BIOPOLYMER PHYSICS
AND PHYSICAL CHEMISTRY

Structure of α-α-Hairpins with Short Connections in Globular Proteins

E. V. Brazhnikov and A. V. Efimov

Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region, 142290 Russia;
E-mail: efimov@protres.ru

Received June 4, 2000

Abstract—The analysis of conformations of more than 100 α-α-hairpins with closely packed helical segments and connections up to four amino acid residues in length was carried out. Five types of the connections were revealed, and their φ and ψ values on the Ramachandran map were found. Each type of α-α-hairpins was shown to have a unique sequence pattern for hydrophobic and hydrophilic residues.

Key words: conformational analysis, protein structure, α-α-hairpins

INTRODUCTION

A number of distinct arrangements of secondary structure elements of the specifically folded polypeptide chain are widespread in proteins. These are called super-secondary structures or structural motifs [1]. In globular proteins, there are both simple structural motifs consisting of two consecutive regular regions of the polypeptide chain and complex ones incorporating three regions and more. Among complex motifs, the β-α-β motif, the Rossmann fold, the abcd unit, the ABCD structure, and the 3β-corner are best characterized. The most widespread simple motifs are α-α- and β-β-hairpins, α-α- and β-β-corners, and L- and V-shaped structures consisting of two α-helices. In proteins some motifs are right-handed only, while others can acquire both the right- and left-handed forms as dependent on mutual arrangement of polypeptide chain regions [2–5].

α–α-Hairpins are abundant in globular proteins, some of which consist predominantly of these motifs [6–8]. The hairpins can differ in length of the α-helices, as well as in the length and conformation of their connections, i.e., regions between the helices. The Ω angle between the α-helix axes can also vary within a limited range. A connection pertains to the disordered part of the molecule, and in principle, its every residue can have a conformation from an allowed region on the Ramachandran map.

To describe a connection, the method proposed by one of us earlier [9] is used. Currently, this method is widely used for description of the structure of irregular regions of globular proteins [10–15].

A stereochemical analysis of α-α-hairpins was reported earlier [16], but then the scanty experimental data allowed confirmation of only some results of the analysis by examples from proteins of known structure. Therefore, the current study was focused on the search in PDB for α-α-hairpins with short connections and a detailed analysis of their structures. As a result, more than 100 α-α-hairpins of nonhomologous proteins were selected, each with a connection of no more than 4 residues. It was shown that these connections can be subclassified into five types characterized by a certain number and conformation of the residues. The relationship between the α-α-hairpin structure and sequence patterns for hydrophobic, hydrophilic, and glycine residues is also discussed.

RESULTS AND DISCUSSION

In proteins, the most abundant α-hairpin is that with the connection conformation $\gamma\alpha_L\beta/\delta$ where β/δ means that this residue can adopt either β or δ conformation (Table 1). In all 46 hairpins of this type, the γ-position is most often occupied by Glu (9 times), Leu (7 times), Thr, and Ala (5 times each), while Gly, Pro, Cys, Asp, and Phe are completely absent. The α_L-position is predominantly occupied by Gly (24 times) and hydrophilic residues with flexible side chains (Asp, Glu, Lys, Asn, Arg), while bulky hydrophobic residues (Phe, Tyr, Trp, Val, Met, Ile) are absent from this position. In contrast, Gly is absent from the β/δ-position, and bulky hydrophobic residues are most rare there too, which confirms the reported results of a stereochemical analysis. Bulky hydrophobic residues are forbidden to occupy the β/δ

Table 1. Conformation of connection $\gamma\alpha_L\beta/\delta$

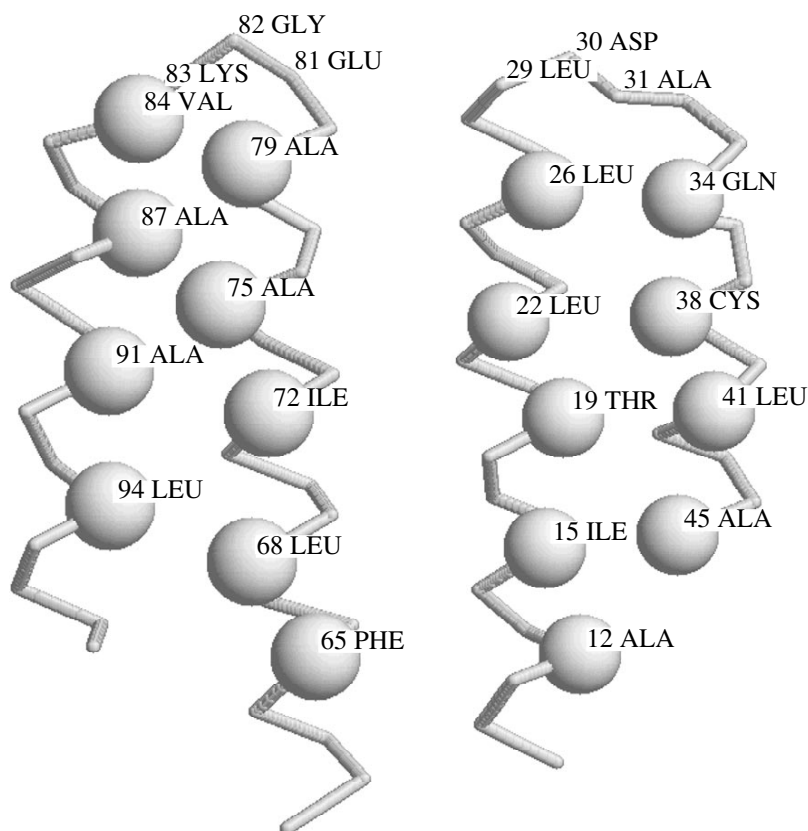| PDB code	Res.	φ	ψ	Res.	φ	ψ	Res.	φ	ψ	Resolution, Å
14PS	A16	−73	20	E17	45	39	R18	−110	88	2.6
14PS	L182	−83	−19	N183	70	137	S184	−122	87	2.6
1A17	A41	−95	15	K42	65	12	D43	−93	79	2.45
1A17	T75	−104	19	E76	60	25	C77	−94	77	2.45
1A17	L109	−84	1	G110	82	−1	K111	−89	78	2.45
1A1W	L70	−73	−14	R71	69	39	R72	−86	−171	NMR
1A59	Y293	−87	1	D294	52	64	R295	−138	71	2.09
1A1Z	K70	−66	−15	R71	73	29	R72	−100	95	NMR
1E2A	N33	−93	7	G34	83	3	D35	−87	73	2.3
1AH7	N103	−87	5	K104	66	26	D105	−98	85	1.5
1AM7	R110	−82	7	G111	95	9	D112	−90	69	2.3
1AUE	E2042	−111	6	R2043	49	45	N2043	−119	98	2.33
1AWC	A16	−76	14	G17	70	41	Q18	−102	78	2.15
1AWC	Y48	−84	−6	G49	80	36	H50	−93	90	2.15
1AWC	E81	−83	−10	G82	82	17	H83	−84	81	2.15
1AWC	H114	−100	15	N115	56	35	H116	−89	77	2.15
1AWC	N147	−98	−2	G148	69	21	N149	−96	89	2.15
1B0J	T283	−113	31	G284	64	32	R285	−128	35	2.5
1B89	V21	−67	−21	S22	60	53	N23	−100	68	2.6
1B8J	A332	−69	−6	A333	59	45	N334	−123	96	1.9
1BCF	W35	−38	−9	G36	96	7	L37	−88	79	2.9
1BD8	R18	−81	−14	G19	79	22	D20	−83	90	1.8
1BD8	T84	−95	0	G85	83	18	F86	−88	82	1.8
1BD8	R149	−78	−15	G150	80	22	A151	−86	78	1.8
1BDJ	A704	−75	−17	Q705	63	37	D706	−84	103	2.68
1BED	L79	−93	2	E80	54	24	V81	−100	27	2.1
1BF3	E319	−119	10	G320	60	22	R321	−95	55	2.2
1BGF	E74	−79	−10	K75	57	40	N76	−87	92	1.45
1BGJ	E319	−77	−3	G320	56	35	R321	−117	70	3.0
1BI7	E88	−93	−5	G89	82	9	F90	−84	83	3.4
1BI7	L121	−63	−11	G122	89	27	H123	−86	71	3.4
1BKE	H247	−52	−17	G248	90	6	D249	−88	106	3.15
1BL5	M367	−66	−21	G368	106	1	W369	−92	57	2.5
1BLX	E117	−92	−9	G118	98	12	H119	−90	97	1.9
1BLX	T84	−81	−18	G85	91	13	F86	−88	98	1.9
1BPO	Q442	−77	−31	G443	96	31	R444	−104	123	2.6
1BUO	L109	−74	−13	E110	58	56	I111	−108	98	1.9
1BXL	K157	−97	27	E158	55	52	M159	−131	56	NMR
1QSP	E53	−70	−13	K54	52	47	N55	−92	110	2.7
256B	E81	−94	8	G82	79	10	K83	−93	71	1.4
103L	Q123	−82	8	K124	50	35	R125	−90	80	1.9
1COL	T127	−116	1	G128	67	17	N129	−106	110	2.4
1ERC	Q9	−84	−5	C10	76	45	V11	−139	107	NMR
1ERD	S11	−84	−11	C12	72	11	E13	−87	80	NMR
1ROP	L29	−90	−1	D30	57	38	A31	−94	89	1.7
1WAS	N143	−101	36	G144	63	−20	N145	−61	52	2.7
Average value		−84	−2		71	29		−98	76	
Residue conformation		γ			α_L			β/δ		

Note: Res., amino acid residue.

Fig. 1. α-α-Hairpins with connections of the type $\gamma\alpha_L\beta/\delta$. On the left, a left-handed hairpin from cytochrome B562 (PDB code 256B); on the right, a right-handed hairpin from ROP-protein (PDB code 1ROP). Hydrophobic side chains forming minimal hydrophobic clusters of α-helices are shown as gray spheres.

position to avoid main chain dehydration at the N terminus of the second helix [16].

This conformation of connections can occur both in the right- and left-handed hairpins (Fig. 1). In spite of the similarity of polypeptide backbone conformations, right- and left-handed hairpins differ much in the sequence patterns for hydrophobic, hydrophilic, and glycine residues (see [16] and Fig. 6). Also, it is noteworthy that in these hairpins the α-helices are packed "side by side" in accordance with the principle of close packing [17], i.e., the "knobs" of the minimal

Table 2. Conformation of connection $\gamma\varepsilon$

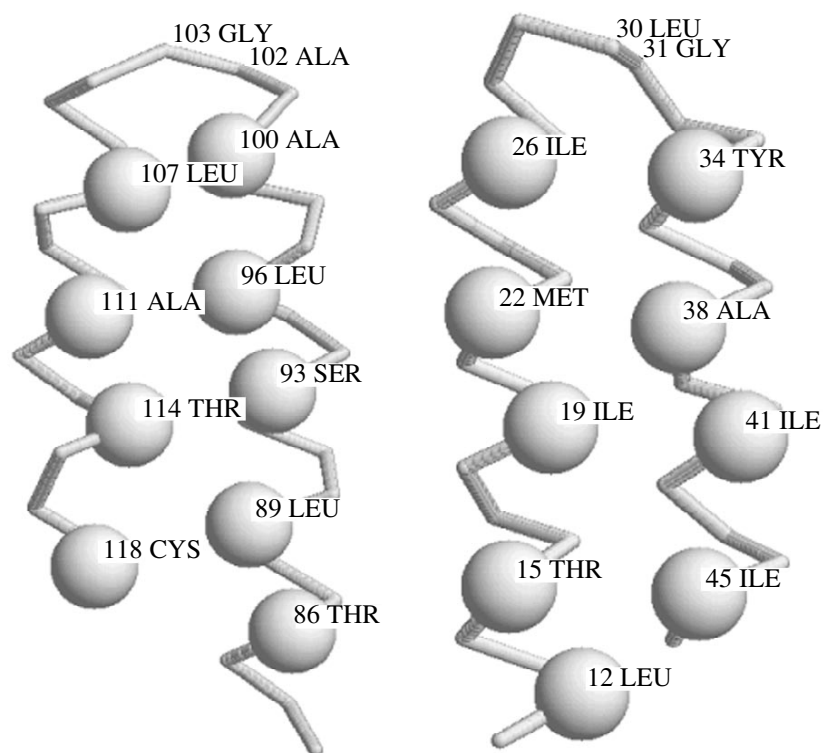| PDB code	Res.	φ	ψ	Res.	φ	ψ	Resolution, Å
1AG1	I198	−93	−73	G199	178	152	2.36
1BKA	L320	−63	−35	G321	96	139	2.4
1BLF	L320	−72	−37	G321	83	147	2.8
1LNF	T296	−132	−3	G297	97	141	1.7
1COL	L30	−98	−16	G31	98	176	2.4
2CCY	A102	−80	−29	G103	135	169	1.67
Average value		−90	−32		114	154	
Residue conformation		γ			ε		

Note: Res., amino acid residue.

Fig. 2. α-α-Hairpins with connections of the type γε. On the left, a left-handed hairpin from cytochrome C (PDB code 2CCY); on the right, a right-handed hairpin from colicin A (PDB code 1COL). For designations, see Fig. 1.
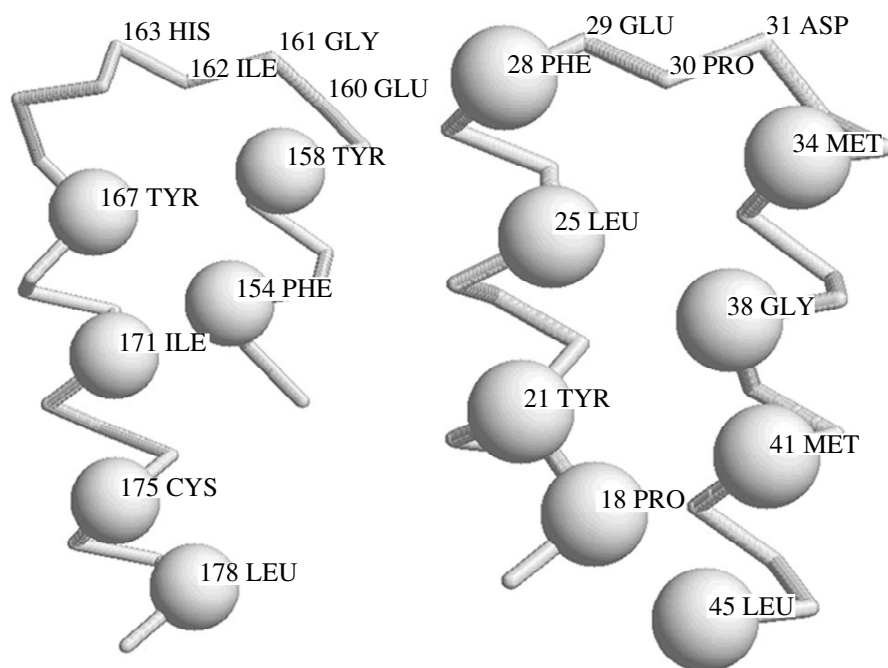

Fig. 3. α-α-Hairpins with connections of the type γα$_L$ββ. On the left, a left-handed hairpin from citrate synthase (PDB code 2CTS); on the right, a right-handed hairpin from uteroglobin (PDB code 2UTG). For designations, see Fig. 1.

Table 3. Conformation of connection γα$_L$ββ

PDB code	Res.	φ	ψ	Res.	φ	ψ	Res.	φ	ψ	Res.	φ	ψ	Resolution, Å
1A18	V23	−70	−11	G24	95	30	V25	−80	133	G26	−60	−143	2.4
1A6D	Q116	−67	−25	G117	119	12	V118	−78	135	H119	−74	131	3.6
1A26	D722	−99	8	G723	64	49	G724	−77	171	S725	−55	146	3.0
1A2D	V23	−65	−33	G24	114	13	V25	−67	150	G26	−75	139	2.4
1A80	E175	−104	4	Q176	46	48	A177	−62	140	S178	−75	178	1.7
1AB0	V23	−69	−22	G24	110	22	V25	−77	128	G26	−60	150	1.9
1AH6	A110	−72	−17	G111	120	−3	A112	−51	145	D113	−73	135	1.8
1AIS	L1123	−86	−13	K1124	66	33	L1125	−62	132	P1126	−69	161	2.1
1AL6	E160	−82	−10	G161	95	20	I162	−49	142	N163	−66	139	1.85
1AMZ	A254	−91	2	L255	82	17	S256	−65	147	D257	−70	165	1.8
1AUM	E175	−99	4	Q176	46	46	A177	−63	145	S178	−78	177	3.0
1AWP	L61	−77	−14	G62	73	66	H63	−72	150	S64	−77	166	2.0
1B5M	V61	−115	−8	G62	60	113	H63	−12	140	S64	−67	180	2.7
1BOU	W64	−105	5	N65	46	53	L66	−86	163	T67	−76	156	2.2
1BM8	A43	−77	3	N44	70	28	F45	−86	137	A46	−71	160	1.71
1BRX	V101	−93	9	D102	57	62	A103	−79	163	D104	−85	169	2.3
1BU2	F66	−102	−7	E67	50	48	L68	−82	161	D69	−44	146	3.0
1BYQ	A124	−106	8	G125	102	6	A126	−68	151	D127	−77	138	1.5
1UTG	F28	−96	−7	E29	53	61	P30	−62	140	D31	−64	170	1.34
2IFB	M21	−75	−8	G22	71	28	I23	−84	144	N24	−70	154	2.0
1PMP	L23	−92	−16	G24	106	6	V25	−63	139	G26	−92	149	2.7
2CTS	E160	−99	25	G161	41	63	I162	−67	163	H163	−72	126	2.0
Average value		−88	−6		76	37		−76	146		−70	153	
Residue conformation	γ			α$_L$			β			β			

Note: Res., amino acid residue.

hydrophobic cluster of one helix enter the "holes" of the hydrophobic cluster of the other helix (Fig.1).

The connection γε is the shortest and hence the most sterically constrained one. This type of connection is rare: only six of them were detected (Table 2). In all cases, the ε-residue is invariably glycine, while the γ-position is occupied by Leu, Ala, Thr, and Ile, as in the hairpin with connection γα$_L$β/δ. In this structure the alternation of hydrophobic and hydrophilic residues provides a complementary "knobs-into-holes" packing of α-helices. The left- and right-handed hairpins differ in the sequence of key residues (Fig. 2 and also Fig. 6).

The connection γα$_L$ββ was detected in 22 proteins (Table 3). For this connection type, no peculiar features of the γ-position were found. It can be occupied both by hydrophobic and by hydrophilic residues. The α$_L$-residue is Gly for 13 hairpins, while in others this position is occupied by residues with flexible side chains. The first β-residue is almost invariably hydrophobic. The second β-position is always occupied by hydrophilics or residues with a short side chain (Gly, Ala, Pro). Fig. 3 and 6 show that sequence patterns for the key hydrophobic and hydrophilic residues are different for left- and right-handed hairpins.

Hairpins with connections of the type γβγβ were found in 18 proteins (Table 4). In this hairpin, hydrophobic aromatic residues are absent from the first γ-position that is predominantly occupied by either hydrophilics or small hydrophobics. The first β-resi-

Table 4. Conformation of connection γβγβ

PDB code	Res.	φ	ψ	Res.	φ	ψ	Res.	φ	ψ	Res.	φ	ψ	Resolution, Å
1A5T	T304	−18	78	G305	−25	103	I306	−18	−34	N307	−118	80	2.2
1A6J	A142	−94	−25	A143	−49	143	Q144	−114	−3	S145	−138	153	2.35
1A37	V132	−74	−45	A133	−48	170	A134	−145	−67	G135	−68	165	3.6
1A3A	H133	−121	−6	T134	−64	152	T135	−111	8	S136	−111	122	1.8
1A7V	T95	−103	−16	I96	−90	99	K97	−108	−23	D98	−153	180	2.3
1AGR	K100	−101	11	I101	−76	140	K102	−116	−33	S103	−93	155	2.8
1AI3	A305	−71	−46	L306	−51	161	R307	−114	−56	S308	−77	115	1.8
1B1H	K455	−104	−1	V456	−114	153	A457	−118	41	D458	−134	128	1.8
1B3U	R20	−92	−3	N21	−69	166	E22	−101	−22	D23	−82	100	2.3
1BKA	D117	−90	−24	L218	−107	95	S219	−38	−61	D220	−99	132	2.4
1BL0	E74	−105	6	S75	−165	172	N76	−112	15	E77	−51	121	2.3
1BLF	N217	−106	−23	L21	−128	90	P219	−56	−39	E220	−94	148	2.8
BUK	L580	−68	−19	S581	−61	157	D582	−113	−6	S583	−133	145	2.5
1B9J	K455	−98	−6	V456	−110	155	A457	−117	42	D458	−142	126	1.8
256B	K19	−97	0	A20	−61	145	D21	−104	−18	N22	−147	178	1.4
2HMQ	Q37	−74	40	A38	−157	110	D39	−79	79	N40	−127	177	1.6
1HC2	Q92	−111	45	C93	−88	152	K94	−79	−46	E95	−146	170	3.2
2MHR	R37	−70	−38	D38	−150	95	N39	−79	68	S40	−112	166	1.7
Average value		−89	−4		−90	136		−96	−9		−112	124	
Residue conformation	γ			β			γ			β			

Note: Res., amino acid residue.

due is almost invariably hydrophobic, but there may occur polar residues with short side chains (Asp, Asn, Ser, Thr). Glycine and bulky hydrophobic residues are absent from the second β-position which, as a rule, is occupied by charged residues and only once by Gly. This type of hairpins is represented by left-handed ones only. Fig. 4 also shows a hairpin structurally similar to the right-handed one, but it is seen that the complementary hydrophobic stripe includes 3 hydrophilic residues that participate in the interaction by their hydrophobic portions.

Connections of the type γββ were detected in 15 proteins (Table 5). The β-positions are occupied by structurally different residues other than Gly. The first β-residue is invariably hydrophobic. At this position Ser or Thr are sometimes grouped with hydrophobics due to their ability to interact with main chain atoms. The second β-position must be occupied by either hydrophilic residues or those with small side chains. However, in two cases this residue is Phe, which

seems to be energetically disadvantageous [16]. In most cases, this structure is formed by short helices. The sequence patterns for hydrophobic and hydrophilic residues are shown in Figs. 5 and 6.

The overall schematic sequence patterns for hydrophobic and hydrophilic residues of all α-α-hairpin types discussed above are presented in Fig. 6. As seen, a certain pattern is characteristic of each type, although there is difference between right- and left-handed hairpins within the same type.

The current study deals with α-α-hairpins having short (no longer than 4 residues) connections. We found that there exist only five types of such hairpins, which is in good agreement with the results of stereochemical analysis reported earlier [16] and findings by other authors [10–15] on the basis of statistical analysis.

In connections of all these five types, the initial residues have a γ-conformation and their average φ and ψ values are practically the same. A slight difference in
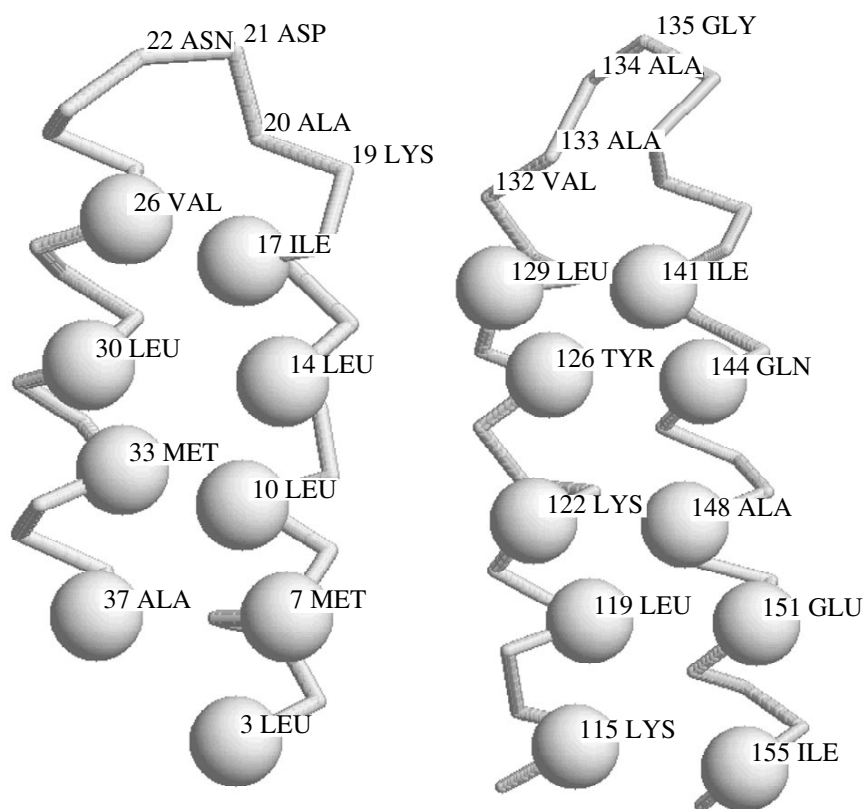
Fig. 4. α-α-Hairpins with connections of the type γβγβ. On the left, a left-handed hairpin from cytochrome B562 (PDB code 256B); on the right, a right-handed hairpin from 14-3-3 protein ZETA (PDB code 1A37). For designations, see Fig. 1.
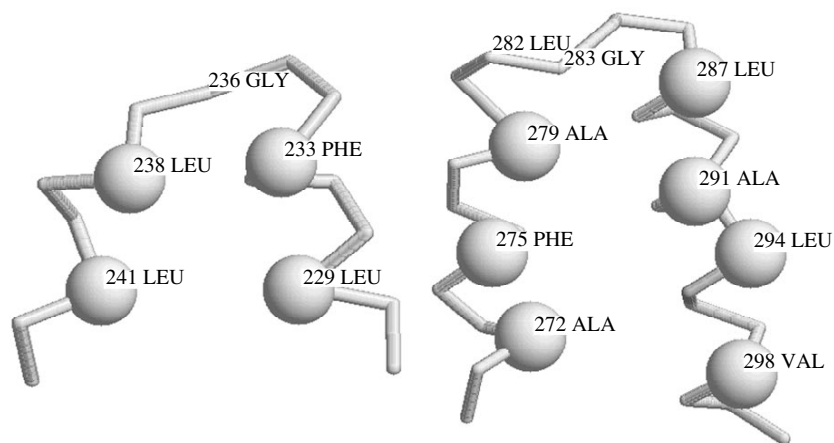

Fig. 5. α-α-Hairpins with connections of the type γββ. On the left, a left-handed hairpin from alpha-amylase (PDB code 1AQH); on the right, a right-handed hairpin from citrate synthase (PDB code 1AJ8). For designations, see Fig. 1.

these values reported for a hairpin with the connection γε is explained by the fact that in this case the connection residues are sterically constrained. Also, it should be noted that in a hairpin with the conformation γβγβ, the average φ and ψ values for the first and second β-residues are almost equal. The average values for the β-conformation vary slightly from connection to connection, which is not surprising since the β-region is
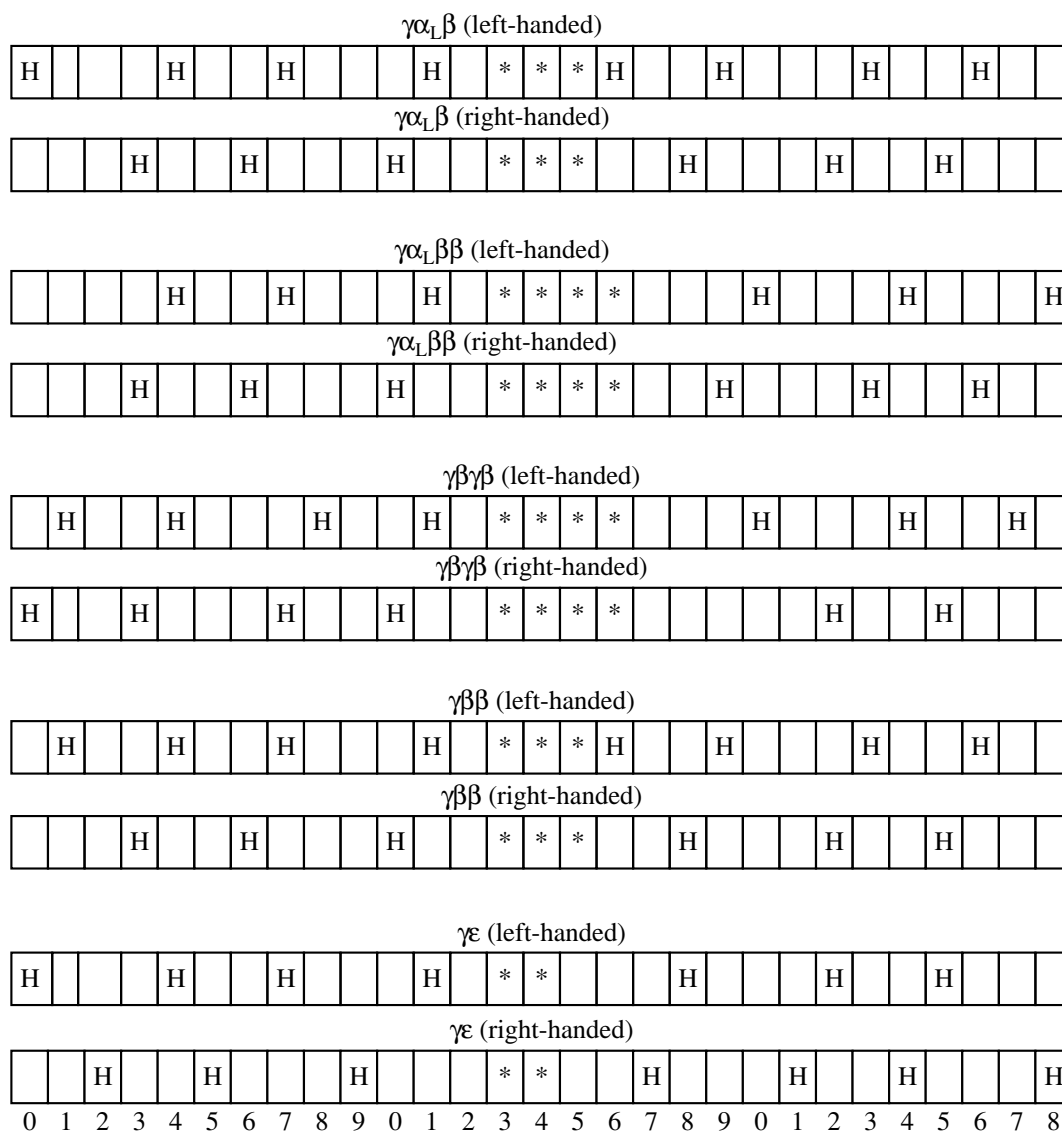
$\gamma\alpha_L\beta$ (left-handed)

$\gamma\alpha_L\beta$ (right-handed)

$\gamma\alpha_L\beta\beta$ (left-handed)

$\gamma\alpha_L\beta\beta$ (right-handed)

$\gamma\beta\gamma\beta$ (left-handed)

$\gamma\beta\gamma\beta$ (right-handed)

$\gamma\beta\beta$ (left-handed)

$\gamma\beta\beta$ (right-handed)

$\gamma\epsilon$ (left-handed)

$\gamma\epsilon$ (right-handed)

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8

Fig. 6. Sequence patterns for hydrophobic and hydrophilic residues in ideal α-α-hairpins. (∗) denotes connection; H, hydrophobic residues forming the necessary hydrophobic cluster; open square, hydrophilic residues. Residue conventional numbers are given below.

the largest on the Ramachandran map. In a hairpin with the connection conformation $\gamma\alpha_L\beta/\delta$, the φ and ψ angles of the last residue most often fall into the δ-region.

Globular proteins are closely packed, and they have a hydrophobic core surrounded by a polar shell; consequently, protein α-hairpins must be formed in accordance with these principles. For example, one face of the hairpin must be hydrophobic and the other hydrophilic when the hairpin is located in the protein surface. With the entire hairpin buried in a hydrophobic core, its both faces can be hydrophobic. To form a hydrophobic surface on one face, the α-helices should be packed in the "side by side" manner [17] like

pieces of a jigsaw puzzle. To form the necessary or minimal hydrophobic cluster, each α-helix should have a certain alteration of hydrophobic and hydrophilic residues. Figs. 1 to 5 present amino acid residues forming the necessary hydrophobic clusters (gray spheres). In most cases, for long helices in particular, "knobs" in clusters of one helix correspond to "holes" of the other, thereby providing a close packing of the structure.

The results of α-α-hairpin stereochemical analysis [16] and the hairpin features described here suggest that this structural motif can fold independently and serve as a nucleus or a structural unit during protein folding. In protein modeling the α-α-hairpins can be

Table 5. Conformation of connection γββ

PDB code	Res.	φ	ψ	Res.	φ	ψ	Res.	φ	ψ
1AE9	S194	−102	17	S195	−97	167	P196	−59	148
1AF2	Q11	−98	2	L12	−79	160	A13	−64	147
1AJ8	L282	−114	−18	G83	−77	162	D163	−64	131
1AL7	K16	−115	−29	L17	−74	155	P18	−71	146
1ALN	Q11	−84	−1	L12	−84	158	A13	−63	143
1AQH	N235	−117	0	G236	−12	−151	S237	−118	137
1ASS	144	−89	0	S45	−75	147	D46	−67	135
1AVC	E420	−83	−17	L421	−109	159	S422	−154	161
1B3U	A40	−48	−35	L41	−92	152	G42	−75	174
1BER	F136	−106	−46	L137	−77	156	D138	−96	171
1BHD	K204	−70	−12	M205	−77	162	S206	80	160
1BHP	R19	−129	−7	G20	−119	165	A21	−77	161
1AXN	E148	−108	4	T149	−142	169	S150	−155	−180
Average value		−97	−8		−85	137		−92	132
Residue conformation		γ			β			β	

Note: Res., amino acid residue.

used as starting structures capable of growing into ones of a higher order.

REFERENCES

1. Efimov, A.V., *Structure*, 1994, vol. 2, pp. 999–1002.
2. Sternberg, M.J.E. and Thornton, J.M., *J. Mol. Biol.*, 1976, vol. 105, pp. 367–382.
3. Efimov, A.V., *Prog. Biophys. Mol. Biol.*, 1993, vol. 60, pp. 201–239.
4. Efimov, A.V., *Protein Structure by Distance Analysis*, Bohr, H. and Brunak, S., Eds., Amsterdam-Oxford-Washington-Tokyo-Osaka-Kyoto: IOS Press and Ohmsha, 1994, pp. 187–200.
5. Efimov, A.V., *Bioorg. Khim.*, 1997, vol. 23, pp. 243–250.
6. Batchelor, A.H., Piper, D.E., Charles de la Brousse, F., McKnight, S.L., and Wolberger, C., *Science*, 1998, vol. 279, pp. 1037–1041.
7. Venkataramani, R., Swaminathan, K., and Marmorstein, R., *Nature Struct. Biol.*, 1998, vol. 5, pp. 74–78.
8. Groves, M.R., Hanlon, N., Turowski, P., Hemmings, B., and Barford, D., *Cell*, 1999, vol. 96, pp. 99–110.
9. Efimov, A.V., *Mol. Biol.*, 1986, vol. 20, pp. 250–260.
10. Oliva, B., Bates, P.A., Querol, E., Aviles, F.X., and Sternberg, M.J.E., *J. Mol. Biol.*, 1997, vol. 266, pp. 814–830.
11. Donate, L.E., Rufino, S.D., Canard, L.H.J., and Blundell, T.L., *Protein Sci.*, 1996, vol. 5, pp. 2600–2616.
12. Rufino, S.D., Donate, L.E., Canard, L.H.J., and Blundell, T.L., *J. Mol. Biol.*, 1997, vol. 267, pp. 352–367.
13. Kwasigroch, J.-M., Chomilier, J., and Mornon, J.-P., *J. Mol. Biol.*, 1996, vol. 259, pp. 855–872.
14. Getha, V. and Munson, P.J., *Protein Sci.*, 1997, vol. 6, pp. 2538–2547.
15. Aurora, R. and Rose, G.D., *Protein Sci.*, 1998, vol. 7, pp. 21–38.
16. Efimov, A.V., *Protein Eng.*, 1991, vol. 4, pp. 245–250.
17. Efimov, A.V., *FEBS Lett.*, 1999, vol. 463, pp. 3–6.