ELSEVIER

Biophysical Chemistry 111 (2004) 95–103

Biophysical Chemistry

www.elsevier.com/locate/bpc

Characteristic features of amino acid residues in coiled-coil protein structures

M. Michael Gromiha[a,*], David A.D. Parry[b]

[a] *Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), Aomi Frontier Building 17F, 2-43 Aomi, Koto, Tokyo 135-0064, Japan*

[b] *Institute of Fundamental Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand*

Received 5 March 2004; received in revised form 6 May 2004; accepted 6 May 2004
Available online 28 May 2004

Abstract

Detailed analyses of protein structures provide an opportunity to understand conformation and function in terms of amino acid sequence and composition. In this work, we have systematically analyzed the characteristic features of the amino acid residues found in α-helical coiled-coils and, in so doing, have developed indices for their properties, conformational parameters, surrounding hydrophobicity and flexibility. As expected, there is preference for hydrophobic (Ala, Leu), positive (Lys, Arg) and negatively (Glu) charged residues in coiled-coil domains. However, the surrounding hydrophobicity of residues in coiled-coil domains is significantly less than that for residues in other regions of coiled-coil proteins. The analysis of temperature factors in coiled-coil proteins shows that the residues in these domains are more stable than those in other regions. Further, we have delineated the medium- and long-range contacts in coiled-coil domains and compared the results with those obtained for other (non-coiled-coil) parts of the same proteins and non-coiled-coil helical segments of globular proteins. The residues in coiled-coil domains are largely influenced by medium-range contacts, whereas long-range interactions play a dominant role in other regions of these same proteins as well as in non-coiled-coil helices. We have also revealed the preference of amino acid residues to form cation–π interactions and we found that Arg is more likely to form such interactions than Lys. The parameters developed in this work can be used to understand the folding and stability of coiled-coil proteins in general.

Keywords: *B*′-factor; Hydrophobicity; Inter-residue contacts; Coiled coils; Conformational parameters; Cation–π interactions

1. Introduction

A classical coiled coil is a bundle of right-handed α-helices coiled about one another to form a left-handed superhelix. The hallmark of this structure is the distinctive packing of apolar amino acid side chains in the core of the bundle, in which a residue from one helix packs into a space surrounded by four side chains of the facing helix. Recently, the structures of several proteins containing coiled-coil domains have been solved at high resolution and these have been reviewed by Lupas [1].

Several experimental and theoretical investigations have been carried out to understand the structure and function of coiled coils and to predict the domains in proteins that have such a conformation. It has recently been shown, for example, that α-helical coiled-coil domains are ubiquitous in most of the members of collagen superfamily [2] and have an important role in assembly. Lupas et al. [3] have assessed those residues most likely to lie in coiled coils by comparing their flanking sequences with those from known coiled-coil proteins. Berger et al. [4] have also proposed a method, this time based on pairwise residue correlations, for discriminating α-helices in coiled coils from those not present in this conformation. In another approach, Wolf et al. [5] implemented a multidimensional scoring approach for identifying and distinguishing trimeric and dimeric coiled coils. Further, various other widely used methods for predicting the coiled-coil domains have been analyzed in detail by Lupas [6].

The packing of amino acid residues in a hydrophobic core plays an important role in stabilizing all protein structures. In the case of coiled coils, however, there is an underlying seven-residue (heptad) periodicity in the se-

* Corresponding author. Tel.: +81-3-3599-8046; fax: +81-3-3599-8081.
E-mail address: michael-gromiha@aist.go.jp (M.M. Gromiha).

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.05.001

quence that results in a characteristic packing of the apolar residues along the hydrophobic axis of the rope-like structure thus formed. For sequences displaying a discontinuity in their heptad repeats, Brown et al. [7] and Strelkov and Burkhard [8] have analyzed the consequential local distortion of the coiled-coil geometry. Liu et al. [9] investigated the role of core packing by determining the conformation and stability of coiled-coil mutants. Further, the stability of coiled coils has been studied with conformational sampling and energy minimization, molecular dynamics simulations and structural analysis of coiled coils [10–12]. Recently, Smith et al. [13] modelled α-helical coiled-coil interactions by taking in to account the full three-dimensional geometry of the molecules.

Structural analysis of coiled-coil proteins can provide an important insight into their folding stability. In our earlier work, we have systematically analyzed the characteristic features of amino acid residues in different folding types of globular and membrane proteins. These analyses were subsequently used successfully to predict the secondary structure, solvent accessibility, folding rates and stability of similar proteins [14–22].

In our current work, we have extended our studies to include coiled-coil proteins. Similarities and differences between coiled coils and globular/membrane proteins can readily be identified and characterized. From these analyses, we have derived numerical indices for each of the amino acid residues relating to their properties, conformational state, surrounding hydrophobicity and flexibility. As expected we found that the residues Ala, Glu, Lys, Leu and Arg all show a high preference for coiled-coil domains. Less expected, however, the results that the hydrophobic activity of residues in coiled-coil domains was significantly less than for residues in non-coiled-coil domains, and that coiled-coil domains are more stable than other regions in the same class of proteins. Further, the role of medium- and long-range interactions in stabilizing coiled coils has been delineated and the influence of cation–π interactions in coiled-coil structures has been estimated.

2. Materials and methods

2.1. Data set

A database of coiled-coil proteins was derived using published three-dimensional structures [1]. For each protein, only one structure was selected from the Protein Data Bank (PDB) for subsequent analysis. The criteria for selection were as follows:

(i) where a structure had been determined using both X-ray diffraction and NMR the former data set was preferred,
(ii) the X-ray structure having the highest resolution was used,
(iii) where only an NMR structure was available the minimum average structure was used, and
(iv) priority was given to proteins lacking ligands and mutants.

These criteria resulted in a set of 22 coiled-coil protein structures. Further, we have included two other structures, 1D7M and 1FE6, which have been reported recently. All 24 structures have been solved by X-ray diffraction methods. The PDB codes of the coiled-coil proteins used in the present study are, 1BGW, 1BMF, 1COS, 1D66, 1D7M, 1FE6, 1FOS, 1G6N, 1GCB, 1GCL, 1GCM, 1GRJ, 1HTM, 1HUP, 1KLN, 1LBI, 1MOF, 1PYI, 1ROP, 1RTM, 1SRS, 1SRY, 2TMA and 2ZTA. Details about the residues within the coiled-coil domains of each coiled-coil protein were obtained from Lupas [1]. Further, the residues in the coiled-coil domains have been verified with reference to the sequences in SWISS-PROT [23] and their respective three-dimensional structures in PDB [24]. Table 1 lists the coiled-coil domains in each of the proteins studied along with the number of residues in these domains. The total number of residues in each proteins used in these analyses are also listed.

2.2. Development of conformational parameters

A conformational parameter set for the 20 amino acid residues in coiled-coil proteins has been developed as follows: we have both computed the frequency of occur-

Table 1
List of coiled-coil domains in 24 coiled-coil proteins

PDB	N_t	N_{cc}	Coiled-coil domains
1bgw	680	44	1006–1027, 1127–1148
1bmf	122	108	1–44, 209–272
1cos	29	29	1–29
1d66	57	15	50–64
1d7m	101	101	243–343
1fe6	52	52	1–52
1fos	60	37	162–198
1g6n	200	21	112–132
1gcb	452	11	51–61
1gcl	31	31	1–31
1gcm	32	32	1–32
1grj	151	56	8–37, 46–71
1htm	114	66	40–105
1hup	141	21	88–108
1kln	595	49	549–572, 622–646
1lbi	296	19	338–356
1mof	53	33	46–78
1pyi	88	19	76–94
1rop	56	49	4–28, 32–55
1rtm	149	30	74–103
1srs	84	19	161–179
1sry	421	84	26–58, 67–98, 164–182
2tma	284	270	10–279
2zta	31	31	1–31

N_t: total number of residues; N_{cc}: number of residues in coiled-coil domains.

rence of amino acid residues in a coiled-coil domain (f_c) and in the protein as a whole (f_t). The conformational parameter (Conf) is calculated using the equation:

$$\mathrm{Conf}_i = (f_c)_i/(f_t)_i \quad (1)$$

where i represents each of the 20 amino acid residues. These parameters have been normalized using the ratio between the total number of residues in the coiled-coil domains and all those in the considered proteins.

2.3. Computation of surrounding hydrophobicity

The surrounding hydrophobicity for a residue was computed by the procedure described by us earlier [14,25,26]. Firstly, the residues in a protein molecule are represented by their α-carbon atoms. Using the C_α coordinates, a sphere of radius 8 Å is then fixed around each residue and the residues occurring in this volume are identified. It has been shown that the influence of each residue over the surrounding medium extends effectively only up to 8 Å [27] and this limit is sufficient to characterize the hydrophobic behavior of amino acid residues [25] and to accommodate both the local and nonlocal interactions [21]. Further, an 8 Å limit has been used in other studies, including that involving the folding rate of two-state proteins [19,28], protein stability upon mutations [17], thermal stability [18] and the transition state structures of two-state protein mutants [29]. Residues surrounding a given residue are assigned their respective hydrophobic indices [30,31] as obtained from thermodynamic experiment. The surrounding hydrophobicity of the central residue H_j, is thus taken to be the sum of these hydrophobic indices:

$$H_j = \Sigma\ n_{ij}h_i \quad (2)$$

where n_{ij} is the total number of surrounding residues of type i around the jth residue of the protein, and h_i is the experimental hydrophobic index of residue type i [30,31]. The computation was repeated for all the amino acid residues in all proteins. The surrounding hydrophobicity is the average parameter set for each of the 20 amino acid residues.

2.4. Estimation of medium- and long-range contacts

For each residue, we have computed the composition of residues that surround it within a sphere of radius 8 Å as described above. The composition of surrounding residues is analyzed in terms of the location at the sequence level. Contributions from residues less than three residues apart are treated as short-range contacts, those three or four residues apart are defined as medium-range contacts and those greater than four residues apart are treated as long-range contacts [21,32]. Short-range contacts mainly depict the influence of sequence and hence we have concentrated our studies on an analysis of medium- and long-range contacts.

2.5. Temperature factor of residues

The temperature factor of each residue in our data set of coiled-coil proteins has been evaluated by determining the average value of the normalized B-factors (i.e. B'-factors) for all of the atoms in that residue [22,33]. The use of B'-factors is justified because B-factor values in protein crystal structures vary not only as a result of genuine physical differences but also from the refinement strategies used [34,35]. Hence, it is difficult to compare B-factors from independent structures (there are usually large variations in their values even when X-ray structures obtained at high resolution are compared [33]). The B'-factor for an atom is calculated using the following equation:

$$B' = (B - \langle B\rangle)/\sigma \quad (3)$$

where B is the temperature factor given in the PDB file for the corresponding atom; and $\langle B\rangle$ and σ are the mean and standard deviation, respectively, for the temperature factors corresponding to the protein atoms.

2.6. Cation–π interactions in coiled-coil protein structures

We have obtained information about the residues that are involved in cation–π interactions in coiled-coil proteins using the program CAPTURE, developed by Gallivan and Dougherty [36], and which is available at http://capture.caltech.edu. This method employs an energy-based criterion to delineate cation–π interactions in protein structures and has been widely used for the analysis of cation–π interactions [37,38].

3. Results and discussion

3.1. Conformational parameters for the 20 amino acid residues in coiled-coil domains

The conformational parameters have been computed using Eq. (1) and the numerical indices for the coiled-coil domains are presented in Table 2. We found that the residues Ala, Glu, Lys, Leu and Arg have an above average preference for coiled-coil domains. This is in line with earlier data for two- and three-stranded coiled coils [39,40]. This result shows that the preference of residues in coiled coils is a complex combination of hydrophobic, positively and negatively charged residues. The aromatic residues show less preference for coiled coils whereas these residue types are among the most highly preferred for membrane spanning α-helices and strands in membrane proteins [15,41,42]. Further, the Chou–Fasman [43] method shows a high propensity for Phe and Trp in α-helical segments. Interestingly, Arg and Ser have a higher preference for α-helical coiled coils than for helical segments in globular proteins as well as those in globular and membrane proteins. Pro has the

Table 2
Conformational parameter for the 20 amino acid residues

Residue	N_{cc}	N_{oth}	N_t	Conf
Ala	127	237	364	1.22
Asp	62	162	224	0.97
Cys	4	43	47	0.30
Glu	187	237	424	**1.54**
Phe	17	104	121	0.49
Gly	23	188	211	0.38
His	16	63	79	0.71
Ile	69	194	263	0.91
Lys	133	214	347	1.34
Leu	180	317	497	1.26
Met	27	72	99	0.95
Asn	40	136	176	0.79
Pro	7	136	143	**0.17**
Gln	52	128	180	1.01
Arg	86	167	253	1.19
Ser	69	166	235	1.02
Thr	51	167	218	0.82
Val	45	174	219	0.72
Trp	4	37	41	0.34
Tyr	28	110	138	0.71

N_{cc} and N_{oth} are, respectively, the number of residues in coiled-coil domains and other regions, and N_t is the total number of residues in coiled-coil proteins.
The lowest and highest conformational parameters are bold.

lowest preference in coiled-coil proteins (0.17). It is, of course, well known that Pro is not a favored residue in an α-helical conformation unless located within the first turn of the α-helix [41,43–45].

We have further analyzed the relationship between the propensity of each amino acid residue in coiled coils and non-coiled-coil helices. Kawashima and Kanehisa [46] developed a database of amino acid indices (numerical values for the 20 amino acid residues) of various physical–chemical, energetic and conformational properties of amino acids. This (AAindex) database contains a collection of 437 amino acid indices including 18 helical propensities reported by various methods in the literature. We have computed the correlation between the helical propensity in a coiled coil (obtained in this work) and each of the helical propensities in non-coiled-coil helices (available in AAindex database), and the results are presented in Table 3. We observed that the correlation lies in the range of 0.5 to 0.7, indicating the similarities and differences between the propensity of each amino acid residue in coiled-coil and non-coiled-coil helices. This result is reasonable that the similarities might be due to the fact that both the segments are in α-helical conformation and the differences are due to coiled coils. As an example, we have plotted in Fig. 1 the propensity of 20 amino acid residues in coiled coils and typical Chou–Fasman parameters [43]. In this figure, the straight line indicates the correspondence between these two propensity scales. The residues, Glu, Leu, Asp, Thr, Tyr, Gly, Ile, Gln, Ala, Asn and Lys have similar behaviors in both coiled coil and non-coiled-coil helical segments whereas Trp, Phe and Met have significantly higher propensity for non-coiled-coil helices than for coiled coils.

Table 3
Correlation coefficient between propensity scale in coiled-coils and non-coiled-coil helical segments

Propensity scale	r	Reference
Beghin and Dirkx	0.54	[50]
Chou and Fasman	0.66	[43]
Geisow and Roberts	0.60	[51]
Isogai et al.	0.67	[52]
Kanehisa and Tsong	0.72	[53]
Levitt	0.63	[54]
Maxfield and Scheraga	0.68	[55]
Nagano	0.53	[56]
Palau et al.[a]	0.61	[57]
Palau et al.[b]	0.69	[57]
Prabhakaran	0.63	[58]
Qian and Sejnowski	0.62	[59]
Robson and Suzuki	0.63	[60]
Tanaka and Scheraga	0.68	[61]
O'Neil and DeGrado	0.58	[62]
Blaber et al.	0.55	[63]
Koehl and Levitt	0.59	[64]
Burgess et al.	0.54	[65]

[a] Data set of 44 proteins.
[b] Data set of 33 proteins.

3.2. Hydrophobic character of amino acid residues in coiled-coil proteins

Hydrophobicity plays an important role in stabilizing coiled-coil proteins [47,48]. The surrounding hydrophobicity indices for each of the 20 amino acid residues in coiled-coil domains and other regions of coiled-coil proteins have been

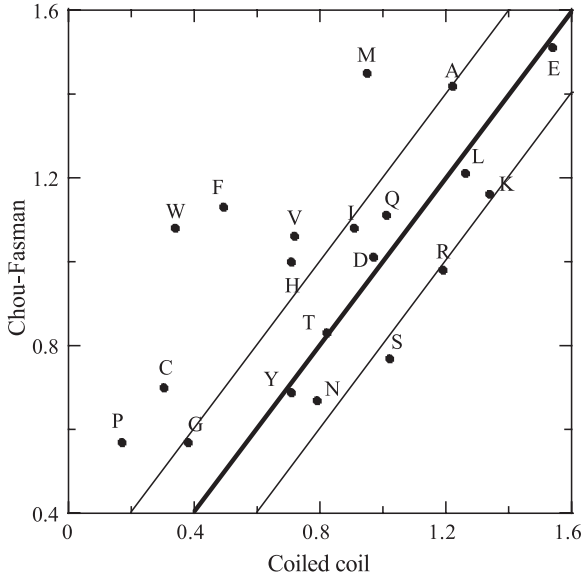

Fig. 1. Relationship between the propensities of 20 amino acid residues in coiled-coil domains and non-coiled-coil helical segments, reported in Chou and Fasman [43]. Solid line indicates the correspondence between these two propensity scales. Two thin lines accommodate the residues, which have a propensity difference of less than 0.2.

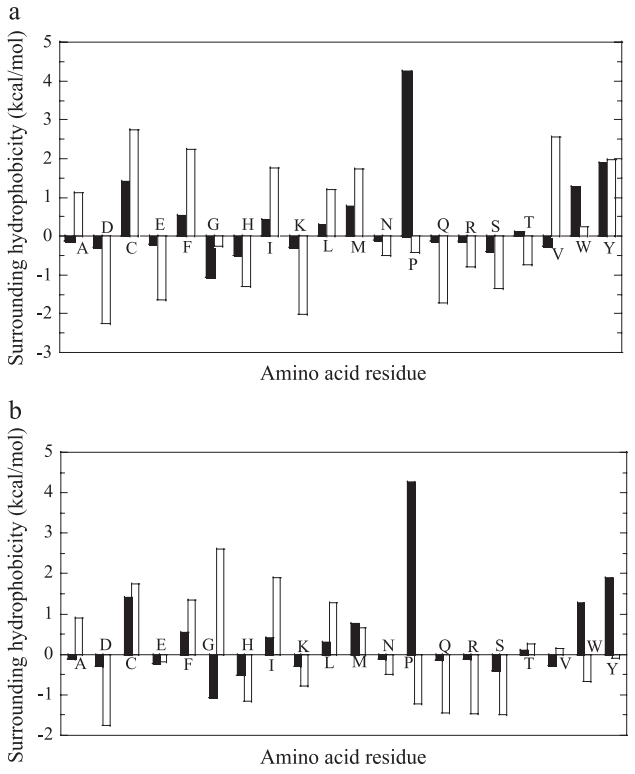

Fig. 2. (a) Surrounding hydrophobicity of the 20 amino acid residues in coiled-coil domains (shaded bars) and other regions (unshaded bars) of coiled-coil proteins. (b) Surrounding hydrophobicity of the 20 amino acid residues in coiled-coil domains (shaded bars) and non-coiled-coil helical segments (unshaded bars).

computed using Eq. (2) and the results are displayed in Fig. 2a. In this figure, the surrounding hydrophobicity of each residue was subtracted from the average value for comparison (10.3 and 12.8 kcal/mol, respectively, in coiled-coil domains and other regions). The former value is similar to that seen for the aqueous part of membrane proteins [42]. From Fig. 2a, we observed that the deviation of the surrounding hydrophobicity for all the amino acid residues (except Gly and Pro) from their average value is less in coiled-coil domains than for other regions of the coiled-coil proteins.

Residue-wise analysis shows that Pro has the highest surrounding hydrophobicity in coiled-coil domains whereas in other regions, it has a hydrophobicity value lower than average. In contrast, the aromatic residues, Phe, Trp and Tyr have higher surrounding hydrophobicity than average in both the coiled-coil domain and in other regions. Hydrophobic residues, Ala and Val have reduced hydrophobicity in coiled-coil domains. In contrast, Thr gained hydrophobicity in a coiled-coil conformation with a value higher than average.

Ponnuswamy et al. [49] computed the surrounding hydrophobicity of the 20 amino acid residues in helical segments of globular proteins as described in Materials and methods. The comparison between the surrounding hydrophobicity of amino acid residues in coiled-coil domains and non-coiled-coil helical segments is shown in Fig. 2b. We observed that Glu, Met and Thr have similar behaviors in both coiled-coil domains and non-coiled-coil helical segments. The hydrophobic character of Asp, His, Lys, Asn, Gln, Arg and Ser are less than average in both types of helices. However, the hydrophobic activity of Asp, Gln and Arg are highly reduced in non-coiled-coil helical segments. Cys, Phe, Ile and Leu have higher surrounding hydrophobicities than average in both coiled-coil and non-coiled-coil helical segments. Interestingly, Ala, Gly and Val are less hydrophobic in coiled coils and more hydrophobic in non-coiled-coil helical segments than their respective averages. Pro, Trp and Tyr have significantly higher surrounding hydrophobicities than average in coiled-coil domains but are less hydrophobic in non-coiled-coil helical segments.

3.3. Short-, medium- and long-range contacts in coiled-coil proteins

We have computed the short, medium- and long-range contacts in coiled-coil domains of our data set and the results are presented in Table 4. We observed that the average number of short-range contacts is close to four in all the coiled-coil domains. The average number of medium-range contacts varies from three to four in all these domains except in 1GCB. Considering the long-range contacts, 1GCB has an average of 3.6 contacts whereas the residues in other proteins have the average contacts between zero and two. Several proteins (67%) have less

Table 4
Medium- and long-range contacts in coiled-coil domains of 24 proteins

PDB	N_{cc}	$\langle N_s \rangle$	$\langle N_m \rangle$	$\langle N_l \rangle$
1bgw	44	4.000	3.886	1.591
1bmf	108	3.889	3.722	1.472
1cos	29	3.793	3.517	0.000
1d66	15	3.800	3.133	0.133
1d7m	101	3.941	3.861	0.020
1fe6	52	3.885	3.692	0.038
1fos	37	3.919	3.811	0.000
1g6n	21	4.000	3.905	0.476
1gcb	11	4.000	0.455	3.636
1gcl	31	3.806	3.548	0.000
1gcm	32	3.813	3.563	0.063
1grj	56	4.000	3.732	1.571
1htm	66	3.955	3.848	0.576
1hup	21	3.857	3.571	0.238
1kln	49	4.000	3.694	1.939
1lbi	19	3.947	3.526	0.158
1mof	33	3.909	3.727	0.121
1pyi	19	4.000	3.842	0.474
1rop	49	3.980	3.735	1.306
1rtm	30	3.967	3.733	0.200
1srs	19	4.000	3.789	0.684
1sry	84	4.000	3.845	1.381
2tma	270	3.993	3.978	0.026
2zta	31	3.806	3.548	0.000
Total	1227			
Average		3.945	3.759	0.599

N_{cc}: number of residues in coiled-coil domains; $\langle N_s \rangle$, $\langle N_m \rangle$ and $\langle N_l \rangle$ are, respectively, average number of short-, medium- and long-range contacts.

Table 5
Average number of medium- and long-range contacts for the 20 amino acid residues in coiled-coil domains and other parts of coiled-coil proteins

Residue	$\langle N_{mcc}\rangle$	$\langle N_{moth}\rangle$	$\langle N_{mdiff}\rangle$	$\langle N_{lcc}\rangle$	$\langle N_{loth}\rangle$	$\langle N_{ldiff}\rangle$
Ala	3.75	2.35	− 1.40	0.69	4.11	3.42
Asp	3.81	1.80	− 2.01	0.48	2.50	2.02
Cys	4.00	1.95	**− 2.05**	2.00	5.77	**3.77**
Glu	3.83	1.95	− 1.88	0.21	2.32	2.11
Phe	3.76	2.05	− 1.71	1.00	4.48	3.48
Gly	3.43	1.76	− 1.67	0.43	3.72	3.29
His	3.88	1.89	− 1.99	0.81	3.19	2.38
Ile	3.83	1.85	− 1.98	0.64	4.55	**3.91**
Lys	3.87	2.00	− 1.87	0.40	2.07	1.67
Leu	3.74	2.21	− 1.53	0.84	3.82	2.98
Met	3.63	2.17	− 1.46	1.15	4.19	3.04
Asn	3.75	2.00	− 1.75	0.30	2.85	2.55
Pro	2.86	1.29	− 1.57	1.71	3.40	1.69
Gln	3.81	2.20	− 1.61	0.46	2.48	2.02
Arg	3.67	2.28	− 1.39	0.45	3.01	2.56
Ser	3.70	1.85	− 1.85	0.46	3.09	2.63
Thr	3.69	1.65	− 2.04	0.86	3.67	2.81
Val	3.62	1.48	**− 2.14**	0.82	5.68	**4.86**
Trp	3.50	1.81	− 1.69	1.75	4.41	2.66
Tyr	3.82	1.62	**− 2.20**	1.50	4.88	3.38

$\langle N_{mcc}\rangle$ and $\langle N_{moth}\rangle$ are average number of medium-range contacts in coiled-coil domains and other regions, respectively. $\langle N_{mdiff}\rangle = \langle N_{moth}\rangle - \langle N_{mcc}\rangle$.
$\langle N_{lcc}\rangle$ and $\langle N_{loth}\rangle$ are average number of long-range contacts in coiled-coil domains and other regions, respectively. $\langle N_{ldiff}\rangle = \langle N_{loth}\rangle - \langle N_{lcc}\rangle$.
The topmost three differences in both medium- and long-range contacts are shown in bold.

than one long-range contact and 33% of the proteins have one to two long-range contacts.

On average, the residues in coiled-coil domains have an average of 8.3 contacts/residue and the average short, medium- and long-range contacts are, respectively, 3.95, 3.76 and 0.60 per residue. On the other hand, residues in other regions of coiled-coil proteins have an average of 9.5 contacts/residue. The contribution due to short, medium- and long-range contacts are, 3.97, 1.93 and 3.56, respectively, per residue. We noticed that the average number of medium-range contacts is appreciable in coiled-coil domains while that of long-range contacts are less in these domains.

We have compared the average short, medium- and long-range contacts in coiled-coil domains and non-coiled-coil helical segments. We observed that the contribution due to short and medium-range contacts are similar in both types of proteins and the number of long-range contacts in non-coiled-coil helical segments is higher than those in coiled-coil domains. The average number of short, medium- and long-range contacts in non-coiled-coil helices are, respectively, 3.98, 3.52 and 2.09, and the total number of contacts is 9.6 per residue.

3.4. Medium- and long-range contacts for the 20 amino acid residues

The inter-residue contacts between all pairs of amino acid residues in both coiled-coil domains and other regions of coiled-coil proteins have been computed and the average number of contacts for each of the 20 amino acid residues is presented in Table 5. We observed that each amino acid residue in the coiled-coil domain has an average of three to four medium-range contacts whereas residues in other regions have one to three contacts. Cys and Pro have the highest and lowest number of medium-range contacts, respectively, in coiled-coil domains. In other regions, Ala has the highest number of medium-range contacts followed by Arg. However, Tyr, Trp and Cys are the three residues, which display the greatest difference between the average number of medium-range contacts in a coiled coil and in other regions.

Considering the long-range contacts, each residue in the coiled-coil domain has an average of zero to two contacts and in the other regions, each residue has an average of three to five contacts. Cys and Val have the highest number of contacts in a coiled coil and other regions, respectively. Val, Ile and Cys are the residues which show the greatest difference between the number of long-range contacts between the residues in coiled-coil domains and other parts of coiled-coil proteins.

We have also calculated the medium- and long-range contacts for the 20 amino acid residues in the helical segments of non-coiled-coil proteins and the results are presented in Table 6. The comparison between the medium-range contacts in coiled-coil and non-coiled-coil helical segments showed that most of the residues have similar levels of contacts in both types of proteins. However, the negatively charged residues, Asp and Glu have more medi-

Table 6
Average number of medium- and long-range contacts for the 20 amino acid residues in coiled-coil domains and non-coiled-coil helical segments

Residue	$\langle N_{mcc}\rangle$	$\langle N_{mhel}\rangle$	$\langle N_{lcc}\rangle$	$\langle N_{lhel}\rangle$
Ala	3.75	3.53	0.69	2.43
Asp	3.81	3.26	0.48	1.24
Cys	4.00	3.85	2.00	3.53
Glu	3.83	3.32	0.21	1.30
Phe	3.76	3.62	1.00	2.50
Gly	3.43	3.43	0.43	2.78
His	3.88	3.67	0.81	1.51
Ile	3.83	3.71	0.64	2.62
Lys	3.87	3.49	0.40	1.53
Leu	3.74	3.68	0.84	2.62
Met	3.63	3.55	1.15	2.76
Asn	3.75	3.54	0.30	1.55
Pro	2.86	2.72	1.71	1.62
Gln	3.81	3.56	0.46	1.46
Arg	3.67	3.53	0.45	1.67
Ser	3.70	3.38	0.46	1.68
Thr	3.69	3.42	0.86	1.81
Val	3.62	3.63	0.82	2.76
Trp	3.50	3.74	1.75	2.13
Tyr	3.82	3.60	1.50	2.54

$\langle N_{mcc}\rangle$ and $\langle N_{mhel}\rangle$are average number of medium-range contacts in coiled-coil domains and non-coiled-coil helices, respectively.
$\langle N_{lcc}\rangle$ and $\langle N_{lhel}\rangle$ are average number of long-range contacts in coiled-coil domains and non-coiled-coil helical segments, respectively.

Table 7
Normalized temperature factors for the 20 amino acid residues in coiled-coil proteins

Residue	B'_{cc}	B'_{other}	$B'_{cc} - B'_{other}$
Ala	− 0.11	− 0.21	0.10
Asp	0.03	0.41	− 0.38
Cys	− 0.31	− 0.18	− 0.13
Glu	0.25	0.41	− 0.16
Phe	− 0.42	− 0.38	− 0.04
Gly	− 0.17	0.00	− 0.17
His	− 0.15	0.13	− 0.28
Ile	− 0.45	− 0.12	− 0.33
Lys	0.07	0.39	− 0.32
Leu	− 0.31	− 0.30	− 0.01
Met	− 0.06	0.09	− 0.15
Asn	− 0.07	− 0.01	− 0.06
Pro	0.91	− 0.06	0.97
Gln	− 0.01	0.28	− 0.29
Arg	0.28	0.33	− 0.05
Ser	0.02	0.15	− 0.13
Thr	− 0.43	− 0.15	− 0.28
Val	− 0.20	− 0.23	0.03
Trp	− 0.73	− 0.51	− 0.22
Tyr	0.05	− 0.44	0.49

B'_{cc} and B'_{other} are, respectively, the normalized temperature factor for the residues in coiled coils and other parts of coiled-coil proteins.

um-range contacts in coiled-coil domains than in non-coiled-coil helices. On the other hand, all the residues except Pro have less long-range contacts in coiled-coil domains than non-coiled-coil helical segments.

We have performed statistical tests to verify the significance of the results reported in the present study and we observed that the average values obtained in the analyses are within 95% confidence level.

3.5. Temperature factors for the amino acid residues in coiled-coil proteins

The normalized B'-factors for all the 20 amino acid residues in coiled-coil domains and other parts of coiled-coil proteins have been computed and the results are presented in Table 7. We observed that in coiled-coil domains Trp, Ile and Phe are highly stable whereas Pro is not. In other regions of coiled-coil proteins, the aromatic residues have the highest stability. Comparing the B'-factors of amino acid residues in coiled-coil domains and in other parts of coiled-coil proteins, we found that 16 residues were more stable in coiled-coil domains than in other regions.

3.6. Cation−π interactions in coiled-coil proteins

We have delineated the cation−π interactions in coiled-coil proteins using the program CAPTURE [36]. We observed that most of the potential cation−π interactions are not in the coiled-coil domains. Among the 40 cation−π interactions calculated for coiled-coil proteins, only one pair lies within the coiled-coil domain, 34 are in other regions and 5 lie between the residues in coiled coils and other regions. The fewer occurrences of cation−π interaction pairs in coiled-coil domains might be due to the lower preference of aromatic residues in coiled-coil domains (Table 2). We have further noticed that the Arg plays a dominant role in the contribution of cation−π interactions as observed also in globular and membrane proteins [36,38].

4. Conclusions

Detailed studies using the known three-dimensional structures of coiled-coil proteins have provided new insight on the folding and stability of these structures. The residues Ala, Glu, Lys, Leu, and Arg have a clear preference for coiled-coil domains, and this is in contrast to that for the aromatic residues. The hydrophobic character of amino acid residues in coiled-coil domains is less than that in other regions. However, Pro has a significantly high surrounding hydrophobicity when in a coiled-coil domain. The analysis of temperature factors revealed that 16 of the 20 amino acid residues are more stable in a coiled-coil conformation than they are in other regions of the proteins. The residues in coiled-coil domains are dominated by medium-range interactions whereas the long-range interactions are dominant in other regions as well as in non-coiled-coil helical segments. The results obtained in the present study will be useful for understanding the folding and stability of coiled-coil proteins and for predicting the occurrence of these domains in other proteins of unknown structure.

Acknowledgements

MMG would like to thank Drs. Yutaka Akiyama and Makiko Suwa for their encouragement.

References

[1] A. Lupas, Coiled coils: new structures and new functions, Trends Biochem. Sci. 21 (1996) 375–382.
[2] A. McAlinden, T.A. Smith, L.J. Sandell, D. Ficheux, D.A.D. Parry, D.J.S. Hulmes, Alpha-helical coiled-coil oligomerization domains are almost ubiquitous in the collagen superfamily, J. Biol. Chem. 278 (2003) 42200–42207.
[3] A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences, Science 252 (1991) 1162–1164.
[4] B. Berger, D.B. Wilson, E. Wolf, T. Tonchev, M. Milla, P.S. Kim, Predicting coiled coils by use of pairwise residue correlations, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 8259–8263.
[5] E. Wolf, P.S. Kim, B. Berger, MultiCoil: a program for predicting two- and three-stranded coiled coils, Protein Sci. 6 (1997) 1179–1189.
[6] A. Lupas, Predicting coiled-coil regions in proteins, Curr. Opin. Struck. Biol. 7 (1997) 388–393.
[7] J.H. Brown, C. Cohen, D.A.D. Parry, Heptad breaks in α-helical coiled coils: stutters and stammers, Proteins: Struct., Funct., Genet. 26 (1996) 134–145.

[8] S.V. Strelkov, P. Burkhard, Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation, J. Struct. Biol. 137 (2002) 54–64.
[9] J. Liu, W. Cao, M. Lu, Core side-chain packing and backbone conformation in Lpp-56 coiled-coil mutants, J. Mol. Biol. 318 (2002) 877–888.
[10] A.E. Keating, V.N. Malashkevich, B. Tidor, P.S. Kim, Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 14825–14830.
[11] M. Orzechowski, P. Cieplak, L. Piela, Theoretical calculation of the coiled-coil stability in water in the context of its possible use as a molecular rack, J. Comput. Chem. 23 (2002) 106–110.
[12] J. Walshaw, D.N. Woolfson, Extended knobs-into-holes packing in classical and complex coiled-coil assemblies, J. Struct. Biol. 144 (2003) 349–361.
[13] T.A. Smith, P.D. Hempstead, C.C. Palliser, D.A.D. Parry, Modeling alpha-helical coiled-coil interactions: the axial and azimuthal alignment of 1B segments from vimentin intermediate filaments, Proteins 50 (2003) 207–212.
[14] P.K. Ponnuswamy, M.M. Gromiha, Prediction of transmembrane helices from hydrophobic characteristics of proteins, Int. J. Pept. Protein Res. 42 (1993) 326–341.
[15] M.M. Gromiha, R. Majumdar, P.K. Ponnuswamy, Identification of membrane spanning beta strands in bacterial porins, Protein Eng. 10 (1997) 497–500.
[16] M.M. Gromiha, S. Selvaraj, Importance of long-range interactions in protein folding, Biophys. Chemist. 77 (1999) 49–68.
[17] M.M. Gromiha, M. Oobatake, H. Kono, H. Uedaira, A. Sarai, Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations, Protein Eng. 12 (1999) 549–555.
[18] M.M. Gromiha, Important inter-residue contacts for enhancing the thermal stability of thermophilic proteins, Biophys. Chemist. 91 (2001) 71–77.
[19] M.M. Gromiha, S. Selvaraj, Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction, J. Mol. Biol. 310 (2001) 27–32.
[20] S. Ahmad, M.M. Gromiha, A. Sarai, Real value prediction of solvent accessibility from amino acid sequence, Proteins 50 (2003) 629–635.
[21] M.M. Gromiha, S. Selvaraj, Inter-residue interactions in protein folding and stability, Prog. Biophys. Mol. Biol. (2004). doi:10.1016/j.pbiomolbio.2003.09.003.
[22] M.M. Gromiha, G. Pujadas, C. Magyar, S. Selvaraj, I. Simon, Locating the stabilizing residues in (α/β)8 barrel proteins based on hydrophobicity, long-range interactions and sequence conservation, Proteins: Struct. Funct. Bioinform. 55 (2004) 316–329.
[23] B. Boeckmann, A. Bairoch, R. Apweiler, M.-C. Blatte, A. Estreicher, E. Gasteiger, M.J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, Nucleic Acids Res. 31 (2003) 365–370.
[24] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[25] P. Manavalan, P.K. Ponnuswamy, Hydrophobic character of amino acid residues in globular protein, Nature 275 (1978) 673–674.
[26] P.K. Ponnuswamy, Hydrophobic characteristics of folded proteins, Prog. Biophys. Mol. Biol. 59 (1993) 57–103.
[27] P. Manavalan, P.K. Ponnuswamy, A study of the preferred environment of amino acid residues in globular proteins, Arch. Biochem. Biophys. 184 (1977) 476–487.
[28] D.A. Debe, W.A. Goddard, First principles prediction of protein folding rates, J. Mol. Biol. 294 (1999) 619–625.
[29] M.M. Gromiha, S. Selvaraj, Important amino acid properties for determining the transition state structures of two-state protein mutants, FEBS Lett. 526 (2002) 129–134.
[30] Y. Nozaki, C. Tanford, The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale, J. Biol. Chem. 246 (1971) 2211–2217.
[31] D.D. Jones, Amino acid properties and side-chain orientation in proteins: a cross correlation approach, J. Theor. Biol. 50 (1975) 167–183.
[32] M.M. Gromiha, S. Selvaraj, Influence of medium and long range interactions in different structural classes of globular proteins, J. Biol. Phys. 23 (1997) 151–162.
[33] S. Parthasarathy, M.R. Murthy, Analysis of temperature factor distribution in high-resolution protein structures, Protein Sci. 6 (1997) 2561–2567.
[34] H. Frauenfelder, G.A. Petsko, Structural dynamics of liganded myoglobin, Biophys. J. 32 (1980) 465–483.
[35] D. Ringe, G.A. Petsko, Study of protein dynamics by X-ray diffraction, Methods Enzymol. 131 (1986) 389–433.
[36] J.P. Gallivan, D.A. Dougherty, Cation–pi interactions in structural biology, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 9459–9464.
[37] M.M. Gromiha, S. Thomas, C. Santhosh, Role of cation–pi interactions to the stability of thermophilic proteins, Prep. Biochem. Biotechnol. 32 (2002) 355–362.
[38] M.M. Gromiha, Influence of cation–pi interactions in different folding types of membrane proteins, Biophys. Chemist. 103 (2003) 251–258.
[39] J.F. Conway, D.A.D. Parry, Structural features in the heptad substructure and longer range repeats of two-stranded α-fibrous proteins, Int. J. Biol. Macromol. 12 (1990) 328–334.
[40] J.F. Conway, D.A.D. Parry, Three-stranded α-fibrous proteins: the heptad repeat and its implications for structure, Int. J. Biol. Macromol. 13 (1991) 14–16.
[41] M.M. Gromiha, A simple method for predicting transmembrane alpha helices with better accuracy, Protein Eng. 12 (1999) 557–561.
[42] M.M. Gromiha, M. Suwa, Variation of amino acid properties in all-beta globular and outer membrane protein structures, Int. J. Biol. Macromol. 32 (2003) 93–98.
[43] P.Y. Chou, G.D. Fasman, Prediction of the secondary structure of proteins from their amino acid sequence, Adv. Enzymol. 47 (1978) 45–148.
[44] D.J. Barlow, J.M. Thornton, Helix geometry in proteins, J. Mol. Biol. 201 (1988) 601–619.
[45] M.M. Gromiha, P.K. Ponnuswamy, Prediction of protein secondary structures from their hydrophobic characteristics, Int. J. Pept. Protein Res. 45 (1995) 225–240.
[46] S. Kawashima, M. Kanehisa, AAindex: amino acid index database, Nucleic Acids Res. 28 (2000) 374.
[47] N.J. Greenfield, S.E. Hitchcock-DeGregori, The stability of tropomyosin, a two-stranded coiled-coil protein, is primarily a function of the hydrophobicity of residues at the helix–helix interface, Biochemistry 34 (1995) 16797–16805.
[48] J.R. Litowski, R.S. Hodges, Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity, J. Biol. Chem. 277 (2002) 37272–37279.
[49] P.K. Ponnuswamy, M. Prabhakaran, P. Manavalan, Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins, Biochim. Biophys. Acta 623 (1980) 301–316.
[50] F. Beghin, J. Dirkx, A simple statistical method to predict protein conformations, Arch. Int. Physiol. Biochim. 83 (1975) 167–168.
[51] M.J. Geisow, R.D.B. Roberts, Amino acid preferences for secondary structure vary with protein class, Int. J. Biol. Macromol. 2 (1980) 387–389.
[52] Y. Isogai, G. Nemethy, S. Rackovsky, S.J. Leach, H.A. Scheraga, Characterization of multiple bends in proteins, Biopolymers 19 (1980) 1183–1210.
[53] M.I. Kanehisa, T.Y. Tsong, Local hydrophobicity stabilizes secondary structures in proteins, Biopolymers 19 (1980) 1617–1628.

[54] M. Levitt, Conformational preferences of amino acids in globular proteins, Biochemistry 17 (1978) 4277–4285.
[55] F.R. Maxfield, H.A. Scheraga, Status of empirical methods for the prediction of protein backbone topography, Biochemistry 15 (1976) 5138–5153.
[56] K. Nagano, Logical analysis of the mechanism of protein folding: I. Predictions of helices, loops and beta-structures from primary structure, J. Mol. Biol. 75 (1973) 401–420.
[57] J. Palau, P. Argos, P. Puigdomenech, Protein secondary structure, Int. J. Pept. Protein Res. 19 (1981) 394–401.
[58] M. Prabhakaran, The distribution of physical, chemical and conformational properties in signal and nascent peptides, Biochem. J. 269 (1990) 691–696.
[59] N. Qian, T.J. Sejnowski, Predicting the secondary structure of globular proteins using neural network models, J. Mol. Biol. 202 (1988) 865–884.
[60] B. Robson, E. Suzuki, Conformational properties of amino acid residues in globular proteins, J. Mol. Biol. 107 (1976) 327–356.
[61] S. Tanaka, H.A. Scheraga, Statistical mechanical treatment of protein conformation: 5. A multistate model for specific-sequence copolymers of amino acids, Macromolecules 10 (1977) 9–20.
[62] K.T. O'Neil, W.F. DeGrado, A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids, Science 250 (1990) 646–651.
[63] M. Blaber, X.J. Zhang, B.W. Matthews, Structural basis of amino acid alpha helix propensity, Science 260 (1993) 1637–1640.
[64] P. Koehl, M. Levitt, Structure-based conformational preferences of amino acids, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 12524–12529.
[65] A.W. Burgess, P.K. Ponnuswamy, H.A. Scheraga, Analysis of conformations of amino acid residues and prediction of backbone topography in proteins, Isr. J. Chem. 12 (1974) 239–286.