Protein Engineering vol.14 no.5 pp.321–328, 2001

Side-chain conformations in 4-α-helical bundles

Vasiliki E.Fadouloglou[1,2], Nicholas M.Glykos[2] and Michael Kokkinidis[1,2,3]

[1]Department of Biology, University of Crete, P.O. Box 2208, GR-71409 Heraklion and [2]Foundation for Research and Technology–Hellas, Institute of Molecular Biology and Biotechnology (IMBB), P.O. Box 1527, GR-71110 Heraklion, Crete, Greece

[3]To whom correspondence should be addressed. E-mail: kokkinid@imbb.forth.gr

The distribution of the χ_1, χ_2 dihedral angles in a dataset consisting of 12 unrelated 4-α-helical bundle proteins was determined and qualitatively compared with that observed in globular proteins. The analysis suggests that the 4-α-helical bundle motif could occasionally impose steric constraints on side chains: (i) the side-chain conformations are limited to only a subset of the conformations observed in globular proteins and for some amino acids they are sterically more constrained than those in helical regions of globular proteins; (ii) aspartic acid and asparagine occasionally adopt rotamers that have not been previously reported for globular or helical proteins; (iii) some rotamers of tyrosine and isoleucine are predominantly or exclusively associated with hydrophobic core positions (*a*, *d*); (iv) mutations in the hydrophobic core occur preferentially between residue types which among other physicochemical properties also share a predominant rotamer.

Keywords: 4-α-helical bundles/χ-dihedral angles/rotamer/side chain

Introduction

The conformation of side chains is an essential feature of protein architecture. Consequently, knowledge of the factors that affect the side-chain conformations is significant, both for the understanding of protein folding and for the successful design of mutated proteins.

Side chains in proteins prefer certain conformations as shown by the non-uniform distribution of the χ-dihedral angles (Janin *et al.*, 1978). The preferred conformations correspond to energy minima that are generally represented by three regions of χ_1 (around 60°, 180° and –60°; Janin *et al.*, 1978). Analyses of the distribution of χ-dihedral angles by two groups (James and Sielecki, 1983; Ponder and Richards, 1987) led to the definition of a rotamer as a dense cluster of points in the χ-angle space; furthermore, a rotamer library was developed based upon 19 well-determined protein structures. The relationship between secondary structure and side-chain conformations was subsequently investigated (McGregor *et al.*, 1987; Summers *et al.*, 1987). These studies revealed that the rotamer preferences of side chains are strongly affected by the secondary structure. A significant correlation between the backbone φ,ψ values and the side-chain dihedral angles was also found (Dunbrack and Karplus, 1993). Based on a significantly enlarged data set, Schrauber *et al.* (Schrauber *et al.*, 1993) further refined the original rotamer library for globular proteins introduced by Ponder and Richards (Ponder and Richards, 1987). The influence of backbone conformations was taken into account by grouping the rotamer distributions for each amino acid according to several secondary structure-based classes. Furthermore, the term 'rotamericity' of an amino acid was introduced (Schrauber *et al.*, 1993), defined as the ratio of the total number of occurrences of the specific amino acid in any of the possible rotamers to the total number of occurrences of this amino acid in the sample.

In the present study, the role of the geometric constraints posed by a specific topology to the side-chain dihedral angles was investigated. As a model for protein topology, the 4-α-helical bundle motif was used. This simple, recurrent tertiary motif consists of four α-helices packed against each other in an antiparallel manner at an angle of about 20° (Figure 1a) (Weber and Salemme, 1980; Cohen and Parry, 1986). The α-helices are usually connected together with loop regions; alternatively, the bundle is formed as an assembly of helices belonging to different polypeptide chains, as is the case with the ColE1 Rop protein (Banner *et al.*, 1987; Presnell and Cohen, 1989; Harris *et al.*, 1994). The amino acid sequences of the helices follow a specific pattern of hydrophilic and hydrophobic residues of the type $(a,b,c,d,e,f,g)_n$ (Crick, 1953). This pattern is repeated every seven residues (heptads). Positions *a* and *d* form the core of the bundle and are generally occupied by hydrophobic amino acids (Figure 1b).

The amino acid frequencies for the seven topologically distinct positions of the heptad repeat revealed highly specific relationships between topology and sequence preferences (Paliakasis and Kokkinidis, 1992). These preferences reflect the constraints imposed by topology to the amino acid sequence. A question of interest is whether steric hindrances posed by the 4-α-helical bundle topology are reflected not only on the pattern of amino acid sequence but also in the conformations of side chains as expressed in terms of χ-angles. In this study, the distribution of χ_1, χ_2 dihedral angles found in a sample of 12 4-α-helical bundles is compared with the conformational preferences of side chains observed in globular proteins or specifically in α-helices. However, this work can only provide qualitative information since the small size of the sample makes it difficult to draw definite conclusions.

Materials and methods

To analyze the conformations of the χ-dihedral angles, an initial sample of 443 4-α-helical bundle structures (326 of which were lysozyme molecules) was obtained from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977; http://pdb-browsers.ebi.ac.uk//) and the FSSP database (Holm and Sander, 1996; http://www2.ebi.ac.uk/dali/fssp/fssp.html). To avoid bias due to sequence homologies, a subset of 12 structures with <25% sequence identity was used (Table I). The analysis was restricted to α-helical residues only (588 out of a total of 741 residues in the sample). Gly, Pro and Ala residues were excluded. Assignment of the heptad positions (*a*–*g*) in the

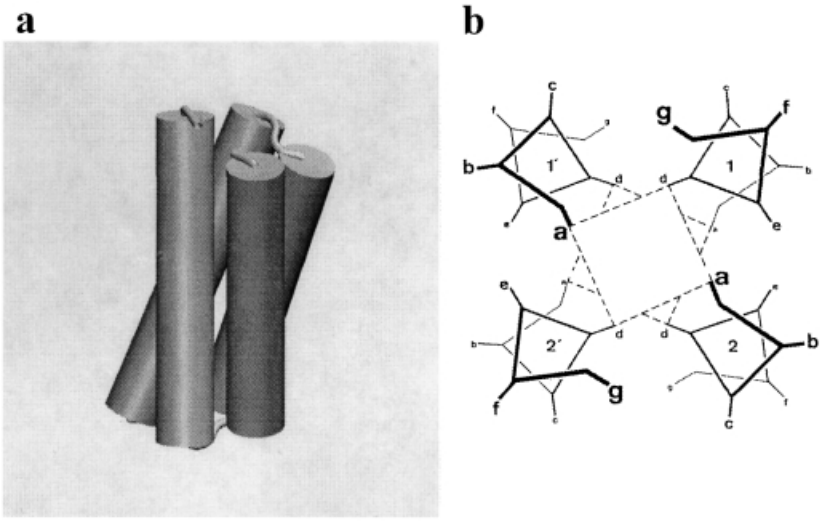

Fig. 1. (**a**) Schematic representation of the 4-α-helical bundle motif. Each cylinder represents an α-helix. (**b**) Transversal section of a canonical 4-α-helical bundle. The positions of the heptad repeat are marked with the letters *a*–*g*. The hydrophobic packing interactions among the residues of the core are illustrated with dotted lines.

Table I. The 4-α-helical bundles used for the analysis of the side-chain conformations

Protein	PDB entry	Helices that form the bundle (number and type of residue)				Resolution (Å)
		(I)	(II)	(III)	(IV)	
Cytochrome *c′*	1cpq	17L–28A	36A–47L	88M–99A	107F–118C	1.7
Apolipoprotein E_3	1nfn	26W–40V	56V–71L	101Q–115L	141L–155L	1.8
(Apo)ferritin	1rcc	12C–30Y	54R–72R	99A–117A	125M–144L	2.4
LIF[a]	1lki	26I–44Y	83L–101Q	111L–129V	162G–180F	2.0
Cytochrome b_{562}	256b	3L–17I	26V–40A	65F–79A	87A–101Y	1.4
Myohemerythrin	2mhr	21L–35C	44L–58E	73H–87L	96V–111D	1.7
Aspartate receptor	2asr	53L–64R	93A–104Y	123I–134L	154T–165F	2.3
Lysozyme	1dyg	95R–102M	114F–121L	129A–136S	146A–153F	2.1
Rop[b]	1rop	5E–29L	31A–56F	5E–29L	31A–56F	1.7
TMV coat protein	2tmv	23L–35F	41R–52W	76L–87F	121I–132L	2.9
R-HU-GCSF [c]	1rhg	17C–31L	78L–92L	103L–117I	154A–168L	2.2
Interleukin 2	3ink	14L–24I	53L–63L	86I–96L	118L–128I	2.5

[a]Leukemia inhibitor factor.
[b]The ColE1 Rop protein forms an α-helical bundle at the level of the dimer.
[c]Granulocyte colony-stimulating factor.

sequences of our sample was carried out manually by inspecting their structures with the program O (Jones *et al.*, 1991). This assignment was generally unambiguous because of the high regularity of the motif. Side-chain dihedral angles were also determined with the program O. Owing to the high variability of the torsion angles beyond χ_2, the analysis was restricted to χ_1, χ_2 only. For the classification of χ_1, χ_2 combinations, the conventions of Schrauber *et al.* (1993) were used, i.e. a side-chain conformation was assigned to a specific rotamer if the dihedral angles did not deviate by more than 20° from the values reported for this rotamer. To investigate whether the tightly packed hydrophobic cores impose special constraints to side chains, each amino acid type was analyzed according to its topological position in the bundle using a classification of residues into two groups. The first group consisted of internal residues (*a* and *d* positions) while the second group comprised more exposed residues (positions *b*, *c*, *e*, *f* and *g*).

Furthermore, the constraints imposed by 4-α-helical bundle cores on the evolution of their sequences were examined from the perspective of rotamer conservation. More specifically, the internal positions (*a* and *d*) in a series of aligned sequences of homologous 4-α-helical bundles were compared and the pattern of amino acid substitutions was examined from the aspect of rotamer conservation. To perform this analysis, representative sets of homologous sequences for eight of the proteins of our sample were identified using the Swiss Prot data bank (Bairoch and Apweiler, 1997; see also Table II) and the FASTA program (Pearson and Lipman, 1988). For each of the eight resulting groups, the sequence of the PDB structure was used as a reference. Within each group, the sequence alignment program ALIGN (http://www2.igh.cnrs.fr/bin/align-guess.cgi) was used to perform the pairwise alignments of all sequences with the reference sequence. Sequence homology combined with the characteristics of the heptad pattern allowed an unambiguous identification of the equivalent *a*, *d* positions within each group.

Results

Owing to the small size of the sample (see Materials and methods section), only qualitative information can be obtained

Table II. The eight groups of homologous sequences and their Swiss Prot entries

Protein	Organism or strain	Swiss Prot entry	Identity[a] (%)
Cytochrome *c′*	*Rhodobacter capsulatus*	CYCP_RHOCA	–
	Paracoccus sp.	CYCP_PARSP	27.4
	Rhodospirillum fulvum	CYCP_RHOFU	30.1
	Rhodocyclus gelatinosus	CYCP_RHOGE	32.1
	Rhodocyclus tenuis	CYCP_RHOTE	28.6
(Apo)ferritin	*Rana catesbeiana* (lower subunit)	FRI3_RANCA	–
	Rana catesbeiana (higher subunit)	FRI1_RANCA	64.2
	Homo sapiens (heavy chain)	FRIH_HUMAN	58.2
	Equus caballus (light chain)	FRIL_HORSE	50.3
	Homo sapiens (light chain)	FRIL_HUMAN	48.6
	Rattus norvegicus (light chain)	FRIL_RAT	48.1
Leukemia inhibitory factor (LIF)	*Mus musculus*	LIF_MOUSE	–
	Bos taurus	LIF_BOVIN	77.3
	Homo sapiens	LIF_HUMAN	79.8
	Rattus norvegicus	LIF_RAT	90.6
Myohemerythrin	*Themiste zostericola*	HEMM_THEZO	–
	Phascolopsis gouldii	HEM1_PHAGO	68.9
	Nereis diversicolor	HEMM_NERDI	60.8
Aspartate receptor	*Ecscherichia coli*	MCP2_ECOLI	–
	Salmonela typhimurium	MCP2_SALTY	78.7
Phage T4 lysozyme	*Bacteriophage T4*	LYCV_BPT4	–
	Bacteriophage T4	VG05_BPT4	43.2
	Cygnus atratus	LYG_CYGAT	20.3
TMV coat protein	TMV *Vulgare*	COAT_TMGMV	–
	TMV strain U2	COAT_TMV06	70.9
	TMV strain 06	COAT_TMVHR	98.1
	TMV strain Holmes Ribgrass (HR)	COAT_TMV	44.9

[a]Sequence identity between the first protein (which is the protein of known structure) and the other proteins of the group.

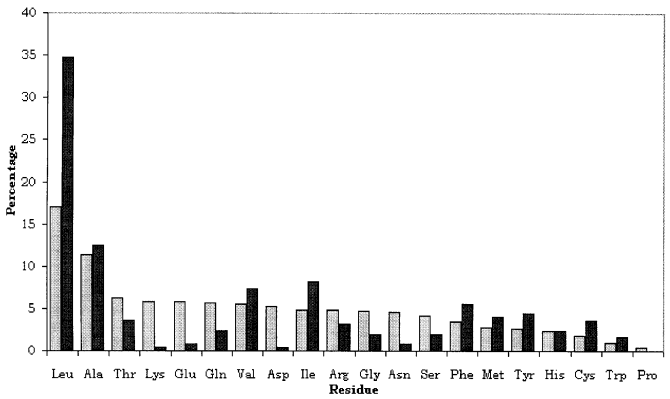

Fig. 2. Composition of the sample of 4-α-helical bundles. Gray bars, over all amino acids; black bars, composition of the hydrophobic cores (positions *a* and *d*).

by this work. Nevertheless, as far as we can assert by estimated standard deviations, our results (rotamer preferences, etc.) are generally consistent with the statistics of the much larger sample of globular proteins performed by other authors (Schrauber *et al.*, 1993).

Amino acid composition of the sample

The amino acid composition of the helical parts of the proteins used in this work is presented in Figure 2. Compared with an earlier analysis (Paliakasis and Kokkinidis, 1992) it shows only minor differences, i.e. (i) a higher occurrence of Leu and (ii) an increase in the Leu to Ala ratio. The composition of internal (*a*, *d*) positions (Figure 2) shows a clear predominance of Leu, Ala, Ile and Val residues, which agrees well with the preferences found by Paliakasis and Kokkinidis.

Rotamer preferences

The clusters of χ_1, χ_2 dihedral angles found for the side chains of our sample are presented in Table III. The most striking feature of this distribution for the majority of amino acid types is that a large fraction of the side chains belong to a single rotamer. Ile, Val, Thr and Cys are extreme examples, where the dominant cluster contains at least five times more members than the next one. On the other hand, Leu and Arg have two densely populated rotamers whereas Ser and Glu do not show pronounced preferences for a particular rotamer.

Table III also includes the rotamers reported for globular proteins by Schrauber *et al.* (Schrauber *et al.*, 1993) and Ponder and Richards (Ponder and Richards, 1987), and also those which are most frequently found in α-helical regions of globular proteins (in bold in Table III). Comparison between our sample and globular proteins shows that, as a rule, the preferred side-chain conformations of 4-α-helical bundles are limited to a small subset of the rotamers found in globular proteins. Furthermore, the side-chain conformations of Tyr, Met, Thr and Cys appear to be more constrained in our sample compared with α-helical regions of globular proteins because there is a strong preference for one out of several possible 'helical rotamers'. As shown in Table III, at least the 50% of the side-chain population of the above amino acid types adopt just one of the rotamers found in α-helical regions. A novel rotamer with $\chi_1 = 178°$ and $\chi_2 = 64°$ which to our knowledge has not been reported earlier, was found for Asp (see Table III and Figure 3a). Similarly, in the case of Asn, there are indications of a novel cluster with $\chi_1 = -170°$ and $\chi_2 = 59°$ (Table III and Figure 3b).

A possible interpretation of these findings could be that the 4-α-helical bundle motif imposes special constraints on side-

Table III. Summary of the rotamers found in 4-α-helical bundles and rotamers of globular proteins reported by other authors

Amino acid	Number	%	Rotamers for 4-α-helical bundles		Rotamer for globular proteins				
					Schrauber *et al.* (1993)[b,c]			Ponder & Richards (1987)	
			χ_1	χ_2	χ_1	χ_2	(%)	χ_1	χ_2
Phe	5	19	-74	84	**-69**	**98**	(32)	-66.3	94.3
(76.9%, 26)[a]	11	42	180	75	**-177**	**83**	(28)	-179.2	78.9
					62	89	(12)	66.0	90.7
	4	15	-66	136	-65	155	(5)		
					-80	200	(4)		
								-71.9	-0.4
Tyr	5	26	-70	107	**-65**	**98**	(34)	-66.5	96.9
(84.2%, 19)	11	58	176	78	**180**	**75**	(26)	-179.7	71.9
					59	90	(12)	63.3	89.1
					-74	167	(3)		
								-67.2	-1.0
His	1	6	-73	-91	**-61**	**-84**	(22)	-62.8	-74.3
(76.4%, 17)	4	24	179	66	**-182**	**75**	(13)		
	1	6	-76	78	**-70**	**97**	(12)	-69.8	96.1
	4	24	-172	-86	**-173**	**-94**	(7)	-175.2	-87.7
	3	18	-72	169	**-75**	**173**	(6)		
					68	-85	(5)	67.9	-80.5
					62	97	(4)	48.0	85.9
								-177.3	100.5
Ile	24	67	-67	170	**-62**	**-192**	(48)	-60.9	168.7
(83.3%, 36)	1	3	68	170	63	-190	(12)	61.7	163.8
	4	11	-62	-60	-58	-63	(10)	-59.6	-64.1
					-171	-192	(8)	-166.6	166.0
	1	3	-155	54	-163	67	(2)	-174.8	72.1
					61	100	(1)		
					-72	100	(1)		
Met	10	50	-74	180	**-72**	**-182**	(29)	-78.3	-174.7
(95.0%, 20)	2	10	-66	-59	**-62**	**-71**	(21)	-64.5	-68.5
	5	25	179	-176	**-177**	**-180**	(16)	178.9	179.0
					64	-178	(6)		
	2	10	-175	64	-169	77	(6)		
Leu	50	40	-72	171	**-64**	**-183**	(49)	-64.9	176.0
(87.3%, 126)	41	32	-179	61	**-177**	**65**	(24)	-176.4	63.1
					-96	38	(4)		
	11	9	-104	20	-108	31	(4)		
					-99	50	(4)		
	5	4	-123	-155	-143	-148	(2)		
					-116	-172	(2)		
					57	74	(1)	44.3	60.4
					-77	-54	(1)		
	4	3	-160	157				-165.3	168.2
Val	34	83	171		**177**		(67)	173.5	
(92.7%, 41)	3	7	-61		-60		(20)	-63.4	
	1	2	66		59		(5)	69.3	
					120		(2)		
Thr	39	85	-61		**-59**		(44)	-59.7	
(95.7%, 46)	5	11	56		**64**		(42)	62.7	
					-173		(7)	-169.5	
Cys	10	77	-71		**-65**		(56)	-65.2	
(92.3%, 13)	2	15	-156		**-175**		(24)	-179.6	
					64		(12)	63.5	
Ser	7	23	51		**69**		(40)	64.7	
(80.0%, 30)	8	27	-71		**-67**		(26)	-69.7	
	9	30	-177		**-177**		(20)	-176.1	
Asp[d]	22	56	-73	170				-68.3	154.3
(79.5%, 39)	2	5	-149	3				-169.1	3.9
	2	5	63	174		–		63.7	2.4
	6	15	178	64					

Table 3 continued

Amino acid	Number	%	Rotamers for 4-α-helical bundles χ_1	χ_2	Rotamer for globular proteins: Schrauber *et al.* (1993)[b,c] χ_1	χ_2	(%)	Ponder & Richards (1987) χ_1	χ_2
Asn[d]	16	48	-75	160				-68.3	143.2
(60.6%, 33)						–		-177.1	1.3
	4	12	-85	116				-67.2	128.8
								63.9	173.2
								-174.9	23.2
								63.6	53.8
	6	18	-174	65					
Gln	18	43	-70	176	**-67**	**-181**	(29)	-66.7	-178.5
(76.2%, 42)	4	10	-168	176	**-174**	**-183**	(14)	-174.6	-177.7
	4	10	-73	–67	-64	-72	(13)	-58.7	-63.8
	5	12	-170	71	**-176**	**69**	(8)	-179.4	67.3
					57	-179	(6)	70.8	-165.6
	1	2	-65	66	-73	81	(3)		
					-167	-90	(1)		
								-51.3	-90.4
								167.5	70.9
Glu	17	40	-74	171	**-68**	**-181**	(25)	-69.6	-177.2
(85.7%, 42)	11	26	177	175	**-176**	**-180**	(18)	-176.2	175.4
	7	17	-74	–66	**-66**	**-61**	(10)	-64.6	-69.1
					-68	82	(5)	-55.6	77.0
					65	-180	(4)	69.8	-179.0
					54	-84	(3)	63.0	-80.4
	1	2	-176	50	-171	63	(3)	-173.6	70.6
					-159	-84	(1)		
Arg	10	29	-72	176	**-71**	**-178**	(33)	-67.6	176.9
(71.4%, 35)	9	26	-174	–177	**-173**	**-180**	(16)	-174.1	-178.6
	3	9	-63	–59	-59	-82	(7)	-67.0	-71.7
					68	-177	(6)	80.0	175.6
	3	9	-179	67	-177	70	(4)	178.2	69.5
								57.1	82.8
								-76.9	54.2
Lys	17	40	-67	179	**-69**	**-181**	(25)	-68.9	-178.4
(76.2%, 42)	9	21	-178	–178	**-174**	**-178**	(18)	-172.1	175.3
	4	10	-65	–63	-56	-67	(7)	-58.1	-74.9
					69	-175	(4)	71.5	-174.3
	1	2	178	62	-185	72	(4)	173.4	83.4
					-155	-93	(1)		
	1	2	-89	82	-83	93	(1)		
								-175.8	-63.9
								-104.0	74.6

[a]Rotamericity, in %, as defined by Schrauber *et al.* (Schrauber *et al.*, 1993) and total number of occurrences of each amino acid in the sample. Owing to the small size of its sample, Trp has been omitted from the analysis.
[b]Rotamers observed by Schrauber *et al.* (Schrauber *et al.*, 1993) for globular proteins are listed in order of decreasing frequency of occurrence. Their frequencies are given in parentheses.
[c]Rotamers preferred by α-helical residues are given in bold.
[d]In this table, all the χ_2 angles for Tyr, Asp and Asn have been reduced in the range 0–180° for symmetry reasons.
[e]Schrauber *et al.* did not find specific rotamers for Asp and Asn (Schrauber *et al.*, 1993).

chain conformations. If this is indeed so, one would expect these constraints to be more pronounced for residues of the tightly packed hydrophobic core. For this reason, the residues occupying the hydrophobic core positions *a* and *d* of the heptad repeat were treated separately. For each amino acid type, the residues were distinguished in two groups according to their location in (*a*, *d*) and (*b*, *c*, *e*, *f*, *g*) positions and the χ_1, χ_2 plots were obtained. Inspection of these diagrams showed that for Ile and Tyr residues the two groups are segregated to different rotamers (Figure 4a and b). Two clusters which appear to be better compatible with internal residues (*a* and *d* positions) were found for these cases. These clusters. with $\chi_1 = –62°$, $\chi_2 = –60°$ for Ile and $\chi_1 = –70°$, $\chi_2 = 107°$ for Tyr, correspond to rotamers which are also present in the globular proteins, but which have not been reported yet in connection with protein cores.

Another useful indicator for assessing the influence of position-specific effects to the side-chain conformations is the rotamericity (Schrauber *et al.*, 1993). For each amino acid type, the rotamericities were calculated separately for the internal and external positions (data not shown). A tendency of the internal residues for higher rotamericity values is observed compared with the external positions.

Correlation between frequency of natural mutations and rotamer preferences in the hydrophobic core of the 4-α-helical bundles

Natural mutations in the hydrophobic core of 4-α-helical bundles were analyzed from the aspect of rotamer conservation

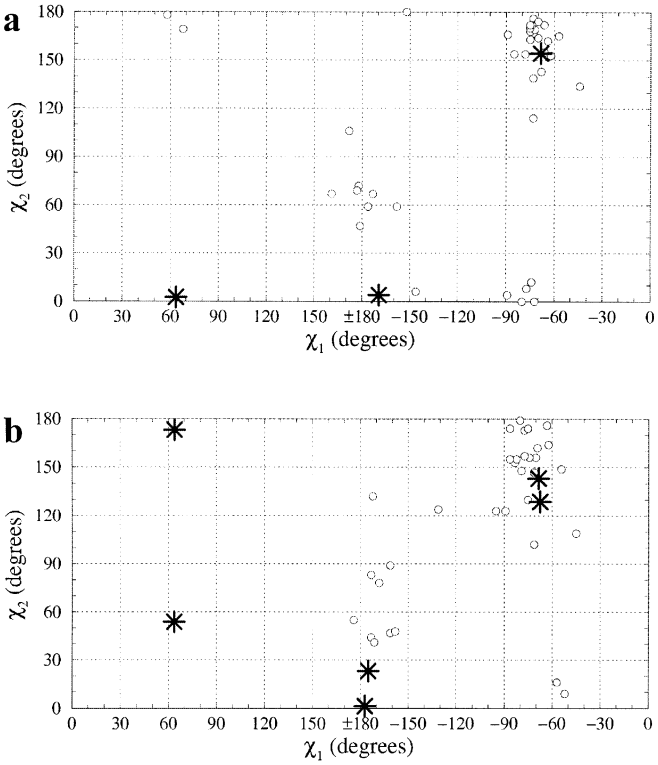

Fig. 3. Distribution of the χ_1, χ_2 dihedral angles for (**a**) Asp and (**b**) Asn residues. The circles represent the data points of our sample. Stars mark the positions of the rotamers listed in the Ponder and Richards rotamer library (Ponder and Richards, 1987) [no values were reported for Asp and Asn by Schrauber *et al.* (Schrauber *et al.*, 1993)]. The symmetry around the χ_2 angle is taken into account. Graphs were prepared using the program Xmgr (http://plasma-gate.weizmann.ac.il/Xmgr/).

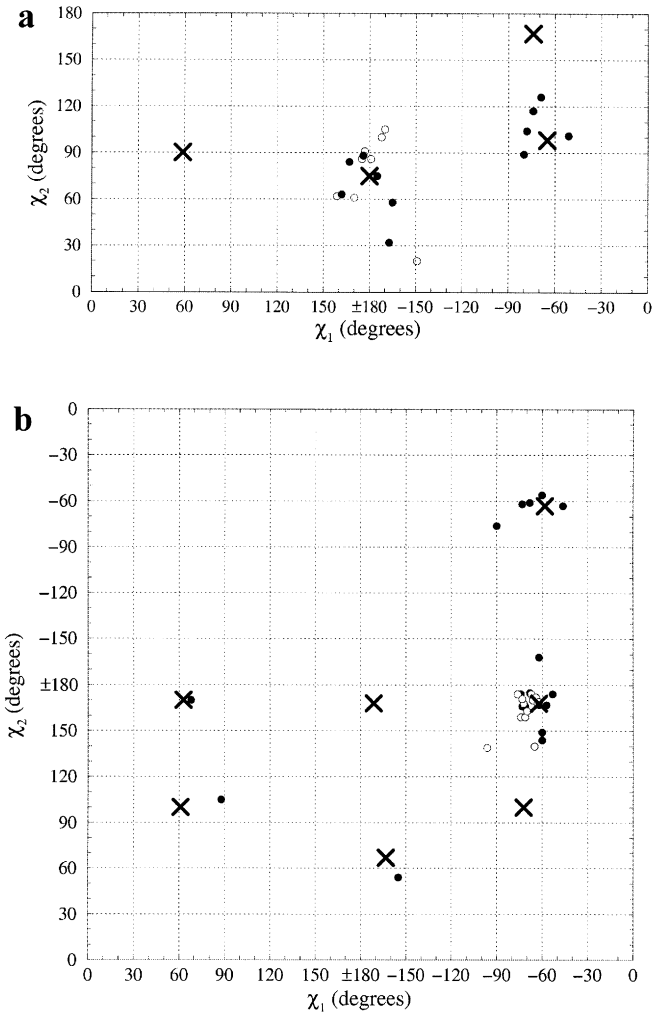

Fig. 4. χ_1, χ_2 angle distribution for (**a**) Tyr and (**b**) Ile residues. Internal positions (*a*, *d*) are indicated by black circles, semi-exposed or exposed residues are presented as white circles. Crosses denote rotamers for globular proteins (Schrauber *et al.*, 1993). For Tyr the symmetry around the χ_2 angle is taken into account.

[rotamer conservation in proteins has been reported (Summers *et al.*, 1987) and forms the basis of molecular modeling by homology]. In this analysis, the sequence alignments described in the Materials and methods section were examined and the pattern of sequence variation at the *a* and *d* positions was studied. From a total of 226 pairs of aligned amino acids in the core, 81 substitutions were found. In Table IV the observed substitutions and the corresponding frequencies for the residues commonly found in *a* and *d* positions are presented. An interesting observation is the relatively high frequency of Val→Leu and Met→Leu substitutions. It is worth noting that only four types of substitution account for 40% of the total observed (Met→Leu, Val→Leu, Leu→Ile and Leu→Val). Comparison of Tables III and IV shows that substitutions occur overwhelmingly between residue types that have their major rotamers in common. For example, the main rotamer of Ile and Met coincides with one of the main rotamers of Leu. As an exception to the above observations, Phe→Ile substitution occurs between residues that do not share any common rotamer while some relatively rare substitutions (e.g. Met→Phe, Tyr→Asp and Tyr→Met) occur between residues which do not share a major rotamer but have at least one less populated χ_1, χ_2 cluster in common. For example, the dominant rotamer of Tyr and Phe ($\chi_1 = 176°$, $\chi_2 = 78°$ and $\chi_1 = 180°$, $\chi_2 = 75°$, respectively) coincide with one of the 'rare' rotamers of Met ($\chi_1 = -175°$, $\chi_2 = 64°$).

Discussion

The conclusions of this work can be summarized as follows. (i) In addition to the known influence of α-helical secondary

Table IV. Substitutions observed in *a* and *d* positions of 4-α-helical bundles

From	To[a]	N[b]	F (%)[b]	S (%)[c]	From	To[a]	N[b]	F (%)[b]	S (%)[c]
Leu (75.0)	Ile	7	25.9	8.6	Tyr (66.7)	Phe	3	37.5	3.7
	Val	6	22.2	7.4		Leu	2	25.0	2.5
	Met	4	14.8	4.9		Met	2	25.0	2.5
	Thr	2	7.4	2.5		Asp	1	12.5	1.2
	Other	8	29.6	9.9	Phe (60.9)	Leu	4	44.4	4.9
Val (36.0)	Leu	11	68.8	13.6		Ile	2	22.2	2.5
	Ile	3	18.8	3.7		Val	1	11.1	1.2
	Tyr	2	12.5	2.5		Tyr	1	11.1	1.2
Met (29.2)	Leu	9	52.9	11.1		Asp	1	11.1	1.2
	Phe	4	23.5	4.9	Ile (81.8)	Leu	2	50.0	2.5
	Val	1	5.9	1.2		Val	2	50.0	2.5
	Gln	1	5.9	1.2					
	Ile	1	5.9	1.2					
	Other	1	5.9	1.2					

The percentage of conservation of each amino acid is given in parentheses.
[a]Substitutions that occur between two residue types with a frequency <5% and also substitutions that occur from a residue type to Ala or to Trp are classified in the category 'Other'.
[b]Absolute number (N) and frequency (F) in % of observed substitutions per residue type in the sample of aligned sequences.
[c]Fraction of the total substitutions observed in the sample.

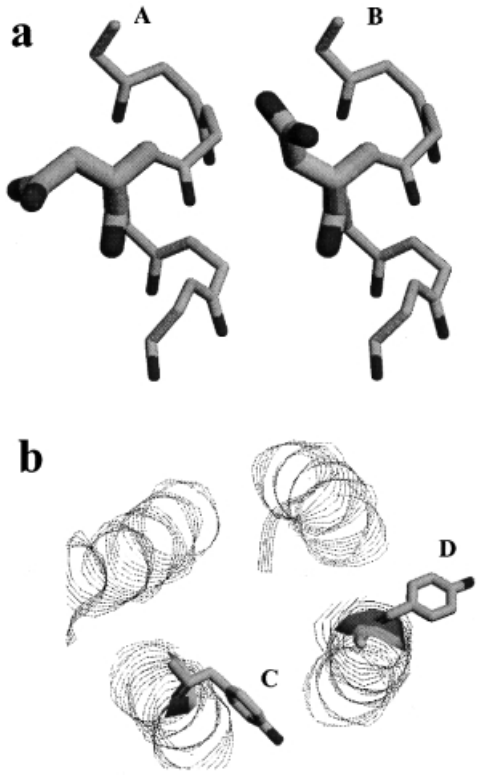

Fig. 5. Examples of side-chain conformations. (**a**) Asp residues (1nfn). A represents a motif-specific conformation (χ_1 = 177°, χ_2 = 69°) and B a non-specific conformation (χ_1 = –62°, χ_2 = 113°). (**b**) Tyr residues (1rhg). C represents a position-dependent rotamer (χ_1 = 159°, χ_2 = 62°) and D a position-independent rotamer (χ_1 = –78°, χ_2 = –76°). The figure was prepared using the programs O (Jones *et al.*, 1991) and Rasmol (Collaborative Computational Project Number 4, 1994).

structure on the side-chain conformations (Schrauber *et al.*, 1993), the 4-α-helical bundle topology probably imposes on some residues additional constraints and restricts the permitted conformations overwhelmingly to a unique rotamer. (ii) The side chains of aspartic acid and probably asparagine occasionally adopt a conformation that is associated with a novel rotamer. (iii) For isoleucine and tyrosine, rotamers were found which appear to be more compatible with residues in the hydrophobic core. (iv) Natural mutations in *a* and *d* positions of the 4-α-helical bundles tend to occur between amino acids which among other properties have their main rotamers in common.

Previous studies have demonstrated that the secondary structure significantly affects the side-chain conformations by restricting them to a subset of those found in globular proteins (McGregor *et al.*, 1987; Schrauber *et al.*, 1993). This is confirmed and reinforced by our analysis, which in addition showed that for some amino acids (Tyr, Met, Thr, Cys) the preference of specific side-chain conformations is much more pronounced even when compared to α-helices. This could be an effect of the 4-α-helical bundle topology, which imposes additional constraints to side chains limiting the permitted conformations to a subset of those adopted by α-helices.

Generally, the majority of side chains in the 4-α-helical bundles do not adopt novel or unusual conformations; side chains are clustered in some of the already known regions (from the globular proteins) of the χ_1, χ_2 space. A notable exception to this statement is the Asp residue, several occurrences of which have been classified as a novel rotamer (Figure 5a). This result was unexpected given that Asp is not subject to the specific constraints of the motif, since it systematically occurs in external or semi-buried positions. Indications of a new rotamer are also present for Asn. An analysis of all Asp and Asn residues in our sample shows that their temperature factors are fairly low and very close to or even lower than the average temperature factor of the structure to which they belong. Thus, the observed behavior is probably not an artifact due to an increased side-chain flexibility. The novel rotamers are not associated with a particular position of Asp/Asn residues (e.g. capping residues) or a particular backbone conformation (all Asp/Asn residues studied have helical ϕ,ψ angles). Detailed inspection and comparison with the known rotamers of the globular proteins, showed that the novel rotamer places the side chain in an optimal position relative to the protein backbone and the side chains of the neighboring α-helices, so that electrostatic repulsions between charged groups and steric hindrances between the side chains are minimized. However, the size of our sample is too small to go beyond a qualitative discussion of these novel rotamers.

A related observation is that Ile and Tyr exhibit a rotamer that is exclusively occupied by internal residues (positions *a* and *d*). Taking into account our limited data we could only assert that these conformations appear to be less well accepted by external residues; however, they are adopted by a significant fraction of the total number of internal residues (~50% for Tyr and 20% for Ile). For Ile, in particular, this position-specific conformation does not belong to the preferred rotamer of α-helices (Schrauber *et al.*, 1993) and is systematically adopted by residues located at the terminal caps of the bundles. However, when the individual amino acids that exhibit this special conformation were systematically examined in comparison with their local environments, no obvious reason that could provide an explanation for this positional dependence was found. On the other hand, for the case of Tyr it is clear that the position-dependent rotamer arranges the bulky, aromatic group out of the hydrophobic core of the bundle and unfavorable packing of the side chains is prevented (Figure 5b).

Side-chain shape, volume, polarity, packing density and cavity volume have been reported as factors that affect the pattern of substitutions in the protein interior (Bordo and Argos, 1990; Vlassi *et al.*, 1999). Conservation of these structural parameters is important for hydrophobic core mutations. Our present study provides evidence that the pattern of side-chain substitutions in 4-α-helix bundles is also consistent with the conservation of highly populated rotamers. In a previous study, Summers *et al.* (1987) concluded that for structurally and functionally homologous proteins, there is a high probability that the side-chain conformations are conserved. For amino acid substitutions, the conservation of orientation of both Cγ and Cδ atoms, was estimated to be in order of 35–75% (Summers *et al.*, 1987). In the case of 4-α-helical bundles, the frequency of the amino acid substitutions that occur between residues with at least one rotamer in common was estimated to be ~75% and indicates an extensive conservation of rotamers in the core of homologous bundles. This is consistent with recent work by Vlassi *et al.* (1999), which found a pronounced tendency of 4-α-helical bundles to preserve hydrophobic core packing interactions upon mutations. Exceptions to this behavior (e.g. Phe→Ile substitutions), where rotamer conservation is not possible, occur at the end of the bundles. It is reasonable to assume that the constraints of the motif at those positions are weaker.

The aim of this work, as mentioned, was to examine

qualitatively whether steric hindrances posed by the 4-α-helical bundle topology are reflected in the conformations of side chains. Within this framework, it has been shown that a tertiary motif can affect to some extent the conformations adopted by the side chains of some amino acids. This happens first through additional restrictions imposed to the side-chain conformations, second through the formation of novel χ_1, χ_2 clusters and third through some rotamers that appear to be more compatible with internal residues. A natural extension of this work would be a systematic analysis of additional known tertiary motifs. Such a study would provide valuable information for the understanding of protein folding and would find applications in the successful design of mutations and in the homology modeling of new structures.

References

Bairoch,A. and Apweiler,P. (1997) *J. Mol. Med*., **75**, 312–316.
Banner,D.W., Kokkinidis,M. and Tsernoglou,D. (1987) *J. Mol. Biol*., **196**, 657–675.
Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rogers,J.R., Kennard,O., Shimanouchi,T. and Tsumi,M. (1997) *J. Mol. Biol*., **112**, 535–542.
Bordo,D. and Argos,P. (1990) *J. Mol. Biol*., **211**, 975–988.
Cohen,C. and Parry,D.A.D. (1986) *Trends Biochem. Sci.*, **11**, 245–248.
Collaborative Computational Project Number 4 (1994) *Acta Crystallogr.*, **D50**, 760–763.
Crick,F.H.C. (1953) *Acta Crystallogr.*, **6**, 689–697.
Dunbrack,L.R. and Karplus,M. (1993) *J. Mol. Biol*., **230**, 543–574.
Harris,N.L., Presnell,S.R. and Cohen,F.E. (1994) *J. Mol. Biol*., **236**, 1356–1368.
Holm,L. and Sander,C. (1996) *Science*, **273**, 595–602.
James,M.N.G. and Sielecki,A.R. (1983) *J. Mol. Biol*., **183**, 299–361.
Janin,J., Wodak,S., Levitt,M. and Maigret,B. (1978) *J. Mol. Biol*., **125**, 357–386.
Jones,T.A., Zou,J-Y., Cowan,S.W. and Kjeldgaard,M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
McGregor,M.J., Islam,S.A. and Sternberg,M.J.E. (1987) *J. Mol. Biol*., **198**, 295–310.
Paliakasis,C.D. and Kokkinidis,M. (1992) *Protein Eng.*, **5**, 739–748.
Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
Ponder,J.W. and Richards,M. (1987) *J. Mol. Biol*., **193**, 775–791.
Presnell,S.R. and Cohen,F.E. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 6592–6596.
Schrauber,H., Eisenhaber,F. and Argos,P. (1993) *J. Mol. Biol*., **230**, 592–612.
Summers,N.L., Carlson,W.D. and Karplus,M. (1987) *J. Mol. Biol*., **196**, 175–198.
Vlassi,M., Cesareni,G. and Kokkinidis,M. (1999) *J. Mol. Biol*., **285**, 817–827.
Weber,P.C. and Salemme,F.R. (1980) *Nature*, **287**, 82–84.

Received September 14, 2000; accepted February 15, 2001