Protein Engineering vol.15 no.5 pp.353–358, 2002

Occurrence, conformational features and amino acid propensities for the π-helix

M.N.Fodje and S.Al-Karadaghi[1]

Department of Molecular Biophysics, Center for Chemistry and Chemical Engineering, University of Lund, Box 124, SE-221 00 Lund, Sweden

[1]To whom correspondence should be addressed.
E-mail: salam.al-karadaghi@mbfys.lu.se

The most abundant helix type in proteins is the α-helix, accounting for about 31% of amino acid secondary structure states, while the 3_{10}-helix accounts for about 4%. The π-helix appears to be extremely rare and is considered to be unstable. Existing secondary structure definition methods find very few within the Protein Data Bank. Using an improved π-helix definition algorithm to search a non-redundant subset of high-resolution and well-refined protein structures, we found that almost every tenth protein contained a π-helix. This enabled us to show for the first time that the π-helix has structural parameters that are different from the hypothesized model values. It also has distinctive amino acid preferences and it is conserved within functionally related proteins. Features that may contribute to the stability of the π-helical structure have also been identified. In addition to hydrogen bonds, several other factors contribute to the stability of π-helices. The π-helix may have some functional advantages over other helical structures. Thus, we describe cases where the side chains of functionally important residues at every fourth position within a π-helix could be aligned and brought close together in a way that would not be allowed by any other helix type.

Keywords: algorithm/helix propensity/π-helix/secondary structure definition

Introduction

Helices are a major type of secondary structure element found in proteins. Helix types are usually designated as x_y based on the number of residues per turn (x) and the number of atoms in the ring closed by a hydrogen bond (y) (Donohue, 1953). Pauling and Corey first hypothesized the α-helix (3.6_{13}) and the γ-helix (5.1_{17}) structures (Pauling and Corey, 1951). Donohue later considered the possibility of other types of helices (2.2_7, 3_{10}, 4.3_{14} and 4.4_{16}) (Donohue, 1953). Low and Baybutt also suggested the possibility of the 4.4_{16}-helix or π-helix (Low and Baybutt, 1952). The main stabilizing factor for helical structures in polypeptides is repeating hydrogen bonds between main chain carbonyl oxygen (C=O) and amide hydrogen (NH) groups with the α-helix characterized by an ($i \leftarrow i + 4$) pattern, the 3_{10}- and the π-helix by repeating ($i \leftarrow i + 3$) and ($i \leftarrow i + 5$) hydrogen bonds, respectively.

Of all the hypothesized helix types, only the α-helix, 3_{10}-helix and the π-helix have been observed in protein structures. The α-helix is considered to be the most abundant form of secondary structure, accounting for about 31% of amino acid secondary structure states, while the 3_{10}-helix accounts for about 4% (Baker and Hubbard, 1984; Barlow and Thornton, 1988). The π-helix, however, appears to be extremely rare. The rarity of the π-helix has been attributed to its instability due to the following properties: (1) the dihedral angles φ and ψ are unfavorable, lying at the very edge of an allowed minimum energy region of the Ramachandran plot (Ramachandran and Sasisekharan, 1968); (2) the larger radius of the π-helix means that main chain atoms are no longer in van der Waals contact along the helix axis, resulting in a hole too small for a water molecule to fill (Low and Baybutt, 1952); (3) a large entropic cost is required to form a helix in which five residues need to be aligned to permit the ($i \leftarrow i + 5$) hydrogen bond (Rohl and Doig, 1996). A few researchers have, however, found π-helices to be formed during molecular dynamics simulations of peptides (Kovacs *et al.*, 1995; Gibbs *et al.*, 1997; Shirley and Brooks, 1997) with some reports of a transition from α-helix to π-helix structure (Duneau *et al.*, 1996; Lee *et al.*, 2000). This suggests that the π-helix is not as unstable as previously believed.

Several algorithms have been designed to assign secondary structure automatically based on three-dimensional coordinates (Kabsch and Sander, 1983; Richards and Kundrot, 1988; Frishman and Argos, 1995; Labesse *et al.*, 1997). Among these, DSSP (Kabsch and Sander, 1983) and STRIDE (Frishman and Argos, 1995) are the most widely used. While DSSP identifies helices based on the presence of repeating ($i \leftarrow i + n$) hydrogen bonds with n corresponding to 3, 4 and 5 for 3_{10}-, α- and π-helices, respectively, STRIDE makes use of both hydrogen bonds and main chain dihedral angles to define secondary structures. In a recent study, by using DSSP, only nine unique π-helices could be identified in a database of more than 6000 proteins (Weaver, 2000). In a similar search using STRIDE, we could not identify any additional hits (unpublished results). Surprisingly, neither STRIDE nor DSSP could identify a few π-helices that we had previously observed from graphical analysis of three-dimensional structures (Al-Karadaghi *et al.*, 1994, 1997). This led us to believe that helix identification algorithms needed some modification in order to identify π-helices reliably.

In this study, by using a modified π-helix definition algorithm, we show, for the first time, that the π-helix is >10 times more prevalent than previously reported. The results have allowed a detailed analysis of structural features and amino acid sequence preferences of this unique helix type, also shedding light on the forces that stabilize it.

Materials and methods

Composition of database

The CullPDB server (Hobohm *et al.*, 1993) was used to generate a non-redundant database of protein chains from the 21 March 2001 edition of the PDB (Protein Data Bank) based on the following criteria: (1) no two proteins included would

have a sequence identity >30%; (2) all proteins included must have been determined to a resolution at or better than 2.0 Å, hence all proteins determined by NMR spectroscopy were excluded; (3) an additional condition was imposed that the crystallographic *R*-factor be 0.2 or better. The resulting database consisted of 936 protein chains, corresponding to 224 046 amino acid residues.

Identification of π-helices

We have adopted an algorithm that conforms to rule 6.3 of the IUPAC–IUB recommendations (IUPAC–IUB, 1970). The first helical residue is defined as the one whose C=O group is involved in an ($i \leftarrow i + 5$) hydrogen bond, while the last residue is the one whose NH group is involved in an ($i - 5 \leftarrow i$) hydrogen bond within the helix. In order to distinguish a π-helix from a π-turn, which has only one ($i \leftarrow i + 5$) hydrogen bond, we considered the minimal π-helix as consisting of at least two ($i \leftarrow i + 5$) hydrogen bonds.

The criteria for defining hydrogen bonds are similar to those used in DSSP (Kabsch and Sander, 1983). When absent, the positions of main chain amide hydrogen atoms were computed according to the general rules described by Pauling *et al.* (Pauling *et al.*, 1951). The electrostatic interaction potential was calculated as described by Kabsch and Sander (Kabsch and Sander, 1983), although partial charges were assigned according to the CNS parameters (Brunger *et al.*, 1998). A hydrogen bond was assigned if the calculated electrostatic interaction free energy was <-0.5 kcal mol^{-1}. In cases where more than one hydrogen bond was possible among the ($i \leftarrow i + 3$), ($i \leftarrow i + 4$) and ($i \leftarrow i + 5$) hydrogen bonds, the bond with the most favorable interaction energy was considered.

The algorithm was implemented using the C++ programming language (this can be obtained from the authors on request). The program was then used to test for the occurrence of π-helices in the database of selected protein chains. All π-helices identified were visually confirmed using the program SPDBViewer (Guex and Peitsch, 1997).

Solvent-accessible surface area, volume and helix parameters

The solvent-accessible surface area of each residue was calculated using the accessibility subroutine in the program STRIDE (Frishman and Argos, 1995). The fraction of exposed surface area was taken as the ratio of the accessible area to the total surface area. The total surface area of each amino acid was taken from Chothia (Chothia, 1975). Molecular volumes and surface areas of polypeptide fragments were calculated using the program SPDBV (Guex and Peitsch, 1997). Helix geometry parameters such as the number of residues per turn, unit rise and unit twist were computed using the program HELANAL (Bansal *et al.*, 2000).

Statistical methods

The distribution of each amino acid type was calculated within the 932 protein chains and all identified π-helices. To elucidate position-specific sequence patterns, we also analysed the residue distributions at seven positions (π1–π7) within the π-helices and five positions on either side of the π-helices, (+1 to +5) on the C-terminal side and (–5 to –1) on the N-terminal side. The number seven was chosen because a majority of the π-helices were seven residues long. The χ^2 test was used to determine if the distribution of amino acid content of π-helices was significantly different from that of the database used. For a system with 19 degrees of freedom, a χ^2 value >43.82 is required to conclude with 99.9% confidence that observed differences are not by chance, while a χ^2 value of 30.144 is required for 95% confidence. Using the database distribution as a reference, the binomial test (Siegal and Castellan, 1998) was used to determine the significance of the distributions at each position within and around π-helices. A position with a binomial probability of ≤0.05 was considered to be highly significant while one with ≤0.1 or less was considered to be marginally significant. The overall propensities and their standard deviations for each amino acid type to occur in π-helices, and the positional propensities for the 17 positions mentioned above, were calculated as described previously (Chou and Fasman, 1974).

Results and discussion

π-Helices are more common than previously believed

We have developed a modified algorithm for the recognition of π-helices. Its main feature is an energy-based classification of hydrogen bonds, after which a preference is given to the bond with the highest energy. In contrast, in DSSP and in STRIDE preference is always given to the α-helix-forming ($i \leftarrow i + 4$) hydrogen bonds when more than one is possible, even if the π-helix-forming ($i \leftarrow i + 5$) hydrogen bonds are stronger. Using our algorithm to search the database of 932 high-resolution and well-refined protein structures, 116 helices were flagged as potential π-helices, 104 of which could be confirmed by visual inspection. Several proteins possessed more than one π-helix with the methane monooxygenase hydroxylase from *Methylococcus capsulatus* (PDB 1MTY) having up to eight. Of the 224 046 amino acids in the database, 728 were found in these π-helices, corresponding to a mere 0.3% of the total. The observed π-helices ranged in size from 7 to 13 amino acid residues, the most common length being seven, which corresponds to 1.5 turns (Table I).

Our results represent the largest number of π-helices ever reported. Using both STRIDE (Frishman and Argos, 1995) and DSSP (Kabsch and Sander, 1983) to search the same database, we could only identify six and nine π-helices, respectively. Our method is therefore more sensitive for the definition of π-helices from protein coordinates. This algorithm differs from those used previously (Kabsch and Sander, 1983) by its ability to compare the energy of all possible hydrogen bonds within a structure in order to define its type.

π-Helices have distinct conformational features

The mean values of the dihedral angles (φ, ψ) of all π-helices observed were found to be at (–76°, –41°) with standard deviations (σ_ϕ, σ_ψ) = (25, 24). These values are significantly different from the proposed model values of (–57°, –70°) (Low and Grenville-Wells, 1953; Ramachandran and Sasisekharan, 1968; Creighton, 1993), but similar to the values of (–78°, –41°) reported for a π-helix in alcohol dehydrogenase (Al-Karadaghi *et al.*, 1994) and (–77°, –54°) for π-helices formed during molecular dynamics simulations (Lee *et al.*, 2000). A position-dependent analysis of the dihedral angles in the helices showed a distinct pattern. Thus, for the π4 and π5 positions, mean (φ, ψ) values were (–96, –26) and (–97, –51), respectively (Figure 2a). These features imply that any method of secondary structure definition from three-dimensional coordinates that makes use of the hypothesized main chain dihedral angle values for the π-helix will perform poorly.

The helical geometry calculated from all observed π-helices with nine or more residues revealed an average unit rise of

Table I. Position and length of identified π-helices[a]

PDB	Position	Hb	PDB	Position	Hb	PDB	Position	Hb
1DYS	A112–118	2	1EYZ	A119–125	2	1DQZ	A97–103	2
1ELK	A95–101	2	1B16	A104–110	2	1MTY	B140–150	4
1EGU	A292–298	2	9GAF	A186–192	2	1MTY	B247–266	5
1EGU	A441–447	2	1B25	A479–485	3	1MTY	B297–304	3
1KVE	B177–183	2	1YGE	261–267	2	1MTY	D185–191	2
1THG	424–430	2	1YGE	494–506	6	1MTY	D202–214	6
2SCP	A56–62	3	1YGE	684–690	2	1MTY	D306–318	7
1DZ4	A150–156	2	1LST	126–132	3	1MTY	D379–385	2
1MUC	A70–76	2	1LST	165–171	2	1HFE	S71–77	2
1QMG	A349–355	3	1C3P	A97–103	2	1CXP	C291–297	2
1QMG	A490–502	4	1D8D	A343–349	2	1C7S	A641–647	2
1EVY	A257–263	2	1CYD	A104–110	2	1C7S	A801–807	2
1FQA	A279–285	2	1GAI	150–156	2	1LML	155–161	2
1E3A	A138–146	3	1B5Q	A74–80	2	1D3Y	A253–259	2
1EI5	A177–183	3	1DEK	A137–145	4	1QOY	A26–32	2
1SUR	131–137	2	1BXK	A98–104	2	2OLB	A301–308	3
1QH3	A154–160	2	1DQS	A142–148	2	1SVF	A171–177	2
1A8D	53–59	2	1DC1	A98–104	2	5CSM	A233–239	2
1A8E	124–130	2	1I0H	A26–32	2	1FDS	111–117	2
1SMD	27–33	3	1EWF	A181–187	2	1ONE	A67–73	2
1F3A	A126–132	2	1CB8	A267–273	2	1PHN	A107–113	2
1D3G	A37–43	2	2EBN	257–263	2	1FUR	A155–161	2
1EOK	A111–118	3	1FP3	A273–279	2	1FUR	A383–389	2
1QLM	A88–94	2	1DOZ	A265–274	3	1DXR	C277–283	2
1QGW	C105–111	2	1EA5	A396–402	2	1DXR	H27–33	2
1NCI	A40–46	2	1EA5	A522–528	2	1DQA	A733–740	3
1DK8	A242–249	3	4PAH	325–331	2	1BDB	112–118	2
1EL4	A44–51	4	1C3W	A213–219	2	1DXR	L129–135	2
1EK6	A105–111	2	1HVB	A183–189	3	1DXR	M156–162	2
1MRO	A313–324	5	1FRP	A276–282	2	1DJ0	A81–87	2
1A8I	488–495	3	2HMQ	A101–107	2	1G8K	A181–187	2
1QH8	A63–72	4	1UOK	393–399	2	1G8K	A242–248	2
1BDM	A217–223	2	7A3H	A146–152	2	1FSW	A174–180	3
1BG6	297–303	2	1F24	A140–146	2	1YAC	A114–120	2
1F24	A214–220	3	1COJ	A26–33	3			

[a]PDB, Protein Databank code; Position, chain identifier, start and end positions for π-helix; Hb, number of ($i \leftarrow i + 5$) type hydrogen bonds within π-helix.

1.2 Å, with 4.4 residues per turn and an average unit twist of 83°. These values are very similar to the hypothesized model values of 1.1 Å unit rise and 4.4 residues per turn. A close similarity of the overall geometry of π-helices is demonstrated in Figure 1b. A superposition of the structures of 11 seven-residue π-helices, shown on the figure, gave an average root mean square deviation (r.m.s.d.) of 0.13 Å (0.013 nm) for the main chain atoms.

The main characteristic of π-helices is the repeating (*ii + 5*) hydrogen bonding pattern. We could identify regions with 2, 3, 4 and 5 consecutive (*ii + 5*) hydrogen bonds within the helices, although 2 hydrogen bonds were the most common. Stretches with more than 5 consecutive bonds could be observed in a less restrictive database (results not shown). In most cases, the C=O group of the first π-helix residue was found to be involved in a bifurcated hydrogen bond with the NH groups of both residues $i + 4$ and $i + 5$, although the ($i \leftarrow i + 4$) bond was weaker (Figure 1b). When only two consecutive hydrogen bonds were observed ($i \leftarrow i + 5$ and $i + 1 \leftarrow i + 6$), the three middle C=O groups (residues $i + 2$, $i + 3$ and $i + 4$) were not involved in a hydrogen bond with any main chain groups and remained exposed to solvent. Similarly, with three consecutive hydrogen bonds, two C=O groups (residues $i + 3$ and $i + 4$) were exposed and with four hydrogen bonds, only one C=O group (residue $i + 4$) was not a hydrogen bond acceptor (Figure 1a). Regions with ($i \leftarrow i + 5$) bonds alternating with non-hydrogen-bonded C=O groups ($i = 1, 3, 5$) were also observed to form a π-helix conformation, as judged from structural parameters. Support for such cases has been added to the algorithm.

π-Helices have a characteristic amino acid distribution

Detailed analysis of sequence–structure relationships within different secondary structure types has been at the focus of many research groups and has often been aimed at revealing amino acid preferences for different secondary structure types. In some cases, the preferences for different positions within a secondary structure type have been analyzed (Kumar and Bansal, 1998). However, such studies on π-helices have been impossible until now.

The distributions of individual amino acids within the observed π-helices, and within the whole database used in the analysis, are listed in Table II. It can be seen that the distribution for π-helices is significantly different ($\chi^2 = 92.51$) from the distribution within the database, as revealed by the χ^2-test at 99% confidence limit. From the overall values for propensities, it appears that in π-helices aromatic and large aliphatic amino acids are generally preferred, while small amino acids are avoided. Thus, Ala, Gly and Pro seem to be avoided in favor of Ile, Leu, Tyr, Trp, Phe, His and Asn. The observed propensities did not show any correlation (correlation coefficient $r = 0.24$) with those for the π-helix (Pace and Scholtz, 1998), demonstrating that the π-helix has unique residue preferences.

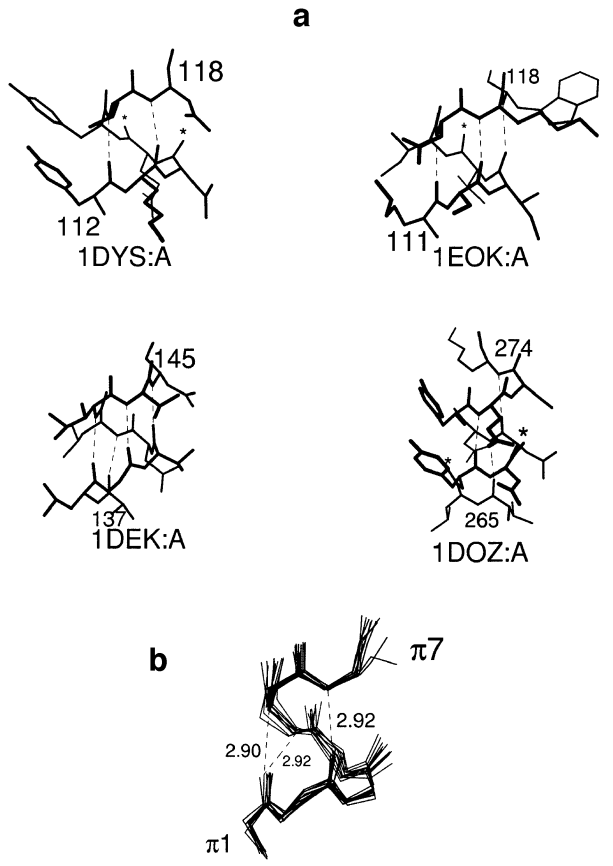


Fig. 1. (**a**) Examples of observed π-helices with different numbers and distributions of (i + 5) hydrogen bonds. PDB codes are shown below each helix. Non-hydrogen-bonded C=O groups are marked with asterisks. (**b**) Structure superposition of the first seven residues of 11 π-helices. Only the backbone atoms are shown. Hydrogen bond distances are in Å. The bifurcated hydrogen bond at the π1 position is also shown.

Table II. Distribution of amino acids[a]

π-Helix				Database	
AA	N/O	P/C	PP	N/O	P/C
Ala	44	6.04	0.70	19290	8.65
Arg	32	4.40	0.95	10367	4.65
Asn	49	6.73	1.47	10218	4.58
Asp	38	5.22	0.87	13351	5.99
Cys	12	1.65	1.17	3134	1.41
Gln	20	2.75	0.73	8359	3.75
Glu	43	5.91	0.96	13774	6.18
Gly	37	5.08	0.64	17812	7.99
His	23	3.16	1.39	5062	2.27
Ile	51	7.01	1.29	12143	5.44
Leu	90	12.36	1.44	19146	8.58
Lys	38	5.22	0.91	12867	5.77
Met	14	1.92	0.91	4691	2.10
Phe	39	5.36	1.34	8902	3.99
Pro	4	0.55	0.12	10557	4.73
Ser	36	4.95	0.84	13158	5.90
Thr	31	4.26	0.74	12768	5.72
Trp	20	2.75	1.80	3414	1.53
Tyr	45	6.18	1.68	8219	3.68
Val	62	8.52	1.20	15808	7.09
Total	728	–	–	223040	–

[a]AA, amino acid; N/O, number of occurrences of this amino acid; P/C, percentage composition; PP, calculated propensity of amino acids within π-helices.

Table III. Position-specific properties of amino acid distribution within and around π-helices

Position	χ^2 residual	Acc.[a]	*B* factor	φ (°)	ψ (°)
π – 4	22.32	24	18	–62	2
π – 3	19.47	25	17	–68	–17
π – 2	37.53	22	17	–66	–18
π – 1	16.61	20	16	–64	–26
π1	53.00	17	16	–65	–33
π2	33.07	14	16	–65	–48
π3	27.29	24	17	–73	–38
π4	84.80	26	17	–94	–29
π5	44.61	22	17	–101	–54
π6	56.68	17	17	–69	–48
π7	17.74	19	17	–62	–37
π + 1	332.80	31	17	–61	–15
π + 2	44.28	24	17	–63	–9
π + 3	25.86	24	17	–61	–3
π + 4	18.12	32	18	–65	–1

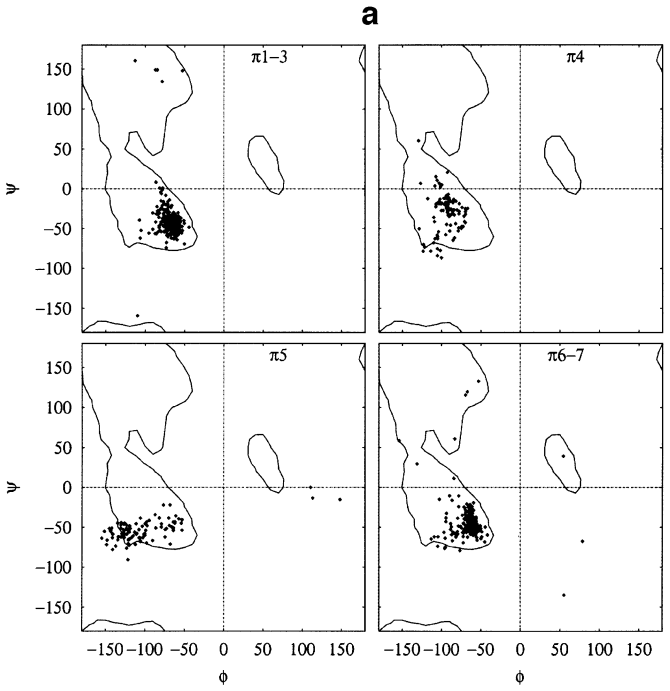
[a]Acc., mean percentage of accessible surface area at each position.

Position-specific preferences within π-helices are presented in Table III and Figure 3. In the analysis, only positions with a significant distribution according to the χ^2-test at 95% confidence level were considered. It can be seen that large amino acids such as Phe, Trp, Tyr, Ile, Leu and Met show high propensities for being located at the beginning and at the end of the helix. The preference for bulky amino acids is probably due to the more favorable van der Waals interactions between the side chains, which may be one of the factors contributing to the stabilization of the helix. Analysis of residue preferences at other positions shows that polar residues such as Asn, Glu, Thr and Ser are preferred. Asn, in particular, shows an extremely high propensity for the middle position. As noted above, the C=O groups at these positions are most often not involved in any main chain to main chain hydrogen bonds. However, interactions of the polar side chains with solvent could compensate for this, which may serve as another factor contributing to helix stability. It can also be seen that the average percentage solvent accessibility is slightly higher at the middle of the helix than at the beginning and end positions (Table III).

Residues at positions before and after the π-helix did not show any statistically significant sequence preferences (Figure 3), except for the +1 position, at which Pro had an extremely high propensity. It has been reported previously that the occurrence of Pro leads to the breakage of at least two adjacent hydrogen bonds (Richardson and Richardson, 1988). In the case of the π-helix, the residue at the +1 position would normally be involved in a hydrogen bond with the residue at position π3 in a seven-residue π-helix. Additionally, the pyrolidine ring of Pro forces the C=O group of the residue at position π4 to point away from the helix axis, leading to a second non-hydrogen-bonded C=O group at this position (Figure 2b). Although it has been suggested that the presence of a proline residue promotes the formation of a π-turn (Rajashankar and Ramakumar, 1996), it appears also to be the limiting factor in the size of the π-helices we have observed. The low propensity of Pro within π-helices and the high propensity at the +1 position (Figure 3) can only be interpreted to mean that Pro leads to the termination of the π-helix.

Stabilizing factors in π-helices

As noted above, the π-helix has been considered to be unstable, a primary reason for its rarity in protein structures. Our



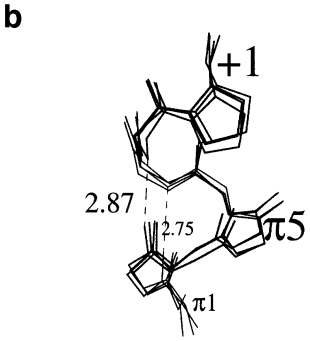


Fig. 2. (**a**) Ramachandran ϕ–ψ plots for the first seven positions within π-helices. The plots for the first three positions and the last two are combined. The 'allowed' region (Kleywegt and Jones, 1996) is outlined. (**b**) Structural superposition of four π-helices with proline at the +1 position, showing the contribution of the pyrrolidine ring to the conformation of the π4 C=O conformation.

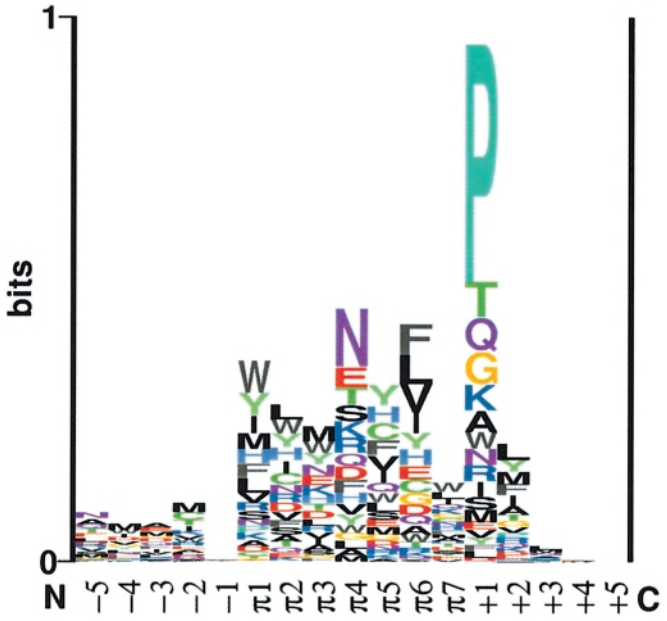


Fig. 3. Logo plot (Schneider and Stephens, 1990) of the amino acid distribution at 17 positions within and around π-helices. The height of each letter is proportional to its frequency at that position and the height of the stack is proportional to its information content, a measure of its conservation. Each stack is ordered according to frequency with the most frequent at the top. Hydrophobic residues are colored in black, neutral in purple, acidic in red, basic in blue, proline in cyan and glycine in orange. Aromatic residues are pale.

analysis suggests several factors, which arise from the particular conformation and amino acid content, that may contribute to the stability of π-helices. Thus, the conformation of a π-helix with 4.4 residues per turn and unit twist of 83° implies that every fourth residue will be in almost azimuthal position with respect to the first. This, together with the fact that the unit rise of a π-helix (1.2 Å) is lower than for either the α (1.5 Å) or 3_{10} (2.0 Å) helices, means that the side chains in a π-helix will be at the closest distance from each other, as compared with other helix types. Steric repulsion arising from this has been suggested to be one of the reasons causing instability of the π-helix (Low and Grenville-Wells, 1953). However, our results point to the contrary. Extensive interactions between side chains, mostly of the van der Waals type, but also aromatic ring stacking and a few electrostatic interactions, appear to be a stabilizing factor (Figure 1a). This correlates well with the observation that aromatic and bulky aliphatic amino acids are more prevalent around the beginning and end of a 1.5-turn π-helix. An additional effect of the alignment of helix residues will be an alignment of main chain groups involved in hydrogen bonds, resulting in shorter and stronger bonds as compared with α- and 3_{10}-helices.

The dihedral angles (–57°, –70°) are thought to be unfavorable, lying at the very edge of an allowed minimum energy region of the Ramachandran plot. The angles (–76°, –41°), however, lie within the minimum energy region, further supporting the stability of the π-helix. These values are in agreement with the free energy minimum observed for a decaalanine π-helix (–77°, –56°) during molecular dynamics simulations in water (Mahadevan *et al.*, 2001). In this study, the π-helical conformation was suggested to be as stable as the corresponding α-helical conformations.

The central axial hole hypothesized for the π-helix is still present despite the significant differences in dihedral angles. However, a factor that may help to 'fill in' the hole could be the tendency of Pro to be at position +1. This causes one or two main chain C=O groups in each turn to be tilted away from the helix axis with the oxygen pointing outwards (Figure 2b), while the NH group of the next residue will be oriented with the hydrogen pointing towards the helix axis. Even within the longer π-helices, one or two residues in every turn had large negative ϕ values. A conformation of this type has also been suggested earlier to contribute to the stability of π-helices (Blundell and Zhu, 1995). This conformation also enables the C=O group to form a hydrogen bond to solvent, an additional stabilizing interaction as discussed for α-helices (Blundell *et al.*, 1983).

Among the factors contributing to the stability of protein structures are entropic effects. A comparison of volume and surface area of a π-helix with those of an α-helix (calculated for a 22-residue polyalanine ideal π- and α-helix) shows that

the π-helix occupies almost 10% less volume and surface area, as computed by the program SPDBV (Guex and Peitsch, 1997). This would result in a more favorable solvent entropic effect for π-helices. This may also compensate for the entropic effect required to align four residues for a single turn to be formed, as opposed to only three in the case of the α-helix.

Functional insights

The functional importance of π-helices is emphasized by their conservation within functionally related proteins, despite low sequence identity. The chelatase family of proteins is a good example since they share <12% overall sequence identity. Three-dimensional structures are available for ferrochelatases from *Bacillus subtilis* and humans (PDB code 1DOZ and 1HRK, respectively) (Al-Karadaghi *et al.*, 1997; Lecerof *et al.*, 2000; Wu *et al.*, 2001) and one cobalt chelatase from *Salmonella typhimurium* (PDB 1QGO) (Schubert *et al.*, 1999). These enzymes catalyze the insertion of metals (Fe^{2+} for ferrochelatase and Co^{2+} for cobalt chelatase) into protoporphyrin and precorin-2, to produce heme and cobalt-precorin, respectively. The structure of ferrochelatase from *B.subtilis*, which was part of the database used in this study, was flagged to contain a π-helix at positions 265–274. The other two structures, 1HRK and 1QGO, were reported to possess a π-helix at positions 342–357 and 207–213, respectively. Aligning these structures places the π-helices at exactly the same position, about 10 Å from the active site. In *B.subtilis* ferrochelatase, a hydrated magnesium complex has been found bound to the residues of the π-helix. In cobalt chelatase, the authors placed a sulfate moiety at a similar position. However, in the human ferrochelatase no ligand has been observed at this position. A superposition of the π-helices from these three proteins shows that the acidic residues are aligned along the edge of the helix and lead to the active site, the last residue being an invariant glutamate in ferrochelatases and a histidine in cobalt chelatase. These residues outline the proposed path taken by the metal ion substrate into the active site (Al-Karadaghi *et al.*, 1997; Lecerof *et al.*, 2000). This alignment of residues can only be achieved with a π-helix, which permits the retention of an acidic residue at every fourth position. In a 3_{10}-helix acidic residues at every third position will also be aligned with respect to the first, but the unit rise per residue (2.0 Å) will presumably result in too great a distance between side chains. This, in turn, will affect their ability to coordinate a ligand. Another example can be seen within the dioxygenase (1YGE) where histidine residues (H494, H499, H504) within a π-helix are aligned to form a metal binding site. Hence it is evident that the π-helix possesses features that make it a unique structural element where functionally important residues need to be aligned.

The alignment of π-helix residues will also lead to the alignment of main chain dipoles resulting in a slightly higher value of the total dipole moment compared with other helix types of the same length. The functional role of the helix dipole has been extensively reviewed before (Hol *et al.*, 1978; Hol, 1985). Our analysis shows that there is a correlation between π-helices and specific binding sites within proteins. Thus, among the 104 observed cases, 40 π-helices were directly involved in an interaction with a ligand. The functional significance of a few of the π-helices found has been discussed previously (Weaver, 2000). A few cases have been observed in which the π-helix dipole could be used in metal binding. In an ion transport protein (PDB 1A8E), the ferric ion-binding site is located at the N-terminus of the π-helix. Also in an oxidoreductase (PDB 1QH8), two iron–sulfur clusters are bound to the N-terminus of a π-helix.

Acknowledgements

This work was supported by grants from the Swedish Natural Science Research Council to S.A. We are grateful to Anders Liljas, Sven Hovmöller and Gerard Kleywegt for helpful discussions.

References

Al-Karadaghi,S., Cedergren-ZepperZauer,E.S. and Hovmöller,S. (1994) *Acta Crystallogr.*, **D50**, 793–807.
Al-Karadaghi,S., Hansson,M., Nikonov,S., Jonsson,B. and Hederstedt,L. (1997) *Structure*, **5**, 1501–1510.
Baker,E.N. and Hubbard,R.E. (1984) *Prog. Biophys. Mol. Biol.*, **44**, 97–179.
Bansal,M., Kumar,S. and Velavan,R. (2000) *J. Biomol. Struct. Dyn.*, **17**, 811–819.
Barlow,D.J. and Thornton,J.M. (1988) *J. Mol. Biol.*, **201**, 601–619.
Blundell,T.L. and Zhu,Z.Y. (1995) *Biophys. Chem.*, **55**, 167–184.
Blundell,T., Barlow,D., Borkakoti,N. and Thornton,J. (1983) *Nature*, **306**, 281–283.
Brunger,A.T. *et al.* (1998) *Acta Crystallogr.*, **D54**, 905–921.
Chothia,C. (1975) *Nature*, **254**, 304–308.
Chou,P.Y. and Fasman,G.D. (1974) *Biochemistry*, **13**, 222–244.
Creighton,T.E. (1993) *Proteins : Structures and Molecular Properties*. Freeman, San Francisco, pp. 182–186.
Donohue,J. (1953) *Proc. Natl Acad. Sci. USA*, **39**, 470–478.
Duneau,J.P., Genest,D. and Genest,M. (1996) *J. Biomol. Struct. Dyn.*, **13**, 753–769.
Frishman,D. and Argos,P. (1995) *Proteins*, **23**, 566–579.
Gibbs,N., Sessions,R.B., Williams,P.B. and Dempsey,C.E. (1997) *Biophys. J.*, **72**, 2490–2495.
Guex,N. and Peitsch,M.C. (1997) *Electrophoresis*, **18**, 2714–2723.
Hobohm,U., Scharf,M. and Schneider,R. (1993) *Protein Sci.*, **1**, 409–417.
Hol,W.G. (1985) *Prog. Biophys. Mol. Biol.*, **45**, 149–195.
Hol,W.G., van Duijnen,P.T. and Berendsen,H.J.C. (1978) *Nature*, **273**, 443–446.
IUPAC–IUB (1970) *Biochemistry*, **9**, 3471–3479.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Kleywegt,G.J. and Jones,T.A. (1996) *Structure*, **4**, 1395–1400.
Kovacs,H., Mark,A.E., Johansson,J. and van Gunsteren,W.F. (1995) *J. Mol. Biol.*, **247**, 808–822.
Kumar,S. and Bansal,M. (1998) *Proteins*, **31**, 460–476.
Labesse,G., Colloc'h,N., Pothier,J. and Mornon,J.P. (1997) *Comput. Appl. Biosci.*, **13**, 291–295.
Lecerof,D., Fodje,M., Hansson,A., Hansson,M. and Al Karadaghi,S. (2000) *J. Mol. Biol.*, **297**, 221–232.
Lee,K.H., Benson,D.R. and Kuczera,K. (2000) *Biochemistry*, **39**, 13737–13747.
Low,B.W. and Baybutt,R.B. (1952) *J. Am. Chem. Soc.*, **74**, 5806.
Low,B.W. and Grenville-Wells,H.J. (1953) *Proc. Natl Acad. Sci. USA*, **39**, 785–801.
Mahadevan,J., Lee,K. and Kuczera,K. (2001) *J. Phys. Chem. B*, **105**, 1863–1876.
Pace,C.N. and Scholtz,J.M. (1998) *Biophys. J.*, **75**, 422–427.
Pauling,L. and Corey,R.B. (1951) *Proc. Natl Acad. Sci. USA*, **37**, 235–240.
Pauling,L., Corey,R.B. and Branson,H.R. (1951) *Proc. Natl Acad. Sci. USA*, **37**, 205–211.
Rajashankar,K.R. and Ramakumar,S. (1996) *Protein Sci.*, **5**, 932–946.
Ramachandran,G.N. and Sasisekharan,V. (1968) *Adv. Protein Chem.*, **23**, 283–437.
Richards,F.M. and Kundrot,C.E. (1988) *Proteins*, **3**, 71–84.
Richardson,J.S. and Richardson,D.C. (1988) *Science*, **240**, 1648–1652.
Rohl,C.A. and Doig,A.J. (1996) *Protein Sci.*, **5**, 1687–1696.
Schneider,T.D. and Stephens,R.M. (1990) *Nucleic Acids Res.*, **18**, 6097–6100.
Schubert,H.L., Raux,E., Wilson,K.S. and Warren,M.J. (1999) *Biochemistry*, **38**, 10660–10669.
Shirley,W.A. and Brooks,C.L.,III (1997) *Proteins*, **28**, 59–71.
Siegal,S. and Castellan,N.J. (1998) *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York, p. 38.
Weaver,T.M. (2000) *Protein Sci.*, **9**, 201–206.
Wu,C.K., Dailey,H.A., Rose,J.P., Burden,A., Sellers,V.M. and Wang,B.C. (2001) *Nature Struct. Biol.*, **8**, 156–160.

Received September 27, 2001; revised December 21, 2001; accepted January 28, 2002