PROTEINS: Structure, Function, and Genetics 48:571–579 (2002)

Variants of 3_{10}-Helices in Proteins

Lipika Pal,[1] Gautam Basu,[2*] and Pinak Chakrabarti[1*]
[1]*Department of Biochemistry, Bose Institute, Calcutta, India*
[2]*Department of Biophysics, Bose Institute, Calcutta, India*

ABSTRACT **An analysis of the shortest 3_{10}-helices, containing three helical residues and two flanking capping residues that participate in two consecutive i + 3 → i hydrogen bonds, shows that not all helices belong to the classic 3_{10}-helix, where the three central residues adopt the right-handed helical conformation (α_R). Three variants identified are: 3_{10}^{L}-helix with all residues in the left-handed helical region (α_L), 3_{10}^{EL}-helix where the first residue is in the extended region followed by two residues in the α_L conformation, and its mirror-image, the $3_{10}^{E'R}$-helix. In the context of these helices, as well as the equivalent variants of α-helices, the length dependence of the handedness of secondary structures in protein structure is discussed. There are considerable differences in the amino acid preferences at different positions in the various types of 3_{10}-helices. Each type of 3_{10}-helix can be thought to be made up of an extension of a particular type of β-turn (made up of residues i to i + 3) such that the (i + 3)th residue assumes the same conformation as the preceding residue. Distinct residue preferences at i and i + 3 positions seem to decide whether a particular stretch of four residues will be a β-turn or a 3_{10}-helix in the folded structure. Proteins 2002;48:571–579.** © 2002 Wiley-Liss, Inc.

Key words: 3_{10}-helix; β-turn; conformation; handedness in protein structures; sequence preference

INTRODUCTION

The ideal 3_{10}-helix is comprised of three residues per helical turn, with the backbone carbonyl oxygen of residue i hydrogen bonded to the amide proton of residue i + 3, forming a 10-atom ring.[1] 3_{10}-Helical structures are relatively common in proteins comprising as much as 4% of all residues.[2–5] Although typical protein 3_{10}-helices are only three or four residues long, the occurrence of longer 3_{10}-helices is not uncommon.[6] The position-specific occurrence of different residues that can provide stability to these structures has been determined.[2,7,8]

As secondary structures, β-turns are much more prevalent, encompassing on average 25% of the residues[9] and causing a reversal in the direction of the polypeptide chain.[10] A β-turn consists of four consecutive residues (i, i + 1, i + 2, i + 3) which are not part of an α-helix but still has the two terminal C^{α} atoms within a distance of 7 Å.[3,11,12] Although a 4 → 1 type of hydrogen bond linking the CO group of residue i to the NH of residue i + 3 was originally used by Venkatachalam[13] to characterize them, β-turns need not necessarily be stabilized by intra-turn hydrogen bonds.[14] β-Turns have been classified into nine different types (I, I′, II, II′, VIa1, VIa2, VIb, VIII, and IV) based on the values of the dihedral angles ϕ, ψ of the (i + 1)th and (i + 2)th positions in the turn.[3,15] It is convenient to distinguish another category, viz., type III (and the main-chain mirror-image, type III′) occupying very similar regions of ϕ, ψ space as those of type-I β-turns.[11] Overlapping type III β-turns give rise to the ideal 3_{10}-helix (ϕ = −60°; ψ = −30°). Besides the four-residue β-turn, one can also have short loops involving three or five residues (with a 3 → 1 or a 5 → 1 hydrogen bond between the carbonyl oxygen at position i and amide hydrogen at positions i + 2 or i + 4, respectively). The former constitutes a γ-turn[16–18] and the latter an α-turn.[19] Moreover, the conformational features of a π-turn involving a loop of six residues and usually following α-helices have been delineated.[20] Turns are often conserved during evolution and have been implicated in molecular recognition[11] and protein folding.[21] The knowledge of amino acid preferences at individual positions in β-turns and γ-turns have been used for turn prediction.[10,22–30]

As secondary structural elements, turns and helices are separate entities; however, it should be noted that occurring in isolation, the first three residues (i + 1 through i + 3) of an α-helix with a single 5 → 1 (i + 4 to i) hydrogen bond is basically an α-turn.[19] Two such consecutive α-turns give rise to the shortest α-helix of four residues (i + 1 through i + 4). Similarly, two consecutive type-III β-turns, with two consecutive 4 → 1 hydrogen bonds, with the first β-turn defined by residues i to i + 3 and the second composed of residues i + 1 to i + 4, form the shortest 3_{10}-helix (i + 1 through i + 3), as shown in Figure 1. However, no helical structure can be constructed from the repetitive extension of other types of β-turns, primarily because the non-identical ϕ, ψ angles at the two central residues do not allow the extension of helical hydrogen bonding network. In the present work, we show that some types of β-turns (notably II and II′) can be part of variants of 3_{10}-helices. Analysis of the residues in these structures

Grant sponsor: Council of Scientific and Industrial Research.

*Correspondence to: Pinak Chakrabarti, Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700 054, India. E-mail: pinak@boseinst.ernet.inor Gautam Basu, Department of Biophysics, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700 054, India. E-mail: gautam@boseinst.ernet.in

Received 23 January 2002; Accepted 25 March 2002

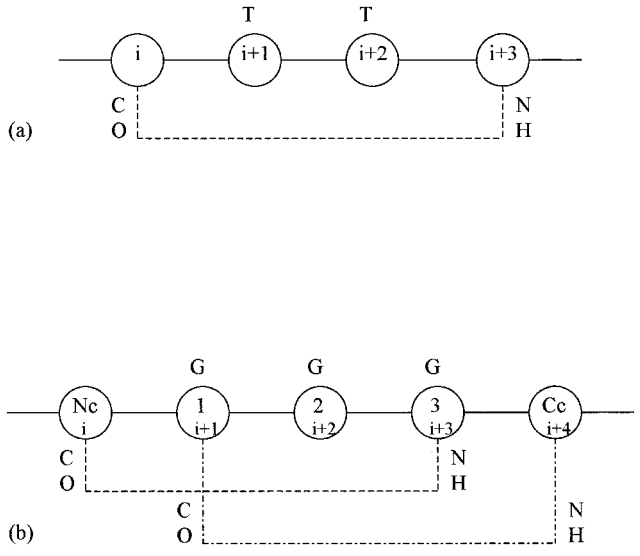


Fig. 1. Schematic representations of **(a)** two-residue (marked TT) β-turn with a 4 → 1 hydrogen bond, and **(b)** three-length (marked GGG) 3_{10}-helix with two consecutive 4 → 1 hydrogen bonds. The two β-turn residues are numbered i + 1 and i + 2, whereas the three 3_{10}-helix residues, flanked by two capping residues Nc and Cc, are numbered i + 1, i + 2, and i + 3.

shows that these have features different from classic 3_{10}-helices, as well as β-turns from which they are made. Distinct preferences are also observed for the closest secondary structures from the two ends. All of this information can be used for modeling and refining prediction of secondary structural elements in proteins.

MATERIALS AND METHODS

The March 1999 culledpdb data set,[31,32] containing 1085 protein chains with less than 30% sequence identity and ≤2.5 Å resolution, was used in this analysis. Secondary structure assignments were made using the Define Secondary Structure of Proteins (DSSP) program[9] which uses the criteria of two or more consecutive (i,i + 3) or (i,i + 4) hydrogen bonds to identify a 3_{10}- or an α-helix, respectively. The DSSP program assigns single-letter conformational codes to individual residues (α-helix: H; 3_{10}-helix: G; β-strand: E; β-bridge: B; bend: S; turns: T). In addition, we assign the code C (coil) to a residue when no conformational code is assigned to a residue in the DSSP output file. A Protein Data Bank (PDB) file is mentioned in the text as the four-letter PDB code (in lowercase letters) followed by the chain identifier (in uppercase letters); when there is no chain ID, an underscore (_) is used.

The propensity of a given residue for a particular attribute was calculated as the ratio of the actual number of observations to the expected number of observations, where the expected number of observations is given by $(n_{xaa}/n_{total}) \times n^*$, where n_{xaa} is the total number of residue type Xaa in the data set, n_{total} is total number of all residues in the data set, and n* is the total number of residues in the data set with the attribute of interest (e.g., helix Nc). In calculating amino acid composition, over-representation was estimated according to Karpen et al.[2] For each amino acid at a specific helix position, a z-value was calculated. If $|z| \geq 1.96$ (5% significance level), the observed number of occurrences was considered to deviate significantly from its expected value. Negative values of z indicate under-representation and positive values indicate over-representation.

Hydrogen-bonded partners were assigned using the program HBPLUS.[33] The nomenclature for helices and their flanking residues is as follows: …N2′— N1′— Nc—1—2—3—Cc—C1′—C2′… where 1 through 3 belong to helix proper, Nc and Cc represent helix capping positions, whereas the primed residues represents residues preceding and succeeding the helix.

RESULTS

Identification of Variants of 3_{10}-Helices From the Distribution of ϕ, ψ Angles

The culledpdb data set consists of 1085 protein chains of which 3.8% (10,086) residues were found to exhibit the 3_{10}-helical conformation. Of these, approximately 70% were part of three-length 3_{10}-helices. We have been interested in the pattern of residue distribution in shortest possible helices, viz., three-length 3_{10}-helices. The backbone ϕ, ψ angles of residues in such structures identified by the commonly used program DSSP[9] are shown in Figure 2(a). Although deviations from average ϕ, ψ angles of ∼−65°, −20° were noted earlier when 3_{10}-helices of various lengths were taken together,[34] by restricting to only three-length helices the angles are found to cluster in four distinct regions, designated as: α_R, corresponding to the right-handed α-helix; α_L, left-handed helix; E, extended conformation and a region (allowed for Gly residues, but normally disallowed for non-Gly residues) E′ related by a center of inversion. The average values of the angles in all the regions are given in Table I. On scrutiny, it was observed that the residues occur in these regions in specific order and this was used to classify these 3_{10}-helices (Table II). With all three residues in the α_R conformation, one obtains the classic 3_{10}-helix, which we call 3_{10}^R, which is the major class. As pointed out in Table I, further analysis of the α_R region showed that the average backbone ϕ angle decreased (−57°, −67°, and −95°) whereas the backbone ψ angle increased (−32°, −16°, and 0°) as the residue position is changed from 1 through 3 in the helix. The 3_{10}^R helix is thus not the ideal 3_{10}-helix composed of two overlapping type-III β-turns (ϕ = −60°, ψ = −30°), rather it can be considered to be formed from overlapping type III/I β-turns. A small number of helices (designated as 3_{10}^L-helices) are found with the opposite hand, where the residues have the α_L conformation. Then there are two mirror-image related categories, 3_{10}^{EL} and $3_{10}^{E'R}$, of nearly equal occupancy and with conformations, which can be designated as $E\alpha_L\alpha_L$ and $E'\alpha_R\alpha_R$, respectively.

It may be noted that the two central residues in type-II and -II′ β-turns have conformations $E\alpha_L$ and $E'\alpha_R$, respectively.[15] An addition of an α_L residue to II and α_R to II′ β-turn leads to 3_{10}^{EL}- and $3_{10}^{E'R}$-helices, respectively. It may be asked to what extent the addition of more residues with α_L or α_R conformation (as the case may be) can be sustained. Table II shows that there is a drastic reduction

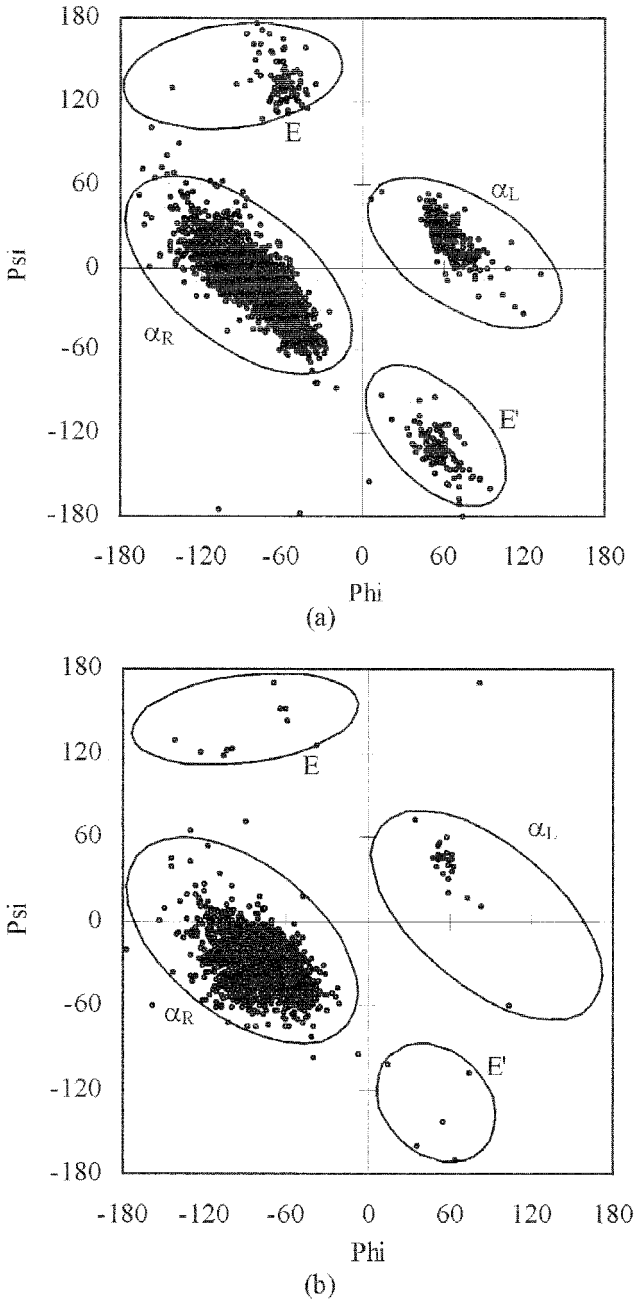


Fig. 2. The distribution of ϕ, ψ angles (degrees) for residues in **(a)** three-length 3_{10}-helices and **(b)** four-length α-helices.

in the number of cases as the 3_{10}-helix gets longer with contiguous α_L residues. A similar dependence on length can be seen when one compares the variant four-length α-helices [Fig. 2(b)] with those from the 3_{10}-helix categories. α-Helices are found in smaller numbers and the numbers get insignificant as the length increases for cases with contiguous α_L residues. To find out the secondary structure type exhibited by other sequences with $E\alpha_L\alpha_L$ or $E'\alpha_R\alpha_R$ type of conformation, the entire data base was searched. Only ~15% sequences with $E\alpha_L\alpha_L$ conformation showed GGG secondary structure, characterizing a three-length 3_{10}-helix, whereas a large fraction (~62%) was found either at the end of a β-strand (secondary structure ETT) or were present as a TTT stretch. For sequences that show the $E'\alpha_R\alpha_R$ conformation, ~35% was present as GGG, ~19% was present as HHH, whereas ~25% showed a stretch of two turn (T) conformations.

TABLE I. Average Backbone Torsion Angles (°) of the Shortest Protein Helices

	Three-length 3_{10}-helices		Four-length α-helices	
Region[a]	ϕ[b]	ψ[b]	ϕ[b]	ψ[b]
α_R	−73 (30)[c]	−15 (31)[c]	−70 (18)	−32 (24)
α_L	65 (15)	22 (15)	61 (13)	35 (25)
E	−61 (15)	135 (15)	−87 (33)	136 (18)
E′	56 (13)	−132 (14)	49 (24)	−137 (30)

[a]Regions are defined in Figure 2.
[b]Standard deviations are in parentheses.
[c]Position specific (from the N-termini) torsion angle values for the α_R region are: 1: $\phi = -57$ (9), $\psi = -32$ (12); 2: $\phi = -67$ (9), $\psi = -16$ (12); 3: $\phi = -95$ (16), $\psi = 0$ (6).

Positional Potentials in Different Three-Length 3_{10}-Helices

We have calculated the propensity and z-value (see Materials and Methods) of 20 amino acid residues to occupy each position in 3_{10}^{EL} and $3_{10}^{E'R}$ helices (Table III). To facilitate comparison, the propensities of residues that are over-represented ($z \geq 1.96$) are plotted in Figure 3 along with the values observed in 3_{10}^{R}-helices. In conformity with what has been observed for 3_{10}-helices in general,[2,7,8] three-length 3_{10}^{R}-helices in our sample prefer Asp, Pro, Asn, and His at the Nc position, Pro at position 1, and acidic groups (Glu and Asp) and Ser at position 2. Preferences for a residue of a particular chemical nature are not strong at position 3, whereas at the Cc position, hydrophobic residues Ile, Leu, Val, Phe, Tyr, and Cys are strongly favored. There are distinct differences in the use of residues in other types of 3_{10}-helices. For example, in contrast to short polar residues occupying Nc in the 3_{10}^{R}-helix, Cys and Gly are highly favored at this position in 3_{10}^{EL}-helices. These helices also have Tyr, in addition to Pro, at position 1. Positions 2 and 3 in 3_{10}^{EL}-helices have an α_L conformation, which is normally accessible by Asp and Asn, besides Gly, and these residues are found predominantly at these positions (Gly is more prominent at position 3). Though some residues have reasonably large propensity values, no residue is over-represented at the Cc position. In $3_{10}^{E'R}$-helices, the favored amino acids are: branched aliphatic residues Thr and Leu at Nc, Gly at position 1 (a conformation of E′ is normally disallowed for a non-Gly residue), Asp at position 2, Asn, His, Tyr, Lys, and Ala at position 3, and Ala, Ser, and Gly at the Cc position.

The number of 3_{10}^{L}-helices is very low (18 cases) for doing any statistical analysis. In 18 cases of 3_{10}^{L}-helices, Nc is occupied by Asp, Asn, or Gly (eight cases), and Lys is found at Cc in four instances. In the total of 54 positions in the main body of the helices, Gly is found in 24 positions, followed by Asn (six), Phe (three), and Trp (three). Two extreme situations are found in the PDB file 1mroB where all helical positions are occupied by Gly (residue range

TABLE II. Classification and Occurrence of 3_{10}- and α-Helix Variants

Conformation[a]	Variant 3_{10}-helix nomenclature	Total occurrence			
		3_{10}-Helix		α-Helix	
		3-Length	>3-Length	4-Length	>4-Length[b]
$\alpha_R\alpha_R\alpha_R(\alpha_R\ldots)$	3_{10}^{R}	2140	570	806	6247
$\alpha_L\alpha_L\alpha_L(\alpha_L\ldots)$	3_{10}^{L}	18[c]	3	2	—
$E\alpha_L\alpha_L(\alpha_L\ldots)$	3_{10}^{EL}	91	9[d]	9[e]	1[g]
$E'\alpha_R\alpha_R(\alpha_R\ldots)$	$3_{10}^{E'R}$	117	56	4[f]	77

[a]The shorthand nomenclature for the backbone conformation is given in Figure 2.
[b]Only helices with length 5–21 are considered.
[c]In addition, residues 218–220 in 1gnd_ adopt the $\alpha_R\alpha_L\alpha_L$ conformation.
[d]Helices longer than four-length were not observed.
[e]Of these, in five cases, the helix does not start with an E, which is located at the second or third position from Nc.
[f]In addition, residues 50–53 in 2dkb_ adopt the $\alpha_LE'\alpha_R\alpha_L$ conformation.
[g]Residues 39–43 in 1bd0A form such a helix. In yet another example, 1vaoA, residues 519–535 constitute a helix with the first four residues in conformation $E\alpha_L\alpha_L\alpha_L$ and the remainder in α_R conformation with a kink between the two parts.

106–110: Leu-Gly-Gly-Gly-Lys) and 2bbkH where no Gly is involved (39–43: Pro-Ala-His-Phe-Ala).

Comparison of Potentials for the Formation of 3_{10}-Helix and β-Turn

As shown in Figure 1, an additional hydrogen bond attributable to the placement of an extra residue in the right conformation extends a β-turn into a 3_{10}-helix and it would be of interest to see if one can distinguish, based on the types of residues occurring at different positions, whether a given stretch of the polypeptide chain would form an isolated turn or get extended into a 3_{10}-helix. Residues with high propensity values at different positions in various 3_{10}-helices (Table III) are compared with the prominent residues at equivalent positions of β-turns, the positional residue potentials of which have been calculated by Hutchinson and Thornton[15] (Table IV). The structures to be compared are 3_{10}^{R}- and type-I β-turn (ideally one should use type-III β-turn, but because of paucity of data, potentials are not available and instead, we have used type-I β-turns, which are conformationally very similar; in addition we showed earlier that 3_{10}^{R} helices are formed from overlapping type III/I β-turns), 3_{10}^{EL}-helix and type-II β-turn, and $3_{10}^{E'R}$-helix and II′ β-turn. It can be seen that the residue preferences at position 3 of the helices are entirely different from the equivalent i + 3 positions in the turns, and the same is true even for the Nc/i position (except for the first pair). Thus, the two central residues at positions i + 1 and i + 2 determine the type of β-turn, but whether it will remain a β-turn in the final structure or extend to a 3_{10}-helix is controlled by the two adjacent residues.

Secondary Structures on Either Side of 3_{10}-Helix Variants

To see whether 3_{10}-helix variants have any preference to be positional between common secondary structures (helix or β-strand), we looked at the structural features of the Nc and Cc positions in 91 3_{10}^{EL}- and 117 $3_{10}^{E'R}$-helices. For 34 cases (37%) of 3_{10}^{EL}-helices and 16 cases (14%) of $3_{10}^{E'R}$-helices, both the positions are present in C-conformation, as described in Materials and Methods. A count of the number of times a β-strand (E) or H (including G) occurs as the nearest secondary structure from the 3_{10}-helix shows that for the 3_{10}^{EL}-helix, E occurs 62 times (36 times within three residues) before and 62 times (42 times within three residues) after the helix, whereas the corresponding numbers for H are 27 (five times within three residues) and 29 (six times within three residues), respectively. This shows that a 3_{10}^{EL}-helix is more likely to have a strand rather than a helix as a closest neighbor. For $3_{10}^{E'R}$-helices, the numbers of E and H are nearly equal on both the sides {E: 52(33), 47(31); H: 44(33), 57(36)}.

When simultaneous occurrence of an E or H/G conformation on both sides of the helices were analyzed (Fig. 4), it was found that for 3_{10}^{EL}-helices, there are 15 cases [Fig. 4(a)] with an E occurring before the Nc position and an E/H/G occurring after the Cc position (of these, 13 cases have an E following Cc). For five cases [Fig. 4(b)], a stretch of E ends at N1′ position and an E/G stretch starts after the C1′ position (of five cases, four cases have an E after C1′).

Sequence neighbors of $3_{10}^{E'R}$-helices showed a variety of secondary structure combinations: 19 cases were found [Fig. 4(c)] with an α-helix starting at the Cc position whereas a stretch of E or H ends anywhere within four positions before the helix; eight cases were found [Fig. 4(d)] with an α-helix leading directly into a $3_{10}^{E'R}$-helix, which, after a gap of two residues with a nonregular structure, leads into a E/H/G stretch (H stretch for six cases). When a stretch of E ends at the N1′ position of $3_{10}^{E'R}$-helices, two more prominent combinations are observed after the helix: one combination has six cases [Fig. 4(e)] with E/H stretch starting from the Cc position (H stretch for five cases), whereas the other combination has five cases [Fig. 4(f)] with an E stretch occurring after the Cc position as in 3_{10}^{EL}-helices [Fig. 4(a)]. Finally, four cases [Fig. 4(g)] consisted of an α-helix ending at the Nc position and a β-strand starting from the Cc position.

Capping Hydrogen Bond Interaction

We investigated whether the NH groups at Nc, 1 and 2 positions and the CO groups at 2, 3, and Cc positions are involved in any local (within five sequence contiguous

TABLE III. Propensities and z-Values of Residues at Different Positions of Three-Length 3_{10}^{EL}- and $3_{10}^{E'R}$-Helices†

	Nc			1			2			3			Cc		
Residues	No.	P	Z	No.	P	Z	No.	P	Z	No.	P	Z	No.	P	Z
(a) 3_{10}^{EL}-helix															
Ala	7	0.9	−0.3	9	1.2	0.5	1	0.1	−2.5	5	0.7	−1.0	10	1.3	0.9
Cys	6	**4.6**	**4.2**	2	**1.5**	0.6	3	**2.3**	1.5	0	0.0	**−1.2**	1	0.8	−0.3
Asp	5	0.9	−0.2	3	0.6	−1.1	13	**2.4**	**3.3**	11	**2.0**	**2.4**	3	0.6	−1.1
Glu	2	0.4	−1.6	6	1.0	0.1	**1**	0.2	−2.1	2	0.4	−1.6	4	0.7	−0.8
Phe	2	0.6	−0.8	3	0.9	−0.3	**4**	1.1	0.2	3	0.9	−0.3	**1**	0.3	−1.4
Gly	21	**3.0**	**5.5**	2	0.3	−2.0	23	**3.3**	**6.2**	43	**6.1**	**14.1**	10	**1.4**	1.2
His	3	**1.4**	0.6	2	1.0	−0.1	3	**1.4**	0.6	1	0.5	−0.8	2	1.0	−0.1
Ile	3	0.6	−1.0	1	0.2	−1.9	1	0.2	−1.9	1	0.2	−1.9	7	**1.4**	0.9
Lys	2	0.4	−1.5	6	1.1	0.3	3	0.6	−1.1	1	0.2	−1.9	**4**	0.8	−0.6
Leu	4	0.5	−1.4	1	0.1	−2.5	2	0.3	−2.2	1	0.1	−2.5	11	**1.4**	1.2
Met	1	0.5	−0.7	0	0.0	−1.4	1	0.5	−0.7	1	0.5	−0.7	**4**	**2.0**	1.4
Asn	7	**1.7**	1.5	2	0.5	−1.1	18	**4.4**	**7.0**	11	**2.7**	**3.5**	2	0.5	−1.1
Pro	2	0.5	−1.1	20	**4.7**	**7.8**	0	0.0	−2.1	0	0.0	−2.1	0	0.0	−2.1
Gln	2	0.6	−0.8	4	1.2	0.3	1	0.3	−1.3	3	0.9	−0.2	3	0.9	−0.2
Arg	1	0.2	−1.7	2	0.5	−1.2	2	0.5	−1.2	1	0.2	−1.7	5	1.2	0.3
Ser	8	**1.5**	1.1	5	0.9	−0.2	6	1.1	0.2	6	1.1	0.2	7	1.3	0.7
Thr	9	**1.7**	1.7	5	1.0	−0.1	0	0.0	−2.4	0	0.0	−2.4	5	1.0	−0.1
Val	4	0.6	−1.0	8	1.3	0.7	1	0.2	−2.2	1	0.2	−2.2	8	1.3	0.7
Trp	2	**1.5**	0.6	2	**1.5**	0.6	2	**1.5**	0.6	0	0.0	−1.2	2	**1.5**	0.6
Tyr	0	0.0	−1.8	8	**2.4**	**2.7**	6	**1.8**	1.5	0	0.0	−1.8	2	0.6	−0.7
(b) $3_{10}^{E'R}$-helix															
Ala	11	1.1	0.4	1	0.1	−3.0	13	1.3	1.0	16	**1.6**	2.0	23	**2.3**	**4.4**
Cys	1	0.6	−0.5	0	0.0	−1.3	3	**1.8**	1.0	2	1.2	0.3	2	1.2	0.3
Asp	3	0.4	−1.6	6	0.9	−0.4	26	**3.7**	**7.4**	8	1.1	0.4	3	0.4	−1.6
Glu	5	0.7	−0.9	2	0.3	−2.1	10	**1.4**	1.0	5	0.7	−0.9	5	0.7	−0.9
Phe	8	**1.8**	1.6	0	0.0	−2.2	4	0.9	−0.3	5	1.1	0.2	6	1.3	0.7
Gly	8	0.9	−0.4	88	**9.7**	**27.3**	4	0.4	−1.8	4	0.4	−1.8	15	**1.7**	**2.0**
His	2	0.7	−0.4	0	0.0	−1.7	2	0.7	−0.4	5	**1.9**	1.4	0	0.0	−1.7
Ile	6	0.9	−0.2	0	0.0	−2.6	1	0.2	−2.2	2	0.3	−1.8	5	0.8	−0.6
Lys	**4**	0.6	−1.1	2	0.3	−1.9	9	1.3	0.8	10	**1.5**	1.2	5	0.7	−0.7
Leu	17	**1.7**	**2.3**	1	0.1	−3.0	3	0.3	−2.3	8	0.8	−0.7	11	1.1	0.3
Met	4	**1.6**	0.9	2	0.8	−0.4	4	**1.6**	0.9	3	1.2	0.3	2	0.8	**−0.4**
Asn	2	0.4	−1.5	3	0.6	−1.0	1	0.2	−1.9	13	**2.5**	**3.4**	1	0.2	−1.9
Pro	4	0.7	−0.6	0	0.0	−2.4	8	**1.5**	1.1	0	0.0	**−2.4**	0	0.0	−2.4
Gln	2	0.5	−1.2	3	0.7	−0.7	2	0.5	−1.2	4	0.9	−0.2	1	0.2	−1.7
Arg	6	1.1	0.2	2	0.4	−1.6	3	0.5	−1.1	6	1.1	0.2	6	1.1	0.2
Ser	3	0.4	−1.6	4	0.6	−1.2	10	**1.4**	1.1	7	1.0	0.0	15	**2.1**	**3.1**
Thr	15	**2.2**	**3.3**	2	0.3	−1.9	6	0.9	−0.3	5	0.7	−0.7	9	1.3	0.9
Val	7	0.9	−0.4	0	0.0	−3.0	5	0.6	−1.2	5	0.6	−1.2	5	0.6	−1.2
Trp	3	**1.8**	1.0	0	0.0	−1.3	1	0.6	−0.5	1	0.6	−0.5	2	1.2	0.3
Tyr	6	**1.4**	0.9	1	0.2	−1.6	2	0.5	−1.1	**8**	**1.9**	1.9	1	0.2	−1.6

†3_{10}^{EL} and $3_{10}^{E'R}$ helices are defined in Table II. P stands for propensity and Z for z-values of amino acid residues. P > 1.3 and Z > 1.96 are given in bold.

residues) hydrogen bonding. Only the groups at Nc and Cc were found participating in substantial numbers (about 44 and 30%, respectively) in both the 3_{10}^{EL}- and $3_{10}^{E'R}$-helices. In 3_{10}^{EL}-helices (91 cases), the largest grouping has the NH of Nc and CO of Cc hydrogen bonded to each other (19 cases). Of these, 11 cases correspond to Fig. 4(a) with a stretch of E ending at N1′ and another starting from C1′. Among the rest, cases in which H or E stretch is present within four residues on both sides of the helix have: 1. E at Nc with E at Cc, 2. E at Nc with E at C1′, 3. E at N1′ with E at Cc, 4. E at N1′ with E at C3′, and 5. E at Nc with H at C3′. In the majority of these cases, the 3_{10}^{EL}-helices are part of a β-hairpin turn, a representative example of which is shown in Figure 5. Five-residue β-hairpin turns with a single hydrogen bond between the anchoring β-strand residues (in our case Nc and Cc) have been classified as 3:5 β-hairpin turns.[36] From an analysis of 106 β-hairpin turns in proteins, Sibanda et al.[36] had identified G1 β-bulge as a prominent subclass of 3:5 β-hairpin turns. The 3_{10}^{EL}-helix containing 3:5 β-hairpin turn, as identified here, can be considered to be a new subclass of 3:5 β-hairpin turn (conformation of the three central residues in G1 β-bulge is: $\alpha_R\alpha_R\alpha_L$ whereas in 3_{10}^{EL}-helix containing 3:5 β-hairpin turns it is $E'\alpha_L\alpha_L$). Among four cases of 3_{10}^{EL}-helices [Fig. 4(b)], when a β-strand ends at N1′ position and the other β-strand starts from C2′ position, three cases have hydro-

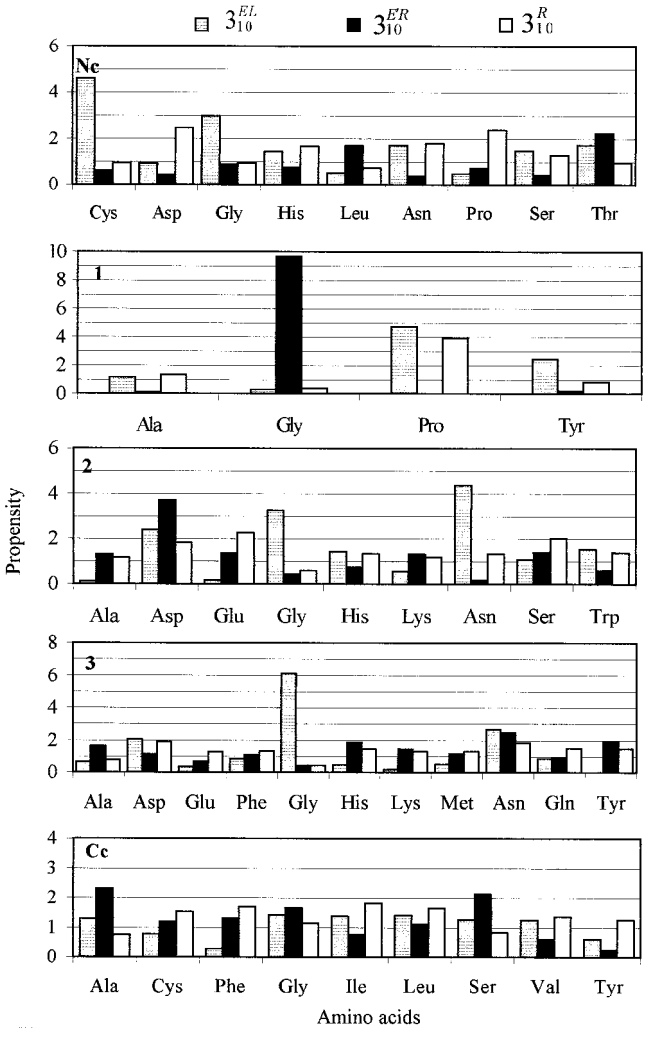


Fig. 3. Histogram of residue propensities in 3_{10}-helix variants as a function of helix position. Only those residues are shown that are over-represented (based on z-values) in at least one of the three 3_{10}-helix variants.

gen bonding between NH of Nc and CO of C1′. This type belongs to the class of 4:6 β-hairpin in accordance to the convention of Sibanda et al.[36] In the case of $3_{10}^{E'R}$-helices, when a β-strand ends at N1′ position and another β-strand starts from C1′ position [Fig. 4(f)], three of five cases have a 3:5 β-hairpin-like hydrogen bonding pattern.

DISCUSSION

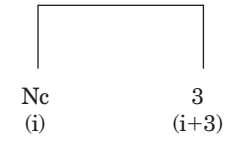
Classification of 3_{10}-Helices

Under IUPAC convention rule 6.3,[37] a residue is considered to be the part of an α-helix (or β-sheet) if either its NH or CO groups are involved in the appropriate hydrogen bond. It does not imply that the ϕ, ψ angles for this residue should be in the appropriate region for α-helix (or β-sheet). Based on the occurrence of two consecutive i + 3 to i hydrogen bonds (Fig. 1) we have identified all cases of three-length 3_{10}-helices from the PDB. The ϕ, ψ distribution (Fig. 2) shows that not all constituent residues have their torsion angles correspond to the canonical conformation in the α_R region of the Ramachandran plot. Based on the ϕ, ψ angles, three variants of the classic 3_{10}-helices can be identified (Table II and Fig. 6). In 3_{10}^{L}-helices, all the residues have α_L conformation. The 3_{10}^{EL}-helices have the first residue in the extended conformation, followed by two residues with α_L-conformation. The last category is the mirror-image related $3_{10}^{E'R}$-helix, where the conformation of the three residues can be specified as $E'\alpha_R\alpha_R$. These helices can also be considered as the extension of two-residue β-turns (corresponding to a stretch of four residues, i to i + 3, where the ϕ, ψ angles of the two central residues specify the turn type), such that the (i + 3) residue has the same conformation as the (i + 2) residue. Thus, the type-III β-turn leads to the classic 3_{10}-helix (the more commonly observed 3_{10}^{R}-helix is initiated by a type-III β-turn followed by a type-I turn), type II to 3_{10}^{EL}, and II′ to $3_{10}^{E'R}$-helices (Table IV).

TABLE IV. Differences in Residue Preferences of 3_{10}-Helix Variants and the Corresponding β-Turns Whose Extension Leads to Their Formation

3_{10}-Helix type / β-Turn type	Nc (i)	1 (i + 1)	2 (i + 2)	3 (i + 3)
(a) 3_{10}^{R}	D, P, N, H, S	P	E, S, D	D, N, Q, Y
I	D, N, H, C, S, P	P, E, S	D, N, T, S	G
(b) 3_{10}^{EL}	C, G, N, S	P, Y	N, G, D	G, N, D
II	P, Y	P, K	G, N	C, K, S
(c) $3_{10}^{E'R}$	T, L	G	D	N, H, Y, A, K
II′	Y	G	N, S, D	T, G

Schematic diagram of the four residues with their conventional labels (in parentheses for β-turns) is shown to the right. The propensities for β-turns are from Hutchinson and Thornton.[15] One-letter amino acid code has been used. Positions where there are differences within a pair of structures are shaded.



Handedness of Secondary Structures

Most structural elements in protein structures, for which handedness can be defined, prefer one enantiomer over the other. Handedness manifests itself in the right-handedness of α-helix, the twist of β-sheet, and the topology of connectivity in parallel β-sheet.[3] Nevertheless, mirror images are observed, although at much lower proportions in short fragments, such as I and I′, II and II′ β-turns. The ratios of the numbers observed in the two categories are 8:1 and 5.5:1, respectively.[31] The 3_{10}-helices can be grouped into two pairs of structures of opposite hands: 3_{10}^{R} and 3_{10}^{L}, 3_{10}^{EL} and $3_{10}^{E'R}$. For three-length helices, the ratios of the numbers observed are 119:1 and 0.78:1, respectively (Table II). Thus, of all pairs of structures of opposite handedness, 3_{10}^{EL}- and $3_{10}^{E'R}$-helices seem to be of comparable occurrence. However, as the length is increased, the number of helices containing α_L conformation is drastically reduced, indicating that it is difficult to have a structure with more than two residues with α_L

conformation. The same conclusion can be arrived at considering α-helices also. Whereas there has been just one reported example of a left-handed α-helix,[38] our survey revealed three cases of four-length α-helices, with an additional nine having the conformation $E\alpha_L\alpha_L\alpha_L$. However, as the length is increased beyond four, we are left with just one example of $E\alpha_L\alpha_L\alpha_L$.

Disallowed Conformation as Part of Regular Secondary Structure

The occurrence of a non-Gly residue in the disallowed region of the Ramachandran plot may be a source of concern while validating a newly determined structure.[39] The conformation E′ (Table II, Fig. 2) found in the $3_{10}^{E'R}$-helix has been identified as the region II in the disallowed region of the Ramachandran plot.[40] Although the occurrence of such conformations is normally indicative of the presence of local strains, it is likely that when they form a part of a larger secondary structure identified here, such conformations can be accepted with a higher degree of confidence.

Sequence Position

	...	N1′	Nc	1	2	3	Cc	C1′	...	Frequency
a)	E		C	3_{10}^{EL}			C	E/H/G		15[a]
b)	E		C				C	C	E/G	5[b]
c)	E/H			$3_{10}^{E'R}$			H			19[c]
d)	H						S	C	E/H/G	8[d]
e)	E		C				E/H			6[e]
f)	E		C				C	E		5[f]
g)	H						E			4

Fig. 4. Schematic diagram showing nearest neighbor residues in helical (G/H) or β-strand (E) conformation at both ends of 3_{10}^{EL}- and $3_{10}^{E'R}$-helices (see Materials and Methods for the designation of all secondary structural states). Comments: [a]Of which 13 cases have only E after the helix; in two cases, Nc position is S, and in one case the Cc position is B (instead of C). [b]Only E in four cases after the helix; Nc position is S in one case, and in two cases, the Cc position is S (instead of C). [c]The nearest E or H residue is anywhere within four positions before the helix; E occurs five times at N1′ position. [d]Two cases have C at Cc position. [e]Five cases have only H after the helix. [f]Two cases have only T and one case has B at the Cc position.

Potentials for 3_{10}-Helix Variants and Their Differences From β-Turns

Different types of 3_{10}-helices are also quite distinct from one another on the basis of amino acid usage, and many of the residues are selected at a given position because they are endowed with the ϕ, ψ angles required at the position. Thus, positions 2 and 3 with α_L conformation in the 3_{10}^{EL}-helix are predominantly occupied by Gly, Asn, and Asp, the preference of Gly being particularly strong at 3 (Fig. 3, Table III). Gly is also the highly favored residue at 1 in $3_{10}^{E'R}$-helices, where an E′ conformation is normally not allowed for a non-Gly residue. Pro has a high propensity value for this position in 3_{10}^{R}- and 3_{10}^{EL}-helices. Two residues with the highest propensity values at the Nc position are Asp and Pro in the 3_{10}^{R}-helix, Cys and Gly in the 3_{10}^{EL}-helix, and Thr and Leu in the $3_{10}^{E'R}$-helix. In general, this position is usually hydrophobic for 3_{10}^{R}-helices and hydrophilic for $3_{10}^{E'R}$-helices. The Cc position is generally hydrophobic for 3_{10}^{R}- and 3_{10}^{EL}-helices (the latter also has Gly), whereas the propensities are quite high for Ala, Ser, and Gly to be at this position in $3_{10}^{E'R}$-helices.

The three-length 3_{10}-helices are a single-residue extension of a particular type of β-turn (specified by the ϕ, ψ angles of the two central residues in a segment of i to i + 3 residues). For instance, type-I (more precisely type-III, of

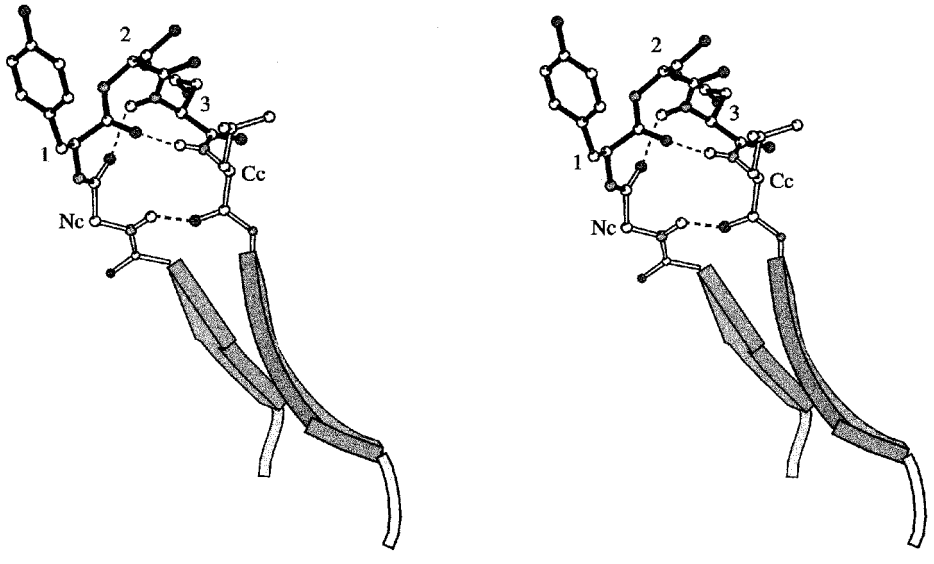


Fig. 5. MOLSCRIPT[35] stereo representation of a β-hairpin structure (residues 147–161 in labrB) containing a 3_{10}^{EL}-helix (residues 152–154). The 3_{10}^{EL}-helix is represented by ball-and-stick model. The usual 3_{10}^{EL}-type hydrogen bonds are in thinner dashed line and the Nc NH to Cc CO hydrogen bond is shown in thicker dashed line. In the ball-and-stick region, only those amide hydrogens are shown that are involved in backbone hydrogen bonding. Carbon atoms are white, oxygen atoms are black, and the amide nitrogen atoms are gray in color.

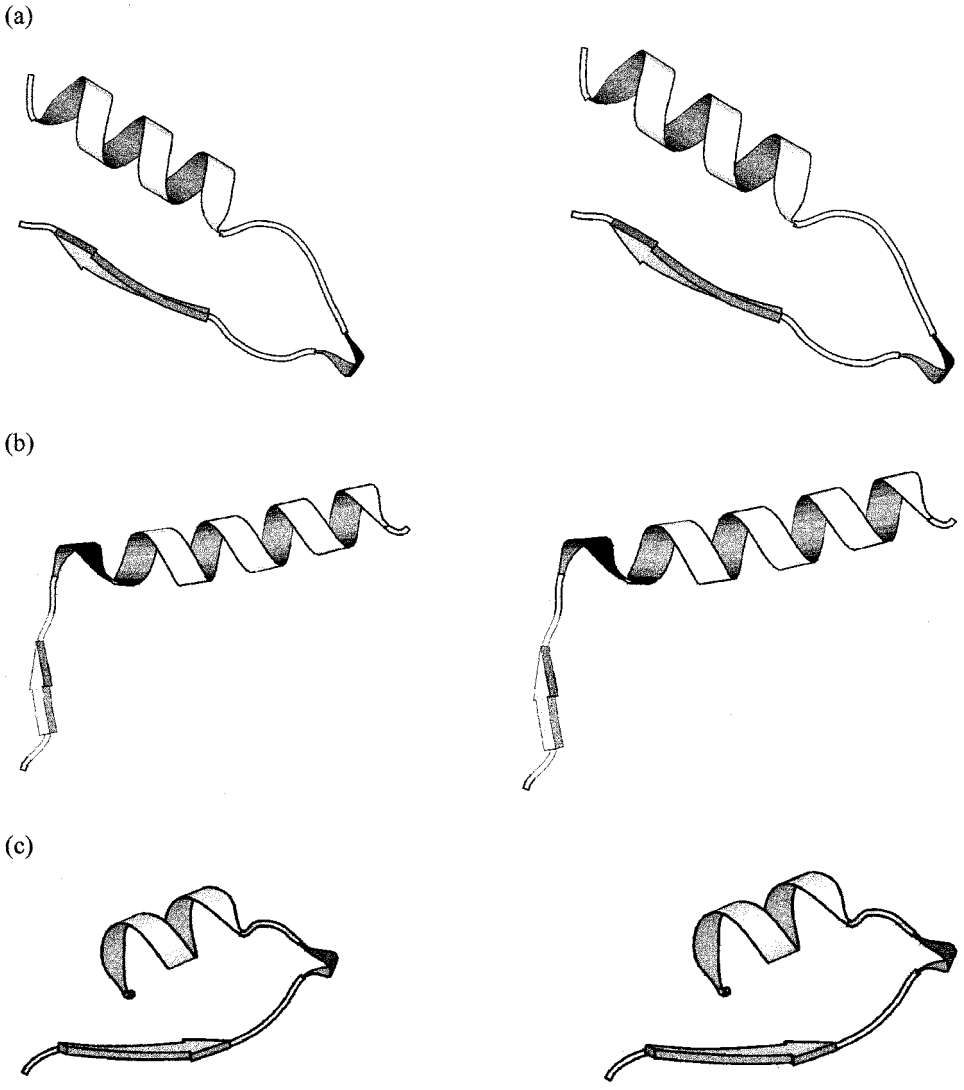


Fig. 6. Cartoon representations of **(a)** 3_{10}^{EL}- (residues 40–42 in 1bkjA), **(b)** $3_{10}^{E'R}$- (residues 61–63 in 3pte_), and **(c)** 3_{10}^{L}- (residues 191–193 in 2bgu_) helices (all in darker shade), shown in stereo and generated by MOLSCRIPT.[35] The secondary structural elements in the immediate environment are also included in the figure.

which there are not many well-characterized examples) β-turn leads to the classic 3_{10}-helix, type II the 3_{10}^{EL}-helix and type-II′ to the $3_{10}^{E'R}$-helix. When we compared the equivalent position within a pair of 3_{10}-helix and β-turn (Table IV), we found that the two types of structures can be distinguished based on residue types that occur at position i and especially at i + 3. This shows that the signature of whether a four-residue segment will be a β-turn or a 3_{10}-helix lies in the two terminal residues.

Secondary Structural Motifs Containing 3_{10}-Helix Variants

Like the super-secondary structural motifs containing the classic 3_{10}-helices,[6] the variants can also be embedded in between different structural elements. The most prominent location of a 3_{10}^{EL}-helix is in β-hairpins (Figs. 4 and 5), which are usually of the class 3:5 according to the notation of Sibanda et al.[36] For $3_{10}^{E'R}$-helices, a considerable number of structures have an α-helix at their C-terminal end [Figs. 4(c) and 6(b)].

Finally, for modeling of protein structures, it is important to identify local structural motifs (sequence of specific conformational angles) and the residue patterns that give rise to them. With increasing size of the available database, it is possible to refine the conformational parameters and the sequences involved in the motifs.[41] In this study, based on ϕ, ψ angles, we have classified 3_{10}-helices into four types and identified the residue characteristics that determine whether a four-residue segment will form a β-turn or a 3_{10}-helix in the folded structure. However, because of the limited size of the data set used in this work, any attempt to quantitatively investigate the predictive power of the sequence rules is prone to errors,[42] and therefore was not attempted. Recent studies indicate that

3_{10}-helices may act as intermediates in α-helix formation.[43] The results presented herein suggest that β-turn and 3_{10}-helices may share similar folding pathways. Although overwhelmingly helices are conformationally repetitive structures, we have identified a number of helices, which have the first residue in a nonstandard conformation. In addition, we have also shown that the population of mirror images of such variants are found to be almost equally populated in proteins, especially for three-length 3_{10}-helices ($3_{10}^{E'R}$ and 3_{10}^{EL}).

ACKNOWLEDGMENTS

We thank the Bioinformatics Centre for computational facilities.

REFERENCES

1. Donohue J. Hydrogen bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci USA 1953;39:470–478.
2. Karpen ME, De Haseth PL, Neet KE. Differences in the amino acid distributions of 3_{10}-helices and α-helices. Protein Sci 1992;1: 1333–1342.
3. Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem 1981;34:167–339.
4. Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984;44:97–179.
5. Barlow DJ, Thornton JM. Helix geometry in proteins. J Mol Biol 1988;201:601–619.
6. Pal L, Basu G. Novel protein structural motifs containing two-turn and longer 3_{10}-helices. Protein Eng 1999;12:811–814.
7. Doig AJ, MacArthur MW, Stapley BJ, Thornton JM. Structures of N-termini of helices in proteins. Protein Sci 1997;6:147–155.
8. Blundell TL, Zhu Z-Y. The α-helix as seen from the protein tertiary structure: a 3-D structural classification. Biophys Chem 1995;55:167–184.
9. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
10. Lewis PN, Momany FA, Scheraga HA. Chain reversals in proteins. Biochim Biophys Acta 1973;303:211–229.
11. Rose GD, Gierasch LM, Smith JA. Turns in peptides and proteins. Adv Protein Chem 1985;37:1–109.
12. Wilmot CM, Thornton JM. Analysis and prediction of the different types of β-turn in proteins. J Mol Biol 1988;203:221–232.
13. Venkatachalam CM. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 1968;6:1425–1436.
14. Nemethy G, Scheraga HA. Stereochemical requirements for the existence of hydrogen bonds in beta-bends. Biochem Biophys Res Commun 1980;95:320–327.
15. Hutchinson EG, Thornton JM. A revised set of potentials for β-turn formation in proteins. Protein Sci 1994;3:2207–2216.
16. Matthews BW. The γ-turn: evidence for a new folded conformation in proteins. Macromolecules 1972;5:818–819.
17. Milner-White EJ, Ross BM, Ismail R, Belhadj-Mostefa K, Poet R. One type of gamma-turn, rather than the other gives rise to chain-reversals in proteins. J Mol Biol 1988;204:777–782.
18. Milner-White EJ. Situations of gamma-turns in proteins: their relation to alpha-helices, beta-sheets and ligand binding sites. J Mol Biol 1990;216:386–397.
19. Nataraj DV, Srinivasan N, Sowdhamini R, Ramakrishnan C. α-Turns in protein structures. Curr Sci 1995;69:434–447.
20. Rajashankar KR, Ramakumar S. π-Turns in proteins and peptides: classification, conformation, occurrence, hydration, and sequence. Protein Sci 1996;5:932–946.
21. Brunet AP, Huang ES, Huffine ME, Loeb JE, Weltman RJ, Hecht MH. The role of turns in the structure of an α-helical protein. Nature 1993;364:355–358.
22. Chou PY, Fasman GD. Prediction of protein conformation. Biochemistry 1974;13:222–245.
23. Chou PY, Fasman GD. Prediction of beta-turns. Biophys J 1979;26: 367–373.
24. Garnier J, Osguthorpe DJ, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 1978;120:97–120.
25. Cohen FE, Abarbanel RM, Kuntz ID, Fletterick RJ. Turn prediction in proteins using a pattern-matching approach. Biochemistry 1986;25:266–275.
26. McGregor MJ, Flores TP, Sternberg MJ. Prediction of beta-turns in proteins using neural networks. Protein Eng 1989;2:521–526.
27. Chou KC, Blinn JR. Classification and prediction of beta-turn types. J Protein Chem 1997;16:575–595.
28. Cai YD, Yu H, Chou KC. Prediction of beta-turns. J Protein Chem 1998;4:363–376.
29. Shephard AJ, Gorse D, Thornton JM. Prediction of the location and type of beta-turns in proteins using neural networks. Protein Sci 1999;8:1045–1055.
30. Guruprasad K, Rajkumar S. β- and γ-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. J Biosci 2000;25:143–156.
31. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
32. Hobohm U, Scharf M, Schneider R. Selection of representative protein data sets. Protein Sci 1993;1:409–417.
33. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
34. Chakrabarti P, Pal D. The interrelationships of side-chain and main-chain conformations in proteins. Prog Biophys Mol Biol 2001;76:1–102.
35. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24: 946–950.
36. Sibanda BL, Blundell TL, Thornton JM. Conformation of β-hairpins in protein structures: a systematic classification with applications to modelling by homology, electron density fitting, and protein engineering. J Mol Biol 1989;206:759–777.
37. IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. J Mol Biol 1970;52:1–17.
38. Soman KV, Ramakrishnan C. Occurrence of a single helix of the collagen type in globular proteins. J Mol Biol 1983;170:1045–1048.
39. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
40. Pal D, Chakrabarti P. On residues in the disallowed region of the Ramachandran map. Biopolymers 2002;63:195–206.
41. Pal D, Chakrabarti P. Cis peptide bonds in proteins: residues involved, their conformations, interactions, and locations. J Mol Biol 1999;294:271–288.
42. Pal L, Basu G. Neural network prediction of 3_{10}-helices in proteins. Indian J Biochem Biophys 2001;38:107–114.
43. Bolin KA, Millhauser GL. α- and 3_{10}: the split personality of polypeptide helices. Acc Chem Res 1999;32:1027–1033.