Structure *14*, 529–542, March 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2005.11.021

Conformational and Sequence Signatures in β Helix Proteins

Prathima Iengar,[1] N.V. Joshi,[2] and Padmanabhan Balaram[1,*]
[1] Molecular Biophysics Unit
[2] Centre for Ecological Sciences
Indian Institute of Science
Bangalore 560012
India

Summary

β helix proteins are characterized by a repetitive fold, in which the repeating unit is a β-helical coil formed by three strand segments linked by three loop segments. Using a data set of left- and right-handed β helix proteins, we have examined conformational features at equivalent positions in successive coils. This has provided insights into the conformational rules that the proteins employ to fold into β helices. Left-handed β helices attain their equilateral prism fold by incorporating "corners" with the conformational sequence P_{II}-P_{II}-α_L-P_{II}, which imposes sequence restrictions, resulting in the first and third P_{II} residues often being G and a small, uncharged residue (V, A, S, T, C), respectively. Right-handed β helices feature mid-sized loops (4, 5, or 6 residues) of conserved conformation, but not of conserved sequence; they also display an α-helical residue at the C-terminal end of L2 loops. Backbone conformational parameters (ϕ,ψ) that permit the formation of continuous, loopless β helices (Perutz nanotubes) have also been investigated.

Introduction

The first crystal structure of a parallel β helix protein to be reported was that of pectate lyase C from *Erwinia chrysanthemi*, a plant pathogenic bacterium (Yoder et al., 1993a). The enzyme, which consisted predominantly of parallel β strands, was described as folding into a new motif; the parallel β strands coiled into a large right-handed cylinder or helix that was called a parallel β helix (Figure 1A). Sternberg and Thornton (1977) have noted in an analysis of the connections between parallel β strands in proteins that, irrespective of the secondary structure in the linking segment (helix, sheet, or coil), the connections are almost always right-handed. They suggest that the twist between parallel β strands is responsible for the connections being right-handed. Richardson (1976) notes that the inherent right-handed twist of β strands and the right-handedness of α-helical segments together drive crossover connections into a right-handed conformation during protein folding. The right-handed, parallel β helix, with right-handed connections between the parallel β strands, corroborated the above observations (Cohen, 1993). However, when the first left-handed, parallel β helix structure of UDP-N-acetylglucosamine acyltransferase was solved, the structure was found to have repeated left-handed connections between the parallel β strands, suggesting that alternative modes of generating parallel β sheet structures were possible (Figure 1B; Raetz and Roderick, 1995). The flat and untwisted nature of the β sheets, the long connections between the parallel β strands, and the absence of helices in the connecting segments were thought to explain why β helices with left- and right-handed connections were possible (Raetz and Roderick, 1995).

Yoder et al. (1993b), from a detailed analysis of two pectate lyases, PelC and PelE, suggest that the stability of the parallel β sheets in these proteins lies in the extensive network of interstrand H bonds, and in the presence of unusual side chain stacks formed by interacting side chains of strand residues, inside and outside the β helix. Crystal structures of a number of other right- and left-handed β helix proteins have been solved over the past 10 years. Cowen et al. (2002), based on an examination of β helix structures, have developed the computer program BetaWrap to predict the right-handed β-helical fold from protein sequence data. β helix proteins have been comprehensively reviewed by Jenkins and Pickersgill (2001). The notable differences between right- and left-handed β helices (henceforth, RBH and LBH, respectively) lie in the length and conformational properties of the segments of polypeptide chain that link adjacent β strands. In LBH, each β-helical turn (or coil) consists of three strand segments that are connected by three usually short connecting loops of length 1–2 residues. Each coil in an RBH also consists of three strand segments connected by three loop segments; however, one loop segment is short and usually 1 residue long, while the other two loop segments are of appreciable length (Figure 2).

Interest in β-helical structures has also been stimulated by an intriguing proposal by Perutz et al. (2002a) that the poly-Gln aggregates in Huntington's disease adopt a β-helical nanotube structure. The model proposed by them, based on the X-ray fiber-diffraction data for poly-Gln, features a polypeptide chain in a continuous extended conformation, wrapping around a cylinder (Figure 1C). The polypeptide chain coils around (like a helix) to form a curved β sheet, or β cylinder or nanotube. The main chain runs approximately perpendicular to the fibril axis (referred to as a "cross-β structure"), while the β sheet H bonds, between adjacent coils of the helix, run parallel to the fibril axis. The nanotube may be formally viewed as a parallel β helix without the connecting loops, a "connectionless" β helix. Such a structure requires at least 20 residues per turn in order to achieve optimum H bonding between adjacent coils. Perutz et al. (2002a, 2002b) suggest that their model may also represent the structure of a variety of amyloid fibrils involved in the pathological deposition observed in a variety of neurodegenerative disorders (Dobson, 2002). It has also been pointed out that the parallel β helix observed in β-helical proteins is consistent with various experimental data on the structure of the amyloid fibril (Jenkins and Pickersgill, 2001; Wetzel, 2002; Pickersgill,

*Correspondence: pb@mbu.iisc.ernet.in

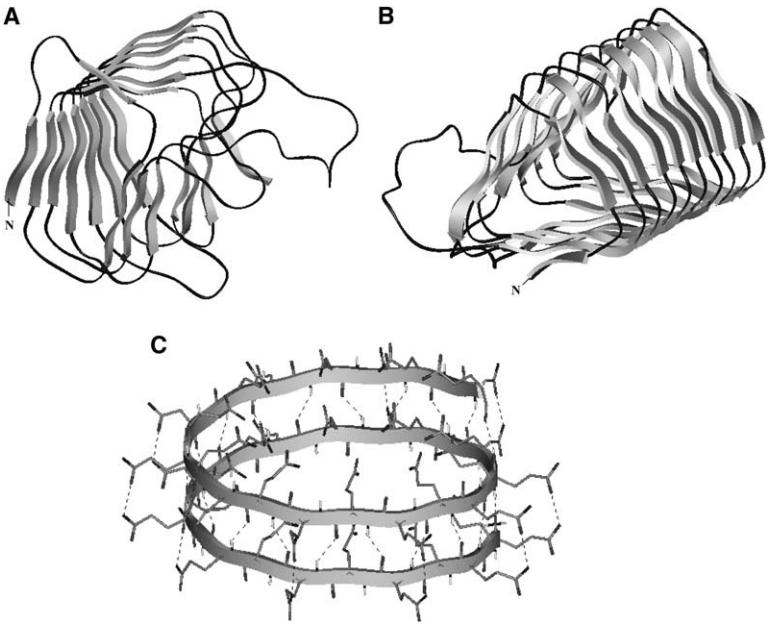

Figure 1. Right- and Left-Handed β Helices vis-à-vis the Connectionless β Helix.

(A) The right-handed β helix of 1AIR (segment 91–267).

(B) The left-handed β helix of 1LXA (segment 1–186).

(C) The Perutz poly-Gln nanotube (residues 1–40) generated by repeating a pair of (ϕ,ψ) values (ϕ_1 = 160°, ψ_1 = −170°; ϕ_2 = −161°, ψ_2 = 168°); one side chain and two backbone H bonds per Gln residue are shown as broken lines. Strands of sheets are shown as gray ribbons, and connecting loops are shown as black tubes.

2003). It has a "cross-β structure" and shows some similarity to the Perutz β-helical nanotube.

In the present study, we have regarded β helix proteins as proteins with a repetitive fold. We have compiled a data set of 14 RBH and 7 LBH and have examined the conformational features of the repeating units, viz., the β-helical coils. We find that LBH fold into their characteristic equilateral prism structure because of a certain conformational sequence occurring at the corners of each β-helical coil. The sequence restrictions arising from this conformational sequence are discussed. RBH are built by a different design and are not subject

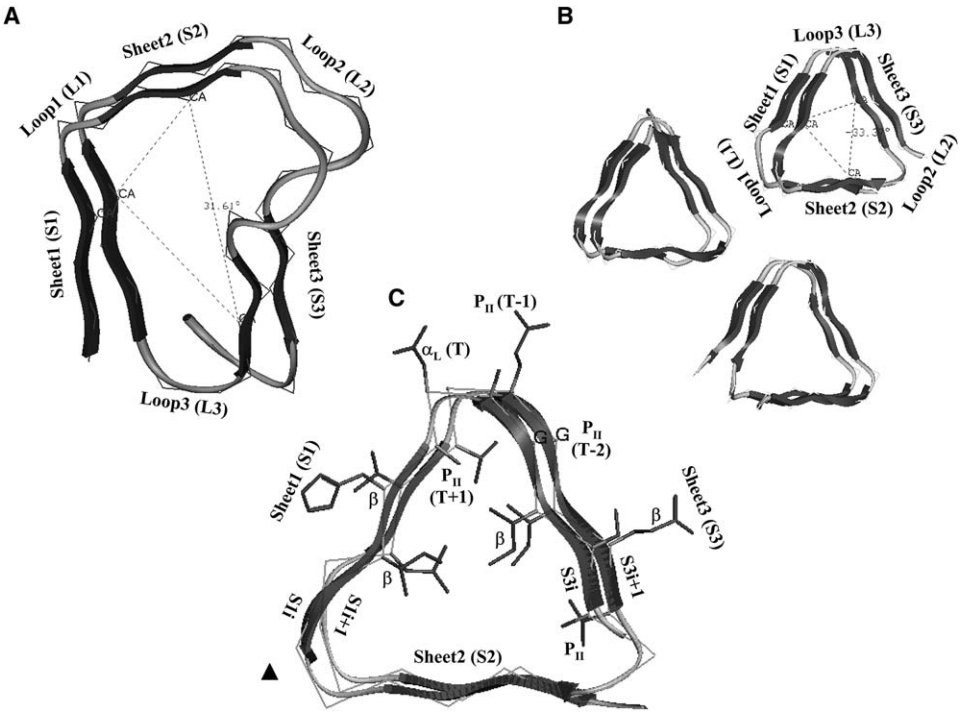

Figure 2. The Repeating Unit of β Helices Is the β-Helical Turn

(A–C) Two β-helical turns of: (A) the RBH 1AIR (segment 204–249), and (B) and (C) the trimer and monomer, respectively, of the LBH 1LXA (segment 17–52). Strands are shown as black ribbons, and connecting loops are shown as gray tubes. In (A) and (B), sheets S1, S2, and S3 and loops L1, L2, and L3 are labeled; a virtual torsion angle defined by the central C^{α} atoms of consecutive strands is indicated by dotted lines and a value. In (C), the 3-fold axis of the trimer is indicated by a black triangle; side chains are shown for two L3 "corners," formed by strands $S3_i$ and $S3_{i+1}$ of sheet S3 with strands $S1_i$ and $S1_{i+1}$ of sheet S1, respectively; (T−2), (T−1), T, and (T+1) residues adopting the P_{II}-P_{II}-α_L-P_{II} conformational sequence and G residues at (T−2) are indicated.

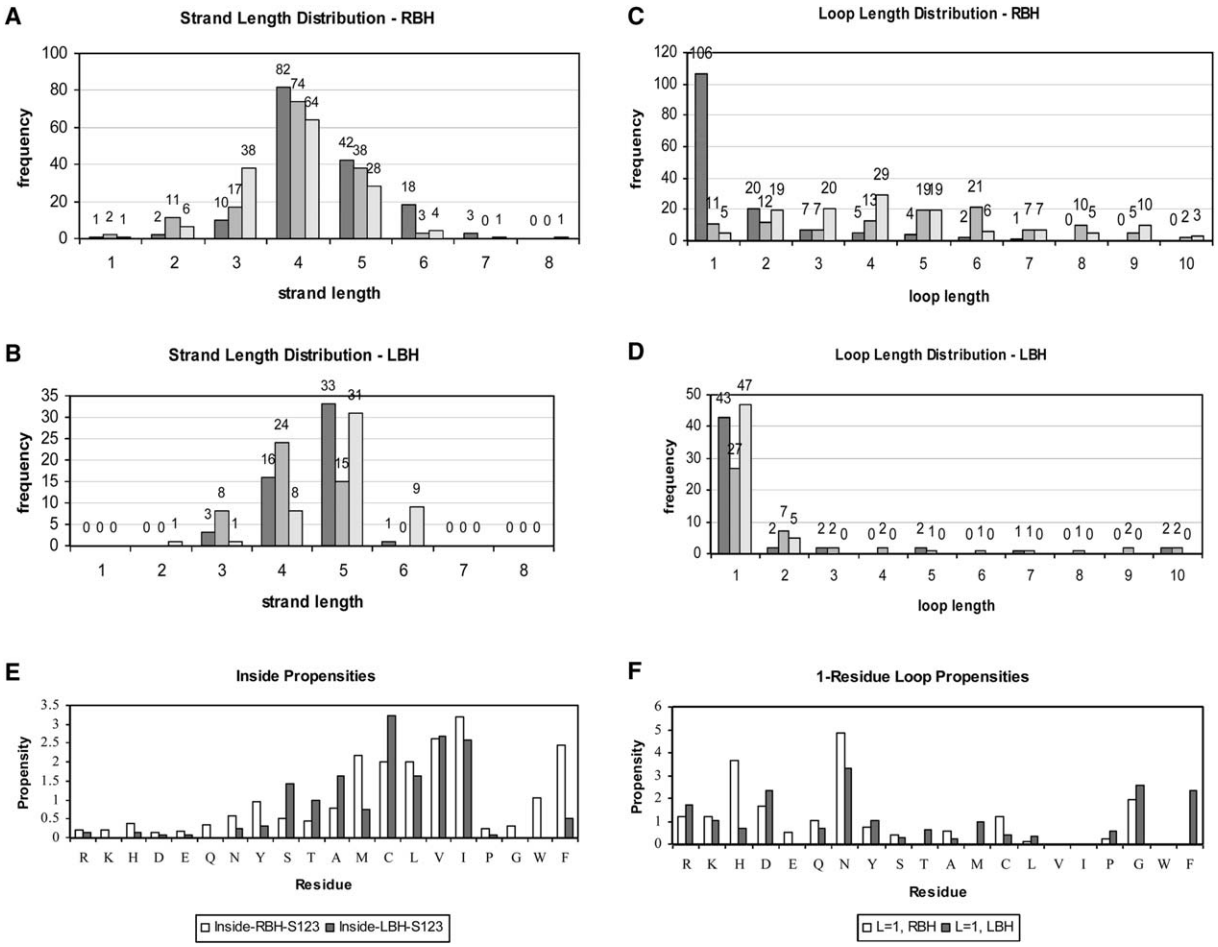

Figure 3. Strand and Loop Length Distributions and Amino Acid Propensities in β Helices

(A–D) Distribution of strand lengths in (A) RBH and (B) LBH and the distribution of loop lengths in (C) RBH and (D) LBH. The number of residues in a strand or loop (strand length or loop length) is plotted versus the frequency with which strands or loops of that length are observed. The three histograms in (A) and (B) correspond to strand length distributions observed in sheets S1, S2, S3, respectively; those in (C) and (D) correspond to loop length distributions observed in loops L1, L2, L3, respectively.

(E and F) Propensities of amino acids to occur (E) in the sheets, with side chains pointing inside the β helices, and (F) in 1-residue loops. The white and gray bars in (E) and (F) are for RBH and LBH, respectively.

to sequence restrictions. The manner in which they fold gives rise to mid-size loops of conserved conformation but not of conserved sequence. Some of these loops as well as differences in construction of RBH and LBH are discussed. Conformational parameters for forming the "connectionless" β helices are also investigated.

Results

Some General Features of β Helices

Sheet and Loop Lengths

In the 14 RBH in our data set, there are 158 sheet1 (S1) strands, 145 sheet2 (S2) strands, and 143 sheet3 (S3) strands. In the 7 LBH in our data set, there are 53 S1 strands, 47 S2 strands, and 50 S3 strands. Figures 3A and 3B show the distribution of strand lengths in sheets 1, 2, and 3 of RBH and LBH, respectively. In all three sheets of RBH, 4-residue strands are the most frequent (Figure 3A). Five-residue strands are the next most frequent in S1 and S2, while, in S3, 3- and 5-residue strands are frequent. In general, strands in S1 are longer (4, 5, or 6 residues long), while shorter, 2- and 3-residue, strands are more commonly observed in S2 and S3. In LBH, S1, S2, and S3 are usually 4 or 5 residues long (Figure 3B). While S1 and S3 are more frequently 5 residues long, S2 is 4 residues long. However, 4-residue S1 strands and 5 -residue S2 strands are also observed. In four out of seven LBH (1LXA, 1MR7, 1OCX, 3TDT), all three sheets are made up of 5-residue strands; in 1HV9 and 1QRE, two sheets are made up of 5-residue strands, while the third (S2) is made up of 4-residue strands; in 1M8N all three sheets are made up of 4-residue strands. Thus, the observed strand length distribution is a result of the different strand lengths observed in the sheets of LBH.

Figures 3C and 3D show the distribution of loop lengths in the L1, L2, and L3 loops of RBH and LBH, respectively. In RBH, L1 loops are predominantly 1-residue loops, with 2-residue loops being the next most frequent. L2 and L3 loops are of varying lengths. Six- and 5-residue L2 loops are the most frequently observed, with 4-, 2-, and 1-residue loops being the next most frequent. Four-residue L3 loops are the most frequently observed, with 3-, 2-, and 5-residue loops being the next most frequent. Some of these loops are

Table 1. Identification of Sheets and Loops for the LBH 1LXA

S1	S1	S1	S1	S1	*L1*	*L1*	*L1 (T)*	S2	S2	S2	S2
A6 **(−76,** **141)** **i**	**F7** **(−99,** **118)** **o**	**V8** **(−120,** **109)** **i**	**H9** **(−61,** **128)** **o**			*P10* *(−44,* *−38)*	*T11* *(−72,* *−23)*	**A12** **(−65,** **160)** **i**	**I13** **(−124,** **106)** **o**	**V14** **(−102,** **130)** **i**	**E15** **(−85,** **152)** **o**
A24 **(−83,** **173)** **i**	**H25** **(−140,** **117)** **o**	**I26** **(−112,** **109)** **i**	**G27** **(−77,** **176)** **o**	**P28** **(−57,** **152)** **o**			*F29* *(47,* *36)*	**C30** **(−102,** **161)** **i**	**I31** **(−133,** **125)** **o**	**V32** **(−123,** **122)** **i**	**G33** **(−67,** **165)** **o**
T42 **(−91,** **147)** **i**	**V43** **(−127,** **135)** **o**	**L44** **(−117,** **96)** **i**			*K45* *(−42,* *−59)*	*S46* *(169,* *160)*	*H47* *(48,* *50)*	**V48** **(−120,** **148)** **i**	**V49** **(−121,** **116)** **o**	**V50** **(−122,** **118)** **i**	**N51** **(−141,** **−177)** **o**
N60 **(−90,** **155)** **i**	**E61** **(−125,** **121)** **o**	**I62** **(−126,** **117)** **i**	**Y63** **(−86,** **150)** **o**	**Q64** **(−34,** **137)** **o**			*F65* *(77,* *12)*	**A66** **(−84,** **156)** **i**	**S67** **(−140,** **107)** **o**	**I68** **(−118,** **124)** **i**	
N90 **(−86,** **160)** **i**	**R91** **(−124,** **107)** **o**	**I92** **(−110,** **111)** **i**	**R93** **(−75,** **170)** **o**	**E94** **(−52,** **146)** **o**			*S95* *(57,* *14)*	**V96** **(−71,** **152)** **i**	**T97** **(−131,** **130)** **o**	**I98** **(−134,** **109)** **i**	**H99** **(−95,** **145)** **o**
N115 **(−97,** **164)** **i**	**L116** **(−127,** **106)** **o**	**L117** **(−105,** **96)** **i**	**M118** **(−67,** **169)** **o**	**I119** **(−43,** **131)** **o**			*N120* *(60,* *17)*	**A121** **(−70,** **145)** **i**	**H122** **(−126,** **129)** **o**	**I123** **(−116,** **114)** **i**	**A124** **(−79,** **171)** **o**
C133 **(−104,** **168)** **i**	**I134** **(−141,** **125)** **o**	**L135** **(−116,** **118)** **i**	**A136** **(−76,** **176)** **o**	**N137** **(−49,** **125)** **o**			*N138* *(70,* *5)*	**A139** **(−64,** **139)** **i**	**T140** **(−130,** **126)** **o**	**L141** **(−105,** **136)** **i**	**A142** **(−83,** **−177)** **o**
A151 **(−66,** **146)** **i**	**I152** **(−130,** **123)** **o**	**I153** **(−116,** **114)** **i**	**G154** **(−47,** **145)** **o**	**G155** **(−66,** **150)** **o**			*M156* *(61,* *29)*	**T157** **(−105,** **139)** **i**	**A158** **(−109,** **127)** **o**	**V159** **(−110,** **132)** **i**	**H160** **(−79,** **137)** **o**
V169 **(−81,** **154)** **i**	**M170** **(−127,** **113)** **o**	**V171** **(−103,** **117)** **i**			*G172* *(−23,* *−82)*	*G173* *(128,* *136)*	*C174* *(68,* *31)*	**S175** **(−120,** **162)** **i**	**G176** **(−131,** **119)** **o**	**V177** **(−111,** **110)** **i**	
V185 **(−95,** **161)** **i**	**I186** **(−108,** **134)** **o**	**A187** **(−122,** **157)** **i**	**Q188** **(−163,** **156)** **o**								

Each row constitutes one complete turn of the β helix. In each turn, residues forming a part of sheets 1, 2, and 3 are in bold and are labeled S1, S2, and S3, respectively, and residues forming loops 1, 2, and 3 are in italics and are labeled L1, L2, and L3, respectively. Loop residues in α_L conformation are labeled "*(T)*"; loops longer than 3 residues are indicated by dotted lines. Below each residue is its (ϕ,ψ) value and, for sheet residues, an "i" or "o" label, which indicates whether the residue side chain points "inside" or "outside" the β helix.

discussed in greater detail in the next section. In LBH, the L1, L2, and L3 loops are most frequently 1-residue loops (Figure 3D).

β-Helical Coils

β helix proteins may be regarded as being generated from large repetitive turns or coils (Figure 1), and each coil is composed of three strand segments and three loop segments (Figure 2; Table 1). Adjacent coils are registered by backbone H bonds formed between strand segments in the three parallel β sheets. There are 135 and 46 RBH and LBH coils, respectively, in our data set. In LBH, the majority of helical turns were 15–20 residues long, and the largest number were 17–18 residues long. In RBH, the majority of turns were 16–34 residues long, and the largest number were 20–23 residues long. In RBH, we can account for the most frequently occurring 23 residues in a coil as follows: 4 residues per strand (Figure 4A), therefore, 12 residues would occur in three strands; a 1-residue L1 loop, a 6-residue L2 loop, and a 4-residue L3 loop would account for 11 residues; thus, together, 23 residues would form a coil in RBH. Likewise, in LBH, assuming 5 residues per strand, there would be 15 residues in three strands (as noted above, four out of seven LBH have all 5 residue strands), and assuming 1 residue in each loop, there would be a total of 15 + 3 = 18 residues to a coil.

S2	L2	L2 (T)	L2	S3	S3	S3	S3	S3	S3	L3	L3 (T)
						M1 (−, 166) o	I2 (−129, 128) i	D3 (−57, 127) o		K4 (−58, −25)	S5 (−85, −10)
E16 (−67, 125) o		G17 (116, −21)			A18 (−63, 126) i	S19 (−105, 123) o	I20 (−121, 124) i	G21 (−69, 165) o	A22 (−57, 146) o		N23 (50, 25)
	P34 (−71, −16)	H35 (−108, 15)			V36 (−86, 149) i	E37 (−136, 110) o	I38 (−110, 120) i	G39 (−69, 169) o	E40 (−54, 143) o		G41 (55, 22)
	G52 (79, −161)	H53 (−96, 82)			T54 (−116, 131) i	K55 (−114, 119) o	I56 (−123, 128) i	G57 (−65, −158) o	R58 (−107, 157) o		D59 (57, 31)
	G69 (94, −2)	…..	T82 (−80, 167)	R83 (−117, 171) o	V84 (−147, 125) i	E85 (−123, 129) o	I86 (−126, 133) i	G87 (−70, −143) o	D88 (−125, 158) o		R89 (68, 25)
R100 (−67, 148) o	G101 (−65, −177)	…..	G107 (59, 18)	L108 (−144, 123) o	T109 (−101, 117) i	K110 (−124, 140) o	V111 (−132, 125) i	G112 (−67, −138) o	S113 (−129, 164) o		D114 (51, 34)
H125 (−53, 139) o		D126 (63, 25)			C127 (−86, 175) i	T128 (−141, 121) o	V129 (−129, 130) i	G130 (−72, −139) o	N131 (−122, 151) o		R132 (56, 40)
G143 (−70, 151) o		H144 (61, 24)			V145 (−90, 144) i	S146 (−122, 124) o	V147 (−120, 128) i	D148 (−99, 167) o	D149 (−44, 135) o		F150 (70, 0)
Q161 (−45, 143) o		F162 (79, −2)			C163 (−79, 154) i	I164 (−118, 136) o	I165 (−124, 107) i	G166 (−48, 160) o	A167 (−70, 149) o		H168 (63, 13)
	A178 (−104, −15)				Q179 (−126, 174) o	D180 (−69, 147) o	V181 (−116, 102) i	P182 (−55, 148) o	P183 (−51, 147) o		Y184 (68, −1)

Virtual Torsion Angles

A virtual torsion angle was defined between the central C^{α} atoms of strands $S1_i$-$S2_i$-$S3_i$-$S1_{i+1}$, where i and i+1 are two successive helical turns or coils of the β helix (Figures 2A and 2B). A total of 133 and 46 such virtual torsion angle values were calculated from the coils of RBH and LBH, respectively. The 133 values for RBH were positive, with a mean of 27.1°, while the 46 values for LBH were negative, with a mean of −32.5°. The virtual torsion angles are opposite in sign in RBH and LBH because of the opposite handedness of the β-helical coils.

Propensity Calculations

Propensities of amino acids to occur in sheets, loops, and inside and outside RBH and LBH were calculated; the most interesting sets are shown in Figures 3E and 3F. The sheet propensities (not shown) are not unusual when compared to the Chou and Fasman (1974) β sheet propensities. β-branched residues (I,V), which have been noted to favor β sheet formation (Nesloney and Kelly, 1996), are preferred in both RBH and LBH. P and G are avoided in RBH and are present, but not preferred, in LBH. Most other residues are present in sheets to varying extents, with hydrophobic residues being preferred over charged and polar-uncharged residues. In LBH, small residues are preferred over larger ones. Figure 3E shows the propensities of amino acids to occur "inside" LBH and RBH. Sheet residues directed toward the interior of β helices constituted the set of inside or "i" residues, and those directed toward the exterior constituted the outside or "o" residues (Experimental Procedures). The figure shows the extreme preference for hydrophobic residues in the interior. This is not surprising, because the packing of hydrophobic residues in the interior would be crucial to the stability of

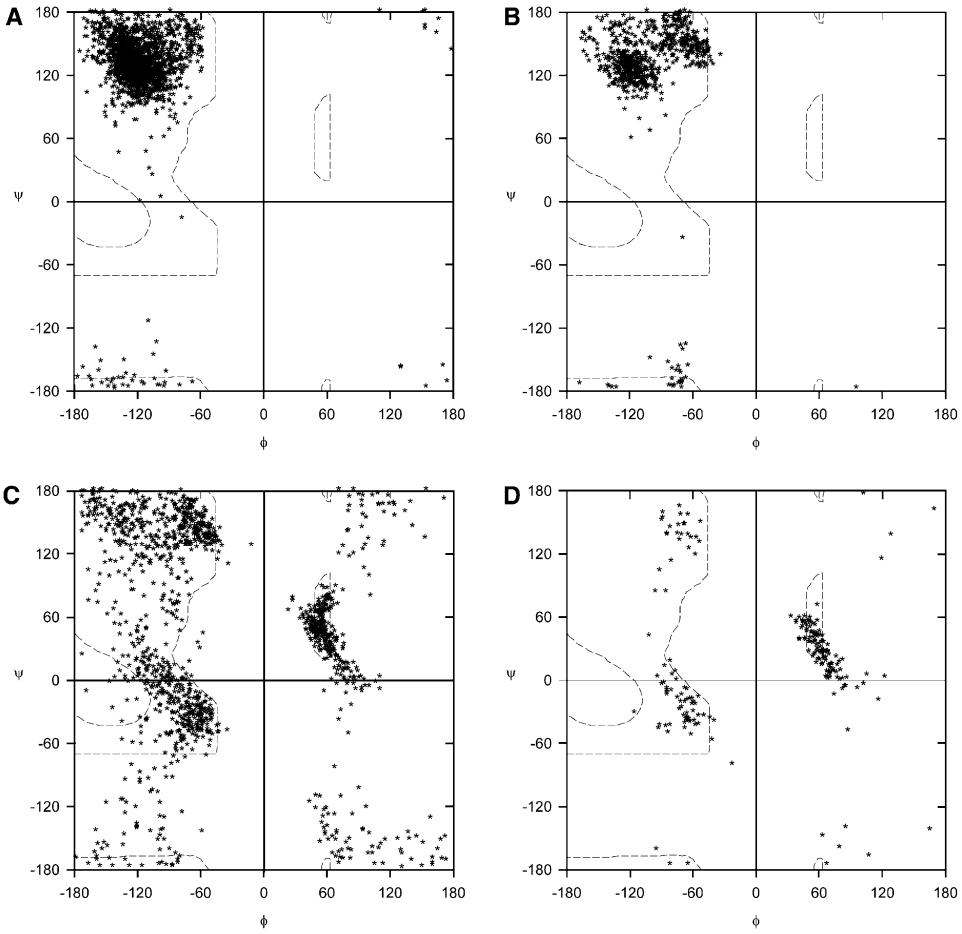

Figure 4. Backbone Torsion Angle or (ϕ,ψ) Distributions for Sheet and Loop Residues in β Helices

(A–D) (ϕ,ψ) plots for sheet residues in (A) RBH (1840 residues) and (B) LBH (683 residues); (ϕ,ψ) plots for residues in loops of size 1–10 in (C) RBH (1306 residues) and (D) LBH (261 residues).

β helices. On the outside of β helices (propensities not shown), C is avoided, and charged and polar-uncharged residues are preferred over hydrophobic residues. Figure 3F shows the propensities of residues to occur in the 1-residue loops. In both RBH and LBH, N is the most preferred residue, with G also being preferred. In RBH, H is also a preferred residue. In loops of size 2–10 (propensities not shown), P and G are the most preferred residues, and charged and polar-uncharged residues are preferred over hydrophobic residues. The loop propensities are consistent with the Chou and Fasman (1974) predictions according to which P, G, N, S, and D are preferred loop residues.

Backbone Conformation

Figures 4A and 4B show the (ϕ,ψ) distributions for sheet residues in RBH and LBH, respectively. As sheet residues in each β helix protein were identified in a semiautomated way, with manual intervention, these (ϕ,ψ) distributions are proof of the reliability with which sheet residues were identified. There are seven and one outliers in the RBH and LBH data sets, respectively. The latter (G362 from 1HV9) was included, as, upon visual inspection, it appeared to be part of a β sheet. The outliers, however, are <0.4% and <0.2% of the total number of sheet residues in the RBH (1840) and LBH (683) data sets, respectively. Henceforth, we will make two distinctions in the β sheet region of the Ramachandran map. β, β region, or β sheet region will refer to conformations in the region $\phi = -100°$ to 180°, ψ = 75° to 180° (i.e., the extended, parallel β sheet conformation, with $[\phi,\psi] \approx [-120°,120°]$). P_{II}, P_{II} region, or polyproline conformation will refer to conformations in the region $\phi = -45°$ to $-100°$, ψ = 120° to 180° (i.e., the polyproline conformation with $[\phi,\psi] \approx [-70°,150°]$). Thus, Figure 4A shows that the backbone torsion angles for the sheet residues of RBH lie predominantly in the β region of the Ramachandran map. On the other hand, for LBH, the (ϕ,ψ) distribution shows two clusters (Figure 4B), one of which occurs in the β region and the other in the polyproline or P_{II} region. This suggests that there is a difference in the nature of the sheets in RBH and LBH. Figures 4C and 4D show (ϕ,ψ) distributions for residues in the RBH and LBH L1, L2, and L3 loops of size 1–10, respectively. While the (ϕ,ψ)s of LBH loops show a preponderance to occur in the α_L region, those of the elaborate loops of RBH occur in the α_L, α_R, bridge, β, and P_{II} regions (α_L region: ϕ = 30° to 100°; $\psi = -30°$ to 100°; α_R region: $\phi = -40°$ to $-90°$, $\psi = -20°$ to $-60°$; bridge region: $\phi = -40°$ to 180°, $\psi = -20°$ to 75°).

As the L1, L2, and L3 loops of LBH and the L1 loops of RBH are predominantly 1-residue loops (Figures 3C and 3D) occurring in an α_L conformation, we decided to

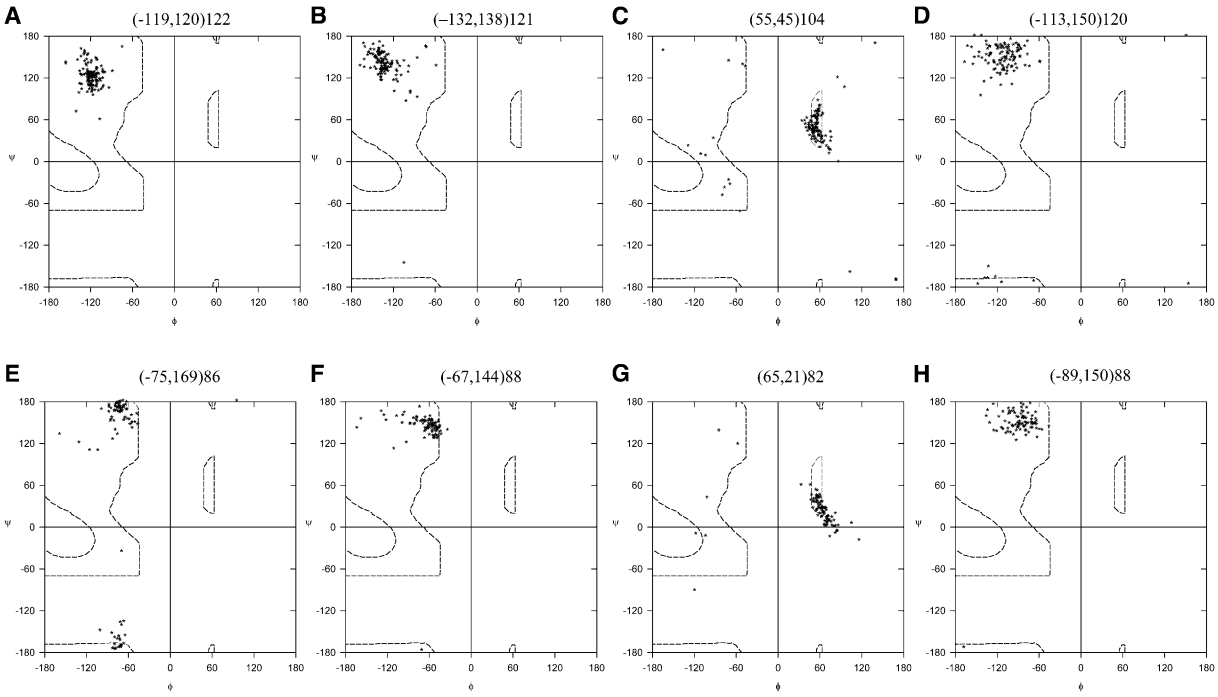

Figure 5. (ϕ,ψ) Plots for Residues Preceding, Including, and Following the 1-Residue Loops of RBH and LBH

(A–H) T is the loop residue. (ϕ,ψ) plots are shown for the (T−2), (T−1), T, and (T+1) residues, respectively, of (A–D) 122 1-residue RBH loops and (E–H) 88 1-residue LBH loops. The mean value and number of data points for the largest cluster are given in each figure. The 1-residue L1, L2, L3 loops of 1M8N (16) and the L2 loops of 1QRE (3) and 1HV9 (8) have been left out in (E)–(H).

examine 1-residue loops in greater detail. In addition to the loop residue, the conformations of 2 residues preceding and 1 residue succeeding the loop residue were examined. The residue forming the loop was labeled T, and 2 residues before and 1 residue after T were labeled (T−2), (T−1), and (T+1), respectively. Figure 5 shows (ϕ,ψ) plots for the (T−2), (T−1), T, and (T+1) residues of 1-residue loops of RBH and LBH. The average (ϕ,ψ) value and the number of points in the largest cluster are given in each figure. In both RBH and LBH, the loop or T residue adopts an α_L conformation, with ϕ and ψ nearly always being positive. Figures 5A–5D show (ϕ,ψ) plots for 122 RBH 1-residue loops. In general, the (T−2), (T−1), T, and (T+1) residues occur in the β, β, α_L, and β conformations, respectively. While the (T−2), (T−1), and (T+1) residues all lie in the β region of the Ramachandran map, the actual region of the map occupied by residues at each position varies, and this is reflected in the varying average values. Figures 5E–5H show the cluster plots for 88 LBH 1-residue loops. The (T−2), (T−1), T, and (T+1) residues occur in the P_{II}, P_{II}, α_L, and P_{II} conformations, respectively. Thus, in LBH, at each "corner," 2 residues preceding and one residue following the loop residue adopt the P_{II} conformation. Sequence constraints resulting from this conformational sequence are discussed in the next section. Figure 5 illustrates that, in proteins with a repetitive fold, the Ramachandran (ϕ,ψ) map can be used to understand protein folding. The repeating unit of the β helix is the β-helical turn or coil (Figure 2). An analysis of backbone torsion angles in structurally equivalent positions in successive coils can provide insights into the conformational rules that these proteins employ to achieve their characteristic repetitive fold. In Figures 5E–5H, (ϕ,ψ)s for the L1, L2, and L3 loops of 1M8N (16) and the L2 loops of 1QRE (3) and 1HV9 (8) have not been plotted. These loops differ from the rest of the LBH loops in that the (ϕ,ψ)s for the (T−2), (T−1), and (T+1) residues do not adopt a P_{II} conformation, but, instead, adopt a β conformation. The three sheets of 1M8N are formed by strictly 4-residue strands, with all strand residues occurring in the β conformation, and with SS bonds at some of the corners serving to covalently hold the sheets together. Thus, the 1M8N β helix and parts of the 1QRE and 1HV9 β helices are built differently when compared to the remaining LBH in the data set. The L1 loops in RBH are true 1-residue loops, as they are preceded and followed by residues in the β conformation. The α_L residue serves to bend the main chain through a right-angle while passing from S1 to S2. In LBH, the P_{II}-P_{II}-α_L-P_{II} sequence causes the main chain to turn through an acute angle (≈60º), resulting in a V-shaped arrangement at the corners. The P_{II}-α_L conformation (−70,150; 50,40) corresponds to a type II β turn.

Loops of sizes 1–10 in RBH and LBH have been sorted and grouped (Experimental Procedures). While the majority of loops were too diverse to be grouped, we found three groups, in RBH, that were interesting. These were a group each of 6-residue L2 loops, 5-residue L2 loops, and 4-residue L3 loops (Figures 6A–6C). These loops were interesting because they were long loops from different RBH that had similar (ϕ,ψ) values. Figure 6A shows a stereoview of the backbone superposition of 12 6-residue L2 loops, and Table 2 gives the mean (ϕ,ψ) values and standard deviations (SD) for the loops. Figures 6B and 6C similarly show stereoviews for 10 5-residue L2

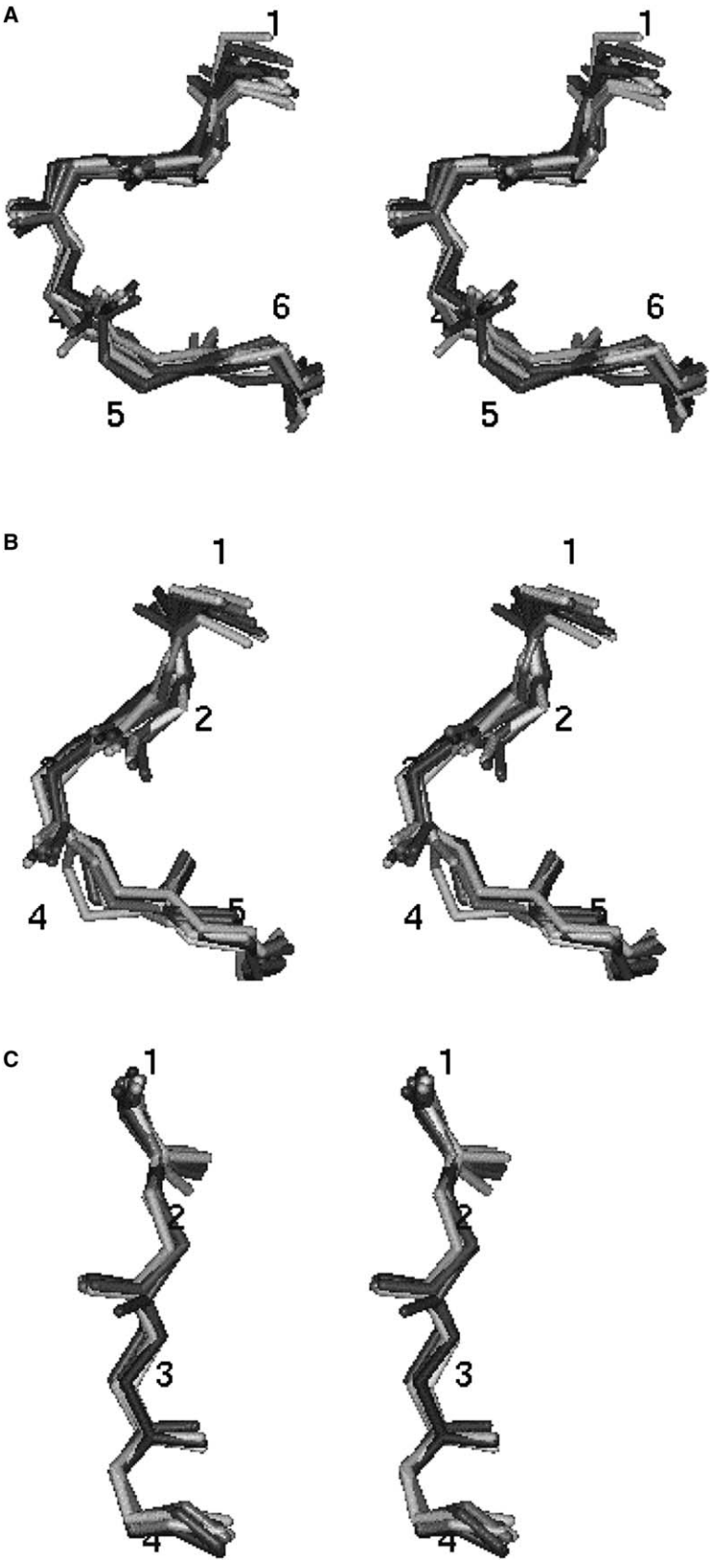

Figure 6. Long Loops of Conserved Conformation in RBH

(A–C) A list of the loops is given in Table 2. Stereoviews showing backbone superpositions of (A) 12 6-residue L2 loops (mean rmsd = 0.88 Å), (B) 10 5-residue L2 loops (mean rmsd = 0.60 Å), and (C) 10 4-residue L3 loops (mean rmsd = 0.53 Å).

Table 2. Mean (ϕ,ψ) Values for the Conformationally Conserved 6-, 5-, and 4-Residue Loops in RBH

	12 6-Residue L2 Loops		10 5-Residue L2 Loops		10 4-Residue L3 Loops	
	Mean	SD	Mean	SD	Mean	SD
ϕ1	54.3	±10.2	56.9	±8.02	51.9	9.26
ψ1	52.3	±15.8	51.5	±19.7	45.8	11.9
ϕ2	−86.1	±20.2	−121	±21.8	−113	17.5
ψ2	133	±11.8	144	±26.3	150	19.5
ϕ3	−69.3	±11.7	−68	±13.0	−129	19.9
ψ3	−32.3	±12.2	−52.7	±18.2	142	14.6
ϕ4	−123	±15.9	−129	±26.2	57.1	5.67
ψ4	−108	±35.1	136	±32.9	56.6	15.5
ϕ5	−101	±39.1	−79.8	±16.3		
ψ5	146	±10.6	−54.9	±17.6		
ϕ6	−77.3	±12.0				
ψ6	−59.4	±17.3				

The mean and SD values are for 12 6-residue L2 loops, 10 5-residue L2 loops, and 10 4-residue L3 loops. The 6-residue L2 loops are: 1air, 236–241; 1ee6, 103–108, 152–157; 1qcx, 173–178; 1rmg, 147–152; 1bhe, 171–176; 1k5ca, 116–121, 250–255; 1tyu, 356–361, 388–393, 435–440, 457–462. The 5-residue L2 loops are: 1air, 187–191; 1jta, 205–209; 1ee6, 126–130, 176–180; 1qcx, 278–282; 1rmg, 245–249, 273–277, 303–307; 1k5ca, 220–224; 1h80a, 297–301. The 4-residue L3 loops are: 1air, 137–140; 1rmg, 70–73, 97–100, 134–137, 157–160; 1bhe, 107–110, 158–161, 208–211; 1tyu, 266–269, 286–289.

loops and 10 4-residue L3 loops, respectively. Only two and three torsion angles in the 6- and 5-residue loops, respectively, show SDs > 25° (Table 2). The small SDs for the (ϕ,ψ)s and the good superpositions indicate that the (ϕ,ψ)s are similar in each group of loops, and that the loops are conformationally conserved. The greatest conformational dissimilarity is seen in Figure 6A between positions 4 and 5 in the 6-residue L2 loops, where the orientation of the peptide unit differs in some of the structures. Considering the peptide plane in the 1BHE loop as the reference plane (the innermost carbonyl group), the largest "flip" from this plane was made by the peptide plane in the 1AIR loop (outermost carbonyl group). The angle between the 1BHE and 1AIR peptide planes is 112°. The other large flip angles are 101° (1K5C) and 94° (1EE6). Thus, while the overall conformations of the loops are similar, localized variations do occur. An examination of the sequences in each group of loops shows that there is no sequence conservation. Thus, each group exhibits conservation of structure in the absence of sequence conservation. It was pointed out above that, in RBH, the L2 loops are usually 6 or 5 residues long and that the L3 loops are 4 residues long (Figure 3C). The 6-, 5-, and 4-residue loops described here are examples of favored RBH L2 and L3 loops. It would thus appear that different RBH proteins (with <20% sequence identity) often use long loops of a certain preferred conformation to fold into their characteristic β-helical structure.

Conformational Sequence for an LBH β-Helical Turn

Average (ϕ,ψ) values were calculated for the sheets of LBH and are given below. To calculate the averages for S1 sheets, for example, all S1 strands in the LBH data set were aligned such that the inside or "i" residues appeared one below the other, as did the outside or "o" residues (shown in Table 1 for a single protein in the data set). The average (ϕ,ψ)s calculated for each sheet, along with the number of residues used to calculate the averages, are given below.

	i	o	i	o	o
S1	(-95,151)	(-120,123)	(-114,118)	(-84,156)	(-59,141)
	52	53	53	49	34
S2	(-95,149)	(-124,123)	(-116,120)	(-107,142)	(-72,132)
	46	47	47	40	15
S3	(-111,145)	(-119,128)	(-121,124)	(-79,173)	(-70,147)
	46	49	50	49	41

The averages indicate that, in strands of all three sheets, the first, fourth, and fifth residues are in a polyproline or P_{II} conformation. The following conformational sequence may be written to describe one β-helical turn of an LBH:

i - o – i – o – o – o – i – o – i – o – o – o – i – o – i – o – o – o

P_{II} - β - β - P_{II} - P_{II} - α_L - P_{II} - β - β - P_{II} - P_{II} - α_L - P_{II} - β - β - P_{II} - P_{II} - α_L

strand strand strand

corner corner

The three strand segments (P_{II}-β-β-P_{II}-P_{II}) are connected by three 1-residue loops in the α_L conformation. As discussed above, the "corners" of LBH proteins are formed by the sequence P_{II}-P_{II}-α_L-P_{II}. Figure 2C shows two β-helical coils; side chains are shown for the strands of sheets S3 and S1 and for the corner between these strands. A regular β strand, with side chains alternately pointing on opposite sides (e.g., inside-outside-inside-outside), is distorted by the occurrence of a P_{II} residue. In LBH strands, the P_{II} residue, which occurs at the (T−2) position of LBH corners, occurs at an "o" or outside position, where the side chain should point outside. However, as a result of the P_{II} conformation, the side chain of the P_{II} residue, instead of pointing outside, points toward the backbone of the next strand in the sheet (i.e., the side chain of a P_{II} residue at the [T−2] position in strand $S3_i$ would point toward the backbone of the corresponding P_{II} residue in strand $S3_{i+1}$, where i and i+1 refer to two successive β-helical turns; see Figure 2C). The side chain of the second P_{II} residue at the LBH corner (the [T−1] residue) points outward, and the side chain of the α_L or T residue again points outward. The side chain of the P_{II} residue immediately after the α_L residue (the [T+1] residue) points inward.

The P_{II} conformation at the (T−2) position imposes steric restrictions as a result of which the residue at this position is often Gly (Figure 2C); as the side chain at (T−2) would point toward the backbone of the next strand in the sheet, long side chains would bump into the backbone of the next strand. Thus, Gly is often preferred at this position. The (ϕ,ψ)s of 88 LBH corners with 1-residue loops have been plotted in Figures 5E–5H. Impressively, G is observed 44 times at the (T−2) position of these corners. A, the second most frequently occurring residue at this position, occurs 11 times. P and E occur five times each. Thus, by far, G is the most preferred residue at the (T−2) position, occurring 50% of the time at the corners. The 88 corners consist of 35 L1, 11 L2, and 42 L3 corners, and G is observed 13, 2, and 29 times, respectively, at these corners. G occurs at the (T−2) position most frequently at the L3 corners. The L3 loops, in the LBH trimers, occur away from the 3-fold axis

and are the shortest and most regular (Figure 2B). G residues are probably present for steric reasons, but they might also serve as a cue for correct folding and registry of the coils of LBH. G is observed least frequently at the L2 corners. The longer L2 loops (Figures 1B and 3D) probably relieve steric strain at the (T−2) position, making the need for G less critical. At the L1 corners (which face the 3-fold axis), G residues are observed at (T−2), but they are observed less frequently compared to the L3 corners. The L1 loops are a little longer and less regular than the L3 loops (Figure 3D), which might again relax the critical need for G. Out of 13 Gs observed at the 35 L1 corners, 6 belong to 1HV9. 1HV9 differs from other LBH because it has 4-residue strands forming S2 and 5-residue strands forming S3 and S1. The short S2 strands perhaps tighten the helical coil, resulting in the need for G residues at the (T−2) positions of the L3 and L1 loops. Other frequently occurring residues at the (T−2) positions of the corners are: L1, A = 6, S = 4; L2, A = N = R = 2; L3, P = 4, A = 3. Conservation of (T−2) G residues is also evident when a comparison is made between homologous β helix proteins. 1KRR and 1OCX share 42% sequence identity, and, hence, only 1OCX has been included in the data set. Residues occurring at the (T−2) positions in the two proteins are:

1KRR–G59, G79, G99, G135, G141, G153, G159, A175;
1OCX–T58, G77, G97, G133, G139, G151, A157, G173.

Conservation of the majority of Gs at (T−2) is evident. The only mutational changes observed are to small residues, G/A, G/T.

The P_{II} residue at (T+1) has its side chain pointing inward, toward the interior of the β helix (Figure 2C). Owing to steric hindrance or the limited space available in the turn region, residues with small, hydrophobic side chains will be preferred at this position. Accordingly, V, A, S, T, and C are the most frequently occurring residues at the (T+1) position in the set of 88 corners in Figures 5E–5H: V = 25, A = 20, S = 10, T = 10, and C = 12. (The number of residues contributed by the L1, L2, and L3 loops to this total are as follows: 35 L1 loops: V = 9, A = 10, S = 5, T = 6, C = 3; 11 L2 loops: V = 3, A = 4, S = 0, T = 0, C = 2; 42 L3 loops: V = 13, A = 6, S = 5, T = 4, C = 7). In Figure 4B, a cluster of β and a cluster of P_{II} sheet residues are observed. The presence of ∼3 P_{II} residues and 2 β residues in each strand of each sheet of LBH explains the occurrence of the two clusters in the figure. The equilateral prism-like appearance of LBH arises from the occurrence of the P_{II}-P_{II}-α_L-P_{II} conformational sequence at the corners.

The sheet averages and consensus conformational sequence for one β-helical turn of 1M8N, a β helix that differs from the remaining LBH, are given below:

S1	(-120,156)	(-120,118)	(-119,126)	(-127,122)
	7	7	7	7
S2	(-125,149)	(-109,114)	(-115,120)	(-117,137)
	6	6	6	6
S3	(-118,163)	(-132,125)	(-126,120)	(-131,143)
	6	6	6	6

E-β-β-β-α_L-E-β-β-β-α_L-E-β-β-β-α_L

E stands for a β conformation with a large ψ value (>135°), e.g., ∼(−120,160). Each α_L residue is followed by an E residue, and no P_{II} residues are observed; all strands in the protein are strictly 4 residues long, and some of the corners are covalently stabilized by SS bonds. Corners of the kind observed in 1M8N are also observed at some of the L2 corners of 1HV9 and 1QRE; this may be an effect of the shorter, 4-residue S2 strands in the β-helical coils. Thus, polypeptide chains in nature have arrived at more than one solution to the problem of generating LBH.

It has been pointed out that the LBH 1LXA has been built from an imperfect hexapeptide repeat (Raetz and Roderick, 1995; Vuorio et al., 1994):

```
          i-2          i   i+1
                          (T-2)T   (T+1)
…(V,A,S,T,C)X(I,V,L)GXX(V,A,S,T,C)X(I,V,L)G…
                  └──────────────────────┘
                       hexapeptide repeat
```

An aliphatic residue (I, V, L) occurs at every sixth position (called i), a small residue (V, A, S, T, C) appears at the (i−2) position, and a G residue often appears at the (i+1) position (Raetz and Roderick, 1995). This motif is arranged around the LBH such that the aliphatic residue occurs at the center of the strand, G occurs at the (T−2) position of an LBH corner, and the small residue (V, A, S, T, C) occurs at the (T+1) position of the corner. As described above, the residue at (T−2) needs to be G because a side chain at this position is likely to bump into the backbone of the next helical turn. A G residue at (T−2) would thus be a sequence constraint arising from conformational considerations. The residue at (T+1) must be small because it occurs in the turn region, which is spatially constrained, and has its side chain pointing inside the β helix. A small residue at (T+1) would be another sequence constraint arising from conformational compulsions. Thus, while the hexapeptide repeat motif has been pointed out before, our analysis provides the stereochemical basis for the rationale behind the choice of residues.

Conformational Sequence for an RBH β-Helical Turn

The average (ϕ,ψ) values for each RBH sheet, along with the number of residues used to calculate the averages, is given below.

	i	o	i	o	i	o	i
S1	(-93,150)	(-112,131)	(-113,139)	(-122,128)	(-119,119)	(-130,134)	(-103,129)
	17	55	146	153	150	144	26
S2			(-104,150)	(-120,130)	(-113,124)	(-121,134)	(-99,146)
			124	140	135	127	45
S3		(-119,143)	(-116,129)	(-120,126)	(-114,118)	(-117,138)	(-118,138)
		23	138	139	135	95	22

The averages indicate that, in general, all sheet residues are in an extended conformation. In RBH, L1 is a 1-residue loop, while L2 and L3 are longer loops. Figures 7A–7D show (ϕ,ψ) maps for the first and last residues of the L2 and L3 loops of RBH. There is a considerable spread of points in the figures because of the way in which loops have been defined and also because of their elaborate and diverse nature. Residues not in registered sheets are loops. Thus, for instance, residues that are a continuation of strands, but are not a part of registered sheet regions, would become loops and occupy the β regions of the map. Nevertheless, despite the spread,

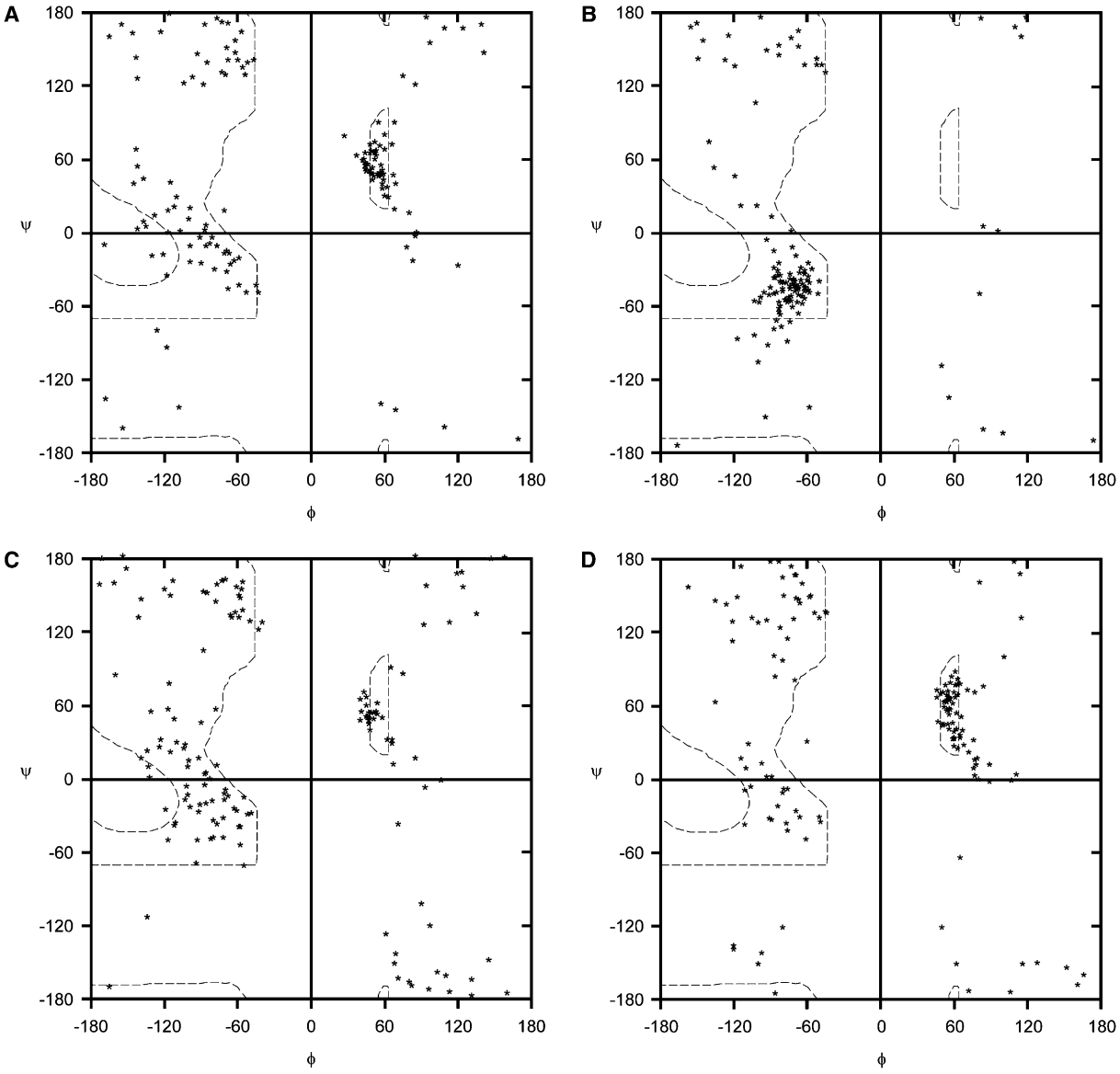

Figure 7. (ϕ,ψ) Plots for the First and Last Residues in the L2 and L3 Loops of RBH

(A–D) (ϕ,ψ) plot for (A) the first residue (N terminus) in the L2 loops, (B) the last residue (C terminus) in the L2 loops, (C) the first residue in the L3 loops, and (D) the last residue in the L3 loops.

clusters of residues give us an idea about the design principles used to generate RBH. The L2 loops often start with an α_L residue that they use to bend S2 (Figure 7A). Interestingly, almost as a rule, the L2 loops end in a right-handed α-helical (α_R) residue (Figure 7B). The α_R residue appears to be the cue for the S3 strand to begin. The conserved α_R conformational feature is unique to RBH and is a point of difference with the LBH. The LBH β-helical coils do not have a similar conserved α_R position. Their loops consist largely of α_L residues (Figure 4D). The α_R residue might be playing an important structural role in modulating the handedness of the β helix and might, in fact, be ensuring the right-handedness of the RBH. The L3 loops in several RBH may be regarded as forming a narrow fourth sheet. In the present study, we have not defined a fourth sheet and, instead, have regarded connections between S3 and S1 as L3 loops. Figure 7C shows that the conformation of the first residue of L3 loops is not well defined and can be an α_L, α_R, bridge, or P_{II} residue. Figure 7D shows that L3 loops usually end in an α_L residue. This residue positions the main chain for forming the S1 strand of the next β-helical turn. Thus, like LBH, RBH also largely use α_L residues to bend their backbone. The α_R residue at the last position of the L2 loops is the only exception. Each RBH β-helical turn may be described by the conformational sequence:

S1 strand-(α_L)-S2 strand-(α_L-xxx-α_R)-S3 strand-(α_L/α_R-xxx-α_L),

where loops are enclosed in parentheses and "xxx" indicates a varying number of residues occurring in varied conformations.

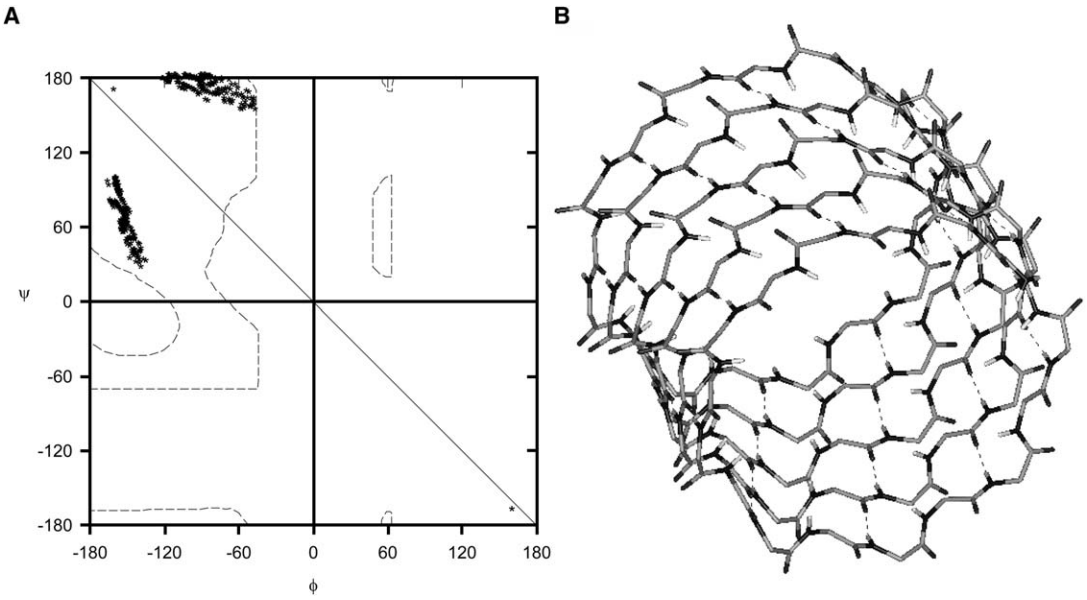

Figure 8. Generating Connectionless β Helices

(A) (ϕ,ψ) pairs that yield β helices with one H bond per residue; (ϕ,ψ)s for the Perutz nanotube (ϕ_1 = 160°, ψ_1 = −170°; ϕ_2 = −161°, ψ_2 = 168°) are shown as two points lying below and above the diagonal.

(B) A β helix generated from a pair of (ϕ,ψ) values taken from (A) (ϕ_1 = −94°, ψ_1 = 178°; ϕ_2 = −150°, ψ_2 = 61°). The β helix is left-handed and has 17.4 residues per turn.

LBH are subject to several conformational and sequence constraints. They are built from a repeating hexapeptide motif, their corners are formed from P_{II}-P_{II}-α_L-P_{II} conformational sequences, they require a G residue at the (T−2) position, and they require a small residue at the (T+1) position of their corners. RBH appear not to be subject to such conformational or sequence constraints and can use a wider variety of residues at each position of their β-helical turns. In this respect, the RBH would appear to be more sophisticated than the LBH. Both RBH and LBH, however, are subject to the constraint that the inside or core of the β helix must be hydrophobic while the outside must be polar (Figure 3E). In both protein families, this is achieved by having segments in which hydrophobic and polar residues alternate. During protein folding, when the sheets are formed, hydrophobic residues are directed inward and hydrophilic residues are directed outward. It is difficult to pinpoint a single conformational feature as being responsible for crafting left- or right-handedness into a single helical turn of the β helix. However, once the handedness (left or right) is determined in the first coil of the protein, it is possible that subsequent coils merely follow suit.

Ideal β Helices

In both RBH and LBH, extended strands are connected by loops to form a β helix. In principle, β-helical structures may also be generated in a continuous, loopless fashion in which the extended polypeptide chain is wound around a cylinder, resulting in the formation of a parallel β helix, with H bonds between adjacent turns of the β helix. Such a connectionless β helix (also described as a nanotube) was first proposed by Perutz et al. (2002a, 2002b) as a model for the poly-Gln aggregates of Huntington's disease (Figure 1C). Their β helix has been generated by repeating a pair of (ϕ,ψ) values: (160°,−170°), (−161°,168°). A notable feature of their β helix nanotube is that the repeating unit consists of 2 residues that adopt an almost fully extended conformation, and that lie on opposite sides of the diagonal bisecting (ϕ,ψ) space (Figure 8A). The conformation ($\phi \approx$ 160°,$\psi \approx$ −170°) is unfavorable for amino acid residues other than Gly, as it occurs in the disallowed region of the Ramachandran map for non-Gly residues.

We have investigated the ideal, connectionless, continuous β helix further to see if alternative forms of the Perutz nanotube can be generated. We tried to calculate all (ϕ,ψ) values in allowed β sheet regions of the Ramachandran map that give rise to the Perutz nanotube with 22 residues per turn, 0.22 Å rise per residue, and two backbone H bonds per residue. A program based on the method by Sugeta and Miyazawa (1967) was used to for the calculations. Using a single (ϕ,ψ) repeat, it was not possible to generate any kind of β helix. By repeating a pair of (ϕ,ψ)s, it was possible to generate β helices with 22 residues per turn and 0.22 Å rise per residue, but with only one backbone H bond per residue. Figure 8A shows the (ϕ,ψ) pairs that give rise to such β helices. There are two clusters in the map; a pair of (ϕ,ψ)s, one from each cluster, when repeated over and over again gives rise to a β helix such as the one shown in Figure 8B, with only one backbone H bond per residue. It should, in principle, be possible to generate ideal β helices with two H bonds per residue by repeating a set of three or a set of four (ϕ,ψ)s (which would permit greater variation in the [ϕ,ψ] values as compared to repeating a pair of [ϕ,ψ]s) or by introducing nonplanarity at the peptide unit. Efforts in this direction are in progress.

We also tried to generate β helices manually, by using the software InsightII and by assigning pairs of (ϕ,ψ) values. We found that β helices with two H bonds per residue could be generated by using a pair of (ϕ,ψ) values such as (−150°,122.5°), (150°,−140°), where (ϕ_1,ψ_1) and (ϕ_2,ψ_2) lie on opposite sides of the diagonal in the Ramachandran map and in diagonally opposite quadrants. Such a β helix would be accessible only to Gly residues, as both points would not lie in the allowed β sheet region of the Ramachandran map for non-Gly residues. Such a helix generated from "mirror" values on opposite sides of the diagonal would be reminiscent of the $\pi_{(L,D)}$ helix of the linear peptide antibiotic gramicidin A, which has an alternating sequence of L and D residues (Urry, 1971; Ramachandran and Chandrasekharan, 1972). The β helix generated by Perutz et al. (2002a, 2002b) (160°,−170°; −161°,168°; Figures 8A and 1C) is actually an LD helix.

(ϕ,ψ)s for the Perutz et al. (2002a, 2002b) poly-Gln β helix model and also for the (ϕ,ψ) pairs in Figure 8A do not lie in densely populated β regions of the Ramachandran map. Instead, they lie at the edges of the allowed β regions. One might expect smaller probabilities for such regions of the map to be populated in real proteins. The β helix proteins, on the other hand, form their parallel β sheets, and also the loops that bend their β sheets, by using conformations that occur in well-populated regions of Ramachandran (ϕ,ψ) space. Thus, β helices as observed in proteins would appear to be more attractive as models for amyloid deposits (Wetzel, 2002). The RBH model may be preferred over the LBH one because the former enjoys greater freedom in the amino acid sequence it can incorporate, and also because stretches of sequence, in an amyloid-forming protein, that are incompatible with the formation of sheets can be extruded out as irregular loops.

Experimental Procedures

Abbreviations

Abbreviations used in this article are as follows: RBH, right-handed β helix (helices); LBH, left-handed β helix (helices); P_{II}, polyproline conformation; α_L, left-handed α-helical region; α_R, right-handed α-helical region; β, β-region; S1, sheet1; S2, sheet2; S3, sheet3; L1, loop1; L2, loop2; L3, loop3; SD, standard deviation; rmsd, root mean square deviation; aa, amino acid.

Data Set

The data set of β helix protein crystal structures was selected by using the SCOP database (Murzin et al., 1995). There were 94 β helix structures in the PDB at the time of this study. Initially, a set of 25 RBH and 9 LBH was compiled and included representatives from each species in each SCOP superfamily. From this set, proteins that had a distorted β sheet (1EA0) or that were regularly SS bonded (1EZG) were removed. A ClustalW alignment (Thompson et al., 2000) of the remaining RBH and LBH was then carried out. Only those proteins that showed <20% sequence identity to all other proteins in the set were retained, because our aim was to use a representative data set of nonhomologous β helix proteins. As far as possible, structures with the best resolution were included. The final data set consisted of 14 RBH and 7 LBH. The 14 RBH (PDB code, enzyme name [chain and segment], resolution) are: 1AIR, pectate lyase C, 2.2 Å; 1JTA, pectate lyase A (A), 1.8 Å; 1EE6, pectate lyase (A), 2.30 Å; 1QCX, pectin lyase B (A), 1.70 Å; 1RMG, rhamnogalacturonase A, 2.0 Å; 1BHE, polygalacturonase, 1.9 Å; 1K5C, endopolygalacturonase I (A), 0.96 Å; 1DBG, chondroitinase B (A), 1.70 Å; 1H80, iota-carrageenase (A), 1.6 Å; 1GQ8, pectin methylesterase (A), 1.75 Å; 1TYU, tailspike endorhamnosidase, 1.8 Å; 1DAB, virulence factor P.69 pertactin (A), 2.50 Å; 1HF2, bacterial cell-division inhibitor (A), 2.2 Å; 1K4Z, C-terminal domain of adenylyl cyclase-associated protein (A), 2.30 Å. The 7 LBH in the data set are: 1LXA, UDP N-acetylglucosamine o-acyltransferase (1–188), 2.6 Å; 3TDT, tetrahydrodipicolinate N-succinyltransferase (101–274), 2.0 Å; 1OCX, maltose o-acetyltransferase (A), 2.15 Å; 1MR7, streptogramin A acetyltransferase (A:30–158), 1.8 Å; 1HV9, UDP-N-acetylglucosamine pyrophosphorylase (A:252–452), 2.10 Å; 1QRE, carbonic anhydrase (A), 1.46 Å; 1M8N, antifreeze protein (A), 2.45 Å.

Partitioning of the Data Set into Sheets and Loops

Backbone torsion angles, (ϕ,ψ), for each β helix protein in the data set were calculated by using the software InsightII (Accelrys Inc., 1999). Each β helix consists of three parallel β sheets, whose strands are connected by loops (Figure 1). The sheets and loops are visually discernible from the C^α-trace of the protein displayed on the InsightII window. Each β helix was partitioned into sheet and loop regions in a semiautomated way. Sheets were first identified by using (ϕ,ψ) values in combination with visual inspection of the C^α-trace of the protein. Using the C^α-trace display, all residues forming a sheet (e.g., sheet1) were visually identified. The sheet was identifiable owing to the registry between the strands. (ϕ,ψ)s for the identified sheet residues were checked; residues were accepted as sheet residues if their (ϕ,ψ) values were in the β sheet region of the Ramachandran map (ϕ = −45° to 180°; ψ = 60° to −170°; Ramachandran et al., 1963; Ramakrishnan and Ramachandran, 1965; Ramachandran and Sasisekharan, 1968). Occasionally, residues with marginally deviating (ϕ,ψ) values were also accepted if, upon visual inspection, they were clearly found to form a part of the sheet. The strands were trimmed based on strand registry, i.e., if a strand in the sheet was substantially longer than the remaining registered strands, it was trimmed, and only registered residues were included as part of the sheet. Sheets 1, 2, and 3 for each β helix in the data set were thus identified. Sheet1 (or S1), sheet2 (or S2), and sheet3 (or S3) residues identified for an LBH, 1LXA, are given in Table 1.

Once the sheet residues in β helices had been identified, the residues that were not in sheets were regarded as loops. One complete β-helical turn (or coil) of a β helix consists of alternating strand and loop segments that appear in the order: S1 β strand-L1 loop-S2 β strand-L2 loop-S3 β strand-L3 loop-on to the S1 β strand of the next β-helical coil (Figure 2). While the L1, L2, and L3 loops range in size from 1 to >100 residues, in the present analysis, only loops of sizes 1–10 were studied. 1-, 2-, 3-, …, 10-residue loops were separated out, independently for RBH and LBH. Loops of each size (e.g., all 6-residue L1, L2, and L3 loops in RBH) were then further sorted into groups, such that members of a group had similar (ϕ,ψ)s. The majority of loops of sizes 2–10 were too diverse to be grouped. However, some interesting groups were found for 4-, 5-, and 6-residue loops, and these are discussed.

In Table 1, each row constitutes one complete turn of the β helix. The table shows three blocks of sheet residues, one each for sheets 1, 2, and 3. Going down the different rows of a block (the S3 block, for e.g.), one encounters the different strands of the corresponding parallel β sheet (sheet3). The sheets were labeled 1, 2, and 3 as follows. In all RBH, the sheet whose strands ended in 1-residue α_L loops was identified as S1; the sheet immediately after the α_L loops was labeled S2; the third sheet was labeled S3 (Figure 2A). In the trimeric LBH, loops located closest to the 3-fold axis were labeled L1 (Figure 2B; coordinates for LBH trimers were obtained by using the PQS server [Henrick and Thornton, 1998]). Strands N-terminal to L1 constituted the S1 sheet, and strands C-terminal to L1 constituted the S2 sheet. This nomenclature allows us to take advantage of the 3-fold axis of the trimer to label loops that occupy similar locations in the different trimeric LBH the same way. For 1M8N, which is not a trimer, the first sheet encountered from the N terminus was labeled S1, followed by S2 and S3. The orientation of residue side chains in the three sheets was also examined. If a side chain pointed toward the interior of the β helix, it was labeled as "i," and if it pointed toward the outside of the β helix, it was labeled "o" (Table 1). The "i-o" assignment was done by visual inspection of the side chain orientation in the C^α-trace of each β helix protein by using the software spdbview (Guex and Peitsch, 1997). Thus, each β helix protein in the data set was partitioned into sheets and loops, and the side chain orientation for each sheet residue was labeled, as shown for 1LXA in Table 1. While analyzing the conformational features of

sheets and loops, only complete β-helical coils were considered. Coils formed from antiparallel β strands were left out (e.g., 1M8N: A:102–121).

Amino Acid Propensity Calculation

Chou and Fasman (1974) propensity values were calculated to examine if different amino acid residues were preferred in different regions of β helices (sheets, loops, inside and outside β helices). The propensity of an amino acid (aa) to occur in any subset of the data set (e.g., sheets) was calculated as follows:

$$\text{Propensity, P, of an aa to occur in subset} = \frac{(\text{Number of the aa in subset})/(\text{Number of all aa in subset})}{(\text{Number of the aa in dataset})/(\text{Number of all aa in dataset})}, \quad (1)$$

$P \geq 1.4$ was taken to indicate preference, $P \leq 0.6$ to indicate avoidance, and $0.6 < P < 1.4$ to indicate neutrality (neither preference nor avoidance).

Acknowledgments

We are grateful to anonymous reviewers for their detailed and constructive comments, which have helped to improve our manuscript. We thank Sir Alan Fersht, Medical Research Council Centre for Protein Engineering, for the coordinates of the Perutz nanotube model. Thanks to S. Aravinda and Ramesh Sistla for help with plotting the (ϕ,ψ) maps. P.I.'s work is supported by a grant from the Department of Science and Technology under the Women Scientists Scheme.

Received: July 11, 2005
Revised: October 6, 2005
Accepted: November 5, 2005
Published online: March 14, 2006

References

Accelrys, Inc. (1999). InsightII Version 2000 Molecular Modeling System (San Diego: Accelrys, Inc.).

Chou, P.Y., and Fasman, G.D. (1974). Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochemistry *13*, 211–222.

Cohen, F.E. (1993). The parallel β-helix of pectate lyase C: something to sneeze at. Science *260*, 1444–1445.

Cowen, L., Bradley, P., Menke, M., King, J., and Berger, B. (2002). Predicting the β-helix fold from protein sequence data. J. Comput. Biol. *9*, 261–276.

Dobson, C.M. (2002). Getting out of shape. Nature *418*, 729–730.

Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-Pdbviewer: an environment for comparative protein modeling. Electrophoresis *18*, 2714–2723.

Henrick, K., and Thornton, J.M. (1998). PQS: a protein quaternary structure file server. Trends Biochem. Sci. *23*, 358–361.

Jenkins, J., and Pickersgill, R. (2001). The architecture of parallel β-helices and related folds. Prog. Biophys. Mol. Biol. *77*, 111–175.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Nesloney, C.L., and Kelly, J.W. (1996). Progress towards understanding β-sheet structure. Bioorg. Med. Chem. *4*, 739–766.

Perutz, M.F., Finch, J.T., Berriman, J., and Lesk, A. (2002a). Amyloid fibers are water-filled nanotubes. Proc. Natl. Acad. Sci. USA *99*, 5591–5595.

Perutz, M.F., Pope, B.J., Owen, D., Wanker, E.E., and Scherzinger, E. (2002b). Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid β-peptide of amyloid plaques. Proc. Natl. Acad. Sci. USA *99*, 5596–5600.

Pickersgill, R.W. (2003). A primordial structure underlying amyloid. Structure *11*, 137–138.

Raetz, C.R.H., and Roderick, S.L. (1995). A left-handed parallel β helix in the structure of UDP-N-acetylglucosamine acyltransferase. Science *270*, 997–1000.

Ramachandran, G.N., and Chandrasekharan, R. (1972). Conformation of peptides containing both L- and D-residues: Part I - Helical structure with alternating L and D residues with special reference to the LD-ribbon and LD-helices. Indian J. Biochem. Biophys. *9*, 1–11.

Ramachandran, G.N., and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. Adv. Protein Chem. *23*, 283–438.

Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J. Mol. Biol. *7*, 95–99.

Ramakrishnan, C., and Ramachandran, G.N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. Biophys. J. *5*, 909–933.

Richardson, J.S. (1976). Handedness of crossover connections in β-sheets. Proc. Natl. Acad. Sci. USA *73*, 2619–2623.

Sternberg, M.J.E., and Thornton, J.M. (1977). On the conformation of proteins: the handedness of the connection between parallel β-strands. J. Mol. Biol. *110*, 269–283.

Sugeta, H., and Miyazawa, T. (1967). General method for calculating helical parameters of polymer chains from bond lengths, bond angles and internal-rotation angles. Biopolymers *5*, 673–679.

Thompson, J.D., Plewnial, F., Thierry, J.-C., and Poch, O. (2000). Rapid and reliable global multiple alignments of protein sequences detected by database searches. Nucleic Acids Res. *28*, 2919–2926.

Urry, D.W. (1971). The gramicidin A transmembrane channel: a proposed $\pi_{(L,D)}$ helix. Proc. Natl. Acad. Sci. USA *68*, 672–676.

Vuorio, R., Härkönen, T., Tolvanen, M., and Vaara, M. (1994). The novel hexapeptide motif found in the acyltransferases LpxA and LpxD of lipid A biosynthesis is conserved in various bacteria. FEBS Lett. *337*, 289–292.

Wetzel, R. (2002). Ideas of order for amyloid fibril structure. Structure *10*, 1031–1036.

Yoder, M.D., Keen, N.T., and Jurnak, F. (1993a). New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. Science *260*, 1503–1507.

Yoder, M.D., Lietzke, S.E., and Jurnak, F. (1993b). Unusual structural features in the parallel β-helix in pectate lyases. Structure *1*, 241–251.