

ω -Helices

Purevjav Enkhbayar¹, Bazartseren Boldgiv², and Norio Matsushima^{3,*}

¹Department of Biophysics and Bioinformatics, ²Department of Ecology, School of Biology and Biotechnology, National University of Mongolia, Ulaanbaatar-210646/377, Mongolia. ³Division of Biophysics, Center for Medical Education, Sapporo Medical University, Sapporo, Hokkaido 060-8556, Japan

ABSTRACT

The ω -helix is a modification of the α -helix. The α -helix has 3.6 residue in a single turn (3.6_{13} -helix), while the ω -helix (4.0_{13} -helix) has four residue per turn, although both are characterized by consecutive ($i \leftarrow i + 4$) hydrogen bonds. Until now the ω -helix has been identified only in synthetic polypeptides of derivatives of aspartic acid and cysteine. However, recent analysis of The Protein Data Bank by a total least-squares method for fitting a helix to data points (called the HELFIT program) demonstrated that right-handed ω -helices actually occur in proteins. In this review we describe structural feature of ω -helices in synthetic polypeptides and proteins and address functional roles of many ω -helices identified in proteins including *Chromobacterium violaceum* phenylalanine hydroxylase.

KEYWORDS: α -helix, ω -helix, 3_{10} -helix, π -helix, HELFIT, parahelix, non-planarity, aromatic-aromatic interactions

ABBREVIATION

PDB: Protein Data Bank

INTRODUCTION

Helices are a major type of secondary structure found in proteins. Helix types are usually designated

by n_m , where n is the number of residues per turn and m the number of atoms contained in the ring joined by a hydrogen bond [1]. The helices have intra strand hydrogen-bonds; a β -strand is also a 'helix' with $n = 2$. The α -helix has 3.6 residue in a single turn and $m = 13$. Thus, the α -helix is designated 3.6_{13} -helix, while the 3_{10} -helix and π -helix are designated 3.0_{10} -helix and 4.4_{16} -helix, respectively (Figure 1) [1-3]. The α -helix is characterized by consecutive ($i \leftarrow i + 4$) hydrogen bonds; the 3_{10} -helix and π -helix, by ($i \leftarrow i + 3$), and ($i \leftarrow i + 5$) H-bonds, respectively. Helix parameters - pitch (P), residues per turn (n), and radius (r) - of their canonical helices are summarized in Table 1.

The right-handed α -helix, 3_{10} -helix, and π -helix have been observed in protein structures consisting of only L-amino acids [4-12]. The α -helix is the most ubiquitous helical secondary structure (32% of residues) [4]. The 3_{10} -helix and π -helix account for 4% and 0.3%, respectively [5-7, 10]. Also left-handed α -helix, 3_{10} -helix, and polyproline II helix have been recognized [13-18]. In the 'transitional α -helix to 3_{10} -helix there is often a bifurcated H-bond(s) [19]. The ω -helix is a modification of the α -helix (Figure 1). The ω -helix has four residue per turn, and it, as well as the α -helix, is characterized by consecutive ($i \leftarrow i + 4$) H-bonds; thus it is also designated as a 4.0_{13} -helix. Until now both right-handed and left-handed ω -helix has been identified only in synthetic polypeptides of derivatives of aspartic acid and of cysteine [20-23]. We recently found that right-handed ω -helices also occur in proteins [24].

*Corresponding author
matusima@sapmed.ac.jp

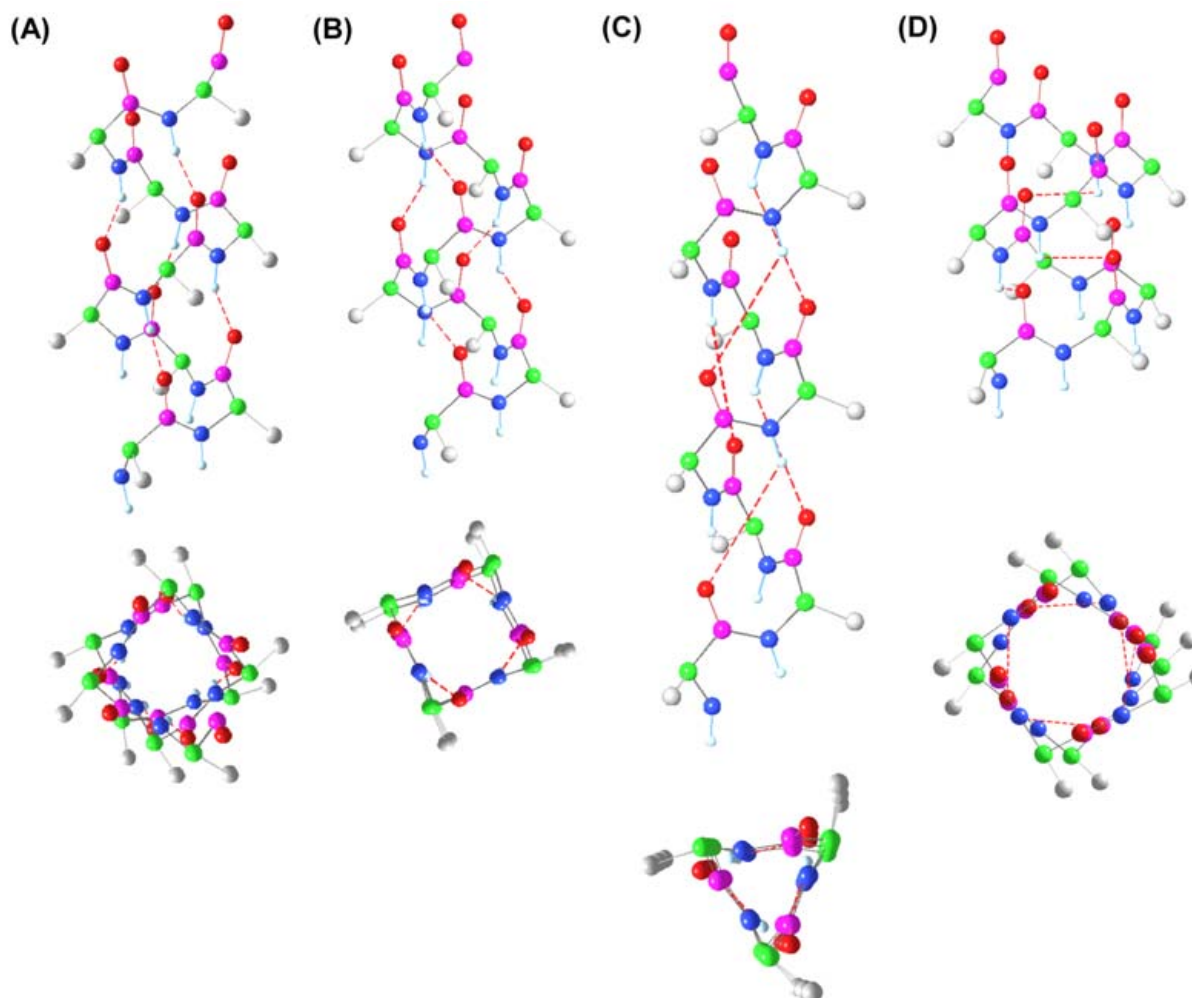


Figure 1. Canonical helical conformations. Side views and views along the axes of a helix in a canonical α -helix (A), canonical ω -helix (B), canonical 3_{10} -helix (C), and canonical π -helix (D), which have right handedness. α -Carbons are colored green, β -carbons grey, carboxyl-carbons purple, oxygens red, nitrogens blue, and hydrogens of the -NH group light blue. Dotted lines indicate hydrogen bonds between atoms in the backbone of polypeptide. The PDB files of helices were generated using PYMOL and their figures were prepared with MatLab.

Table 1. Helix parameters of canonical α -helix, ω -helix, 3_{10} -helix and π -helix with right handedness.

	ϕ^a ($^\circ$)	ψ^a ($^\circ$)	P^a (\AA)	n^a	Δz^a (\AA)	r^a (\AA)	V_c^a (\AA^3)
Canonical α -helix ^b	-67	-44	5.2	3.67	1.42	2.4	26.0
Canonical ω -helix ^c	-62.8	-54.9	5.30	4.0	1.325	2.54	26.8
Canonical 3_{10} -helix ^d	-49	-26	5.8	3.0	1.93	1.9	21.9
Canonical π -helix ^e	-57.1	-69.7	5.16	4.40	1.15	2.68	25.9

^a ϕ, ψ backbone dihedral angle; P , helix pitch; n , residue per turn; Δz , helix rise per residue; r , helix radius. $V_c = \pi r^2(\Delta z)$.

^bPauling *et al.* ^cBradbury *et al.*, torsion angle, $\omega = -175.4^\circ$. ^dPerutz. ^eDonohue, and Low and Grenville-Wells.

We have to consider critically a helical segment of polypeptide, because, in proteins, the helical length is not infinite. A helical segment of polypeptide chain may be defined either in terms of the backbone dihedral angles, ϕ and ψ (rule 6.2) or in terms of helical symmetry and H-bond arrangement (rule 6.3) [25]. A regular helix is potentially of infinite length, with all ϕ identical and all ψ identical. The ϕ , ψ of a canonical, right-handed α -helix are (-67° , -44°) that lie on the $n = 3.67$ line (Table 1). By rule 6.2 the first and last residues of the helical segment are those that have ϕ , ψ equal to those that define the helix [25]. In terms of the H-bond arrangement, rule 6.3, the first helical residue is taken as the first whose CO group is H-bonded to NH along the helix; the last helical residue is the last whose NH is H-bonded to CO along the helix [25]. By rule 6.3 the residue preceding the first (by rule 6.2) may have its CO in proper orientation to form H-bonds. The residue following the last (by 6.2) may have its NH in proper orientation to form H-bonds, thereby enabling the formation of an additional H-bond at each end. These rules prescribe no definitions for irregular helical segments. Here we employ rule 6.3. The first and the last residues of the helices have been called N-cap (N_c) and C-cap (C_c), respectively [24, 26]. By these conventions the residues of a helix six residues long ($N = 6$) would be labeled $-N_c, 2, 3, 4, 5, C_c$.

In this review first we describe structural features of ω -helices in synthetic polypeptides and proteins. Secondly, we address the role of ω -helices identified in regard to protein function.

ω -Helix in synthetic polypeptides

The ω -helix in synthetic polypeptides has been observed based on the x-ray diffraction patterns of their fibers. They are two esters of poly(L-aspartic acid) - poly(β -benzyl-L-aspartate) and poly(β -p-chlorobenzyl-L-aspartate) - and poly-S-benzyl-thio-L-cysteine - in which their side chains are $-\text{CH}_2\text{-COO-CH}_2\text{-C}_6\text{H}_5$, $-\text{CH}_2\text{-COO-CH}_2\text{-C}_6\text{H}_4\text{Cl}$ and $-\text{CH}_2\text{-S-S-CH}_2\text{-C}_6\text{H}_5$, respectively [20-22]. The ω -helix can be also detected from its characteristic infrared spectrum [27, 28] and ^{13}C NMR [29].

When heating solid films of poly(β -benzyl-L-aspartate) to 160°C and cooling, the X-ray diffraction

pattern was indexed on a tetragonal cell with $a = b = 13.85\text{\AA}$ and $c = 5.30\text{\AA}$. Only the fourth layer, together with the tetragonal packing, strongly indicates an ω -helix [20]. A model structure was proposed with a consideration of the infrared dichroism; it is a left-handed helix with dihedral angles of $\phi = 62.8^\circ$, $\psi = 54.9^\circ$, $\omega = 175.4^\circ$, which requires non-planarity in the peptide groups (Table 1) [20, 30]. Bradbury *et al.* [20] suggested that the energy necessary for the left-handed ω -helix of poly(β -benzyl-L-aspartate) comes from the close packing of the side chains benzene rings into four stacks tetragonally disposed around the 4.0_{13} helical core. Thus, poly(β -phenethyl L-aspartate) has been inferred to exist in a form similar to a right-handed ω -helix [28]. A right-handed ω -helix has also been inferred for poly(β -p-chlorobenzyl-L-aspartate) [22] and poly-S-benzyl-thio-L-cysteine [21].

The ω -helix content and helix sense have been studied in copolymers of L-aspartate esters [31, 32]. Poly(β -n-propyl L-aspartate) in chloroform solution undergoes a transition from right-handed to the left-handed helix form upon heating [31]. Although methyl L-aspartate forms a left-handed α -helix similar to benzyl-L-aspartate, the introduction of methyl L-aspartate residues prevents the formation of the ω -helix. Copoly(β -alkyl-L-aspartate- β -benzyl-L-aspartate) with a low degree of alkylation, in which the alkyl groups are ethyl, propyl, butyl, hexyl, nonyl, dodecyl, or stearyl, adopt the ω -helix at high temperatures of 50°C , while the copolyaspartates are in the α -helix at 20°C [32]. The helix-helix transition of poly(β -benzyl-L-aspartate) monolayers at air-liquid interfaces has been studied by the external reflectance Fourier transform infrared method [33].

In summary, synthetic polypeptides of derivatives of aspartic acid and cysteine prefer a left-handed or right-handed ω -helix and show a transition in structure from α -helix to ω -helix with increasing temperature in solution and in the solid state. These observations appear to be due to stacking interactions of the benzene rings of their side chains.

ω -Helices in proteins

We developed a software package called HELFIT that is a total least-squares method for fitting a

helix to data points (<http://biology.num.edu.mn/HELFIT/>) [34]. The HELFIT program simultaneously determines all of the helical parameters-helix axis, radius (r), pitch (P), and handedness with high accuracy. A parameter, p , is calculated

$$p = \frac{rmsd^h}{(N-1)^{1/2}} \quad (1)$$

where $rmsd^h$ is the root mean square distance from the best-fit helix to data points and N is the number of data points (non-hydrogen atom coordinates such as α -carbons (C_α 's) in proteins and amino acids); in a helix over its length. The p parameter estimates the regularity of helix independent of the length of helix. The minimum number required for the analysis is four contiguous C_α 's [34]. α -Helices and 3_{10} -helices with $p \leq 0.10$ Å are satisfactorily regular, because the p values indicate that the deviation of helical parameters such as r are less than 5% [26]. ω -Helices in proteins were successfully identified by the HELFIT program [24].

A structural analysis of 866 protein chains with low sequence identity ($\leq 20\%$) and solved at better than 1.6-Å resolution by the HELFIT program found a total of 1,496 regular α -helices 6-9 residues long with $p \leq 0.10$ Å [24]. Statistical analyses indicate that the frequency distribution of the helices versus residues per turn is bimodal at very high significance levels; the average residue per turn (\bar{n}) of the two distributions is $\bar{n}_\alpha = 3.6$ (corresponding to typical α -helix) and $\bar{n}_\omega = 4.0$ (corresponding to ω -helix), respectively [24]. In addition to $n \approx 4.0$, the other helical parameters - helix pitch (P) and helix radius (r) - were considered taking the ω -helical parameters observed in the synthetic polypeptides [24]. The regular α -helices with $3.85 < n < 4.15$, $5.0\text{Å} < P < 5.7\text{Å}$, and $2.45\text{Å} < r < 2.65\text{Å}$ were selected as regular ω -helices. 7.2% of the proteins and 0.2% of the 182,284 amino acids were regarded as right handed ω -helices (Figure 2 and Table 2) [24]. Sixty-two identified ω -helices are only short pieces that are

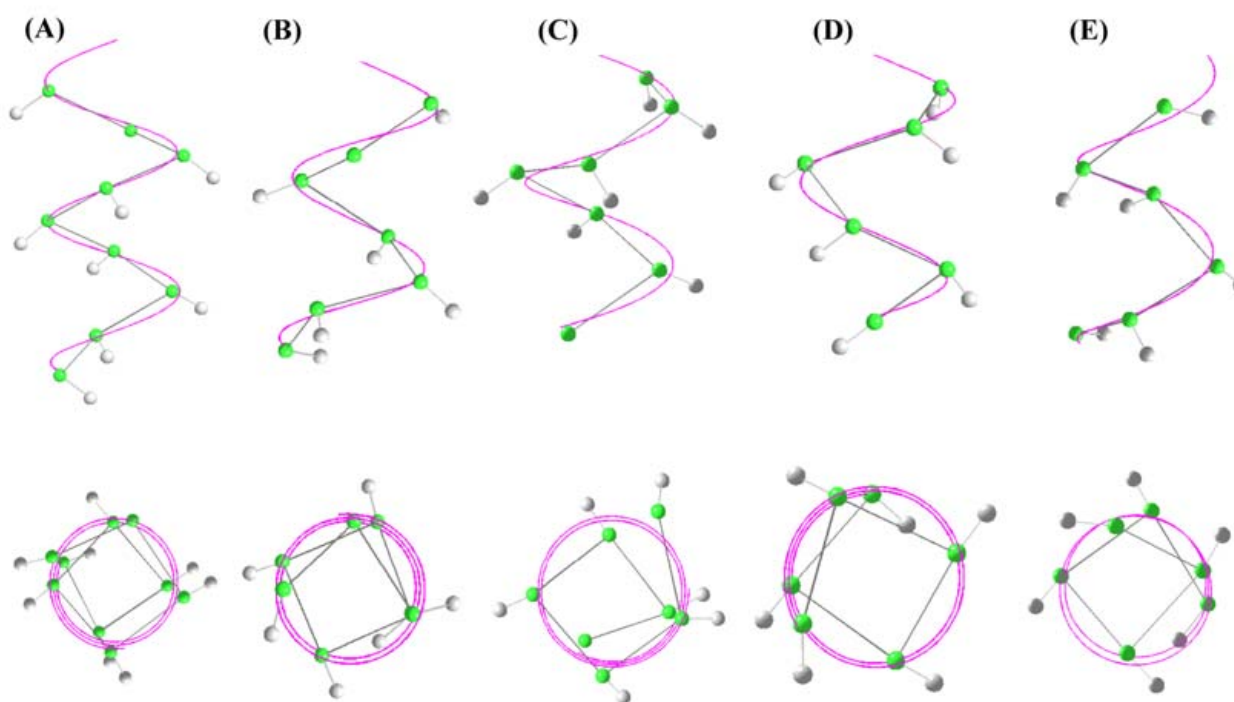


Figure 2. The best fitted-lines by the HELFIT program for ω -helices in five proteins. (A) [1LTZ]; (B) [1FP2]; (C) [1M55]; (D) [1XD3]; (E) [1S1D]. Side views and views along the axes of a helix in the five ω -helices. Best fitted lines by HELFIT are colored purple, α -carbons green and β -carbons grey. Their figures were generated using MatLab.

Table 2. Position and helix length (N) of identified ω -helices in proteins [24].

Length	PDB	Position	Length	PDB	Position
$N=9$	1LTZ	A_135-143	$N=6$	1C1D	A_283-288
$N=9$	1SQS	A_52-60	$N=6$	1RUT	X_56-61
$N=8$	1P0Z	A_84-91	$N=6$	1KQF	A_92-97
$N=8$	1QXY	A_54-61	$N=6$	1CXQ	A_123-128
$N=8$	1JET	A_298-305	$N=6$	1VKP	A_126-131
$N=8$	1N55	A_95-102	$N=6$	1PA7	A_92-97
$N=8$	1U7G	A_62-69	$N=6$	1XD3	A_38-43
$N=7$	1MC2	A_1016-1023	$N=6$	1VL1	A_171-176
$N=7$	1FP2	A_120-126	$N=6$	1NXM	A_41-46
$N=7$	1S3E	A_405-411	$N=6$	1RW1	A_32-37
$N=7$	1PO5	A_220-226	$N=6$	1UUY	A_476-481
$N=7$	1EB6	A_147-153	$N=6$	1FT5	A_167-172
$N=7$	1H80	A_50-56	$N=6$	1RKI	A_57-62
$N=7$	1J34	C_413-419	$N=6$	1T6C	A_48-53
$N=7$	1O8X	A_126-132	$N=6$	1ARB	_244-249
$N=7$	1UBK	S_112-118	$N=6$	2ENG	_81-86
$N=7$	1XMT	A_76-82	$N=6$	1U4G	A_221-226
$N=7$	1GPI	A_77-83	$N=6$	1RV9	A_238-243
$N=7$	1G5A	A_104-110	$N=6$	1S1D	A_238-243
$N=7$	1M55	A_148-154	$N=6$	1GWE	A_309-314
$N=7$	1UKK	A_33-39	$N=6$	1HBN	B_407-412
$N=6$	1QFT	A_7-12	$N=6$	1GQI	A_545-550
$N=6$	1THF	D_224-229	$N=6$	7ODC	A_276-281
$N=6$	1XSQ	A_9-14	$N=6$	1RYL	A_29-34
$N=6$	1RQW	A_154-159	$N=6$	1H1N	A_175-180
$N=6$	1LC0	A_163-168	$N=6$	1IKP	A_6-11
$N=6$	2AYH	_189-194	$N=6$	1V0W	A_189-194
$N=6$	1UWK	A_344-349	$N=6$	1MFM	A_133-138
$N=6$	1MJ4	A_55-60	$N=6$	1T9H	A_148-153
$N=6$	1IV3	A_56-61	$N=6$	1VJU	A_301-306
$N=6$	1A4I	A_257-262	$N=6$	1GK9	A_142-147

6 to 9 residues long. The average helix length is 6.4 residues, which corresponds to 1.6 turns. This length is comparable to the most common length of 3_{10} -helices and π -helices which are five (corresponding to 1.7 turns) [6, 26] and seven (1.5 turns) [10], respectively. The average helical parameters are $P = 5.25 (\pm 0.22 \text{ \AA})$, $n = 3.98 (\pm 0.10)$ and $r = 2.51 (\pm 0.06 \text{ \AA})$. The average values of backbone dihedral angles (ϕ , ψ) were found to be $(-75^\circ, -34^\circ)$ with standard deviations $(\sigma_\phi, \sigma_\psi) = (10^\circ, 12^\circ)$.

Are ω -helices in proteins parahelices?

The ω -helices in proteins show several structural features [24]. The average ϕ , ψ 's of ω -helices observed shift systematically from $\sim -56^\circ$, $\sim -42^\circ$ to $\sim -93^\circ$, $\sim -21^\circ$ for helices 6, 7, and 8 residues long. Significant departures from trans-planarity ($\omega \neq 180^\circ$) were observed in most of the ω -helices; the magnitude of $\Delta\omega = \text{sign}(\omega) - 180.0^\circ$, $|\Delta\omega|$, is $4.1 (\pm 3.7^\circ)$ on average. Moreover, energy minimizations suggest that the ω -helix is less stable than the canonical α -helix; they were

performed for regular, “average” right handed ω -helices and regular, “average” right handed α -helices at each helix length of oligo-Ala with hydrogen atoms, in which the ϕ , ψ 's and ω 's of “average” helices are set to observed position-specific, average values of the respective helix. The values of $\Delta E_T = E_T$ (“average” ω -helix) - E_T (“average” α -helix), in which E_T is total energy per residue, increase with increasing helix length. Consequently, the ω -helices in proteins are imperfect. It is inherently unfavorable for the ω -helices to form long, regular helices. Such helices are also observed in 3_{10} -helices in proteins; these are called parahelices [26]. We can conclude that both ω -helices and 3_{10} -helices in proteins are parahelices.

ω -Helices in proteins show amino acid preference for Asp, Cys, Gly, Ser, His, Tyr and Phe, while Gln and Met seem to be avoided [24]. Significantly, the preference for Asp and Cys is closely related with the experimental results that synthetic polypeptides of derivatives of these amino acids prefer ω -helices [24].

α -Helices are characterized by the repeating ($i \leftarrow i+4$) hydrogen-bond pattern, as noted. The electrostatic interaction energy between H-bonding groups was calculated by the DSSP program [35]. The interaction energy was used for the identification of main chain, hydrogen bonds; the cutoff value used for H-bond definition is -0.5 kcal/mol. The energies of ($i \leftarrow i + 4$) H-bonds for both ω -helices and α -helices observed in proteins as well as those of the canonical right-handed ω -helix and α -helix for comparison. The energy (E_H) of the first H-bond between N_c and 4 is lower than that of the last E_H between C_c-4 and C_c for both $N = 6, 7, 8$ and 9 ω -helices and the corresponding regular α -helices. Contrary to expectation, the first E_H for all of the ω -helices is significantly lower than that for the corresponding, regular α -helices.

In α -helices, the first four NH groups and the last four C=O groups necessarily lack intrahelical H-bonds. Instead, it is well known that such groups are often capped by alternative H-bond partners, as reviewed by Aurora and Rose [36]. Seven distinct capping motifs were identified - three at the N-terminus and four at the C-terminus

including the Schellman motif [37]. As expected, additional hydrogen bonds are observed between the backbone NH or C=O groups of residues on the ω -helix and the side chains of residues on the ω -helix. Furthermore, the side chains of residues on the ω -helix form H-bonds with residues outside the ω -helix or with bound water. Such additional hydrogen bonds are frequently observed in Ser (70%) and His (66%), according to the observation that these two residues are more prevalent. The conformation of the ω -helix with 4.0 residues per turn and unit twist of 90° implies that every fourth residue is in azimuthal position. Thus, the aromatic (i) - aromatic ($i + 4$) interaction does occur frequently (71%). In addition, the aromatic (i) - aromatic ($i + 1$) interaction also occur frequently (83%). This accords with the observation that aromatic amino acids are more prevalent. In the protein structures these aromatic side chains frequently interact in a herringbone conformation.

ω -Helix structure/function relationship in proteins

Among the 62 identified ω -helices, 21 (21/62 = 34%) appear to be important for the functions of the proteins [24]. Of these 21 ω -helices 17 are directly involved in an interaction with a ligand or form an active site in the enzymes and four have neighboring residues that interact directly with a ligand. The remaining four ω -helices are in homodimers and make a contact surface for dimerization. The following observations appear to suggest that ω -helices are flexible. This flexibility would be important for the functions of the proteins.

Coordination sites of metal ions

In four cases [1LTZ, 1KQF, 1PO5, and 1J34,], the side chains of His, Cys or γ -carboxyglutamic acid (Gla) in ω -helices form a site for coordination of metal ions such as iron, magnesium or calcium ions.

Phenylalanine hydroxylase (PheOH) is non-heme, iron(III)-dependent enzyme that catalyzes the hydroxylation of the aromatic L-phenylalanine (L-Phe) to L-tyrosine (L-Tyr) utilizing the cofactors pterin and dioxygen (O_2) [38]. Three crystal structures of PheOH from *Chromobacterium violaceum* reveal the presence of the longest ω -helix. The ω -helix is located on 135-

DVFHDLFGHV-144 ($N=10$) in both the apo (no iron bound) form (CvPheOH-apo) [1LTU] and in the oxidized Fe(III) form (CvPheOH-Fe(III)) [1LTV], and on 135-DVFHDLFGH-143 ($N=9$) in

the 7,8-dihydro-Lbiopterin plus Fe(III)-bound form (CvPheOH-Fe(III)-7,8-BH₂) [1LTZ] (Figure 3A and Table 3) [38]. Two ω -helix residues of His138 and His143 are included in active site. The iron atom

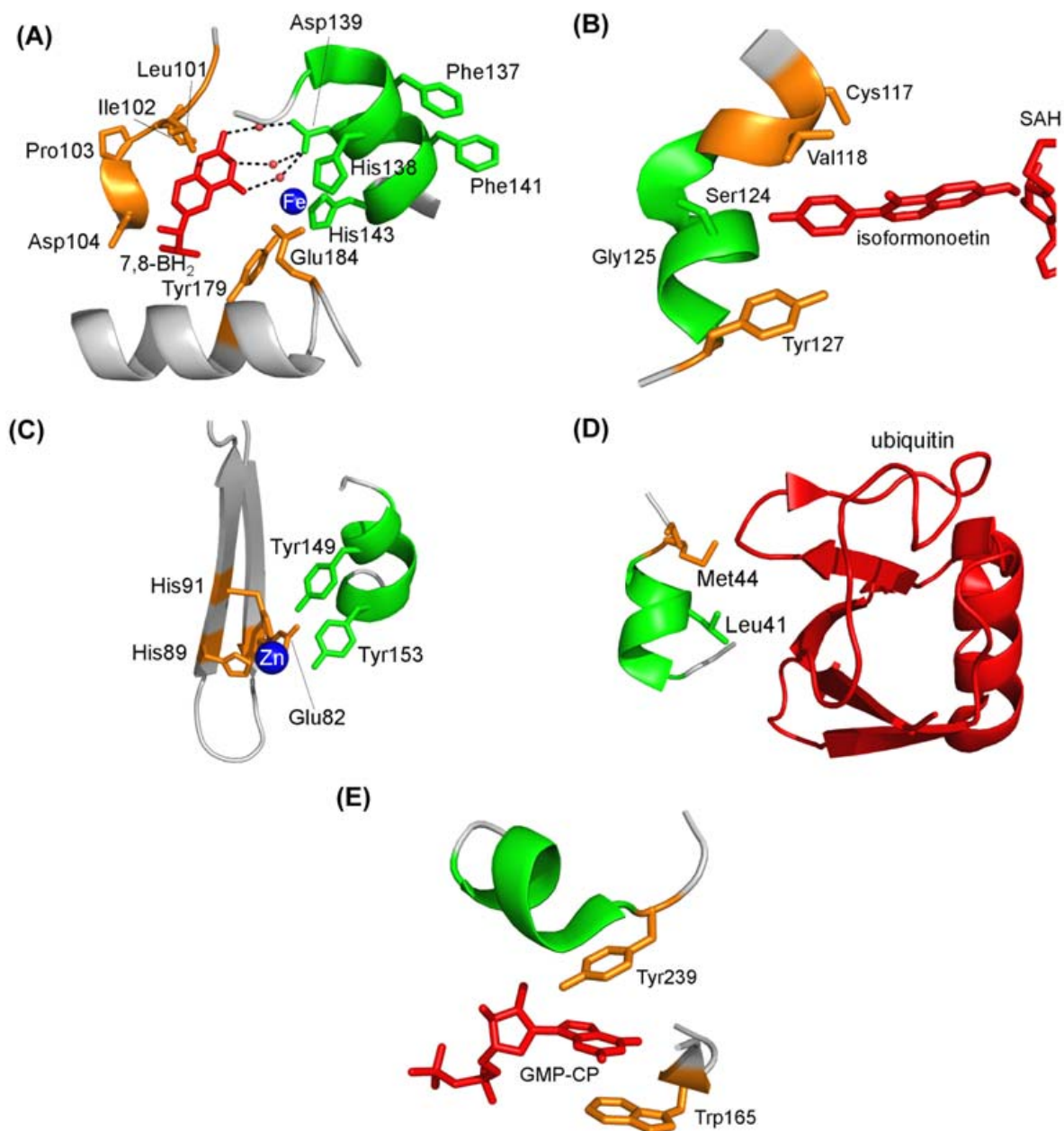


Figure 3. ω -Helices in five proteins. (A) phenylalanine hydroxylase [1LTZ]; (B) isoflavone O-methyltransferase (IOMT) [1FP2]; (C) the N-terminal domain of Replication (Rep) protein from Adeno-associated virus (AAV) [1M55]; (D) the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate 3 [1XD3]; (E) human apyrase complexed with the nucleotide guanosine α/β -methylene diphosphate (GMP-CP) [1S1D]. The ω -helices are colored green, metal ions blue, ligands or ubiquitin red, residues (that interact with the ω -helices, the metal ions or the ligands) brown, and others gray. All figures were prepared with PYMOL.

Table 3. Helix parameters of ω -helices identified in phenylalanine hydroxylase (PheOH) from human, *Chromobacterium violaceum*, and *Colwellia psychrerythraea*.

	PDB	Resolution (Å)	Position	Sequence	Helix parameters ^a				
					P (Å)	n	Δz (Å)	r (Å)	p (Å)
N=10	1LTU	1.74	A_135-144	DVFHDLFGHV	5.43	4.06	1.34	2.51	0.10
N=10	1LTV	2.00	A_135-144	DVFHDLFGHV	5.31	4.06	1.31	2.52	0.07
N=9	1LTZ	1.40	A_135-143	DVFHDLFGH	5.17	4.01	1.29	2.64	0.09
N=9	1J8U	1.50	A_282-290	DICHELLGH	5.42	3.96	1.37	2.52	0.09
N=9	1DMW	2.00	A_282-290	DICHELLGH	5.45	3.96	1.37	2.53	0.04
N=9	2V27	1.50	A_119-127	DIFHEIFGH	5.43	3.93	1.38	2.49	0.09

^a P , helix pitch; n , residue number per turn; Δz , helix rise per residue; r , helix radius; p , $(rmsd^h)/(N-1)^{1/2}$ where $rmsd^h$ is the root mean square distance from the best-fit helix to data points and N is the helix length.

has an octahedral coordination geometry, with ligation by His138, His143, Glu184, and three water molecules. The pteridine NH2, N3, and O4 atoms form hydrogen bonds to Asp139 in the ω -helix through three bridging water molecules. Homologous PheOH's from human and from *Colwellia psychrerythraea* also form an ω -helix with 282-DICHELLGH-290 [1J8U] and with 119-DIFHEIFGH-127 [2V27], which are the corresponding segment of CvPheOH (Table 3) [39, 40].

Formate dehydrogenase consists of three subunits (α -, β - and γ -subunits) and functions as a major electron donor during anaerobic respiration, when nitrate is used as an electron acceptor. The α -subunit that forms the active site includes one [4Fe-4S] cluster and molybdopterine-guanine dinucleotides [1KQF] [41]. The α -subunit contains an ω -helix with 92-CPKAGAG-97. Cys92 at N_c is bonded to the [4Fe-4S] cluster.

Cytochrome P450 is a group of heme-thiolate monooxygenases. Cytochrome P450 2B4 crystallizes as a dimer [42]. Residues 213-230 form close intermolecular hydrophobic interaction, and an intermolecular bond occurs between His226 and the heme iron [1PO5] [42]. His226 is located at N_c in an ω -helix with 220-FSGFLKH-226.

Factor IX is a pivotal participant in early stages of the blood coagulation cascade. Factor IX binds to a phospholipid membrane through a γ -carboxyglutamic acid (Gla) domain at the N-terminus. A Gla domain consists of 10-13 γ -carboxyglutamic acid residues and Ca²⁺ ions for stabilization of the active conformations for membrane binding [43].

Residues of 413-NLYRYCK-419 within the Gla domain form a ω -helix where Y is γ -carboxyglutamic acid [1J34] [43]. Two ω -helix residues -Y415 and Y417- coordinate Mg²⁺ and Ca²⁺, respectively.

Binding sites of substrates or ligands

Isoflavone O-methyltransferase (IOMT) is an S-adenosyl-l-methionine (SAM) dependent plant natural product methyltransferase in *Medicago sativa* (alfalfa). The crystal structure of IOMT in complex with the products S-adenosyl-l-homocysteine (SAH) and isoformononetin (4'-hydroxy-7-methoxyisoflavone) was refined to 1.4 Å [1FP2] [44]. Residues 117-127 with CVLDPTLSGSY form the substrate binding site, in which residues 120-126 adopt an ω -helix (Figure 3B). The Ser124 and Gly125 residues as well as Cys117, Val118 and Tyr127 bind directly with the substrate, 4'-dihydroxyisoflavone.

The structure of 2-C-methyl-2,4-cyclodiphosphate (MECDP) synthase that binds Mg²⁺ and the substrate, 2-phospho-4-(cytidine-5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME2P) has been determined [45]. The active sites of MECDP synthase are three large symmetry-related cavities stretching across two subunits of the trimer. The ribose ring of the substrate is held by Asp56' and Gly58' (prime indicating the neighboring subunit). As noted, the side-chain carboxyl hydrogen atom bonds to the main-chain N atom of Leu59. Asp56 and Gly58 are located at N_c and N_c+2 in an ω -helix with 56-DIGLLF-63, respectively [1IV3].

Triosephosphate isomerase (TIM) is a dimeric enzyme that catalyzes the interconversion of

D-glyceraldehyde 3-phosphate (DGAP) and dihydroxyacetone phosphate (DHAP) during glycolysis. The structure of *Leishmania mexicana* TIM complexed with the transition state analog, 2-phosphoglycolate (PGA), has been determined [46]. In the active site, PGA is bound in two conformations. In both of them, the O11 atom of PGA is hydrogen-bonded to O ϵ 2 of Glu167. In the predominant conformation (70% occupancy), O12 of PGA is at a hydrogen-bonding distance to N ϵ 2 of Lys13 and N ϵ 2 of His95; whereas in the minor conformation (30% occupancy), it points further down and is weakly hydrogen-bonded to N ϵ 2 of Asn11. The His95 residue is located at N ϵ in an ω -helix with 95-**HSERRTY**-102 [1N55].

Pyridoxal-5'-phosphate (PLP) dependent enzymes catalyze a broad range of reactions, resulting in bond cleavage at C α , C β , or C γ carbons of D and L amino acid substrates. Ornithine decarboxylase (ODC) is a PLP-dependent enzyme that controls a critical step in the biosynthesis of polyamines, small organic polycations whose controlled levels are essential for proper growth [47]. ODC contains an ω -helix with 276-**GRYYVA**-281. An ω -helix residue -Arg277- contributes a hydrogen bond to the phosphate via its N ϵ , while the N ϵ nitrogens are held in place by a salt bridge to Asp332. This salt bridge restrains Arg277 while interconnecting the barrel and sheet domains. A tightly bound water molecule, interacting with a carbonyl (Phe238) and a mainchain nitrogen (Tyr278), is found buried beneath the phosphate, hydrogen-bonded to O3P [7ODC].

LMO4 consists of two tandem LIM domains that are specialized zinc-fingers that bind two zinc atoms per domain. The LIM domains bind the LIM interaction domains that bind the LIM interaction domain (LID) of the nuclear adaptor protein LIM domain-binding protein-1 (Ldb1). The crystal structure of the LMO4:Ldb1 reveals that the proteins form a rod-shaped complex with Ldb1-LID bound in an extended manner across the entire length of both LIM domains of LMO4 to form a tandem β -zipper [48]. The LMO4 structure contains an ω -helix with 56-**QLGDIG**-61 in the first LIM domain [1RUT]. The side chains of Leu57 (an ω -helix residue) and Ile322 interact in Ldb1. Also the backbone -NH group of

Gly58 (an ω -helix residue) and the side chain of Asp59 (an ω -helix residue) form a hydrogen bond and electrostatic interactions, respectively, with the side-chain of Arg324.

Sulfite oxidase catalyzes the terminal step in the degradation of the sulfur-containing amino acids cysteine and methionine. The N-terminal cytochrome b₅ domain has a bis-imidazole-coordinated heme, which has an axial histidine coordination [49]. The heme ligand is additionally stabilized by a large aromatic side chain of Trp59 in the human cytochrome b₅ domain. Trp59 is within the ω -helix with 55-**LEPFWA**-60 [1MJ4]. Furthermore, the bulky residues Phe58 and Tyr62 block solvent access.

Cnx1 is involved in molybdenum cofactor biosynthesis and consists of the E-domain and the G-domain. The high-resolution crystal structures of the Cnx1 G domain (Cnx1G) in complex with molybdopterin and with adenylated molybdopterin (molybdopterin-AMP) has been determined [46]. Cnx1G contains a ω -helix with 476-**SDTVSA**-481 [1UUY]. Ser476 at N ϵ is in close proximity to the adenosine or the pyrophosphate bond [46].

H(A16-M) is an endo-1,3-1,4-P-D-glcans 4-glucanohydrolase (1,3-1,4- α -glucanase, lichenase). These enzymes cleave P-1, 4- glycosidic linkages that are adjacent to β -1,3-glycosidic linkages in mixed-linked glucans. The structure of *Bacillus* H(A16-M) contains an ω -helix with 189-**VDDWLG**-194 [2AYH] [50]. In the modeled β -glucan hexamer in the channel of H(A16-M), sugar residues interact with the protein through stacking type van der Waals' contact between Trp192 (an ω -helix residue) and the fifth glucan [50].

Active sites of enzymes

Replication (Rep) protein from Adeno-associated virus (AAV) is essential for viral replication and integration. Its amino terminal domain possesses site-specific DNA binding and endonuclease activities required for replication initiation and integration. The Rep endonuclease active site is established by the location of the active site tyrosine, Tyr153 in AAV-2 (Tyr156 in AAV-2) (51). The active site tyrosine is located in ω -helix of 148-**GYIPAYL**-154 (Figure 3C) [1M55] [51]. Two ω -helix residues of Tyr149 and Try153

forms an aromatic (*i*) - aromatic (*i* + 4) interaction. Moreover, His89 and His91 contribute two of the four ligands to a zinc ion which is located 5.7Å from the hydroxyl group of Tyr153; other two ligands are a side-chain carboxylate oxygen from Glu82 and another from Asp24.

Penicillin G acylase (PGA) hydrolyzes the side chain of penicillin G and related β -lactam antibiotics releasing 6-amino penicillanic acid (6-APA). The crystal structure of the α -subunit of PGA having two subunits (α and β) reveal the presence of an ω -helix with 142-MANRFS-147 [1GK9] [52]. In PGA complexed with the inhibitor penicillin G sulphoxide (PGSO), two ω -helix residues, Arg145 and Phe146, form the active site with Ser1, Phe71 and As241 in the β -subunit [1GM9] [52]. McVey *et al.* [52] proposed that these two aromatic residues - Phe146 in the α -subunit and Phe71 in the β -subunit- are the main contributors to substrate recognition and binding and indicated that Arg145 and Phe146 in the α -subunit and Phe71 in the β -subunit can adopt alternative positions, in response to the presence of ligands.

Human Cu, ZnSOD is a homodimeric enzyme of 153 residues per subunit that catalyzes the dismutation of two superoxide radicals to molecular oxygen and hydrogen peroxide. The crystal structure of the monomeric human SOD mutant F50E/G51E/E133Q reveal that residues 133-QSTKTG-138 form a ω -helix [1MFM]. The active site channel is part of residues 52-60 and residues 130-143 [53]. Thus, the ω -helix is involved in the active site channel. Residues 130-143 provide the electrostatic potential to drive the substrate to the reaction site [54].

Osmotically inducible protein C (PsmC) functions as a dimer. In the dimer structure the active site consists of Cys57 and Cys123 from one monomer and an Arg37 and Glu47 from the other [55]. Arg37 is within an ω -helix with 33-SYPSRFE-39 [1UKK]. Both residues are almost completely buried.

Phospholipase D catalyses the hydrolysis of phospholipids (normally phosphatidylcholine) into orthophosphatidic acid and polar head group (normally choline). Phospholipase D from *Streptomyces sp.* Strain PMF contain a ω -helix with 189-WKDDYL-194 [1V0W]. An ω -helix residue - Tyr193 - likely forms a part of active site region [56].

ω -helices on contact surface for dimerization

Proteins having ω -helices sometimes form homodimers. In three cases [1XD3, 1NXM, and 1VJU], the ω -helices form part of the contact surfaces for the dimerizations.

The crystal of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate has two complexes per asymmetric unit [57]. The UCH-L3 structure contains a ω -helix with 38-DPELLS-43 [1XD3] (Figure 3D) [57]. Leu41 (an ω -helix residue) and Met44 make contact with ubiquitin in only one of the two complexes. This observation indicates that the ω -helix is flexible. The crystal of dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase (RmlC) forms a homodimer. The part of an ω -helix with 41-QKEKML-46 interacts directly to each other [1NXM] [58]. Hypothetical protein from *Leishmania major* Lmaj006828 contains an ω -helix with 301-YFLTKR-36 [1VJU]. The monomers interact with each other through their respective ω -helices.

Ligand-binding sites in neighboring residues from ω -helices

In four cases [1S1D, 1H1N, 1GQI and 1LC0], neighboring residues from ω -helices interact directly with a ligand.

Apyrase hydrolyzes ATP and ADP equally well. Human apyrase complexed with the nucleotide guanosine α/β -methylene diphosphate (GMP-CP), a substrate analog, contains a ω -helix with 240-SEKDDE-245 [1S1D] (Figure 3E) [59]. The concentrically stacked rings of Tyr239 at N_c-1 (preceding Ser240 at N_c) and Trp165 cage the guanine ring of GMP-CP. In addition, the side chain of Tyr239 is oriented so as to allow its hydroxyl group to make bifurcated hydrogen bonds with the guanine ring and ribose of the ligand (OH-N3 = 2.9 Å and OH-2'OH = 2.5 Å).

Endoglucanase function as endohydrolysis of 1,4- β -D-glucosidic linkages in cellulose, lichenin and cereal β -D-glucans [1H1N]. The major endoglucanase from *Thermoascus aurantiacus* contains an ω -helix at 175-TWTNVN-180. An aromatic residue, Trp174 at N_c-1 (preceding Thr175 at N_c) is possibly involved in the binding of substrate [60].

α -Glucuronidases remove the α -1,2 linked 4-O-methyl glucuronic acid from xylans. In the structure

of the α -glucuronidases from *Pseudomonas cellulose*, 540-RADWT-544 and 545-AVYYHR-550R form a 3_{10} -helix and ω -helix, respectively [1GQI] [61]. Trp543 at Nc-2 forms a hydrophobic platform for the xylosyl unit at the nonreducing end upon binding xyloligosaccharide.

Biliverdin-IX α -reductase (BVR) reduces the γ -methene bridge of the open tetrapyrrole, biliverdin IX α , to bilirubin with the concomitant oxidation of a NADH or NADPH cofactor. In the structure, residues 163-168 form a ω -helix and residues 168-180 form an α -helix [1LC0]. Ser170 at C_c+2 and Arg171 at C_c+3 probably bind to the substrate [62].

CONCLUDING REMARKS

ω -Helices account for 0.2% of residues in proteins. The ω -helices are 6 to 9 residues long and show amino acid preference for Asp and Cys. There are systematic, position-specific shifts in the backbone dihedral angles and non-planarity of the peptide groups. The ω -helices in proteins often have additional hydrogen bonds of main chain and side chains of ω -helix residues and aromatic-aromatic interactions of the side chains of ω -helix residues. The ω -helices, as well as 3_{10} -helices, in proteins are not appropriately defined in terms of average, uniform dihedral angles; both are parahelices. Functional roles for many ω -helices in proteins have been identified.

ACKNOWLEDGMENTS

The authors thank Dr. Robert H. Kretsinger of the University of Virginia for his critical reading and valuable comments.

REFERENCES

1. Donohue, J. 1953, Proc. Natl. Acad. Sci. USA, 39, 470.
2. Pauling, L., Corey, R. B., and Branson, H. R. 1951, Proc. Natl. Acad. Sci. USA, 37, 205.
3. Perutz, M. F. 1951, Nature, 167, 1053.
4. Barlow, D. J. and Thornton, J. M. 1988, J. Mol. Biol., 201, 601.
5. Karpen, M. E., de Haseth, P. L., and Neet, K. E. 1992, Protein Sci., 1,1333.
6. Pal, L. and Basu, G. 1999, Protein Eng., 12, 811.
7. Pal, L., Chakrabarti, P., and Basu, G. 2003, J. Mol. Biol., 326, 273.
8. Vieira-Pires, R. S. and Morais-Cabral, J. H. 2010, J. Gen. Physiol., 136, 585.
9. Weaver, T. M. 2000, Protein Sci., 9, 201.
10. Fodje, M. N. and Al-Karadaghi, S. 2002, Protein Eng., 15, 353.
11. Riek, R. P. and Graham, R. M. 2010, J. Struct. Biol., 173, 153.
12. Cooley, R. B., Arp, D. J., and Karplus, P. A. 2011, J. Mol. Biol., 404, 232.
13. Novotny, M. and Kleywegt, G. J. 2005, J. Mol. Biol., 347, 231.
14. Hovmoller, S., Zhou, T., and Ohlson, T. 2002, Acta Crystallogr. Section D-Biol. Crystallogr., 58, 768.
15. Hollingsworth, S. A., Berkholtz, D. S., and Karplus, P. A. 2009, Protein Science, 18, 1321.
16. Martin-Pastor, M., De Capua, A., Alvarez, C. J. P., Diaz-Hernandez, M. D., Jimenez-Barbero, J., Casanueva, F. F., and Pazos, Y. 2010, Bioor. Med. Chem., 18, 1583.
17. Pal, L., Basu, G., and Chakrabarti, P. 2002, Proteins, 48, 571.
18. Adzhubei, A. A. and Sternberg, M. J. E. 1993, J. Mol. Biol., 229, 472.
19. Preissner, R., Egner, U., and Saenger, W. 1991, FEBS Letters, 288, 192.
20. Bradbury, E. M., Brown, L., Downie, A. R., Elliott, A., Fraser, R. D., and Hanby, W. E. 1962, J. Mol. Biol., 5, 230.
21. Fraser, R. D., Macrae, T. P., and Stapleton, I. W. 1962, Nature, 193, 573.
22. Takeda, Y., Iitaka, Y., and Tsuboi, M. 1970, J. Mol. Biol., 51, 101.
23. Fraser, R. D. B. and MacRae, T. P. 1973, Conformation in Fibrous Proteins. New York and London, Academic Press.
24. Enkhbayar, P., Boldgiv, B., and Matsushima, N. 2010, Protein Journal, 29, 242.
25. IUPAC-IUB Commission Biochemical, Nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. 1970, J. Mol. Biol., 52, 1.
26. Enkhbayar, P., Hikichi, K., Osaki, M., Kretsinger, R. H., and Matsushima, N. 2006, Proteins, 64, 691.
27. Bradbury, E. M., Crane-Robinson, C., Goldman, H., and Rattle, H. W. E. 1968, Biopolymers, 6, 851.

28. Hashimoto, M. and Arakawa, S. 1967, *Bull. Chem. Soc. Jpn.*, 40, 1698.
29. Saito, H., Tabeta, R., Ando, I., Ozaki, T., and Shoji, A. 1983, *Chem. Lett.*, 11, 1437
30. Takeda, Y. 1975, *Biopolymers*, 14, 891.
31. Bradbury, E. M., Carpenter, B. G., and Stephens, R. M. 1968, *Biopolymers*, 6, 905.
32. Tsujita, Y. 1988, *Biophys. Chem.*, 31, 11.
33. Riou, S. A., Hsu, S. L., and Stidham, H. D. 1988, *Biophys. J.*, 75, 2451.
34. Enkhbayar, P., Damdinsuren, S., Osaki, M., and Matsushima, N. 2008, *Comput. Biol. Chem.*, 32, 307.
35. Kabsch, W. and Sander, C. D1983, *Biopolymers*, 22, 2577.
36. Aurora, R. and Rose G. D. 1998, *Protein Sci.*, 7, 21.
37. Schellman, C. 1980, *The α L conformation at the ends of helices*. New York: Elsevier/North-Holland.
38. Erlandsen, H., Kim, J. Y., Patch, M. G., Han, A., Volner, A., Abu-Omar, M. M., and Stevens, R. C. 2002, *J. Mol. Biol.*, 320, 645.
39. Andersen, O. A., Flatmark, T., and Hough, E. 2001, *J. Mol. Biol.*, 314, 279.
40. Leiros, H. K. S., Pey, A. L., Innselset, M., Moe, E., Leiros, I., Steen, I. H., and Martinez, A. 2007, *J. Biol. Chem.*, 282, 21973.
41. Jormakka, M., Tornroth, S., Byrne, B., and Iwata, S. 2002, *Science*, 295, 1863.
42. Scott, E. E., He, Y. A., Wester, M. R., White, M. A., Chin, C. C., Halpert, J. R., Johnson, E. F., and Stout, C. D. 2003, *Proc. Natl. Acad. Sci. USA*, 100, 13196.
43. Shikamoto, Y., Morita, T., Fujimoto, Z., Mizuno, H. 2003, *J. Biol. Chem.*, 278, 24090.
44. Zubieta, C., He, X. Z., Dixon, R. A., and Noel, J. P. 2001, *Nat. Struct. Biol.*, 8, 271.
45. Kishida, H., Wada, T., Unzai, S., Kuzuyama, T., Takagi, M., Terada, T., Shirouzu, M., Yokoyama, S., Tame, J. R., and Park, S. Y. 2003, *Acta Crystallogr. Section D-Biol. Crystallogr.*, 59, 23.
46. Kursula, I. and Wierenga, R. K. 2003, *J. Biol. Chem.*, 278, 9544.
47. Kern, A. D., Oliveira, M. A., Coffino, P., and Hackert, M. L. 1999, *Structure with Folding & Design.*, 7, 567.
48. Deane, J. E., Ryan, D. P., Sunde, M., Maher, M. J., Guss, J. M., Visvader, J. E., and Matthews, J. M. 2004, *EMBO J.*, 23, 3589.
49. Rudolph, M. J., Johnson, J. L., Rajagopalan, K.V., and Kisker, C. 2003, *Acta Crystallogr. Section D-Biol. Crystallogr.*, 59, 1183.
50. Hahn, M., Keitel, T., and Heinemann, U. 1995, *Eur. J. Biochem.*, 232, 849.
51. Hickman, A. B., Ronning, D. R., Kotin, R. M., and Dyda, F. 2002, *Mol. Cell*, 10, 327.
52. McVey, C. E., Walsh, M. A., Dodson, G. G., Wilson, K. S., Brannigan, J. A. 2001, *J. Mol. Biol.*, 313, 139.
53. Ferraroni, M., Rypniewski, W., Wilson, K. S., Viezzoli, M. S., Banci, L., Bertini, I., and Mangani, S. 1999, *J. Mol. Biol.*, 288, 413.
54. Sines, J. J., Allison, S. A., McCammon, J. A. 1990, *Biochemistry*, 29, 9403.
55. Ramachandran, G. N. and Sasisekharan, V. 1968, *Adv. Protein Chem.*, 23, 283.
56. Leiros, I., McSweeney, S., and Hough, E. 2004, *J. Mol. Biol.*, 339, 805.
57. Misaghi, S., Galardy, P. J., Meester, W. J., Ovaa, H., Ploegh, H. L., Gaudet, R. 2005, *J. Biol. Chem.*, 280, 1512.
58. Dong, C. J., Major, L. L., Allen, A., Blankenfeldt, W., Maskell, D., and Naismith, J. H. 2003, *Structure*, 11, 715.
59. Dai, J. Y., Liu, J., Deng, Y. Q., Smith, T. M., and Lu, M. 2004, *Cell*, 116, 649.
60. Van Petegem, F., Vandenberghe, I., Bhat, M. K., and van Beeumen. 2002, *Biochem. Biophys. Res. Commun.*, 296, 61.
61. Nurizzo, D., Nagy, T., Gilbert, H. J., and Davies, G. J. 2002, *Structure*, 10, 547.
62. Whitby, F. G., Phillips, J. D., Hill, C. P., McCoubrey, W., and Maines, M. D. 2002, *J. Mol. Biol.*, 319, 1199.