ELSEVIER

International Journal of Biological Macromolecules 35 (2005) 55–62

INTERNATIONAL JOURNAL OF
Biological
Macromolecules
STRUCTURE, FUNCTION AND INTERACTIONS

www.elsevier.com/locate/ijbiomac

Structural analysis of residues involving cation-π interactions in different folding types of membrane proteins

M. Michael Gromiha*, Makiko Suwa

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

Received 28 June 2004; received in revised form 1 December 2004; accepted 2 December 2004

Abstract

Cation-π interactions play an important role to the stability of protein structures. In our earlier work, we have analyzed the influence and energetic contribution of cation-π interactions in three-dimensional structures of membrane proteins. In this work, we investigate the characteristic features of residues that are involved in cation-π interactions. We have computed several parameters, such as surrounding hydrophobicity, number of long-range contacts, conservation score and normalized *B*-factor for all these residues and identified their location, whether in the membrane or at surface. We found that the cation-π interactions are mainly formed by long-range interactions. The cationic residues involved in cation-π interactions have higher surrounding hydrophobicity than their average values in the whole dataset and an opposite trend is observed for aromatic residues. In transmembrane helical proteins, except Phe, all other residues that are responsible for cation-π interactions are highly conserved with other related protein sequences whereas in transmembrane strand proteins, an appreciable conservation is observed only for Arg. The analysis on the flexibility of residues reveals that the cation-π interaction forming residues are more stable than other residues. The results obtained in the present study would be helpful to understand the role of cation-π interactions in the structure and folding of membrane proteins.

Keywords: Cation-π interactions; Membrane protein; Conservation score; Surrounding hydrophobicity; Flexibility; *B*-factor; Long-range contacts

1. Introduction

Protein structures are stabilized with various non-covalent interactions, such as hydrophobic, electrostatic, hydrogen bonds and van der Waals interactions [19,38]. In addition, the cation-π interaction between aromatic rings and positively charged groups is recognized as an important non-covalent interaction in structural biology [8,9]. Gallivan and Dougherty [9] proposed an energy-based criterion for identifying the cation-π interactions and analyzed the distribution of these interactions in a set of globular protein structures. Recently, cation-π interactions are observed to be an important factor for understanding the thermal stability of thermophilic proteins [6,26]. Further, the importance of this interaction has been stressed by several investigators in determining the helicity of α-helical peptides [44], folding of polypeptides [5], etc. On the other hand, cation-π interactions are suggested to play a role in the stability of protein–DNA complexes and the specificity of protein–DNA recognition [39,48].

The cation-π interaction in protein structures is mainly based on the distance between the contacting (aromatic and positive charged) atoms and the interaction energy between them [9,13]. In membrane proteins, it has been reported that the aromatic residues have higher preference to be in membrane than in aqueous part [11,24]. Further, residues involving cation-π interactions have significant number of inter-residue contacts in the membrane part of outer membrane proteins [22].

Abbreviations: TMH, Transmembrane helical; TMS, Transmembrane strand

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

0141-8130/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijbiomac.2004.12.001

Recently, several theoretical and experimental works have been carried out to understand the importance of cation-π interactions in membrane proteins. Ward et al. [45] systematically analyzed the role of individual residues in the transmembrane domain of acetylcholine receptor and found that the cation-π interaction forming residue, Tyr381 plays key roles in receptor function. Kueltzo and Middaugh [29] suggested that the transport of cations across the lipid bilayer of membrane proteins is facilitated by cation-π interactions with aromatic residues and these interactions play a significant role in the binding of ligands to membrane proteins [40,49]. Aliste et al. [1] showed that the cation-π interactions between the Arg and Trp side chains stabilize a protein/polypeptide in lipid bilayer. Further, the stability of receptor proteins due to the network of cation-π interactions involving Trp and Arg side chains have been reported [33].

In our earlier work, we have analyzed the role of cation-π interactions and their energetic contributions in the two major classes of membrane proteins, such as, transmembrane helical and strand proteins [13]. The transmembrane helical (TMH) proteins consist of α-helices as transmembrane segments and transmembrane strand (TMS) proteins, are having β-strands as their membrane spanning segments [34,41,42,46,47]. In this work, we analyze the characteristic features of residues that are forming cation-π interactions with the aid of several properties, such as, surrounding hydrophobicity, long-range interactions, conservation score, flexibility and the location of residues in the membrane/surface. We found that the long-range interactions mainly contribute to the formation of cation-π interactions. The surrounding hydrophobicity is less for the aromatic residues and high for the positive charged residues. Most of the cation-π interactions forming residues are highly conserved in TMH proteins. Further, the residues responsible for cation-π interactions are more stable than other residues in membrane proteins.

2. Materials and methods

2.1. Database

We have constructed a database of membrane proteins from the information about their three-dimensional structures available in the literature [17,22,47]. These protein structures are non-redundant and are solved at high resolution. Further, all these proteins have multiple spanning helical/strand segments. The Protein Data Bank (PDB) codes of the proteins used in the present study are, for TMH proteins, 1PRC:L, 1PRC:M, 1OCC:A, 1OCC:C, 2BRD, 1E12, 1F88, 1PSS:L, 1PSS:M, 1QLA:C, 1AR1:A and 1BGY:C, and for TMS proteins, 1A0S:P, 1BXW:A, 1BY5:A, 1E54:A, 1EK9:A, 1FEP:A, 1OPF, 1OSM:A, 1PHO, 1PRN, 1QD6:C, 1QJ9:A, 2MPR:A, 2POR and 7AHL:A. The coordinates of all the membrane protein structures have been taken from the PDB [3].

2.2. Location of amino acid residues

The location of each residue in all the membrane proteins has been assigned from the knowledge of their three-dimensional structures. We have followed the illustration of membrane spanning segments given by the crystallographers, who solved the structure. The location of each residue has been classified into membrane spanning helical (H) and strand (S) segments, and loop (L) in the aqueous part (surface). In a TMH (TMS) protein, the residues, which are present in a helical (strand) segment that traverse the membrane are assigned as to be in transmembrane helix (strand) and other residues are assigned to be in surface.

2.3. Surrounding hydrophobicity

The amino acid residues in a protein molecule are represented by their α-carbon atoms and each residue is assigned with the hydrophobicity index obtained from thermodynamic transfer experiments [28,35]. The surrounding hydrophobicity (H_p) of a given residue is defined as the sum of hydrophobic indices of various residues, which appear within 8 Å radius limit from it [31].

$$H_p(i) = \sum_{j=1}^{20} n_{ij} h_j \quad (1)$$

where n_{ij} is the total number of surrounding residues of type j around ith residue of the protein and h_j the experimental hydrophobic index of residue type j in kcal/mol [28,35].

It has been shown that the influence of each residue over the surrounding medium extends effectively only up to 8 Å [30] and this limit is sufficient to characterize the hydrophobic behavior of amino acid residues [31,37] and to accommodate both the local and non-local interactions [16,21,27]. Further, 8 Å limit has been used in several studies, such as to understand the folding rate of two-state proteins [7,18], protein stability upon mutations [25], thermal stability of proteins [12,23] and to determine the transition state structures of two-state protein mutants [20].

2.4. Computation of long-range contacts

For each residue, we computed the residues coming within a sphere of 8 Å radius as described in a previous section. For a given residue, the composition of surrounding residues is analyzed in terms of the location at the sequence level and the contributions from $<\pm 3$ residues are treated as short range contacts, ± 3 or ± 4 residues as medium range contacts and $>\pm 4$ residues are treated as long range contacts [14,16,21]. This classification enables us to evaluate the contribution of long-range contacts in the formation of cation-π interactions.

2.5. Conservation of amino acid residues

We have evaluated the conservation of residues in each protein with the aid of the Consurf server ([10]; http://consurf.tau.ac.il). This server compares the sequence of a PDB chain with the proteins deposited in Swiss-Prot [4] and finds the ones that are homologous to the PDB sequence. The number of PSI-BLAST iterations and the *E*-value cut-off used in all similarity searches were 1 and 0.001, respectively. All the sequences that were found to be evolutionary related with each one of the membrane protein chains in the data set were used in the subsequent multiple alignments. These protein sequence alignments were used to classify the residues in each structure into nine categories: from very variable (score = 1) to highly conserved (score = 9) and the details about the classification of residues have been explained in Glaser et al. [10]. It is worth to mention that, according to the Consurf algorithm, the highest score does not necessarily indicate 100% conservation (e.g. no replacements at all), but rather indicates that this position is the most (or among the most) conserved in the corresponding protein.

2.6. Thermal stability of residues

The thermal stability of each residue is evaluated by the average value of the normalized *B*-factors (i.e. B'-factors) for all the atoms in that residue [36]. The B'-factor for an atom is calculated with the following equation:

$$B'\text{-factor} = \frac{B - \langle B \rangle}{\sigma} \quad (2)$$

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

3. Results and discussions

3.1. Cation-π interactions in membrane protein structures

We have obtained the information about the residues that are involved in cation-π interactions in TMH and TMS proteins using the program, CAPTURE developed by Gallivan and Dougherty [9], available at http://capture.caltech.edu. This method used an energy-based criterion to delineate cation-π interactions in protein structures and it has been widely used for the analysis of cation-π interactions. The details about the number of cation-π interactions in each protein and their energetic contributions have been explained in our earlier article [13]. The list of positively charged and aromatic residues that are forming cation-π interactions in all the TMH and TMS proteins are presented in Table 1 . We noticed that in most cases the geometry of cationic and aromatic side chains is biased to experience favorable cation-π interactions and the role of cation-π interaction is different from conventional non-covalent interactions [15]. Further, the inclusion of cation-π interaction along with conventional hydrogen bonds, electrostatic and hydrophobic interactions provides deep insights to understand/predict the structure and stability of proteins [43]. The characteristic features of these residues have been analyzed using several parameters, such as, (i) the distance of separation between the cationic and π residues (ii) location of cationic and aromatic residues, (iii) surrounding hydrophobicity, (iv) number of long-range contacts, (v) conservation score and (vi) the thermal stability of each cationic and aromatic residues.

3.2. Sequential separation between residues that are forming cation-π interactions

We have calculated the sequential distance between the cationic and aromatic residues for each cation-π interaction and the results are listed in Table 1. We found that in TMH proteins, about 60% of cation-π interactions are influenced by long-range interactions. The marginal contribution of short- and medium-range interactions (20% each) are due to the presence of membrane spanning helices in TMH proteins. It has been reported that these interactions play a vital role in the formation of α-helices [14,16]. In TMS proteins, most of the cation-π interactions (88%) are influenced by long-range interactions. This might be due to the presence of several β-strands in TMS proteins and long-range interactions dominate in this class of proteins [18]. The formation of cation-π interactions between the residues that are far away in the sequence may contribute to the stability of TMS proteins. This study shows that the cation-π interactions are mainly formed by long-range interactions. The role of short and medium-range interactions are minimal although they play an important role in the formation of ion-pairs [2,32].

3.3. Location of cation-π interactions forming residues

The location of cation-π interactions forming residues has been identified and the observations are presented in Table 1. Apparently, there is no specific preference for these residues in membrane or surface. Further, about 60% of cationic residues prefer to be in membrane whereas an opposite trend was noticed for the aromatic residues in TMH proteins. However, this result mainly depends on the assignment of loop boundary for each membrane-spanning segment in TMH and TMS proteins.

3.4. Surrounding hydrophobicity of cation-π interactions forming residues

The computed surrounding hydrophobicity for each cation-π interaction forming residues is included in Table 1. The analysis on the preference of residues at different ranges of surrounding hydrophobicity shows that most of the residues prefer the range of 10–15 kcal/mol, which in-

Table 1
Location, surrounding hydrophobicity, long-range contacts, conservation score and normalized *B*-factor for the residues that are forming cation-π interactions

PDB	Cation						π						D_{seq}
	Residue	Str	H_p	N_l	Cons	B'	Residue	Str	H_p	N_l	Cons	B'	
Transmembrane helical proteins													
1PRC:L	Arg135	H	25.79	6	9	−0.46	Phe160	L	11.26	0	5	−0.83	25
	Arg10	L	8.55	0	9	0.30	Trp25	L	15.68	2	8	−0.30	15
	Arg103	H	19.98	4	9	−0.70	Trp100	H	16.16	3	8	−0.70	3
	Arg257	L	14.82	2	1	1.58	Trp255	L	7.60	1	1	0.24	2
	Lys110	H	13.15	5	6	−0.29	Phe30	L	19.93	5	8	−0.02	80
	Lys8	L	9.78	0	9	−0.25	Tyr9	L	8.05	1	4	−0.77	1
	Lys110	H	13.15	5	6	−0.29	Trp25	L	15.68	2	8	−0.30	85
1PRC:M	Arg245	L	12.13	1	9	−0.48	Phe227	L	4.73	2	6	−0.72	18
	Arg134	H	10.22	0	8	0.61	Tyr50	L	11.57	3	8	0.26	84
	Arg190	L	9.91	1	6	−0.53	Tyr191	L	16.56	6	8	−0.82	1
	Arg130	H	16.39	2	9	0.14	Trp127	H	13.96	4	8	−0.28	3
	Arg134	H	10.22	0	8	0.61	Trp23	L	2.47	1	1	1.09	111
	Lys40	L	16.23	4	4	0.84	Trp37	L	10.67	1	1	2.82	3
1OCC:A	Lys13	H	11.47	3	9	0.66	Trp81	H	21.6	8	9	−0.32	68
1OCC:C	Arg156	H	9.35	2	9	−0.42	Phe225	L	8.47	0	1	(0.69	69
2BRD	Arg175	H	14.62	1	6	−0.14	Phe156	H	14.30	2	4	−0.05	19
	Lys41	H	16.16	1	4	−0.01	Phe42	H	14.40	2	3	−0.29	1
1F88:A	Lys296	L	9.36	1	9	−1.31	Phe293	L	12.38	1	9	−0.87	3
1PSS:L	Arg10	L	10.57	0	7	0.76	Trp25	L	15.68	2	7	1.79	15
	Arg103	H	21.05	5	9	−0.38	Trp100	H	13.28	3	8	0.73	3
	Lys8	L	9.78	0	6	0.94	Tyr9	L	8.05	1	8	−0.36	1
	Lys110	H	16.52	5	6	0.95	Tyr30	L	20.03	5	8	−0.40	80
	Lys110	H	16.52	5	6	0.95	Trp25	L	15.68	2	7	1.79	85
1PSS:M	Arg247	L	11.13	1	9	−0.71	Phe229	L	3.93	2	5	−0.55	18
	Arg136	H	9.74	1	8	0.39	Tyr51	L	12.43	7	7	0.42	85
	Arg132	H	15.74	2	9	−0.02	Trp129	H	13.71	3	6	−0.48	3
	Lys110	H	14.52	2	3	1.45	Trp73	H	20.12	2	4	1.78	37
1QLA:C	Lys115	L	11.81	3	0	−0.96	Trp127	H	11.77	0	0	−0.74	12
1AR1:A	Arg21	L	14.29	0	9	2.49	Trp22	L	17.35	4	8	2.20	1
	Lys354	H	17.21	2	9	−0.55	Trp358	H	21.45	4	9	−0.03	4
	Lys434	H	23.49	8	7	−0.17	Trp532	L	12.31	3	8	−0.13	98
1BGY:C	Arg80	H	23.16	4	9	−0.53	Tyr81	H	19.16	3	9	−0.68	1
	Arg100	H	19.83	4	9	−0.52	Trp31	H	11.91	2	9	−0.59	69
	Lys311	L	4.85	2	9	−0.07	Trp379	L	9.35	2	9	1.27	68
Transmembrane strand proteins													
1A0S:P	Arg187	S	10.90	7	9	−1.22	Phe164	L	17.02	8	8	−0.90	23
	Arg222	S	15.64	7	5	1.70	Phe252	L	11.63	5	8	0.21	30
	Arg190	S	4.28	0	9	−0.92	Tyr213	L	11.12	6	8	−0.16	23
	Arg410	S	11.37	4	1	0.54	Tyr456	S	11.95	5	1	−0.06	46
	Arg187	S	10.90	7	9	−1.22	Trp183	S	14.14	7	9	−0.51	4
	Arg437	L	18.37	6	9	0.25	Trp482	L	11.74	6	9	−0.28	45
1BXW:A	Arg138	S	13.30	8	9	−0.72	Phe40	S	7.32	7	9	−0.75	98
	Arg138	S	13.30	8	9	−0.72	Tyr94	S	14.05	8	9	−0.75	44
	Lys34	L	15.20	9	4	0.20	Tyr72	L	8.70	6	4	0.62	38
	Lys82	S	17.15	9	9	−0.57	Tyr94	S	14.05	8	9	−0.75	12
1BY5:A	Arg81	L	11.17	5	1	−1.03	Phe699	L	17.72	11	5	−1.01	518
	Arg81	L	11.17	5	1	−1.03	Tyr244	L	10.30	4	5	−1.44	163
	Arg93	L	13.33	8	9	−0.64	Tyr541	S	13.03	7	9	−0.82	448
	Arg277	S	9.64	9	7	−1.30	Tyr87	L	5.16	4	1	−1.18	190
	Arg637	S	14.27	9	9	−0.41	Tyr72	L	11.95	7	5	−0.46	555
	Lys249	L	14.30	7	1	−0.13	Phe650	L	7.10	3	4	−0.18	401
	Lys38	L	4.11	2	7	−0.89	Tyr140	L	8.45	1	8	−1.14	102
	Lys154	L	10.17	4	8	−0.99	Tyr124	L	8.14	1	1	−1.29	30
	Lys344	S	22.47	11	1	−0.13	Tyr315	S	22.71	11	1	0.51	29
	Lys526	S	15.23	9	9	−0.17	Tyr493	S	17.12	8	9	0.21	33
	Lys651	L	7.91	4	4	0.06	Tyr643	L	20.08	12	6	0.65	8
	Lys415	L	7.43	2	2	1.15	Trp390	S	14.79	8	1	0.50	25
1E54:A	Arg42	S	15.46	7	0	−0.05	Phe7	S	13.67	5	0	−1.52	35
	Arg203	S	10.65	9	0	−0.81	Phe228	S	12.76	8	0	0.08	5

Table 1 (*Continued*)

PDB	Cation						π						D_{seq}
	Residue	Str	H_p	N_l	Cons	B'	Residue	Str	H_p	N_l	Cons	B'	
	Arg89	S	17.05	8	0	0.22	Tyr141	S	12.15	7	0	0.02	52
	Arg42	S	15.46	7	0	−0.05	Trp56	S	11.89	8	0	−1.16	14
	Arg76	L	15.99	8	0	−1.35	Trp56	S	11.80	8	0	−1.16	20
	Arg119	L	12.50	4	0	0.55	Trp121	L	6.09	3	0	−0.68	2
	Lys256	S	13.80	9	0	0.53	Phe107	L	10.55	4	0	−0.44	149
1EK9:A	Arg390	L	17.97	3	7	−0.97	Phe201	L	7.47	2	2	−0.61	189
	Arg24	L	11.94	4	8	0.30	Tyr98	L	7.29	4	4	−0.33	74
1FEP:A	Arg75	L	9.66	9	9	−0.88	Phe528	S	14.04	9	7	−1.01	453
	Arg283	S	8.57	8	4	0.31	Phe337	L	10.19	6	6	2.25	54
	Arg70	L	5.86	4	1	−0.71	Tyr709	L	11.45	7	9	−0.94	639
	Arg132	L	12.65	7	5	0.21	Tyr133	L	5.52	1	9	−0.09	1
	Arg174	S	12.84	7	6	−1.08	Tyr160	S	15.37	9	5	−1.12	14
	Arg274	L	11.94	6	6	0.88	Trp203	L	10.88	7	5	−0.75	71
	Arg286	S	11.05	8	6	−0.96	Trp113	L	6.17	2	1	−0.46	173
	Lys276	L	11.37	3	4	0.87	Tyr260	L	10.68	9	7	−0.58	16
	Lys276	L	11.37	3	4	0.87	Tyr272	L	3.36	0	8	1.99	4
	Lys649	L	14.71	6	2	0.13	Tyr707	L	18.76	10	1	−0.68	58
	Lys89	L	10.72	1	1	−1.12	Trp306	S	12.02	7	1	−0.86	217
1OPF	Arg132	S	17.60	10	9	−0.99	Tyr102	S	13.14	5	6	−0.86	30
	Lys46	S	14.98	7	7	0.53	Tyr14	S	14.30	7	7	0.01	32
	Lys89	S	16.39	8	8	2.13	Tyr58	S	7.53	8	5	−0.09	31
	Lys219	S	21.07	9	9	0.44	Tyr111	L	11.94	6	4	−0.18	108
1OSM:A	Arg196	S	5.22	7	8	0.18	Phe250	L	8.25	7	1	0.25	54
	Arg132	S	10.75	9	9	−1.00	Tyr102	L	9.54	5	7	−0.97	30
	Lys89	S	13.29	8	8	0.83	Tyr58	S	7.53	8	6	−0.90	31
	Lys314	S	20.91	8	7	−0.31	Tyr294	S	20.40	8	6	−0.08	20
	Lys219	S	19.34	9	9	0.07	Trp111	L	7.95	5	5	0.44	108
1PHO	Arg196	L	4.63	8	8	0.74	Phe250	S	13.19	8	1	0.24	54
	Lys314	L	23.47	8	7	−0.28	Tyr294	L	19.53	8	6	−0.03	20
1PRN	Arg286	S	20.14	9	0	−0.72	Tyr7	S	14.56	7	0	−0.78	279
	Lys50	S	12.94	7	0	−0.29	Tyr117	L	6.59	4	0	−1.00	67
1QD6:C	Arg147	S	1.91	2	0	−0.17	Tyr56	L	9.39	0	0	−0.05	91
	Arg110	S	13.99	4	0	−0.42	Trp97	S	16.99	8	0	−0.87	13
	Arg157	L	22.51	7	0	−0.94	Trp155	L	7.55	9	0	−0.69	2
	Arg157	L	22.51	7	0	−0.97	Trp176	L	20.05	6	0	−0.53	19
	Lys174	S	19.68	7	0	−0.37	Tyr159	L	19.32	8	0	−0.69	15
2MPR:A	Arg109	L	20.06	6	8	−0.71	Phe106	L	11.51	6	6	−0.85	3
	Arg105	S	12.80	8	9	−0.66	Trp101	S	10.37	7	9	−0.73	4
	Arg175	S	16.43	8	9	−0.53	Trp120	L	13.37	6	4	−0.63	55
	Arg370	L	19.81	7	9	−0.56	Trp426	S	15.72	6	7	−0.35	56
	Arg397	L	14.75	3	1	2.00	Trp307	S	7.29	5	7	−0.20	90
	Arg411	L	6.26	5	6	0.07	Trp382	S	9.90	10	3	−0.23	29
	Lys244	L	14.79	6	1	−0.66	Phe117	S	14.29	4	6	−0.72	27
	Lys165	L	9.75	6	1	0.31	Tyr202	L	13.90	6	4	−0.15	37
	Lys229	S	12.91	8	7	−0.36	Tyr227	L	11.57	6	6	−0.22	2
2POR	Lys69	S	17.41	6	0	−0.34	Phe62	S	15.15	7	0	−0.43	7
7AHL:A	Arg200	L	10.10	3	1	1.26	Tyr191	L	15.14	5	8	1.53	9
	Arg253	L	20.17	10	9	−0.28	Trp274	L	6.41	3	1	0.48	21
	Lys273	L	10.25	4	6	0.19	Trp274	L	6.41	3	1	0.48	1

No cation-π interaction has been observed in 1E12 and 1QJ9. Str: location of residues in the membrane spanning helix (H), strand (S) or at surface loops (L). H_p: surrounding hydrophobicity. N_l: number of long-range contacts. Cons: conservation score. B': normalized B'-factor, D_{seq}: sequence distance of separation between cationic and aromatic residues.

cludes the average H_p of each residue obtained from the whole data set of TMH and TMS proteins. We have further examined the percentage of residues, which are higher/lower than the average H_p value and the results are presented in Table 2. We found that both in TMH and TMS proteins, about 70% of Lys prefers to have higher H_p than the average value and there is no preference for Arg. Among the aromatic residues, most of the Phe in TMH proteins (88%) have less H_p than the average (15.33 kcal/mol) whereas there is no priority for the other two residues, Tyr and Trp. This might be due to the occurrence of Phe in the middle of the membrane spanning helices whereas most of the Tyr and Trp occur in the N- and C-terminals of transmembrane helical segments. On the other hand, in TMS proteins, 78–88% of the aromatic residues have less H_p than their respective average values (Table 2). This result indi-

Table 2
Preference of cation-π interaction forming residues to have high/low surrounding hydrophobicity

Residue	Surrounding hydrophobicity							
	TMH				TMS			
	N_{res}	Average (kcal/mol)	High (%)	Low (%)	N_{res}	Average (kcal/mol)	High (%)	Low (%)
Lys	15	11.63	**66.7**	33.3	28	11.07	**75.0**	25.0
Arg	19	13.01	52.6	47.4	45	12.74	51.1	49.9
Phe	8	15.33	12.5	**87.5**	16	11.06	12.5	**87.5**
Tyr	7	15.26	42.9	57.1	37	12.55	21.6	**78.4**
Trp	19	15.18	47.4	52.6	20	10.89	15.0	**85.0**

High and low represent, respectively, the percentage of residues that have higher and lower surrounding hydrophobicity than the average H_p. The values with more than 65% are bold; N_{res}: number of residues.

cates that the cationic groups have higher H_p and the aromatic residues have lower H_p than their average surrounding hydrophobicity values for the formation of cation-π interactions.

3.5. Long-range interactions

Long-range interactions play an important role to the folding and stability of proteins through several non-covalent interactions. Recently, long-range contacts have also been used for predicting the folding rates of proteins [18]. We have calculated the number of long-range contacts for each cation-π interaction forming residues and the results are given in Table 1. We found that 94% of the residues have at least one long-range contact. The details about percentage of residues with different number of long-range contacts are displayed in Fig. 1. We observed that the positively charged residues in TMH proteins have 0–2 long-range contacts whereas the aromatic residues have 1–3 long-range contacts. In TMS proteins, cationic and aromatic residues have 7–9 and 5–8 long-range contacts, respectively. On the other hand, in all-α globular proteins, cationic residues have 0–2 long-range contacts, similar to TMH proteins, and the aromatic residues have 1–4 long-range contacts. In all-β class of globular proteins, we observed that the cation-π interaction forming residues have less number of long-range contacts compared with TMS proteins; positive charged residues have 3–8 contacts and aromatic residues have 4–8 contacts. This result shows that the cation-π interaction forming residues have the tendency of making more number of contacts in TMS proteins than that in all-β globular proteins, which may be important for the stability of TMS proteins.

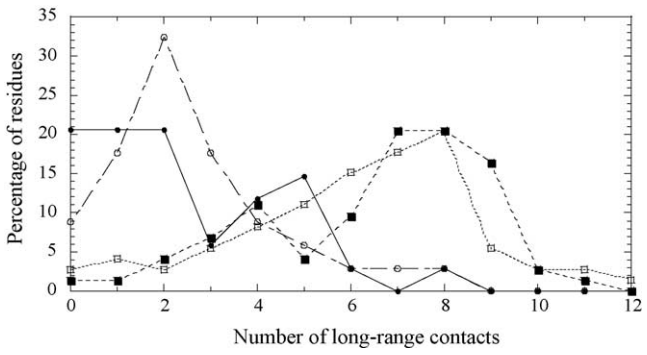

Fig. 1. Number of long-range contacts vs. percentage of residues for positively charged (cation) and aromatic (π) residues in TMH and TMS proteins: (●) positively charged residues in TMH proteins, (○) aromatic residues in TMH proteins, (■) positively charged residues in TMS proteins and (□) aromatic residues in TMS proteins.

Table 3
Average number of long-range contacts for cation-π interaction forming residues in TMH and TMS proteins

Residue	Average number of long-range contacts			
	TMH		TMS	
	All[a]	Cation-π	All[a]	Cation-π
Lys	2.2 ± 1.9	3.1 ± 1.9	5.7 ± 2.5	6.4 ± 3.2
Arg	2.2 ± 1.9	1.9 ± 1.8	6.5 ± 2.6	6.6 ± 3.5
Phe	2.5 ± 2.0	1.8 ± 0.9	5.5 ± 2.6	6.3 ± 2.1
Tyr	3.1 ± 2.4	3.7 ± 1.3	6.6 ± 2.5	6.1 ± 3.0
Trp	2.6 ± 2.3	2.6 ± 1.7	6.0 ± 2.5	6.2 ± 2.2

[a] All shows the average number of long-range contacts in the whole data set of TMH/TMS proteins. Data were taken from Gromiha and Selvaraj [17].

The average number of long-range contacts for Lys, Arg, Phe, Trp and Tyr in the whole data set of TMH and TMS proteins and that for cation-π interaction forming residues are presented in Table 3. We observed that both in TMH and TMS proteins, the cation-π interaction forming Lys residues have higher number of contacts (3.1 and 6.4 average contacts/residue, respectively) compared with the whole data set. Trp and Arg show no significant difference between cation-π interaction forming residues and other residues. In TMH proteins cation-π interaction forming Phe residues have lower number of contacts (1.8 contacts/residue) than other Phe residues (2.5 contacts/residue), while an opposite trend is observed in TMS proteins. This might be due to the difference in the interacting pattern of residues in TMH and TMS proteins.

3.6. Conservation score

We have calculated the conservation score for all the residues in both TMH and TMS proteins. In the case of 1QLAC, 1E54A, 1PRN, 1QD6C and 2POR the number of homologous sequences deposited in Swiss-Prot (three, two, one, four and one, respectively) is lower than the five, which is the number of sequences that the Consurf server [10] needs

to calculate score values. Hence, the analysis has been made with rest of the proteins in the data set. The conservation score has been classified between 1 and 9 and hence the residues with score higher than the average (i.e. 5) are considered to have conservation with other proteins. In TMH proteins, the percentage of residues having conservation score more than 6 are 78.6% for Lys, 94.7% for Arg, 37.5% for Phe, 85.7% for Tyr and 77.8% for Trp. It is interesting that all the cation-π interaction forming residues except Phe are highly conserved among other evolutionary-related protein sequences. On the other hand, only Arg is conserved in the case of TMS proteins. The percentage of residues with conservation score more than 6 is: Lys, 54.2%; Arg, 73.5%; Phe, 58.8%; Tyr, 59.4% and Trp, 35.7%. The difference of conservation score between TMH and TMS proteins might be due to the difference between the pattern of amino acid sequences in the membrane part of TMH and TMS proteins, and it provides more insights about cation-π interaction forming residues in different folding types of membrane proteins.

3.7. Thermal stability of cation-π interaction forming residues

We have examined the thermal stability of cation-π interaction forming residues by means of the normalized *B*-factor (*B*′-factor) calculated using Eq. (2) and the results are reported in Table 1. The percentage of residues at various ranges of *B*′-factors is displayed in Fig. 2. We have also included the result with the control data of all the cationic and aromatic residues in membrane proteins. We found that the cation-π interaction forming residues are more stable than the other residues in membrane proteins. Similar trend is also observed in soluble proteins. However, the *B*′-factors of cation-π interaction forming residues in membrane proteins are more significant than that in globular proteins. The comparison between the *B*′-factors of cation-π interaction forming residues in all-β globular and TMS proteins indicates that 26% of them in TMS proteins have the *B*′-factor less than −1.0 while it is 15% in all-β globular proteins. Similarly, the percentage of cation-π interaction forming residues with negative *B*′-factor is 8% higher in TMH proteins compared with all-α globular proteins. This result indicates that the influence of cation-π interactions is more important in membrane proteins than in globular proteins.

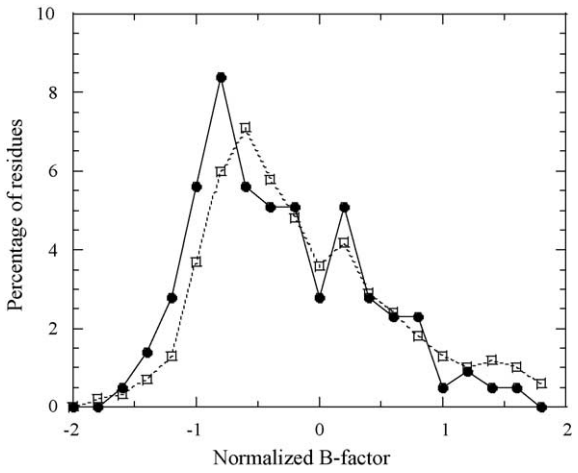

Fig. 2. Percentage of residues at different range of normalized *B*′-factors. (●) Residues that are involved in cation-π interactions (Table 1) and (□) all positively charged and aromatic residues in the dataset of membrane proteins.

3.8. Comparison among TMH, TMS and globular proteins

The comparative analysis of cation-π interactions in TMH, TMS and globular proteins reveals the following insights: (i) cation-π interactions are mainly influenced by long-range interactions in TMS proteins whereas short, medium and long-range interactions play significant roles in TMH and globular proteins, (ii) cation-π interactions forming residues in TMH and all-α class globular proteins have similar number of contacts whereas that in TMS proteins have more number of contacts than all-β class globular proteins, (iii) surrounding hydrophobicity of cation-π interactions forming residues in TMH, TMS and globular proteins are, respectively, 10–15, 10–13 and 7–16 kcal/mol, (iv) Lys, Arg, Trp and Tyr are conserved in TMH and globular proteins whereas only Arg is conserved in TMS proteins and (v) most of the cation-π interactions forming residues in TMH, TMS and globular proteins have negative *B*′-factors, indicating that these residues are more stable than other residues in respective classes of proteins.

4. Conclusions

We have analyzed the characteristic features of cation-π interaction forming residues in TMH and TMS proteins. We found that the cation-π interactions are dominated by long-range interactions and there is a marginal influence by short and medium-range interactions in TMH proteins. The cationic and aromatic residues tend to have lower and higher surrounding hydrophobicity, respectively, than their average values. The cationic and aromatic residues in TMS proteins have higher number of long-range contacts than that in TMH proteins. Most of the cation-π interactions forming residues are highly conserved in TMH proteins. Further, the cation-π interaction forming residues have higher stability than other cationic and aromatic residues. The results obtained in the present study will be helpful to understand the structure and folding of membrane proteins.

Acknowledgment

We thank Dr. Yutaka Akiyama for encouragement.

References

[1] M.P. Aliste, J.L. MacCallum, D.P. Tieleman, Biochemistry 42 (2003) 8976–8987.
[2] D.J. Barlow, J.M. Thornton, J. Mol. Biol. 168 (1983) 867–885.
[3] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucl. Acids Res. 28 (2000) 235–242.
[4] B. Boeckmann, A. Bairoch, R. Apweiler, M.-C. Blatter, A. Estreicher, E. Gasteiger, M.J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, Nucl. Acids Res. 31 (2003) 365–370.
[5] T.P. Burghardt, N. Juranic, S. Macura, K. Ajtai, Biopolymers 63 (2002) 261–272.
[6] S. Chakravarty, R. Varadarajan, Biochemistry 41 (2002) 8152–8161.
[7] D.A. Debe, W.A. Goddard, J. Mol. Biol. 294 (1999) 619–625.
[8] D.A. Dougherty, Science 271 (1996) 163–168.
[9] J.P. Gallivan, D.A. Dougherty, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 9459–9464.
[10] F. Glaser, T. Pupko, I. Paz, R.E. Bell, D. Bechor, E. Martz, N. Ben-Tal, Bioinformatics 19 (2003) 163–164.
[11] M.M. Gromiha, Protein Eng. 12 (1999) 557–561.
[12] M.M. Gromiha, Biophys. Chem. 91 (2001) 71–77.
[13] M.M. Gromiha, Biophys. Chem. 103 (2003) 251–258.
[14] M.M. Gromiha, J. Biophys. Soc. Jpn. 43 (2003) 87–92.
[15] M.M. Gromiha, Polymer 46 (2005) 983–990.
[16] M.M. Gromiha, S. Selvaraj, J. Biol. Phys. 23 (1997) 151–162.
[17] M.M. Gromiha, S. Selvaraj, Int. J. Biol. Macromol. 29 (2001) 25–34.
[18] M.M. Gromiha, S. Selvaraj, J. Mol. Biol. 310 (2001) 27–32.
[19] M.M. Gromiha, S. Selvaraj (Eds.), Recent Research Developments in Protein Folding, Stability and Design, Research Signpost, Trivandrum, India, 2002.
[20] M.M. Gromiha, S. Selvaraj, FEBS Lett. 526 (2002) 129–134.
[21] M.M. Gromiha, S. Selvaraj, Prog. Biophys. Mol. Biol. 86 (2004) 235–277.
[22] M.M. Gromiha, M. Suwa, Variation of amino acid properties in all-β globular and outer membrane protein structures, Int. J. Biol. Macromol. 32 (2003) 93–98.
[23] M.M. Gromiha, A.M. Thangakani, Prep. Biochem. Biotech. 31 (2001) 217–227.
[24] M.M. Gromiha, R. Majumdar, P.K. Ponnuswamy, Protein Eng. 10 (1997) 497–500.
[25] M.M. Gromiha, M. Oobatake, H. Kono, H. Uedaira, A. Sarai, Protein Eng. 12 (1999) 549–555.
[26] M.M. Gromiha, S. Thomas, C. Santhosh, Prep. Biochem. Biotech. 32 (2002) 355–362.
[27] Z. Jiang, L. Zhang, J. Chen, A. Xia, D. Zhao, Polymer 43 (2002) 6037–6047.
[28] D.D. Jones, J. Theor. Biol. 50 (1975) 167–183.
[29] L.A. Kueltzo, C.R. Middaugh, J. Pharm. Sci. 92 (2003) 1754–1772.
[30] P. Manavalan, P.K. Ponnuswamy, Arch. Biochem. Biophys. 184 (1977) 476–487.
[31] P. Manavalan, P.K. Ponnuswamy, Nature 275 (1978) 673–674.
[32] B.W. Matthews, Ann. Rev. Biochem. 62 (1993) 139–160.
[33] T.D. Mulhern, A.F. Lopez, R.J. D'Andrea, C. Gaunt, L. Vandeleur, M.A. Vadas, G.W. Booker, C.J. Bagley, J. Mol. Biol. 297 (2000) 989–1001.
[34] K. Nakai, Adv. Protein Chem. 54 (2000) 277–344.
[35] Y. Nozaki, C. Tanford, J. Biol. Chem. 246 (1971) 2211–2217.
[36] S. Parthasarathy, M.R. Murthy, Protein Sci. 6 (1997) 2561–2567.
[37] P.K. Ponnuswamy, Prog. Biophys. Mol. Biol. 59 (1993) 57–103.
[38] P.K. Ponnuswamy, M.M. Gromiha, J. Theor. Biol. 166 (1994) 63–74.
[39] M. Rooman, J. Lievin, E. Buisine, R. Wintjens, J. Mol. Biol. 319 (2002) 67–76.
[40] J.D. Schmitt, C.G. Sharples, W.S. Caldwell, J. Med. Chem. 42 (1999) 3066–3074.
[41] G.E. Schulz, Curr. Opin. Struct. Biol. 10 (2000) 443–447.
[42] G.E. Schulz, Adv. Protein Chem. 63 (2003) 47–70.
[43] S. Shacham, Y. Marantz, S. Bar-Haim, O. Kalid, D. Warshaviak, N. Avisar, B. Inbal, A. Heifetz, M. Fichman, M. Topf, Z. Naor, S. Noiman, O.M. Becker, Proteins 57 (2004) 51–86.
[44] Z. Shi, C.A. Olson, N.R. Kallenbach, J. Am. Chem. Soc. 124 (2002) 3284–3291.
[45] S.D. Ward, C.A. Curtis, E.C. Hulme, Mol. Pharmacol. 56 (1999) 1031–1041.
[46] S.H. White, W.C. Wimley, Annu. Rev. Biophys. Biomol. Struct. 28 (1999) 319–365.
[47] W.C. Wimley, Protein Sci. 11 (2002) 301–312.
[48] R. Wintjens, J. Lievin, M. Rooman, E. Buisine, J. Mol. Biol. 302 (2000) 395–410.
[49] W. Zhong, J.P. Gallivan, Y. Zhang, L. Li, H.A. Lester, D.A. Dougherty, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 12088–12093.