PROTEINS: Structure, Function, and Bioinformatics 65:75–86 (2006)

Exploring the Environmental Preference of Weak Interactions in $(\alpha/\beta)_8$ Barrel Proteins

S. Chakkaravarthi,[1] M. Madan Babu,[2,3] M. Michael Gromiha,[4] G. Jayaraman,[1*] and R. Sethumadhavan[1*]
[1]*School of Biotechnology and Chemical Engineering, Vellore Institute of Technology, Vellore, India*
[2]*MRC Laboratory of Molecular Biology, Cambridge, United Kingdom*
[3]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland*
[4]*Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan*

ABSTRACT **The environmental preference for the occurrence of noncanonical hydrogen bonding and cation–π interactions, in a data set containing 71 nonredundant $(\alpha/\beta)_8$ barrel proteins, with respect to amino acid type, secondary structure, solvent accessibility, and stabilizing residues has been performed. Our analysis reveals some important findings, which include (a) higher contribution of weak interactions mediated by main-chain atoms irrespective of the amino acids involved; (b) domination of the aromatic amino acids among interactions involving side-chain atoms; (c) involvement of strands as the principal secondary structural unit, accommodating cross strand ion pair interaction and clustering of aromatic amino acid residues; (d) significant contribution to weak interactions occur in the solvent exposed areas of the protein; (e) majority of the interactions involve long-range contacts; (f) the preference of Arg is higher than Lys to form cation–π interaction; and (g) probability of theoretically predicted stabilizing amino acid residues involved in weak interaction is higher for polar amino acids such as Trp, Glu, and Gln. On the whole, the present study reveals that the weak interactions contribute to the global stability of $(\alpha/\beta)_8$ TIM-barrel proteins in an environment-specific manner, which can possibly be exploited for protein engineering applications. Proteins 2006;65:75–86.** © 2006 Wiley-Liss, Inc.

Key words: accessible surface area; amino acid preference; secondary structure; sequential separation; stabilizing residues; TIM-barrel; weak interactions

INTRODUCTION

The stability of a protein is determined by various noncovalent interactions, such as hydrophobic, electrostatic, hydrogen bonds, and van der Waals interactions.[1] Of these, several studies have elucidated that hydrogen bonds have a key role in structure-function relationship of proteins, which includes aiding over all folding, maintaining local structure, facilitating protein ligand recognition, enzymatic activity, protein hydration, and molecular dynamics.[2] Apart from these hydrogen bonds, it is now generally accepted that other weak electrostatic interactions termed noncanonical interactions (NCI), such as those involving C—H···O, C—H···π, and the N—H···π interactions, contribute to structural stability of both small molecules and biological macromolecules.

Even though the occurrence of these NCI was well documented very early in time,[3–5] it was not until recently that their importance was completely understood. Several large-scale studies over the last decade have unambiguously revealed the occurrence of these interactions in crystal structures, revealing the importance of such interactions and therefore reviving interest in studying them in greater detail.[6–18] For instance, the comprehensive study by Steiner and Saenger,[19] has shown that several dipole–dipole interactions other than canonical hydrogen bonds could significantly contribute to decrease the entropy and therefore increase the overall stability of proteins. In another study,[16] the occurrence of C—H···O interactions in protein was analyzed by categorizing them into interactions that involve only the main-chain atoms, side-chain atoms, or both. The same study clearly revealed the role of weak C—H···O interactions in conformational flexibility and in facilitating protein–protein interactions.[16] Various other studies have elucidated the role of NCI in biological macromolecules and have shown that they serve as an additional stabilizing factor in β-sheet,[14] helix termini,[15] helices that contain proline residues,[20] packing of transmembrane helices,[21] collagen,[7] and in DNA.[22] They have also been shown to be important in a variety of functional contexts such as macromolecular recognition,[17,23,24] enzymatic action,[25] and stabilization of secondary structure.[26]

In terms of energetic contribution, theoretical *ab initio* calculations[27–30] have clearly revealed that the energy of these NCI is less than the energy of a conventional hydrogen bond (O/N—H···O═C). For instance, C—H···O

Grant sponsor: National Institutes of Health.

*Correspondence to: G. Jayaraman or R. Sethumadhavan, School of Biotechnology and Chemical Engineering, Vellore Institute of Technology, Vellore-632 014, India. E-mail: raman291068@yahoo.com or sethu_rao@sancharnet.in

Received 29 December 2005; Revised 9 March 2006; Accepted 3 April 2006

Published online 11 July 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21067

interactions may contribute up to 2 kcal/mol, N—H···π interactions contribute up to 3.5 kcal/mol, and C—H···π interactions contribute up to 1 kcal/mol, whereas regular hydrogen bonds may contribute up to 5.5 kcal/mol.[9] Even though the NCI contribute much less in energetic terms, the observation that they can occur as frequently, or sometimes more frequently than regular hydrogen bonds, suggests that they may individually contribute less, but cumulatively provide significant energy for protein stability (M.M.B., unpublished results).

In addition, cation–π interactions between amino acid side-chains are increasingly being recognized as important structural and functional features of proteins and other biomolecules.[31–34] Cation–π interactions can occur between cationic side-chain of either lysine or arginine and the aromatic side-chain of phenylalanine, tyrosine, or tryptophan. The stabilization energy originates in part from electrostatic attraction between the cation (of the basic amino acid residue) and regions of high electron density in π-orbital of the aromatic group, leading to the name cation–π interaction.[35] There are reports of this interaction for their role in the enhancement of stability of thermophilic proteins,[36,37] folding of polypeptides,[38] and the stability of membrane protein.[39,40] The stability and specificity of protein DNA complexes are also reported on the basis of these cation–π interactions.[41,42]

Although previous studies have investigated the occurrence of individual NCI [C^{α}—H···O═C, cation–π and those with aromatic ring systems as hydrogen bond acceptors] in various proteins,[43–45] very few studies have systematically studied the role of these weak interactions in relation to other factors such as amino acid preference, secondary structural elements, solvent accessibility of a particular fold, and stabilizing residues. In this study, we have addressed these issues by analyzing the structures of the $(\alpha/\beta)_8$ barrel proteins (constituted by eight parallel α-strands enveloped by eight α-helices).

The $(\alpha/\beta)_8$ barrel protein is one of the most frequent and regular domain structures of globular proteins.[46–48] This tertiary fold is observed in ~10% of all known enzymes. Despite the similarity in the basic architecture, members of this large family of proteins catalyze very different reactions. Such diversity in function has made this family an attractive target for protein engineering.[49] Several investigations have been performed to understand the principles responsible for the folding and stability of the TIM-barrel fold, in relation to packing of the β-strand residues in the barrel core,[50] overall folding,[51] amino acid clustering patterns,[52] and long-range interactions.[53] Furthermore, the characteristics of TIM-barrel proteins have been reviewed in detail in terms of their structure, folding, function, evolution, distribution, and some of its most remarkable catalytic performances.[54–59]

Although recent studies have focused on the identification of stabilizing residues in TIM-barrel domains,[60] no study has addressed the role and contribution of weak interactions to the overall stability of these TIM-barrel proteins. In this article, we have not only explored the occurrence of the noncanonical and cation–π interactions in TIM-barrel proteins but also systematically investigated the environments in which such interactions occur, and reveal the correlation between the local structural environment and the occurrence of specific types of weak interactions, i.e., the preference in terms of amino acid composition, secondary structural unit, solvent accessibility, and sequential separation.

MATERIALS AND METHODS

Data Set

The data set consisting of 71 TIM-barrel proteins is the same as described by Gromiha et al.[60] The amino acid sequence similarity between any pair of proteins in the database is lower than 20%. The Protein Data Bank (PDB)[61] codes of the TIM-barrel enzymes in our data set are provided in the supplementary Table S-I (http://www.interscience.wiley.com/jpages/0887-3585/suppmat). The complete names of the 71 proteins, along with their respective bibliographic references, are available at http://www.cbrc.jp/~gromiha/tim/proteinlist.html. In case of homo-oligomeric proteins, only one chain has been considered for our analysis.

Identification of Weak Interactions

The type of noncanonical interaction is indicated by a two-letter code in which the first letter indicates the donor atom and the second the acceptor: M, S, and S5 represent the main-chain atom, side-chain atom, and side-chain atom in the five-membered aromatic ring. The different types of NCI involving main-chain to main-chain (MM)-[C—H···O═C], main-chain to side-chain (MS)-[C—H···O═C; C—H···π; N—H···π], main-chain to side-chain five-member aromatic ring (MS5)-[C—H···π], side-chain five-member ring to side-chain (S5S)-[π···π], side-chain to main-chain (SM)-[C—H···O═C], side-chain to side-chain (SS)-[C—H···O═C; C—H···π; N—H···π, π···π], and side-chain to side-chain five-member ring (SS5)-[C—H···π; N—H···π, π···π] were identified using a stand-alone program that identifies each of the aforementioned interactions based on the four different geometric criteria that are defined based on the distances and angles of the atoms under consideration.[62] A few of the interactions could be calculated using the NCI server[62] available at http://www.mrc-lmb.cam.ac.uk/genomes/nci/. In the present investigation, the default parameter values of the NCI server[62] were used to identify these NCI. Occurrence of the canonical hydrogen bonds in the entire TIM-barrel data set were identified using the program HBPLUS.[63]

The number of cation–π interactions in each protein has been calculated using the program CAPTURE developed by Gallivan and Dougherty[31] available at http://capture-.caltech.edu. We have considered only the energetically significant $E_{\text{cat-}\pi} \leq -2$ kcal/mole) interactions in the present study.

Contribution of Individual Amino Acid Residues and Their Sequential Separation

The percentage contribution of each amino acid residue that was identified to participate in the different types of

NCI was calculated as the ratio of the occurrence of a specific amino acid involved in the particular type of noncanonical interaction to the occurrence of the same amino acid in our data set of the 71 TIM-barrel proteins. The amino acid residue pairs involved in the interactions were classified as short-range ($< \pm 3$ residues), medium-range (± 3 or ± 4 residues), or long-range ($> \pm 4$ residues) based on their location in the amino acid sequence.[64,65]

Secondary Structure and Solvent Accessibility

Secondary structure preference and solvent accessibility (ASA) of the amino acid residues are among the key factors that are essential to understand the structure-function relationship of proteins. Secondary structure and accessibility for the amino acid residues were calculated using DSSP.[66] We have systematically analyzed (a) the pattern of preference for the amino acid residues involved in each type of the noncanonical interaction to be present in a particular secondary structure, and (b) their solvent accessibility pattern. The secondary structural units have been classified as helix, strand, turn, and coil in accordance with Heringa and Argos.[67] Solvent accessibility was divided into three classes, 0–20%, 20–50%, and >50%, indicating respectively the least, moderate, and high accessibility of the amino acid residues to the solvent.[68,69] The percentage of an amino acid in a particular ASA class involved in a particular noncanonical interaction was evaluated using the relation

$$\%\mathrm{ASA} = \frac{\mathrm{N_{AA}NCI}}{\mathrm{N_{AA}}} \times 100$$

where N_{AA} NCI and N_{AA} indicate the number of instances a particular amino acid belonging to a specific ASA class is involved in NCI and the total number of instances of the same amino acid (found in that ASA class) in the whole data set, respectively.

Involvement of Stabilizing Residues in Weak Interactions

The stabilizing residues in TIM-barrel proteins have already been identified by Gromiha et al.[60] with certain cutoff values for each measure of stability such as surrounding hydrophobicity, long-range order, stabilizing center, and conservation residues. In the present study, we have identified the frequency occurrence of amino acid residues involved in stabilization and NCI by calculating the ratio of the number of stabilizing and number of noncanonical interaction residues to the total number of amino acids in each protein in the data set. The occurrence ratio of each amino acid residue involved in stabilization to the non-canonical interaction is calculated in order to assess the probability of the amino acids that are predicted to be the stabilizing residue and are also found to involve in the non-canonical interaction.

RESULTS AND DISCUSSION

Weak Interactions in TIM-Barrel Proteins

The representative instances of noncanonical interaction in the TIM-barrel motif of the enzyme fructose 1, 6 bisphosphate aldolase (PDB: 1ado) are given in Figure 1. The interactions of the main-chain to side-chain atoms involving the residue pairs His 80–Phe 79 (NH$\cdots\pi$), Lys 129–Gln 132 (C—H$\cdots$O═C), Gly 28–Phe299 (CH$\cdots\pi$), and that involving only side-chain atoms in Trp 313–Phe57 ($\pi\cdots\pi$) and Trp 313–Tyr 58 ($\pi\cdots\pi$) are indicated. In the complete data set, among the 15 different NCI, MM-CHOC shows the highest contribution of about 35%. Further analysis indicated that the main-chain atoms contribute significantly (about 75%) either as donor or acceptor or both. This global analysis on the data set of TIM-barrel proteins indicates that the distribution of the atoms involved in the noncanonical weak interactions are not random, but are more oriented toward the high incident main-chain atoms. This suggests that the NCI could contribute significantly to the stability of the TIM-barrel proteins. In addition to the main-chain to main-chain interactions, we find high incidence of main-chain to side-chain and side-chain to side-chain interactions in specific TIM-barrel domains containing enzymes. For example, the enzymes diol dehydrogenase (1eex) and monomethylamine methyltransferase (1l2q) indicate significant contribution of the MS-[C—H$\cdots$O═C], MS-[CH$\cdots\pi$], SM-[C—H$\cdots$O═C], SS-[$\pi\cdots\pi$] interactions. Relatively high incidence of MS-[CH$\cdots\pi$] and SS-[$\pi\cdots\pi$] interactions in the majority of the proteins in the data set stresses the importance of aromatic amino acid residues in stabilizing the TIM-barrel scaffold.

Number of Amino Acids in a TIM-Barrel Protein Versus Number of Weak Interactions

It could be contemplated that the observed NCI could be proportional to the total number of amino acids and therefore could be characterized as amino acid independent factor, contributing to the global stability of the enzyme. Similar analysis was performed with respect to cation–π interactions in transmembrane helical (TMH) as well as transmembrane strand (TMS) proteins,[70,71] but the arguments differ in both. Correlation between the number of amino acid residues and the number of interactions is significant in TMS proteins. The present study focuses on the weak interactions, which is much more prevalent in addition to the residue type specific cation–π interactions. The number of amino acid residues was correlated with the number of canonical hydrogen bonds as identified by HBPLUS [Fig. 2(a)], the noncanonical hydrogen bonding interaction [Fig. 2(b)], and cation–π interactions [Fig. 2(c)] in the considered set of TIM-barrel proteins. It could be inferred from the figures [Fig. 2(a–c)] that the correlation of the number of amino acids to the number of nonbonded interactions is different for canonical ($R^2 = 0.76$), noncanonical ($R^2 = 0.55$) hydrogen bonds and cation–π interaction ($R^2 = 0.33$). The lesser regression coefficient in the case of cation–π interactions could be attributable to the low incidence of aromatic amino acid residues in proteins. Relative contribution of the noncanonical and canonical hydrogen bonds to the stability of TIM-barrel domain was also evaluated. The correlation between the number of conventional hydrogen bonding

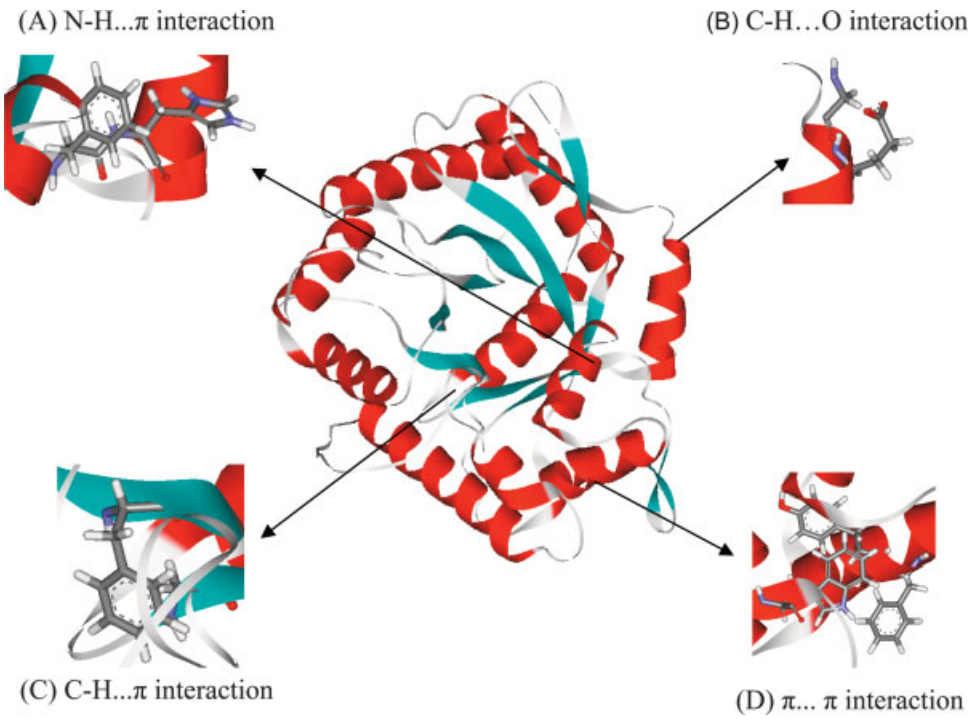

Fig. 1. Representative instances of noncanonical interactions in the enzyme fructose 1,6 bisphosphate aldolase (PDB: 1ado, 1.9 Å resolution) from the glycolytic pathway. **A:** N—H· · ·π interaction involving main-chain N atom of His 80 and side-chain aromatic ring of Phe 79. **B:** C—H· · ·O interactions involving the main-chain C—H of Gly 129 and side-chain O of Glu 132. **C:** C—H· · ·π interaction involving main-chain C—H of Gly 28 and side-chain aromatic ring of Phe 299. **D:** π· · ·π interaction involving the aromatic rings of Trp 313 with Phe 57 and Tyr 58.

and noncanonical hydrogen bonding is given in Figure 2(d). The data (Table I) and Figure 2(d) (slope < 1) indicate that in TIM-barrel proteins, the contribution of the conventional hydrogen bonding is higher than the nonconventional hydrogen bonding interactions.

From Table I it could be inferred that the proteins represented by IDs 2ebn, 1qat, and 1qtw (all belonging to hydrolases) have an identical number of amino acids in their primary structure but display a total of 109, 112, and 87 NCI; 4, 3, and 5 cation–π interactions, respectively. Therefore, in the case of TIM-barrel motifs, less correlation between the number of amino acids and the occurrence of weak interactions exists. Furthermore, the average number of interactions involving only the main-chain atoms is 38 ± 13 whereas the collective contribution of the other interactions involving the side-chain atoms is 68 ± 36. This difference in the number of NCI does indicate that the contribution to the overall stability of TIM-barrel domain is dictated by the amino acid side-chains. Moreover, the measured standard deviation value, especially for the side-chain atoms, implies that the variation in the amino acid composition does contribute to the difference observed in the number of noncanonical interactions. However, the anomalous occurrence of certain types of interactions in specific TIM-barrel domains, in the present data set, cannot be rationalized by mere statistical analysis. For example, the presence of 49 SS–CHOC interactions in malate synthase (1d8c), might be attributed to the local environment of these residues that is essential for the stability/conformation of this enzyme.

Relative Contribution of Amino Acids

If the number of amino acids *per se* does not define the extent of occurrence of NCI, then it is logical to rationalize that the type of amino acid could dictate the number and strength of these interactions. This indirectly would signify the importance of amino acid type/amino acid composition rather than the mere number of amino acid residues present in a given TIM-barrel protein. The relative contribution of each of the amino acid residues as donor and acceptor for each type of interaction was identified and they are tabulated in Table II(a) and II(b), respectively. It could be inferred that only in the MM–CHOC interaction, all the 20 amino acids serve as donor as well as acceptor. As observed in the whole data set (Table I), most of the amino acid residues participate in the weak interactions to a greater percentage through their main-chain atoms which indirectly is dictated by the type of side-chain, irrespective of whether the amino acid is a donor or acceptor. The amino acid residue glycine, which generally induces greater flexibility in the protein chain, seems to have greater apparent involvement in the interaction as indicated by its percentage contribution of 25. This is valid only when glycine is a donor. When glycine is as an acceptor, its contribution is *in par* with other amino acid residues [Table II(a and b)]. This could be attributed to the

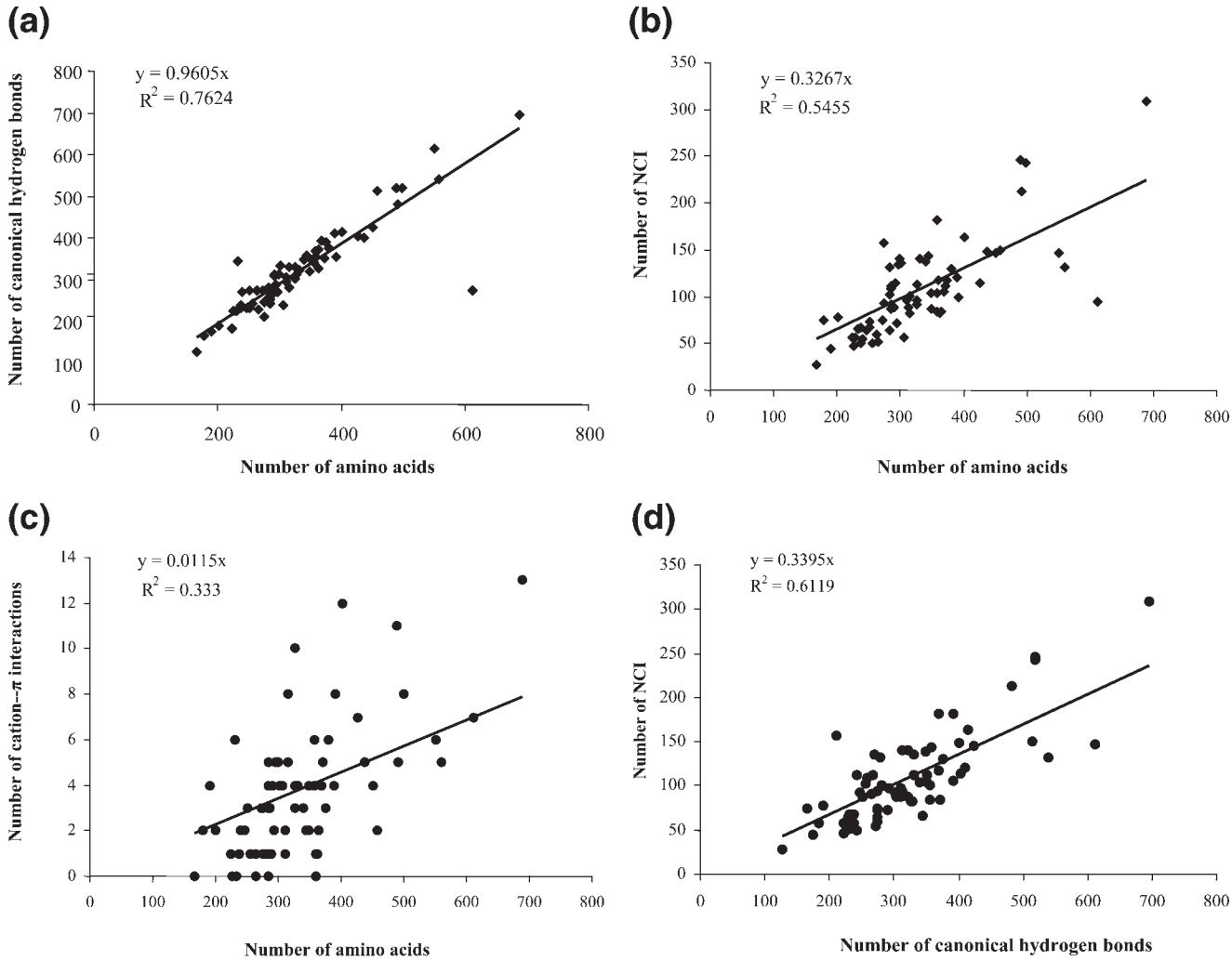

Fig. 2. **a:** Relationship between number of amino acid residues and number of canonical hydrogen bonds in TIM-barrel proteins ($R^2 = 0.76$). **b:** Relationship between number of amino acid residues and number of noncanonical hydrogen bonds in TIM-barrel proteins ($R^2 = 0.55$). **c:** Relationship between number of amino acid residues and number of cation–π interactions in TIM-barrel proteins ($R^2 = 0.33$). **d:** Relationship between number of canonical hydrogen bonds and number of noncanonical interactions in TIM-barrel proteins ($R^2 = 0.61$).

greater probability of the glycine residue to be donor because of the presence of two α protons opposed to one in all the other amino acid residues. It is also interesting to note that the incidence of the aromatic amino acids in NCI is relatively high, with the percentage value as high as 35 for π···π interactions, for both donor and acceptor. The packing of the aromatic amino acid residues with specified geometries is generally considered as one of the major factors contributing to the thermodynamic stability of proteins.[72] In addition, both His and Pro are involved significantly in SM–CHOC interaction. Such CHOC interactions involving His residues have been reported in the active site of serine hydrolases.[25] Similar results were obtained when the amino acid residues that are involved in the NCI were normalized (Supplementary Table S-IIa and S-IIb). We have also estimated the percentage of aromatic and positively charged amino acids that are involved in cation–π interactions in TIM-barrel protein structures. The relative contribution of cationic and aromatic amino acid residues in TIM-barrel proteins is also given in Table II(a and b). We found that percentage contribution of aromatic amino acids in TIM-barrel proteins is more or less similar to that of TMS proteins (Phe 7.5%, Tyr 11.54%, and Trp 19.61%). We also found that, similar to TMS and globular proteins, the percentage contribution of positively charged amino acid Arg (15.17%) is higher than that of Lys (6.17%) in TIM-barrel proteins. Thus, our analysis indicates that the contribution of amino acids toward a particular noncanonical interaction is specific in TIM-barrel proteins.

Distribution of NCI Based on Sequential Separation

The contribution of NCI in TIM-barrel proteins could define either the local or the global stability of the protein. Therefore, to evaluate the contribution of the inter-residual interactions, the sequential distance (short-range: $< \pm 3$ residues, medium-range: ± 3 or 4 residues,

TABLE I. Distribution of Weak Interactions Estimated in TIM-Barrel Proteins

PDB code	No. AA	No. H bonds	Cation–π	MM CHOC	MS CHOC	MS CHPI	MS NHPI	MS5 CHPI	MS5 NHPI	S5S PIPI	SM CHOC	SS CHOC	SS CHPI	SS NHPI	SS PIPI	SS5 CHPI	SS5 NHPI	SS5 PIPI	Total interaction
1a0c	437	400	5	39	2	48	4	1	3	0	8	6	1	1	35	1	0	0	149
1a3x	265	227	0	23	5	8	2	0	0	0	5	3	1	2	3	0	0	0	52
1adl	264	275	1	24	8	9	1	2	0	0	6	3	1	1	3	0	0	2	60
1ado	363	327	1	30	7	16	2	3	1	0	11	3	4	0	4	1	0	0	82
1ads	315	281	8	33	7	13	2	5	2	1	11	7	5	2	10	1	1	0	100
1aq0	612	275	7	38	7	17	0	6	0	0	4	2	4	2	12	1	1	0	94
1b3o	368	238	4	26	6	6	1	0	0	0	7	3	1	1	6	0	0	0	57
1b4k	326	302	3	36	5	17	4	2	1	0	10	3	3	1	10	0	0	0	92
1b54	230	223	6	28	3	12	1	3	1	0	3	2	1	0	3	0	0	0	57
1b5t	275	247	3	30	3	13	0	2	3	0	8	10	2	2	7	0	0	0	80
1bd0	233	344	0	25	2	14	4	5	0	0	5	1	1	0	8	1	0	0	66
1bf6	291	310	4	38	8	15	1	0	0	0	10	3	3	1	9	0	0	0	88
1bpl	201	191	2	14	5	17	2	4	2	0	11	4	2	1	15	0	0	1	78
1bqc	302	332	4	54	10	28	2	9	2	2	8	2	5	0	12	2	0	0	136
1btm	251	231	3	29	8	6	3	1	1	0	8	2	3	0	5	0	0	0	66
1byb	491	482	5	56	13	40	4	11	3	1	13	6	17	2	40	6	1	0	213
1c0d	358	369	6	61	10	24	7	17	7	2	18	7	3	3	21	1	0	1	182
1ceo	332	322	4	39	6	34	1	7	2	3	5	6	6	1	27	0	0	3	140
1cnv	283	255	1	34	7	21	1	6	2	1	7	1	4	1	17	0	0	1	103
1d3g	360	370	0	57	8	24	0	1	0	0	16	4	2	0	5	0	0	0	117
1d8c	688	696	13	73	9	25	0	12	3	1	24	49	3	2	17	2	4	0	224
1dl3	191	174	4	22	2	6	1	2	1	0	4	3	0	0	3	0	0	0	44
1dos	359	355	1	38	2	15	1	3	1	0	5	7	3	2	4	0	1	1	83
1e70	499	520	8	76	9	53	3	8	4	2	13	4	6	1	58	2	1	3	243
1edg	380	377	6	32	4	34	2	10	2	3	6	3	5	4	22	0	1	2	130
1edq	348	351	2	30	5	23	0	3	2	2	8	6	4	1	19	1	0	0	104
1eex	551	613	6	54	20	30	2	3	0	0	10	5	8	3	10	1	1	0	147
1eom	283	279	1	47	8	28	5	6	3	0	11	1	4	2	17	0	0	0	132
1f74	293	313	2	29	11	25	3	1	0	0	6	3	4	1	10	0	0	1	94
1f8m	427	404	7	38	9	20	3	8	2	0	11	4	1	4	13	1	0	0	114
1gg0	275	210	1	39	3	12	2	0	0	0	11	10	18	41	4	0	0	0	140
1gow	489	519	11	62	9	47	4	16	5	2	15	9	14	3	49	7	2	2	246
1gvf	273	274	3	36	3	13	0	2	0	0	7	5	1	1	6	0	1	0	75
1gw1	375	391	3	55	13	30	1	15	4	0	13	3	6	0	37	3	0	1	181
1h7w	312	304	2	43	3	13	1	7	1	0	9	2	0	1	7	0	0	1	88
1hg3	224	184	1	32	6	4	0	1	1	0	10	3	0	0	0	0	0	0	57
1huv	349	321	4	35	8	15	2	7	2	0	7	4	1	0	4	2	0	0	87
1ilw	179	166	2	30	7	12	2	3	0	1	5	6	0	0	9	0	0	0	75
1itq	369	393	4	40	9	18	0	3	2	1	12	2	4	1	12	1	0	0	105
1j5s	451	424	4	29	9	32	2	15	2	1	15	9	8	4	14	4	0	2	146
1jcl	252	275	3	35	7	13	1	1	0	0	3	5	2	0	5	0	0	0	72
1juk	247	230	2	32	3	11	0	0	0	0	5	1	2	1	9	0	0	0	64
1k4g	372	352	5	42	9	21	0	4	1	1	13	5	2	0	14	0	0	0	112
1kb1	364	372	2	34	6	19	1	1	1	0	5	6	0	1	10	0	0	0	84
1kd0	240	271	2	22	4	10	0	2	0	0	2	3	1	4	7	0	0	0	55
1km0	237	232	1	24	3	10	0	0	0	0	7	4	0	0	3	0	0	0	51
1l2q	457	514	2	59	20	25	4	4	1	0	12	8	4	0	11	1	0	1	150
1luc	326	331	4	38	10	20	2	4	1	2	9	1	7	0	17	1	0	1	113
1lwh	391	356	8	30	7	27	4	5	1	0	17	2	4	0	0	1	2	0	100
1nar	289	267	1	36	6	31	1	4	1	0	10	1	1	0	20	0	2	0	113
1onr	316	329	5	26	4	16	1	5	0	0	8	5	5	3	6	2	2	0	83
1pdy	294	289	2	30	5	13	3	4	1	1	5	1	1	0	7	1	0	0	72
1pym	284	275	0	25	3	12	2	4	2	2	5	5	0	0	4	0	0	0	64
1qap	167	126	0	14	2	7	0	1	0	0	2	2	0	0	0	0	0	0	28
1qat	285	242	3	35	6	25	7	4	2	0	10	1	4	1	17	0	0	0	112
1qnr	344	358	2	43	3	22	4	11	3	4	4	1	2	5	34	2	3	2	143
1qo2	238	238	1	36	2	12	0	1	0	0	10	3	1	0	2	0	0	0	67
1qtw	285	251	5	31	5	13	0	5	0	0	9	7	3	0	13	0	1	0	87
1req	559	539	5	38	13	19	5	7	2	0	14	4	7	3	19	0	1	0	132
1smd	402	415	12	49	11	28	1	16	6	1	11	6	2	4	28	0	1	0	164
1tml	286	266	3	40	5	12	0	11	1	0	9	2	0	0	7	2	2	0	91
1ttp	256	242	1	17	4	14	2	0	0	0	5	3	0	0	5	0	0	0	50
1uro	357	340	4	37	6	15	2	3	1	1	18	4	2	5	8	0	2	0	104
2dor	311	293	1	38	8	20	4	0	0	0	9	0	1	0	17	0	0	0	97
2ebn	285	258	4	47	4	25	3	0	0	0	4	4	4	0	17	1	0	0	109
2tmd	340	348	3	42	11	33	1	7	0	1	19	5	2	1	14	2	0	0	138
2tps	226	223	0	25	5	7	0	0	0	0	3	2	0	0	5	0	0	0	47
4ubp	389	411	4	56	7	14	0	4	1	0	23	6	3	1	3	2	0	0	120
5ptd	296	270	5	40	7	24	7	5	1	0	13	3	7	3	24	0	1	0	135
7a3h	300	313	5	51	11	22	1	13	3	1	6	8	3	6	14	0	2	0	141
8ruc	327	311	10	39	7	14	3	6	2	0	7	7	2	2	6	0	2	0	97
Total		22,806	257	2,665	483	1,391	135	342	96	37	653	331	231	132	913	53	35	25	7,522

TABLE II(A). Donor Role of Amino Acid: Percentage Contribution of Different Amino Acids in a Particular Type of Noncanonical Interaction

% Amino acid	Cation–π interactions	MM CHOC	MS CHOC	MS CHPI	MS NHPI	MS5 CHPI	MS5 NHPI	S5S PIPI	SM CHOC	SS CHOC	SS CHPI	SS NHPI	SS PIPI	SS5 CHPI	SS5 NHPI	SS5 PIPI
ALA		8.52	1.62	3.19	*	*	*									
ARG	15.17	8.08	2.11	1.32	*		*					1.93			*	
ASN		9.87	2.06	*	*	*	*					2.51			1.52	
ASP		4.18	*	*	*	*	*									
CYS		13.98	1.43	1.08	1.43		*									
GLN		7.77	0.72	1.31	*	*	*					2.39			*	
GLU		5.86	*	1.14	*	*										
GLY		24.65	4.96	6.98	1.2	*	*									
HIS		11.76	2.94	*	*	*	*		24.45	14.15	10.66			2.02		
ILE		17.2	1.31	*	*	*	*									
LEU		8.73	2.08	*	*	*	*									
LYS	6.17	6.58	2.4	1.63	*	*			10.06	9.83	4.1	1.16		*	*	
MET		11.61	1.07	1.96		*	*									
PHE		15.23	2.21	51.79	3.26	*	*						34.03			1.68
PRO		11.28	4.68	5.07		1.15			30.78	6.88	7.36			2.1		
SER		10.89	2.68	1.64	*	*										
THR		10.74	1.24	*		*										
TRP		8.99	1.69	2.81	1.97	74.16	19.38	9.55					34.83			
TYR		12.66	2.25	48.99	3.91	*	*						26.51			1.07
VAL		15.32	*	1.18	*	*	*									

Blank spaces indicate that the amino acids are not involved in that interaction; *Indicates the <1% contribution of a particular amino acid.

TABLE II(B). Acceptor Role of Amino Acid: Percentage Contribution of Different Amino Acids in a Particular Type of Noncanonical Interaction

% Amino acid	Cation–π interactions	MM CHOC	MS CHOC	MS CHPI	MS NHPI	MS5 CHPI	MS5 NHPI	S5S PIPI	SM CHOC	SS CHOC	SS CHPI	SS NHPI	SS PIPI	SS5 CHPI	SS5 NHPI	SS5 PIPI
ALA		9.56							2							
ARG		10.63							2.81							
ASN		8.07	7.53						3.68	3.14						
ASP		7.55	12.68						4.05	8.7						
CYS		11.83							1.79							
GLN		8.6	7.53						2.51	3.46						
GLU		6.19	9.02						2.29	7.21						
GLY		11.89							3.6							
HIS		13.6							3.31							
ILE		19.88							2.61							
LEU		13.71							2.34							
LYS		10.14							2.63							
MET		12.68							2.5							
PHE	7.50	13.45		62.71	5.78			1.68	2.31		4.94	2.94	35.92			
PRO		8.41							2.1							
SER		9.62							2.83							
THR		10.25							2.48							
TRP	19.61	13.76		25	2.81	85.96	25		4.21		11.79	8.71	33.71	12.36	8.43	6.74
TYR	11.54	13.49		59.29	8.17			2.37	2.37		11.00	3.43	29.23			
VAL		16.75							2.85							

Blank spaces indicate that the amino acids are not involved in that interaction.

and long-range: $> \pm 4$ residues) between the amino acids involved in the various NCI and cation–π interactions was calculated and the result is depicted as percentage occurrence in Figure 3. It is evident that a greater percentage of the interactions is defined by long-range contacts. Short-range interactions dominate only in the case of the main-chain to side-chain interactions involving aromatic amino acid residues. This interaction, specifically involving the aromatic amino acid residues in the present data set, could have a significant role in restricting the dynamics and therefore lower the entropy of the aromatic amino acid residues. The contribution of the medium-range contacts is comparatively less in all the interactions studied. Therefore, all the main-chain to main-chain, side-chain to main-

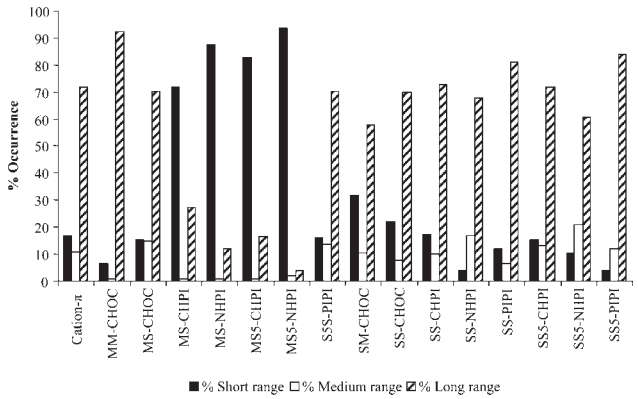

Fig. 3. Percentage occurrence of different noncanonical hydrogen bonding and cation–π interaction ranges in TIM-barrel proteins.

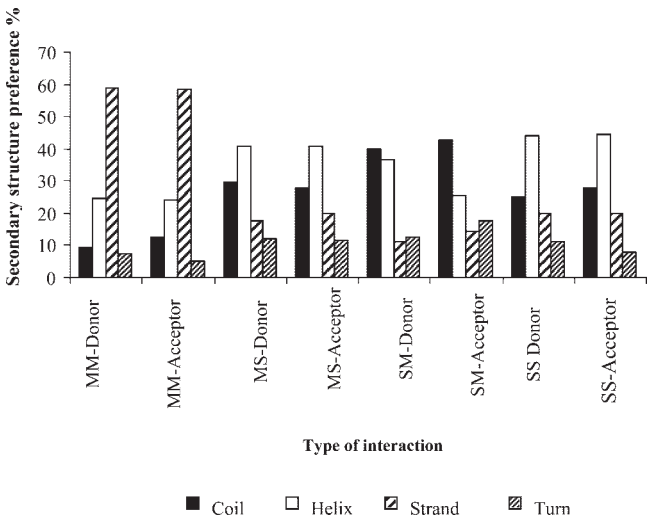

Fig. 4. Percentage of residues in the different secondary structural units that participate in the various types of noncanonical interactions.

chain, and side-chain to side-chains involve only in long-range interactions. Among all the side-chain interactions, SS5-π· · ·π interactions show the highest long-range percentage of occurrence in TIM-barrel proteins. In case of cation–π interactions in TIM-barrel proteins, 17, 11, and 72% of cation–π interactions are contributed by short-, medium-, and long-range interactions. This observation suggested that the majority of the cation–π interactions in TIM-barrel proteins are influenced by long-range contacts as observed in DNA binding proteins.[42] This also indicates that these weak interactions contribute to both local and global conformational stability of TIM-barrel proteins.

Secondary Structure Preference

The occurrence of these weak interactions has been observed at the terminus of the secondary structural units, in particular α-helix and β-sheet.[14,15] These interactions have been proposed to have a definitive role in stabilizing these secondary structural scaffolds of proteins. The propensity of the amino acid residues to favor a particular conformation has been well augmented. Such conformational preference is not only dependent on the amino acid alone but is also dependent on the local amino acid sequence. Therefore, to draw correlation between the occurrences of a particular noncanonical interaction to an amino acid adopting a particular secondary structural fold, we have analyzed the percentage occurrence of the noncanonical interaction in a particular secondary structure, irrespective of the amino acid, and the result is depicted in Figure 4. The amino acid residues (donor and acceptor) involved in main-chain to main-chain interactions prefer the extended conformation. However, in both the main-chain to side-chain and side-chain to side-chain interactions, both donors and acceptors prefer helical conformation whereas only the side-chain to main-chain CHOC interaction prefers to be in a nonhelical and nonextended conformation.

We further analyzed the secondary structure preference for each amino acid that participates in the different types of NCI (Table III). In the whole data set, we did not find any exclusive preference for a particular secondary structure. The majority of the MM–CHOC interactions prefer to occur in the strand, irrespective of the amino acid propensity to adopt a particular secondary structure. To rationalize the existence of these NCI in the strands, we analyzed the total percentage of the secondary structural units in TIM-barrel domains using our data set. However, analysis of the percentage of residues involved in the different secondary structural conformation indicates that the percentage of helices, strands, and turns are 43 ± 7, 15 ± 3, and 13 ± 4, respectively. Thus, the preference of these NCI is dependent on other complex factors such as the positioning and the neighborhood of the amino acid residues, which is necessary for the stability of the strands. Except for the SS–CHOC and SS-π· · ·π, the remaining side-chain to side-chain interactions were found to be not significantly selective (<10% occurrence) to any particular secondary structure. However it is interesting to note that there is a preference for strand in both SS–CHOC interactions that involve the acidic-basic amino acid pairs. Also, the side-chain of the aromatic residues interacts predominantly when they occur in strands. These interacting pairs were found to be in the adjacent strands thereby contributing to the stability of the β-sheet scaffold in TIM-barrel proteins. Fabiola et al.[14] and Derewenda et al.[6] have described that similar cross-strand interactions between MM–CHOC are ubiquitous and therefore could act as an additional stabilizing factor for the β-sheets. The analysis thus indicates that the greater number of NCI occur in the strand although no such correlation could be drawn for amino acids preferring α-helices. We also found that, as in DNA binding proteins,[42] cation–π interaction forming cationic residues prefer to be in β-strands of TIM-barrel proteins. Phe prefers to be in coil, Tyr in turn, and Trp prefer to be in helix. This analysis indicates that, at least in the case of TIM-barrel fold, the weak interactions do not occur at random but have residue-specific preference for a particular secondary structure.

TABLE III. Secondary Structure Preference for the Residues Involved in Weak Interactions

Amino acid	Cation–π		MM CHOC		MS–CHOC		MS–CHPI		MS–NHPI		MS5–CHPI		MS5–NHPI		S5S–PIPI		SM–CHOC		SS–CHOC		SS–CHPI		SS–NHPI		SS–PIPI		SS5–CHPI		SS5–NHPI		SS5–PIPI	
	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc
ALA			S	S	*		*		*		*		*					*															
ARG	S		S	S	*		*		*		*		*					*					*							*			
ASN			S	S	*	S	*		*		*		*					*		*			*							*			
ASP			S	S	*	C	*		*		*		*					*		S													
CYS			S	S	*		*		*		*		*					*															
GLN			S	S	*	S	*		*		*		*					*		*			*							*			
GLU			S	S	*	S	*		*		*		*					*		S													
GLY			S	S	*		S		*		*		*					*															
HIS			S	S	*		*		*		*		*				*	*	S		Eq						*						
ILE			S	S	*		*		*		*		*					*															
LEU			S	S	*		*		*		*		*					*															
LYS	S		S	S	S		*		*		*		*				*	*	S		*		*				*	*					
MET			S	S			*		*		*		*					*															
PHE		C	S	S	*		Eq	Eq	*	*	*		*			*		*				*		*	S	S					*		
PRO			S	S	*		*		*		*						*	*	H		C						*						
SER			S	S	*		*		*		*							*															
THR			S	S	*		*		*		*							*															
TRP		H	S	S	*		*	Eq	*	*	S	S	T	T	*	*		*				T		*	S	C		*		*			*
TYR		T	S	S	*		Eq	S	*	*	*		*					*				H		*	S	S					*		
VAL			S	S	*		*		*		*		*					*															

Don, donor; Acc, acceptor; H, helix; C, coli; S, strand; T, turn; Eq, more or less equally distributed; *Insignificant involvement in a specific secondary structure and blank space shows that the particular amino acid will not participate in that interaction.

TABLE IV. Solvent Accessibility Preference for the Amino Acid Involved in Weak Interaction

Amino acid	Cation–π		MM–CHOC		MS–CHOC		MS–CHPI		MS–NHPI		MS5–CHPI		MS5–NHPI		S5S–PIPI		SM–CHOC		SS–CHOC		SS–CHPI		SS–NHPI		SS–PIPI		SS5–CHPI		SS5–NHPI		SS5–PIPI	
	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc
ALA			B	B	B		B		*		B		*					B															
ARG	B		E	E	E		E		*		*		*					E					E							E			
ASN			B	B	B	B	B		*		*		*					E		B			B							B			
ASP			B	B	E	B	B		P		*		*					B		B													
CYS			B	B	B		B		*		*		*					B															
GLN			B	B	E	B	E		*		*		*					E		E			B							B			
GLU			B	B	E	B	E		*		*							E		E													
GLY			B	B	B		B		B		B		*					B															
HIS			B	B	B		B		*		*		*				B	B	B		B						B						
ILE			B	B	B		B		*		*		*					B															
LEU			B	B	B		B		*		*							B															
LYS	B		E	E	E		E		*		*		*				E	E	E		E		E				E						
MET			B	B	B		B				*		*					B															
PHE		P	B	B	B		B	B	B	B	*					B		B				B		B	B	B					B		
PRO			B	B	B		B				Eq						B	E	B		B						B						
SER			B	B	B		B		B		*							B															
THR			B	B	B		B				*							B															
TRP		E	B	B	B		B	B	B	B	B	B	B	B	B			B				B		B	B	B		B		B			B
TYR		P	B	B	B		B	B	B	B						B		B				B		B	B	B					P		
VAL			B	B	B		B				*							B															

Don, donor; Acc, acceptor; B, buried (0–20% ASA); P, partially buried (20–50% ASA); E, exposed (>50% ASA); Eq, equally distributed. Blank space shows that the particular amino acid does not participate in that interaction; *Less than five interactions involving residues which were deemed to be statistically insignificant.

Solvent Accessibility of Residues Involved in Weak Interactions

We have estimated the solvent accessibility of all residues that are involved in various types of noncanonical and cation–π interactions with the aid of DSSP.[64] The relation between the amino acid residues in these interactions and solvent accessibility is given in Table IV. The solvent accessibility of amino acid residues has been categorized as buried (0–20%), partially buried (20–50%), and exposed (>50%). Amino acid residues that occur in less than five NCI were deemed as statistically insignificant and therefore were not included in the analysis. Most of the other amino acid residues that were involved in NCI prefer to be in the solvent excluded environment, especially when the interaction involves main-chain atoms. The data indicate that the basic amino acid residues prefer to be solvent exposed when they are involved in a noncanonical interaction. However, by analyzing the percentage of cation–π interactions forming residues at various categories of solvent accessibility, we observed that 14.91% of Lys and 25.77% of Arg prefer to be in the buried region. However, the aromatic amino acids Phe and Tyr (13.29% and 15.03%, respectively) prefer to be partially buried, whereas Trp (28.41%) prefers to be in surface. Glu and Gln prefer to be in the protein surface for all types of NCI, except for MM–CHOC interactions. However, as pre-

dicted, the aromatic amino acid residues Phe, Tyr, and Trp that serve as both donor and acceptor in the NCI, prefer to be in the buried environment. This result coincides with the fact that tryptophan residues in C—H· · ·π interaction prefer to be in the interior of protein.[23] Therefore, this analysis indicates that the majority of the amino acid residues prefer to involve in noncanonical interaction only when they are excluded from the solvent. It is quite possible that the competition between the noncanonical interactions among the amino acid residues and the solvation patterns might contribute to the absence of these weak interactions in the solvent exposed amino acid residues, because the contribution to the global energy is much greater in solvation than that of the weak NCI.

Stabilization Residues Versus Weak Interactions

It would also be useful to identify any patterns of correlation between the number of weak interactions in a given TIM-barrel domain and the theoretically predicted stabilizing residues.[60] Therefore, the frequency occurrence of stabilizing and noncanonical interaction involving residues, in the whole TIM-barrel data set, was calculated and is shown in Figure 5. It is obvious from the figure that there is no direct correlation between the frequency of stabilization residues and the frequency of NCI. However, considering the whole data set, of 957 stabilizing residues,[60] 728 residues (76.07%) are involved in the NCI. Interestingly, of 728 noncanonical interaction residues which act as stabilizing residues, about 70% (671 residues) of them are involved in the MM–CHOC interactions (data not shown). Table V gives the percentage contribution of the individual amino acid residues that are involved in NCI and cation–π interactions. It could be observed that the contribution of the long chain polar residues is higher than that of the other amino acid residues. Therefore, this analysis reveals that the weak interactions contribute to the stability of the TIM-barrel domain.

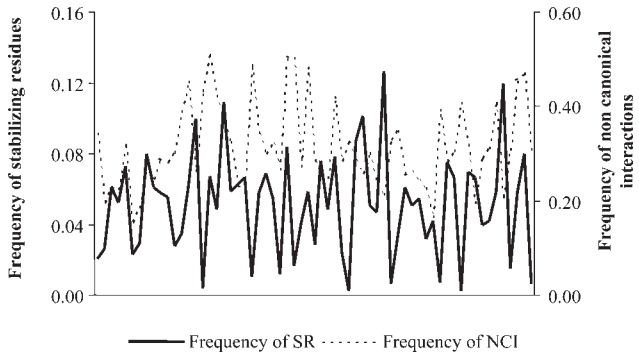

Fig. 5. Frequency of the stabilizing residues and the noncanonical interactions observed for all the TIM-barrel domains in the data set (*x*-axis).

CONCLUSIONS

The present work on the environmental preference of NCI in TIM-barrel proteins clearly indicates that greater contributions to interactions are observed for main-chain to main-chain and those involving aromatic amino acid residues. Moderate correlation between the number of amino acids and number of noncanonical interactions indicates that the amino acid composition/sequence could have a crucial role in dictating the NCI rather than the mere number of amino acids. However, more rigorous analysis is required to strengthen this view. A majority of the main-chain to main-chain, side-chain to main-chain, and side-chain to side-chain pairs involved in a noncanonical interaction tend to be long-range interactions. Second-

TABLE V. Amino Acid Percentage Contribution of Stabilizing Residues in Weak Interactions

Amino acid	Total number of stabilizing residues	Number of stabilizing residues in NCI	% of stabilizing residues in NCI	Number of stabilizing residues in cation–π	% of stabilizing residues in cation–π
ALA	88	55	62.50		
ARG	17	9	52.94	1	5.88
ASN	27	19	70.37		
ASP	17	9	52.94		
CYS	13	9	62.23		
GLN	15	15	100.00		
GLU	28	25	89.29		
GLY	102	80	78.43		
HIS	14	11	78.57		
ILE	125	92	73.60		
LEU	96	68	70.83		
LYS	13	12	92.31	2	15.38
MET	26	15	57.69		
PHE	53	39	73.58	1	1.89
PRO	38	22	57.89		
SER	46	37	80.43		
THR	43	21	48.84		
TRP	14	8	57.14	1	7.14
TYR	28	21	75.00	2	7.14
VAL	154	104	67.53		

Blank spaces indicate the amino acids are not involved in that interaction.

ary structure preference of the noncanonical interaction in TIM-barrel protein shows that about 58% of main-chain to main-chain interaction residues prefer to be in a strand. The preference of a particular secondary structure depends on the interaction type and the amino acid involved. In the TIM-barrel protein data set, the positively charged amino acids, Arg and Lys, involved in noncanonical hydrogen bonding interactions prefer to be in the solvent exposed surface but the aromatic amino acid (Phe, Tyr, and Trp) prefers the buried regions of the protein. Although there is no one-to-one correlation between the occurrence of stabilizing residues and the number of NCI, considering the whole data set of TIM-barrel motifs, we find higher probability of the polar amino acid with long chain to contribute to stability and participate in these weak interactions. Overall, the study does indicate that weak interactions, like the conventional hydrogen bonds, are environment specific and therefore could be a necessary force for both the local and global stability of TIM-barrel proteins.

ACKNOWLEDGMENTS

S.C., G.J., and R.S. thank the management of Vellore Institute of Technology for providing the infrastructure to perform this work. M.M.B. gratefully acknowledges the Intramural Research Program of the National Institutes of Health, USA, for funding his research.

REFERENCES

1. Gromiha MM, Selvaraj S. Recent research developments in protein folding, stability, and design. Trivandrum, India: Research Signpost; 2002.
2. Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984;44:97–179.
3. Wulf OR, Liddel U, Hendricks SB. The effect of ortho substitution on the absorption of the OH group of phenol in the infrared. J Am Chem Soc 1936;58:2287–2293.
4. McPhail AT, Sim GA. Hydroxyl-benzene hydrogen bonding. An X-ray study. Chem Commun 1965:124–125.
5. Sutor DJ. The C—H· · ·O hydrogen bonds in crystals. Nature 1962;195:68–69.
6. Derewenda ZS, Lee L, Derewenda U. The occurrence of C—H· · ·O hydrogen bonds in proteins. J Mol Biol 1995;252:248–262.
7. Bella J, Berman HM. Crystallographic evidence for Cα—H· · ·O═C hydrogen bonds in a collagen triple helix. J Mol Biol 1996;264:734–742.
8. Desiraju GR, Steiner T. The weak hydrogen bond in structural chemistry and biology. Oxford: Oxford University Press; 1999.
9. Steiner T, Koellner G. Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. J Mol Biol 2001;305:535–557.
10. Levitt M, Perutz MF. Aromatic rings act as hydrogen bond acceptors. J Mol Biol 1998;201:751–754.
11. Burley SK, Petsko GA. Aromatic–aromatic interaction—a mechanism of protein-structure stabilization. Science 1985;229:23–28.
12. Toth G, Watts CR, Murphy RF, Lovas S. Significance of aromatic-backbone amide interactions in protein structure. Proteins 2001; 43:373–381.
13. Taylor R, Olga K. Crystallographic evidence for the existence of C—H· · ·O, C—H· · ·N, and C—H· · ·Cl hydrogen bonds. J Am Chem Soc 1982;104:5063–5070.
14. Fabiola GF, Krishnaswamy S, Nagarajan V, Pattabhi V. C—H· · ·O hydrogen bonds in beta sheets. Acta Crystallogr D Biol Crystallogr 1997;53:316–320.
15. Babu MM, Singh KS, Balaram P. A C—H· · ·O hydrogen bond stabilized polypeptide chain reversal motif at the C terminus of helices in proteins. J Mol Biol 2002;322:871–880.
16. Jiang L, Lai L. C—H· · ·O hydrogen bonds at protein–protein interfaces. J Biol Chem 2002;277:37732–37740.
17. Bella J, Humphries MJ. Calpha-H· · ·O═C hydrogen bonds contribute to the specificity of RGD cell-adhesion interactions. BMC Struct Biol 2005;5:4.
18. Singh SK, Babu MM, Balaram P. Registering alpha-helices and beta-strands using backbone C—H· · ·O interactions. Proteins 2003;51:167–171.
19. Steiner T, Saenger W. The ordered water cluster in vitamin B12 coenzyme at 15K is stabilized by C—H· · ·O hydrogen bonds. Acta Crystallogr D Biol Crystallogr 1993;49:592–593.
20. Chakrabarti P, Chakrabarti S. C—H· · ·O hydrogen bond involving proline residues in alpha-helices. J Mol Biol 1998;284:867–873.
21. Senes A, Ubarretxena-Belandia I, Engelman DM. The C—H· · ·O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. Proc Natl Acad Sci USA 2001;98:9056–9061.
22. Ghosh A, Bansal M. C—H· · ·O hydrogen bonds in minor groove of A-tracts in DNA double helices. J Mol Biol 1999;294:1149–1158.
23. Kryger G, Silman I, Sussman JL. Structure of acetylcholinesterase complexed with E2020: implications for the design of new anti-Alzheimer drugs. Structure 1999;7:297–307.
24. Parkinson G, Gunasekera A, Vojtechovsky J, et al. Aromatic hydrogen bond in sequence specific protein DNA recognition. Nat Struct Biol 1996;3:837–841.
25. Derewenda ZS, Derewenda U, Kobos PM. (His) C^{α}—H· · ·O═C hydrogen bond in the active sites of serine hydrolases. J Mol Biol 1994;241:83–93.
26. Armstrong KM, Fairman R, Baldwin RL. The (i, i +4) Phe-His interaction studied in an alanine-based alpha-helix. J Mol Biol 1993;230:284–291.
27. Scheiner S, Kar T, Gu Y. Strength of the C—H· · ·O hydrogen bond of amino acid residues. J Biol Chem 2001;276:9832–9837.
28. Vargas R, Garza J, Dixon DA, Hay BP. How strong is the C—H· · ·O═C hydrogen bond? J Am Chem Soc 2000;122:4750–4755.
29. Levitt M, Perutz MF. Aromatic rings act as hydrogen bond acceptors. J Mol Biol 1998;201:751–754.
30. Duan G, Smith VH Jr, Weaver DF. An ab initio and data mining study on aromatic-amide interactions. Chem Phys Lett 1999;310: 323–332.
31. Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. Proc Natl Acad Sci USA 1999;96:9459–9464.
32. Dougherty DA. Cation-π interaction in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science 1996;271:163–168.
33. Pletneva EV, Laederach AT, Fulton DB, Kostic NM. The role of cation-π interaction in biomolecular association. Design of peptides favoring interaction between cationic aromatic amino acid side-chains. J Am Chem Soc 2001;123:6232–6245.
34. Tatko CD, Waters ML. The geometry and efficacy of cation-π interaction in a diagonal position of a designed α-harpin. Protein Sci 2003;12:2443–2452.
35. Burghardt TP, Juranic N, Macura S, Ajtai K. Cation-π interaction in a folded polypeptide. Biopolymers 2002;63:261–272.
36. Chakravarthy S, Varadarajan R. Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. Biochemistry 2000;41:8152–8161.
37. Gromiha MM, Thomas S, Santhosh C. Role of cation-π interaction to the stability of thermophilic proteins. Prep Biochem Biotech 2002;32:355–362.
38. Shi Z, Olson CA, Kallenbach NR. Cation-π interaction in model alpha-helical peptides. J Am Chem Soc 2002;124:3284–3291.
39. Mulhern TD, Lopez AF, Andrea RJD, Gaunt C, Vandeleur L, Vadas MA. The solution structure of the cytokine-binding domain of the common beta-chain of the receptors for granulocyte-macrophage colony-stimulating factor, interleukin-3, and interleukin-5. J Mol Biol 2000;297:989–1001.
40. Gromiha MM. Influence of cation-π interaction in different folding types of membrane proteins. Biophys Chem 2003;103:251–258.
41. Gromiha MM, Santhosh C, Suwa M. Influence of cation-π interactions in protein–DNA complexes. Polymer 2004;45:633–639.
42. Gromiha MM, Santhosh C, Ahmad S. Structural analysis of cation-π interactions in DNA binding proteins. Int J Biol Macromol 2004;34:203–211.
43. Derewenda ZS, Lee L, Derewenda U. The occurrence of C—H· · ·O hydrogen bonds in proteins. J Mol Biol 1995;252:248–262.

44. Brandl M, Weiss MS, Jabs A, Suhnel J, Hilgenfeld R. C—H· · ·PI-interactions in proteins. J Mol Biol 2001;307:357–377.
45. Steiner T, Koellner G. Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. J Mol Biol 2001;305:535–557.
46. Farber GK, Petsko GA. The evolution of α/α barrel enzymes. Trends Biochem Sci 1990;15:228–233.
47. Branden CI. The TIM barrel—the most frequently occurring folding motif in proteins. Curr Opin Struct Biol 1991;1:978–983.
48. Janecek S, Bateman A. The parallel $(\alpha/\alpha)_8$-barrel: perhaps the most universal and the most puling protein folding motif. Biol Bratislava 1991;51:613–628.
49. Raine ARC, Scrutton NS, Mathews FS. On the evolution of alternate core packing in eightfold α/α barrels. Protein Sci 1994;3: 1889–1892.
50. Lesk AM, Branden CI, Chothia C. Structural principles of alpha/beta barrel proteins: the packing of the interior of the sheet. Proteins 1989;5:139–148.
51. Chou KC, Carlacci L. Energetic approach to the folding of alpha/beta barrels. Proteins 1991;9:280–295.
52. Selvaraj S, Gromiha MM. An analysis of amino acid clustering pattern in $(\alpha/\alpha)_8$ barrel proteins. J Protein Chem 1998;17:407–415.
53. Selvaraj S, Gromiha MM. Importance of long-range interactions in $(\alpha/\alpha)_8$ barrel fold. J Protein Chem 1998;17:691–697.
54. Pujadas G, Palau J. TIM barrel fold: structural, functional, and evolutionary characteristics in natural and designed molecules. Biol Bratislava 1999;54:231–254.
55. Pujadas G. TIM barrel research in the genomic and proteomic era. In: Gromiha MM, Selvaraj S, editors. Recent research developments in protein folding, stability, and design. Trivandrum, India: Research Signpost; 2002. p 37–55.
56. Nagano N, Orengo CA, Thornton JM. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures, and functions. J Mol Biol 2002;321:741–765.
57. Selvaraj S, Gromiha MM. Role of inter-residue interactions in the folding of $(\alpha/\alpha)_8$ barrel proteins. In: Gromiha MM, Selvaraj S, editors. Recent research developments in protein folding, stability, and design. Trivandrum, India: Research Signpost; 2002. p 57–66.
58. Sterner R, Hocker B. Catalytic versatility, stability, and evolution of the $(\alpha/\alpha)_8$-barrel enzyme fold. Chem Rev 2005;105:4038–4055.
59. Anantharaman V, Aravind L, Koonin EV. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. Curr Opin Chem Biol 2003;7:12–20.
60. Gromiha MM, Pujadas G, Magyar C, Selvaraj S, Simon I. Locating the stabilizing residues in $(\alpha/\alpha)_8$ barrel proteins based on hydrophobicity, long-range interactions, and sequence conservation. Proteins 2004;55:316–329.
61. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
62. Babu MM. NCI: a server to identify non-canonical interactions in protein structures. Nucleic Acids Res 2003;31:3345–3348.
63. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
64. Gromiha MM, Selvaraj S. Influence of medium and long range interactions in different structural classes of globular proteins. J Biol Phys 1997;23:151–162.
65. Gromiha MM, Selvaraj S. Inter residue interaction in protein folding and stability. Prog Biophys Mol 2004;86:235–277.
66. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1993;22:2577–2637.
67. Heringa J, Argos P. Strain in protein structures as viewed through nonrotameric side chains. I. Their position and interaction. Proteins 1989;37:30–43.
68. Gilis D, Rooman M. Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials. J Mol Biol 1996;257:1112–1126.
69. Gilis D, Rooman M. Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. J Mol Biol 1997;272:276–290.
70. Gromiha MM. Influence of cation-π interaction in different folding types of membrane proteins. Biophys Chem 2003;103:251–258.
71. Gromiha MM, Suwa M. Structural analysis of residues involving cation-π interactions in different folding types of membrane proteins. Int J Biol Macromol 2005;35:55–62.
72. Hunter CA, Sanders JKM. The nature of π–π interactions. J Am Chem Soc 1990;112:5525–5534.