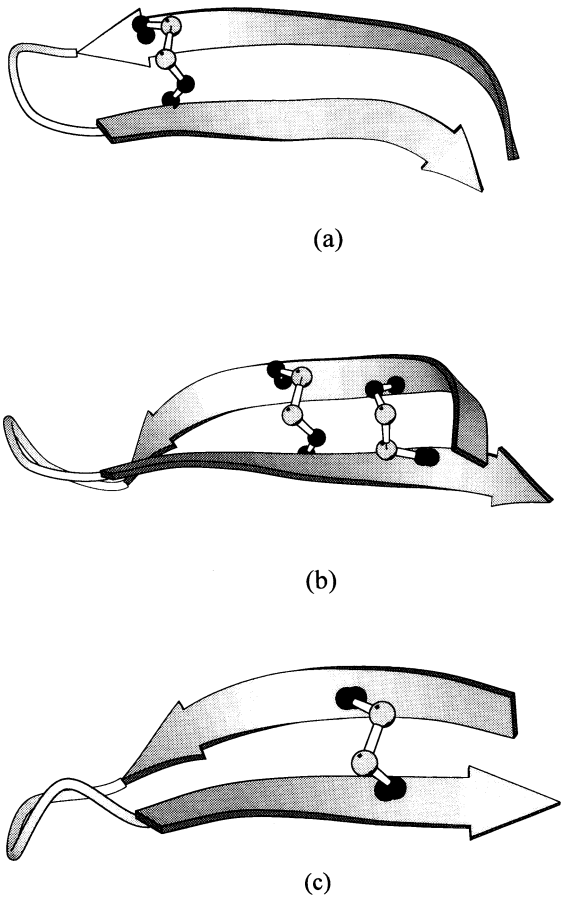

Fig. 11. Examples showing Cys–Cys facing pairs in the strand segments which are covalently linked. (**a**) Neuraminidase (1NSC A 114–132), (**b**) neuraminidase (1NSC A 273–292) and (**c**) diphtheria toxin (1DDT 459–473). The ribbon diagrams were prepared using the program MOLSCRIPT (Kraulis, 1991).

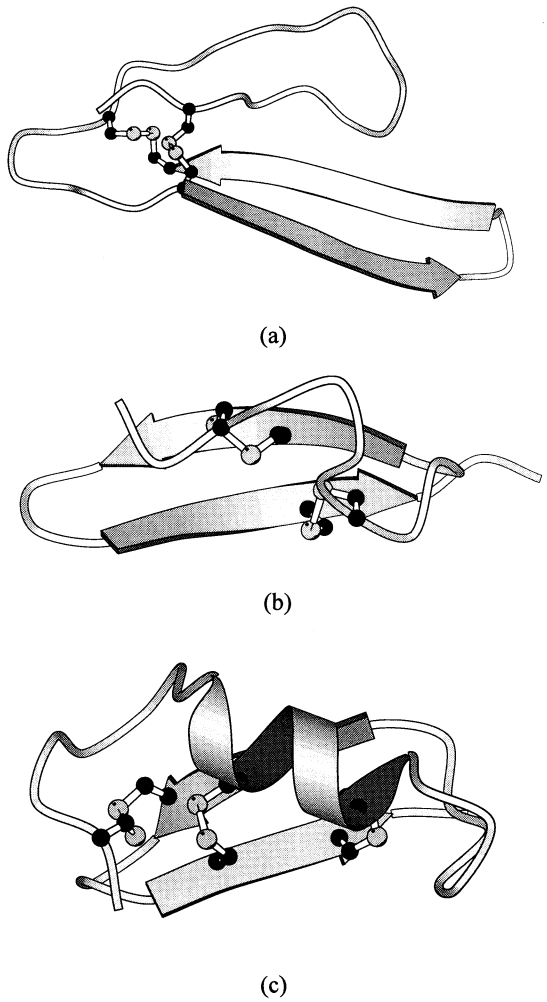

Fig. 12. Examples showing non-covalently linked Cys–Cys facing pairs in the strands. (**a**) Erabutoxin (3EBX 1–41), (**b**). E-Selectin (1ESL 120–146) and (**c**) scorpion neurotoxin (2SN3 15–50). The ribbon diagrams were prepared using the program MOLSCRIPT (Kraulis, 1991).

and L3 positions, while Gly and Asn dominate the L4 position. Residue occurrence at position L4 may again be the consequence of the requirement for α_L conformations in four residue loops which are formed by π-turns. Both Gly and Asn have a high propensity for α_L conformations (Srinivasan *et al.*, 1994). Curiously, the Asn propensity for the L3 position in three residue loops is dramatically lower than that of Gly. This may also be due to a difference in the intrinsic ϕ, ψ (backbone dihedral angles) propensity of Asn and Gly (Munoz and Serrano, 1995; Swindells *et al.*, 1995) since the ϕ, ψ at the L3 position are centered around 85, 7° respectively. The propensity at the L3 position may also be recalculated, in order to establish amino acid preference over and above the conformational preference, using a data set consisting of residues having ϕ, ψ values that fall within the limit of $\pm 45°$ from the mean value (85, 7°). This propensity estimation confirmed that the residue Gly (1.34) is preferred over the residue Asn (0.73) at the L3 position. These amino acid propensities in loops may be useful in constructing consensus loop sequences in *de novo* hairpin design. Also, from these observations, it is easy to choose amino acids for insertion or deletion, to effect a change in the loop size. The remaining 14 amino acids, which include hydrophobic residues and polar residues with bulky side chains, have relatively low propensities for occurrence in loop segments.

Compositional preferences of strands

The propensity of amino acids to occur in the strand segments of β-hairpins has been calculated (Figure 9). High propensities are observed for the β-branched residues Ile, Val and Thr, consistent with their strong preference for β-sheet conformations in proteins (Lifson and Sander, 1979; Minor and Kim, 1994; Smith *et al.*, 1994, Smith and Regan, 1995). [The amino acid propensity for β-strands in β-hairpins can also be estimated using a data set consisting residues in extended conformation and occurring in β-strands with a view to establishing amino acid preference for β-hairpin strands over and above the conformational preference. Such calculation did not show any specific patterns except for the residue Gly, which seems to have a slightly higher propensity to occur in β-strands of short loop hairpins than in β-strands in general (data not shown).] The very high β-hairpin strand preference is noted for the aromatic amino acids, Tyr, Trp and Phe, with Tyr occurring most frequently, suggesting the possibility that specific inter-strand interactions involving aromatic residues may stabilize β-hairpins. It was therefore of interest to examine pairwise interactions of facing residues across β-hairpins (von Heijne and Blomberg, 1977; Lifson and Sander, 1980; Smith and Regan, 1995; Wouters and Curmi, 1995). Figure 10 summarizes

the occurrence of specific facing amino acid pairs in the present data set of 250 proteins. Pair correlations for the residues *i* and *j* were calculated using the equation

$$g_{ij} = \frac{N_{ij}}{E_{ij}}$$

where N_{ij} is the number of times residue *i* and *j* occur as a pair in the data set and E_{ij} is the expected number of pairs of *i* and *j*. E_{ij} is calculated using the equation

$$E_{ij} = \frac{N_i N_j}{N}$$

where, $N_i = \sum_{j=1}^{20} N_{ij}$ and $N = \sum_{\substack{j=1 \\ i=1}}^{20} N_{ij}$

A pair correlation >1 would indicate that the pair occurs more frequently than expected on the basis of random occurrence. The pair correlation values were further examined for specific recognition against non-specific recognition using a statistical significance (χ^2) test, calculated as follows (Lifson and Sander, 1980; Wouters and Curmi, 1995):

$$\chi^2 = (N_{ij} - E_{ij})^2/E_{ij}$$

Table IV compares the results of the present analysis on β-hairpins with an early analysis by Lifson and Sander (1980) and a recent analysis by Wouters and Curmi (1995) which analyzed all antiparallel β-sheets in proteins. The results are compared only for amino acid pairs which show a high pair correlation ($g_{ij} \geqslant 1.5$). The results of the present analysis are in broad agreement with earlier studies. No specific pattern has emerged even for facing pairs of aromatic residues, although the Wouters and Curmi (1995) analysis shows a high preponderance of Phe–Phe pairs at the hydrogen bonding site. Interestingly, all three analyses showed a high pair correlation for Cys–Cys pairs, prompting us to examine the possible occurrence of disulfide-bridged β-hairpins.

β-Hairpin disulfides

Richardson's (1981) analysis stated that 'it is not possible for a disulfide to join neighbouring strands in a β-sheet: any but the closest residues on adjacent strands are too far apart and a closest pair of residues is slightly too close together'. As a consequence, the occurrence of disulfide bridges linking strands within a β-sheet are generally considered as 'unusual' (Xia *et al.*, 1996). An analysis of disulfide bridges in proteins carried out using a 65 protein data set, however, revealed a few examples of disulfide bridge across antiparallel β-strands (Srinivasan *et al.*, 1990). Table V lists the examples of Cys–Cys pairs in β-hairpins observed in the present data set. It is clear that the Cys residues are always placed at the 'non-hydrogen-bonded' position. The constraints of disulfide bridge stereochemistry (Srinivasan *et al.*, 1990) limit the occurrence of Cys–Cys pairs to specific locations on antiparallel strands. Figure 11 shows three examples of disulfide bridging across β-hairpins with one example [neuraminidase (1NSC A 273–292)] having two disulfide bridges across the antiparallel strands. It is important to stress that facing Cys residues occurring at the non-hydrogen-bonded position in six of the nine pairs are disulfide bonded. The remaining three examples which do not form inter-strand disulfides are illustrated in Figure 12. In all three examples the availability of proximal thiols results in the formation of alternative disulfide bridges. Covalent stabilization of *de novo* designed hairpins by appropriately placed disulfide bridges appears to be a viable strategy (Karle, *et al.*, 1988; Sieber and Moe, 1996).

Conclusions

The present analysis has revealed several important conformational and compositional features of β-hairpins which may be of value in peptide design and protein engineering. The design of hairpins with the loop segment ranging from two to four residues appears to be a distinct possibility in view of the strong preference for specific conformational features in short loops. Indeed, the earlier realization by Sibanda and Thornton (1985) that type I′ and type II′ β-turns are frequently found in hairpins has led to the successful design of synthetic peptides that adopt a hairpin conformation, with D-Pro residues acting as a strong conformational determinant (Awasthi *et al.*, 1995), permitting crystallographic characterization (Karle *et al.*, 1996b). The present analysis suggests that two residue loops with longer strand length hairpins nucleated by type I/III turns may also be attractive targets for future design as they are fairly widespread. Furthermore, the rational design of three and four residue loops appears possible in view of the strong preferences for the type I-α_L motif in the former and the α_R-α_R-α_R-α_L motif in the latter. Also, the motifs β-type I′-β, β-type II′-β, β-type I-α_L-β and β-α_R-α_R-α_R-α_L-β have comparable propensities for the nucleation of β-hairpins and β-type I-α_L-β and β-type I′-β have the same sense of twist. The strong preference for specific amino acid residues in the loop segments together with rigid conformational requirements augurs well for the design of consensus loops. The additional possibility of introducing covalent constraints by disulfide bridging across β-hairpins provides a means of locking specific conformations, a feature that has indeed been realized in short synthetic peptides (Karle *et al.*, 1988). The absence of strong preferences of amino acids in strands and of strong pair correlations across strands suggests that a high degree of sequence variability can be built into designed hairpin structures.

Acknowledgements

The work was supported by a grant from the Council of Scientific and Industrial Research, Government of India.

References

Alba,de E., Blanco,F.J., Jimenez,M.A., Rico,M. and Nieto,J.L. (1995) *Eur. J. Biochem.*, **233**, 283–292.
Aurora,R., Srinivasan,R. and Rose,G.D. (1994) *Science*, **264**, 1126–1130.
Awasthi,S.K., Raghothama,S. and Balaram,P. (1995) *Biochem. Biophys. Res. Commun.*, **216**, 375–381.
Baker,E.N. and Hubbard,R.E. (1984) *Prog. Biophys. Mol. Biol.*, **44**, 97–179.
Baldwin,R.L. (1995) *Biophys. Chem.*, **55**, 127–135.
Barlow,D.J. and Thornton,J.M. (1988) *J. Mol. Biol.*, **201**, 601–619.
Bernstein,F.C. *et al.* (1977) *J. Mol. Biol.*, **122**, 535–542.
Betz,S.F., Raleigh,D.P. and DeGrado,W.F. (1993) *Curr. Opin. Struct. Biol.*, **3**, 601–610.
Blanco,F.J., Jimenez,M.A., Herranz,J. Rico,M., Santoro,J. and Nieto,J.L. (1993) *J. Am. Chem. Soc.*, **115**, 5887–5888.
Blanco,F.J., Rivas,G. and Serrano,L. (1994) *Nature Struct. Biol.*, **1**, 584–590.
Bruch,M.D., Dhingra,M.M. and Gierasch,L.M. (1991) *Proteins: Struct. Funct. Genet.*, **10**, 130–139.
Chakrabartty,A. and Baldwin,R.L. (1995) *Adv. Protein Chem.*, **46**, 141–176.
Chou,P.Y. and Fasman,G.D. (1974) *Biochemistry*, **13**, 211–222.
Chou,P.Y. and Fasman,G.D. (1978) *Annu. Rev. Biochem.*, **47**, 251–276.
Constantine,K.L., Mueller,L. Anderson,N.H., Tong,H., Wandler,C.F., Friedrichs,M.S. and Bruccoleri,R.E. (1995) *J. Am. Chem. Soc.*, **117**, 10841–10854.

DeGrado,W.F. (1988) *Adv. Protein Chem.*, **39**, 51–124.
Doig,A.J. and Baldwin,R.L. (1995) *Protein Sci.*, **4**, 1325–1336.
Haque,T.S., Little,J.C. and Gellman,S.H. (1994) *J. Am. Chem. Soc.*, **116**, 4105–4106.
Hutchinson,E.G. and Thornton,J.M. (1994) *Protein Sci.*, **3**, 2207–2216.
Kametkar,S., Schiffer,J.M., Xiong,H., Babik,J.M. and Hecht,M.H. (1993) *Science*, **262**, 1680–1685.
Karle,I.L., Kishore,R., Raghothama,S. and Balaram,P. (1988) *J. Am. Chem. Soc.*, **110**, 1958–1963.
Karle,I.L., Banerjee,A., Bhattacharjya,S. and Balaram,P. (1996a) *Biopolymers*, **38**, 515–526.
Karle,I.L., Awasthi,S.K. and Balaram,P. (1996b) *Proc. Natl Acad. Sci. USA*, **93**, 8189–8193.
Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
Leszczynski,J.F. and Rose,G.D. (1986) *Science*, **234**, 849–855.
Lifson,S. and Sander,C. (1979) *Nature (London)*, **282**, 109–111.
Lifson,S. and Sander,C. (1980) *J. Mol. Biol.*, **139**, 627–639.
Lyu,P.C., Liff,M.I., Marky,L.A. and Kallenbach,N.R. (1990) *Science*, **250**, 669–673.
Lyu,P.C., Zhou,H.X., Jelveh,N. Wemmer,D.E. and Kallenbach,N.R. (1992) *J. Am. Chem. Soc.*, **114**, 6560–6562.
Martin,A.C.R., Toda,K., Stirk,H.J. and Thornton,J.M. (1995) *Protein Engng*, **8**, 1093–1101.
Mathews,B.W. (1972) *Macromolecules*, **5**, 818–819.
Milner-White,E.J. (1990) *J. Mol. Biol.*, **216**, 385–397.
Milner-White,E.J. and Poet,R. (1986) *Biochem. J.*, **240**, 289–292.
Milner-White,E.J., Rose,B.M., Ismail,R., Mostefa,K.B. and Poet,R. (1988) *J. Mol. Biol.*, **204**, 777–782.
Minor,D.L.,Jr and Kim,P.S. (1994) *Nature (London)*, **367**, 660–663.
Munoz,V and Serrano,L. (1995) *Curr. Opin. Biotechnol.*, **6**, 382–386.
Nagarajaram,H.A. (1995) Ph.D. thesis, Indian Institute of Science, Bangalore.
Nagarajaram,H.A., Sowdhamini,R., Ramakrishnan,C. and Balaram,P. (1993) *FEBS Lett.*, **321**, 79–83.
Nataraj,D.V. (1996) Ph.D. thesis, Indian Institute of Science, Bangalore.
O'Neil,K.T. and DeGrado,W.F. (1990) *Science*, **250**, 646–650.
Padmanabhan,S. and Baldwin,R.L. (1994) *Protein Sci.*, **3**, 1992–1997.
Pavone,V. (1988) *Int. J. Biol. Macromol.*, **10**, 238–240.
Presta,L.G. and Rose,G.D. (1988) *Science*, **240**, 1632–1641.
Rajashankar,K.R. and Ramakumar,S. (1996) *Protein Sci.*, **5**, 932–936.
Ramachandran,G.N., Ramakrishnan,C and Sasisekharan,V. (1963) *J. Mol. Biol.*, **7**, 95–99.
Ramakrishnan,C. and Ramachandran,G.N. (1965) *Biophys. J.*, **5**, 909–933.
Ramakrishnan,C. and Soman,K.V. (1982) *Int. J. Pept. Protein Res.*, **20**, 218–237.
Ramirez-Alvarado,M., Blanco,F.J. and Serrano,L. (1996) *Nature Struct. Biol.*, **3**, 604–612.
Richardson,J.S. (1981) *Adv. Protein Chem.*, **34**, 167–330.
Richardson,J.S. and Richardson,D.C. (1988) *Science*, **240**, 1648–1652.
Richardson,J.S. *et al.* (1992) *Biophys. J.*, **63**, 1186–1209.
Ring,C.S., Kneller,D.G., Langridge,R. and Cohen,F.E. (1992) *J. Mol. Biol.*, **224**, 685–699.
Rose,G.D., Gierasch,L.M. and Smith,J.A. (1985) *Adv. Protein Chem.*, **37**, 1–109.
Schellman,C. (1980) In Jaenicke,R. (ed.) *Protein Folding.* Elsevier/North-Holland, New York, pp. 53–61.
Seale,J.W., Srinivasan,R. and Rose,G.D. (1994) *Protein Sci.*, **3**, 1741–1745.
Searle,M.S., Williams,D.H. and Packman,L.C. (1995) *Nature Struct. Biol.*, **2**, 999–1006.
Scholtz,J.M. and Baldwin,R.L. (1992) *Annu. Rev. Biophys. Biomol. Struct.*, **21**, 95–118.
Sibanda,B.L. and Thornton,J.M. (1985) *Nature (London)*, **316**, 170–174.
Sibanda,B.L. and Thornton,J. M. (1991) *Methods Enzymol.*, **202**, 59–82.
Sibanda,B.L. and Thornton,J.M. (1993) *J. Mol. Biol.*, **229**, 428–447.
Sibanda,B.L., Blundell,T.L. and Thornton,J.M. (1989) *J. Mol. Biol.*, **206**, 759–777.
Sieber,V. and Moe,G.R. (1996) *Biochemistry*, **35**, 181–188.
Smith,C.K. and Regan,L. (1995) *Science*, **270**, 980–982.
Smith,C.K., Withka,J.M. and Regan,L. (1994) *Biochemistry*, **33**, 5510–5517.
Soman,K.V. and Ramakrishnan,C. (1986) *Int. J. Biol. Macromol.*, **8**, 89–96.
Sowdhamini,R., Srinivasan,N., Ramakrishnan,C. and Balaram,P. (1992) *J. Mol. Biol.*, **223**, 845–851.
Srinivasan,N., Sowdhamini,R., Ramakrishnan,C. and Balaram,P. (1990) *Int. J. Pept. Protein Res.*, **36**, 147–155.
Srinivasan,N., Sowdhamini,R., Ramakrishnan,C. and Balaram,P (1991) In Balaram,P. and Ramaseshan,S. (eds), *Molecular Conformation and Biological Interactions (G.N. Ramachandran Festschrift).* Indian Academy of Sciences, Bangalore, pp. 59–73.
Srinivasan,N., Anuradha,V.S., Ramakrishnan,C., Sowdhamini,R. and Balaram,P. (1994) *Int. J. Pept. Protein Res.*, **44**, 112–122.
Stewart,D.E., Sarkar,A. and Wampler,J.E. (1990) *J. Mol. Biol.*, **214**, 253–260.
Struthers,M.D., Cheng,R.P. and Imperiali,B. (1996) *Science*, **271**, 342–345.
Swindells,M.B., MacArthur,M.W. and Thornton,J.M. (1995) *Nature Struct. Biol.*, **2**, 596–603.
Venkatachalam,C.M. (1968) *Biopolymers*, **6**, 1425–1436.
von Heijne,G. and Blomberg,C. (1977) *J. Mol. Biol.*, **117**, 821–824.
Wilmot,C.M. and Thornton,J.M. (1988) *J. Mol. Biol.*, **203**, 221–232.
Wouters,M.A. and Curmi,P.M.G. (1995) *Proteins: Struct. Funct. Genet.*, **22**, 119–131.
Xia,Z., Dai,W., Zhang,Y., White,S.A., Boyd,G.D. and Mathews,F.S. (1996) *J. Mol. Biol.*, **259**, 480–501.

Received December 17, 1996; revised June 23, 1997; accepted July 8, 1997