Significance of CH/π Interactions on the Stability of Therapeutic Proteins

Ramanathan K., Shanthi V. and Rao Sethumadhavan*

School of Biotechnology, Chemical and Biomedical Engineering, Bioinformatics Division, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India

Abstract: A computational analysis on the CH/π interactions in a group of 49 therapeutic proteins was investigated. A total of 77 CH/π interactions were observed. The donor atom contribution to CH/π interactions was mainly from aromatic residues. Long-range CH/π interactions are the predominant type of interactions in therapeutic proteins data set. The secondary structure preference, solvent accessibility and stabilization centers of these of CH/π interacting residues were estimated. 73% of the donor residues and 65% of the acceptor residues were highly conserved.

Keywords: CH/π interactions, secondary structure, long range interactions, solvent accessibility, stabilization centers, conservation.

INTRODUCTION

The importance of conventional interactions such as, hydrogen bonds, salt bridges hydrophobicity and other standard interactions in the stabilization of secondary structures [1], protein folding and stability [2, 3] are well established [4-6]. With the recent advances in computational biology one can assess the effect of non-standard interactions on the stability of protein tertiary structure. Among the non-conventional interactions, there is little data on the contribution of CH/π interactions to protein stability. The exothermic dissolution of benzene and similiar compounds (π -electron system: proton acceptor) in chloroform (C–H group: proton donor) was perhaps the origin of an interaction, now known as CH/π interactions, a form of weak hydrogen bond [7]. In 1957, Reeves and Schneider showed by NMR that this interaction was a type of H-bond [8]. Since then, CH/π interactions have been described in a vast number of small molecule systems from simple olefinic and aromatic compounds to complicated clathrates and inclusion complexes. In 1998 Nishio *et al.* published excellent treatise of these observations [9]. In this way CH/π interactions are gradually gaining a lot of importance.

The cases in which CH/π interactions have been described in proteins include the formation of complexes of proteins with special ligands or cofactors such as the heme group [9], pyridoxal-5'-phosphate [10], nucleotides [11, 12], carbohydrates [13] and bound peptides [14], or special geometric circumstances, for instance between neighboring side-chains around a *cis* peptide bond [15]. The importance of this interaction has also been recognized in the design of serine protease inhibitors [16, 17]. There are also recent reviews [18-20] and monographs [9] where the role of CH/π interactions in the structure of chemical and biological macromolecules are described. These interactions also play an important role in the interaction between protein and lipid membranes [11].

These developments motivated us to study the relation between occurrences of CH/π interactions within the protein to the structural stability. To the best of the authors' knowledge, such interactions in therapeutic proteins, which include anticancer antibodies, such as rituximab, used for B-cell non-Hodgkin's lymphoma data set is not yet available. Hence, in this work an effort has been made to collect the information concerning CH/π interactions in the therapeutic proteins data set. In addition, we have chosen only one chain in the therapeutic proteins structure. These represent relatively simpler systems in which all the weaker interactions can be studied in the absence of the effects of a complex quaternary structure and the occurrence of redundancy in the data set. We emphasize that 21 therapeutic proteins in our data set showed a CH/π interactions and hence we accentuate that this investigation is very significant in the sense that, CH/π interactions in therapeutic proteins do play a major role in structural stability of these proteins. It is noteworthy to mention here, that our observations on the results of CH/π interactions was completely different with earlier report on RNA binding protein, in a sense that, only aromatic amino acids takes part in CH/π interactions in therapeutic proteins data set. Ours is the first report on computational investigation on the CH/π interactions in therapeutic proteins.

MATERIALS AND METHODS

Data Set

We have considered a set of 49 therapeutic proteins from the Protein Data Bank [21] for our investigation the details of which are given in Table **1**. According to the structural classification of proteins, 42% of this protein comes under alpha group, 29% comes under beta 11% comes under alpha and beta and remaining18% comes under small proteins in the therapeutic protein data set.

*Address correspondence to this author at the School of Biotechnology, Chemical and Biomedical Engineering, Bioinformatics Division, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India; Tel: +91 4162202522; Fax: +91 4162243092; E-mail: rsethumadhavan@vit.ac.in

1874-1991/09 **2009 Bentham Open**

Table 1. Details of Therapeutic Proteins Used in this Work

PDB ID	Title	Number of Residues
1BML	Complex of the catalytic domain of human plasmin and 2 streptokinase	250
1BMP	Bone morphogenetic protein-7	104
1C4P	Beta domain of streptokinase	137
1CA9	Structure of TNF receptor associated factor 2 in complex 2 with a peptide from tnf-r2	191
1CD9	2:2 complex of g-csf with its receptor	171
1CN4	Erythropoietin complexed with extracellular domains of 2 erythropoietin receptor	217
1DDJ	Crystal structure of human plasminogen catalytic domain	247
1EER	Crystal structure of human erythropoietin complexed to its 2 receptor at 1.9 angstroms	166
1ES7	Complex between bmp-2 and two bmp receptor ia ectodomains	104
1GNC	Structure and dynamics of the human granulocyte colony-2 stimulating factor determined by nmr spectroscopy. loop 3 mobility in a four-helix-bundle protein	175
1HTZ	Crystal structure of tem52 beta-lactamase	263
1L4D	Crystal structure of microplasminogen-streptokinase alpha 2 domain complex	249
1L4Z	X-ray crystal structure of the complex of microplasminogen 2 with alpha domain of streptokinase in the presence cadmium 3 ions	248
1L6X	Fc fragment of rituximab bound to a minimized version of 2 the b-domain from protein a called z34c	241
1M47	Crystal structure of human interleukin-2	122
1M48	Crystal structure of human il-2 complexed with (r)-n-[2-[1-2 (aminoiminomethyl)-3-piperidinyl]-1-oxoethyl]-4-title 3 (phenylethynyl)-l-phenylalanine methyl ester	123
1M49	Crystal structure of human interleukin-2 complexed with sp-2 1985	123
1M4C	Crystal structure of human interleukin-2	229
1N8Y	Crystal structure of the extracellular region of rat her2	592
1N8Z	Crystal structure of extracellular domain of human her2 2 complexed with herceptin fab	214
1NBP	Crystal structure of human interleukin-2 y31c covalently modified at c31 with 3-mercapto-1-(1,3,4,9-tetrahydro-b-carbolin-2-yl)-propan-1-one	121
1PGG	Prostaglandin h2 synthase-1 complexed with 1-(4-iodobenzoyl)-5-methoxy-2-methylindole-3-acetic acid iodoindomethacin), trans model	551
1PGR	2:2 complex of g-csf with its receptor	161
1PW6	Low micromolar small molecule inhibitor of il-2	121
1PY2	Structure of a 60 nm small molecule bound to a hot spot on il-2	117
1QQR	Crystal structure of streptokinase domain b	138
1QRZ	Catalytic domain of plasminogen	246
1QVN	Structure of sp4160 bound to il-2 v69a	123
1R46	Structure of human alpha-galactosidase	390
1R47	Structure of human alpha-galactosidase	390
1REU	Structure of the bone morphogenetic protein 2 mutant l51p	103
1REW	Structural refinement of the complex of bone morphogenetic protein 2 and its type ia receptor	103
1RHG	The structure of granulocyte-colony-stimulating factor and its relationship to those of other growth factors2	145
1RJX	Human plasminogen catalytic domain, k698m mutant	243
1TPG	F1-g module pair residues 1-91 (c83s) of tissue-type plasminogen activator (t-pa) (nmr, 298k, ph2.95, representative structure)	91
1WAQ	Crystal structure of human growth and differentiation factor 5 (gdf-5)	104

(Table 1) contd…..

PDB ID	Title	Number of Residues
1YY8	Crystal structure of the fab fragment from the monoclonal antibody cetuximab/erbitux/imc-c225	213
1YY9	Structure of the extracellular domain of the epidermal growth factor receptor in complex with the fab fragment of cetuximab/erbitux/imc-c225	613
1Z92	Structure of interleukin-2 with its alpha receptor	121
2B5I	Cytokine receptor complex	120
3BMP	Human bone morphogenetic protein-2 (bmp-2)	106
2ERJ	Crystal structure of the heterotrimeric interleukin-2 receptor in complex with interleukin-2	130
2GMF	Human granulocyte macrophage colony stimulating factor	121
2GOO	Ternary complex of bmp-2 bound to bmpr-ia-ecd and actrii-ecd	103
2H62	crystal structure of a ternary ligand-receptor complex of bmp-2	104
2H64	Crystal structure of a ternary ligand-receptor complex of bmp-2	105
2IWG	Complex between the pryspry domain of trim21 and igg fc	207
2OSL	Crystal structure of rituximab fab in complex with an epitope peptide	213
3INK	Unraveling the structure of interleukin-2: reply	122

CH/π Interactions

CH/π interactions are calculated using the program available for this purpose called HBAT [22]. The CH/π interactions considered here were between all possible donor C-H groups in protein structures (C_α-H, C_{ali}-H and C_{aro}-H) and between all side-chain π systems (the aromatic rings of Phe, Tyr, Trp and His). The positions and geometry of donor and acceptor atom are shown in Fig. (**1**). The donor group is represented as C-H and the acceptor is the π system. The distances are usually measured from the centroid (M) ie, centre of the π ring. P1 and P2 are distances from C and H, respectively, to M. P3 is the angle between vectors C-H and H-M while P4 is the angle between the CM and MN. Here N is a normal to the centre of the π ring. The geometry is adapted from earlier work of babu [23]. The CH/π interaction types are represented 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, respectively. We classified the CH/π interactions into four types of CH/π interactions namely, main-chain to side-chain CH/π interactions (MS-CH/π), main-chain to side-chain five member aromatic ring CH/π interactions (MS5-CH/π), side-chain to sidechain CH/π interactions (SS-CH/π) and side-chain to side-chain five member aromatic ring CH/π interactions (SS5-CH/π) [23].

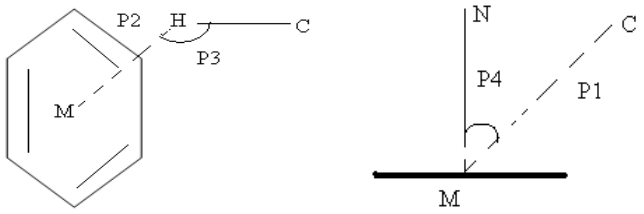

Fig. (1). Parameters such as lengths (P_1 and P_2) and angles (P_3 and P_4) for estimating CH/π interactions.

Secondary Structure and Solvent Accessibility

Secondary structure and solvent accessibility are the two major intermediate steps to understand the structure and function of proteins. Hence a systematic analysis of each CH/π interaction forming residue was performed based on their location in different secondary structures of therapeutic proteins and their solvent accessibility. We obtained the information about secondary structures and solvent accessibility of the proteins using the program DSSP [24]. Solvent accessibility was divided into three classes, buried, partially buried and exposed indicating, respectively, the least, moderate and high accessibility of the amino acid residues to the solvent [25, 26].

Sequential Distance

The CH/π interacting residues coming within a sphere of 8Å was computed as described earlier [27-29]. The residues coming within a sphere of 8Å was computed as described earlier [30]. For a given residue, the comparison of the surrounding residue is analyzed in terms of the location at the sequence level. The contribution from <±4 are treated as short-range contacts, >4 to <±20 as medium-range contacts and >20 are treated as long-range contacts [30]. This classification enables us to evaluate the contribution of short-range, medium-range and long-range contacts in the formation of CH/π interactions.

Stabilization Centers

Stabilization centers are clusters of residues that are involved in medium or long-range interactions [31]. Residue

clusters are identified in protein contact maps where an accumulation of long range interactions are observed. The residues in these cores are called stabilization center (SC) residues, referring to their suspected role in 3D structure stabilization, and are identified as follows. The sequence environment of each residue pair involved in a long range interaction is analyzed. For each such residue pair we locate two additional pairs, one in the N-terminal flanking tetrapeptide and one in the C-terminal tetrapeptide of the original interacting residue pair making the most long range interactions with each other. If the number of interactions of these two triplets, the central interacting residues plus the two additional ones, one on each flanking side is equal to or greater than seven of the possible nine contacts, then the two central residues are accepted as members of a SC. The stabilization centers for the CH/π interacting amino acid residues were computed using the SCide server [32] for computing the stabilization centers.

Conservation Score

We computed the conservation score of CH/π interacting amino acid residues in each protein using the ConSurf server [33]. This server computes the conservation based on the comparison of the sequence of a PDB chain with the proteins deposited in Swiss-Prot [34] and finds the ones that are homologous to the PDB sequence. The number of PSI-BLAST iterations and the *E*-value cutoff used in all similarity searches were 1 and 0.001, respectively. All the sequences that are evolutionarily related with each one of the proteins in the data set were used in the subsequent multiple alignments. Based on these protein sequence alignments the residues are classified into nine categories from highly variable to highly conserved. Residues with a score of 1 are considered highly variable and residues with a score of 9 are considered highly conserved.

Stabilizing Residues

Stabilizing residues were computed using the parameters such as surrounding hydrophobicity, long-range order, stabilization center and conservation score as described by Gromiha [35]. We used the server SRide [35] for this purpose. Conservation score of ≥6 is the cutoff value used to identify the stabilizing residues.

RESULTS

CH/π Interactions

Fig. (**2**) illustrates an SS-5- CH/π interaction in therapeutic protein PDB 1D 1BML (between Trp 685 (CE3) and His 586). The specific pair wise residue involved in CH/π interaction, their position and sequential distances for all the therapeutic proteins studied are given in Table **2**. We found that, in the therapeutic protein data set only aromatic amino acids involved in CH/π interaction. Probably C^{α}-H and C^{aro}-H groups might have more specificity towards other non standard interactions like C-H….O interactions [36]. 82% interactions were SS-CH/π and 18% interactions were SS5-CH/π. The CH/π interactions forming residues distance with respect to its position of C-atom is shown in Fig. (**3**). It was found that, majority 27% of the interactions were found between the residue distances in the range of 4.01 Å to 4.25 Å. The CH/π interactions forming residues distance with respect to its position of H-atom were shown in Fig. (**4**). Of the total 77 interactions, the majority 39% of the interactions were found between the residue distances in the range of 3.76 Å to 4.00 Å. Atom vice contribution to CH/π interactions were studied in the data set. It was observed that, among the donor residues, majority 43% of the interactions from Phe, 31%of the interactions from Trp and remaining 26% of the interactions from Tyr. Among the acceptor residues, 34% of the interactions from Trp, 32% of the interactions from Phe, 18% of the interactions from His and 16% of the interactions from Tyr.

Secondary Structure Preferences

The propensity of the amino acid residues to favor a particular conformation has been well documented. Such conformational preference is not only dependent on the amino acid alone but is also dependent on the local amino acid sequence. We analyzed the secondary structure preference of each amino acid, which participated in all the different types of CH/π interactions namely, SS-CH/π and SS5-CH/π interactions. The secondary structure preference of each of the amino acids involved in all the above said types of CH/π interactions were obtained using DSSP and the results are depicted in Table **3**. We found that, Phe, Trp preferred to be in coil, Tyr preferred to be in strand and His preferred to be in helix.

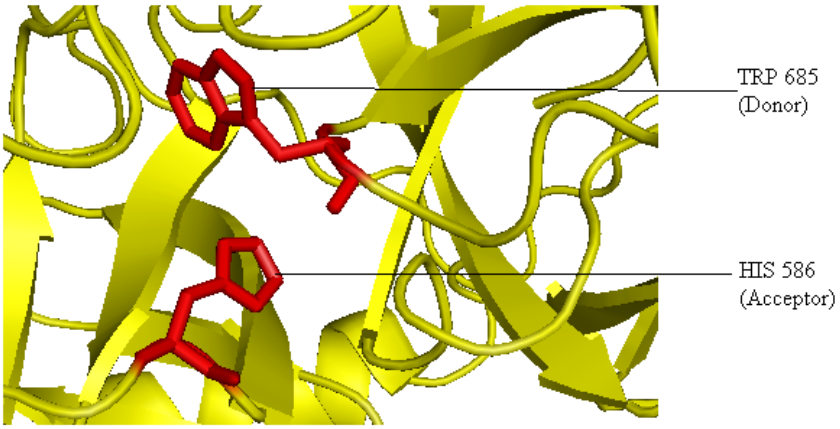

Fig. (2). Pymol view of CH/π interactions in 1BML.

Table 2. CH/π Interaction Forming Residue, Interacting Distance, D_{seq} in Therapeutic Proteins

PDB Code	Donor Residues	Donor Atoms	Acceptor Residues	Distance (CM) in Å	Distance (XM) in Å	D_{seq}	Total Number of Interactions
1BML	W685 Y614	CE3 CD1	H586 F587	4.680 4.709	3.705 3.965	99 27	2
1BMP	-	-	-	-	-	-	NI
1C4P	-	-	-	-	-	-	NI
1CA9	Y382 F381 F354 F410 Y395 F411	CD1 CD1 CE2 CE1 CE1 CE2	Y350 F354 W356 Y395 F410 F426	4.328 3.849 4.964 4.576 4.634 4.294	3.937 3.523 3.986 3.809 3.836 3.407	32 27 2 15 15 15	6
1CD9	-	-	-	-	-	-	NI
1CN4	F81 F81 F81 Y53 Y53 F11 F11 Y156 Y156	CD2 CD2 CE2 CD2 CE2 CE2 CE2 CD2 CE2	W40 W40 Y53 F81 F81 W82 W82 W142 W142	4.181 4.025 4.232 4.024 4.364 4.352 4.310 3.993 4.167	3.408 2.954 3.794 3.411 3.951 3.399 3.291 3.705 3.854	41 41 28 28 28 71 71 14 14	9
1DDJ	W685	CE3	H586	4.186	3.778	99	1
1EER	-	-	-	-	-	-	NI
1ES7	Y42	CE1	H44	4.969	3.886	2	1
1GNC	Y86	CE2	F161	4.516	3.866	75	1
1HTZ	W229 W229	CH2 CH2	W290 W290	3.843 3.734	3.427 2.946	61 61	2
1L4D	-	-	-	-	-	-	NI
1L4Z	W573 W685 W573 W573	CE3 CE3 CH2 CH2	W575 H586 F681 Y753	4.575 4.309 4.622 4.770	3.526 3.948 3.759 3.834	2 99 108 178	4
1L6X	-	-	-	-	-	-	NI
1M47	-	-	-	-	-	-	NI
1M48	-	-	-	-	-	-	NI
1M49	-	-	-	-	-	-	NI
1M4C	-	-	-	-	-	-	NI
1N8Y	-	-	-	-	-	-	NI
1N8Z	F98	CE1	Y36	4.566	3.527	62	1
1NBP	-	-	-	-	-	-	NI
1PGG	Y64 Y355 F356 F220 F407 F292 W100 F210 F407 F395 F200 F200 Y404	CE1 CE1 CD2 CE1 CE2 CE2 CD1 CE2 CD2 CE2 CD2 CE2 CD1	H43 H90 W100 Y147 F200 H204 F356 Y385 F395 Y404 F426 F426 H443	3.476 4.250 4.219 3.494 4.744 3.638 4.594 4.597 3.663 3.747 3.978 3.660 3.734	3.982 3.179 3.208 2.608 3.868 2.938 3.865 3.566 3.698 2.853 3.627 2.896 3.615	21 265 256 73 207 88 256 275 12 9 226 226 39	13

(Table 2). Contd…..

PDB Code	Donor Residues	Donor Atoms	Acceptor Residues	Distance (CM) in Å	Distance (XM) in Å	D_{seq}	Total Number of Interactions
1PGR	-	-	-	-	-	-	NI
1PW6	-	-	-	-	-	-	NI
1PY2	F42	CD1	F44	4.079	3.375	2	5
	F42	CE1	F44	4.294	3.729	2	
	F44	CD1	F117	3.984	3.876	73	
	W121	CD1	F117	3.869	3.178	4	
	W121	CE3	F124	4.160	3.648	3	
1QQR	-	-	-	-	-	-	NI
1QRZ	W685	CE3	H586	4.200	3.428	99	1
1QVN	F42	CD1	F44	4.178	3.462	2	5
	F42	CE1	F44	4.378	3.842	2	
	F44	CD1	F117	3.916	3.685	73	
	W121	CD1	F117	3.907	3.198	4	
	W121	CE3	F124	4.236	3.670	3	
1R46	W47	CE3	H46	4.858	3.894	1	7
	Y184	CD2	W204	4.177	3.861	20	
	Y216	CD1	W226	4.297	3.795	10	
	Y216	CE1	W226	4.117	3.273	10	
	F229	CD1	W245	4.283	3.403	16	
	F229	CD1	W245	3.358	2.665	16	
	F229	CE1	W245	3.860	3.601	16	
1R47	W47	CE3	H46	4.378	3.425	1	7
	Y184	CD2	W204	4.134	3.733	20	
	Y216	CD1	W226	4.454	3.856	10	
	Y216	CE1	W226	4.448	3.670	10	
	F229	CD1	W245	4.735	3.866	16	
	F229	CD1	W245	3.615	2.946	16	
	F229	CE1	W245	4.123	3.882	16	
1REU	-	-	-	-	-	-	NI
1REW	-	-	-	-	-	-	NI
1RHG	-	-	-	-	-	-	NI
1RJX	W573	CE3	W575	4.692	3.633	2	5
	W573	CE3	W575	4.942	3.985	2	
	W761	CE3	H603	4.162	3.831	158	
	W573	CH2	F681	4.319	3.488	108	
	W575	CH2	Y753	4.888	3.976	178	
1TPG	W21	CE3	Y15	4.132	3.657	6	2
	Y33	CD2	H44	4.592	3.624	11	
1WAQ	-	-	-	-	-	-	NI
1YY8	F98	CE2	Y36	4.511	3.455	62	1
1YY9	F156	CE2	F126	4.209	3.536	30	2
	W176	CH2	W140	4.602	3.777	36	
1Z92	-	-	-	-	-	-	NI
2B5I	-	-	-	-	-	-	NI
3BMP	-	-	-	-	-	-	NI
2ERJ	-	-	-	-	-	-	NI
2GMF	-	-	-	-	-	-	NI
2GOO	-	-	-	-	-	-	NI
2H62	Y42	CE1	H44	4.934	3.847	2	1
2H64	-	-	-	-	-	-	NI
2IWG	W381	CD1	Y391	4.101	3.044	10	1
2OSL	-	-	-	-	-	-	NI
3INK	-	-	-	-	-	-	NI

NI: No interactions; D_{seq}: Sequential distance.

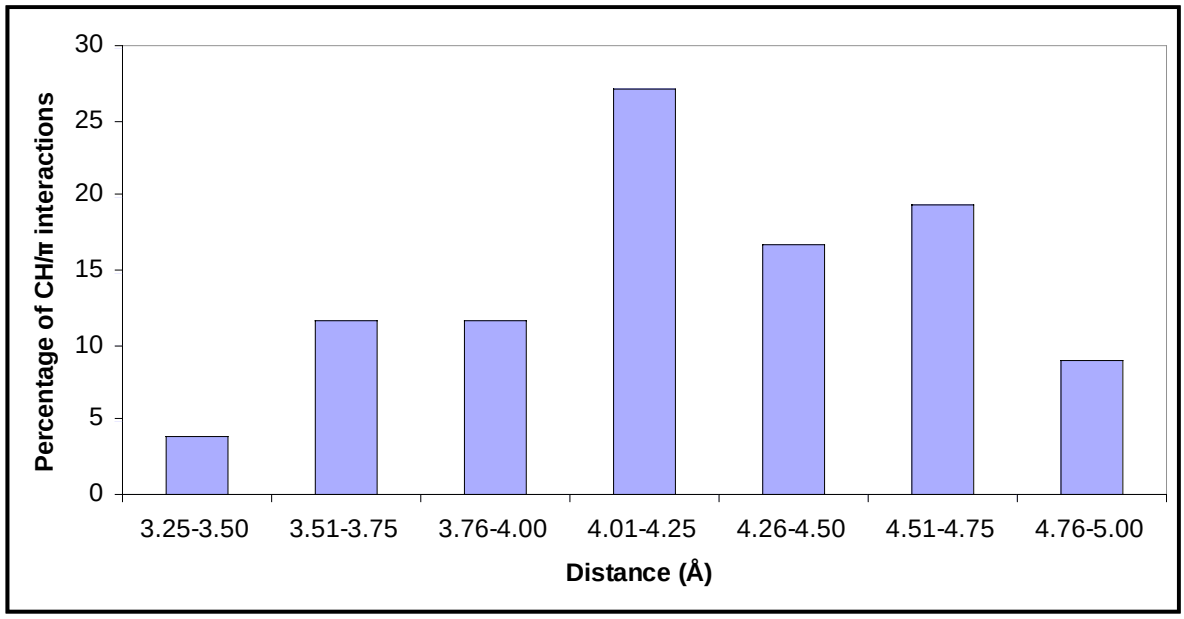

Fig. (3). Distance in (**Å**) between C and π as a function of percentage of C-H/π interactions in therapeutic proteins.

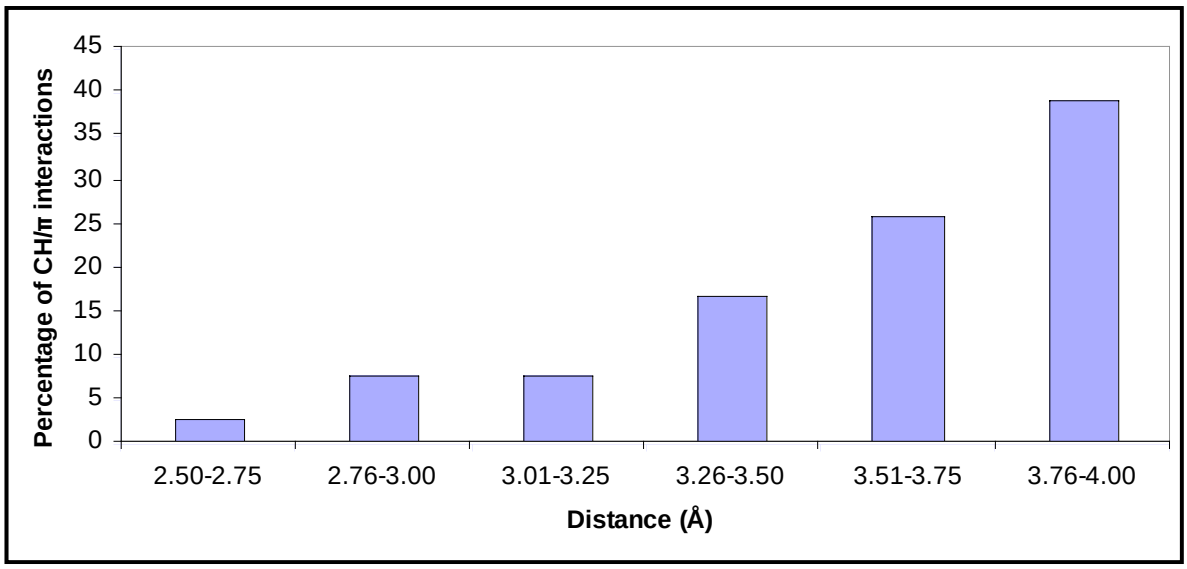

Fig. (4). Distance in (**Å**) between H and π as a function of percentage of C-H/π interactions in therapeutic proteins.

Solvent Accessibility

The relation between the amino acid residues in CH/π interactions and solvent accessibility is depicted in Fig. (**5**). The solvent accessibility of amino acid residues has been categorized as buried, partially buried and exposed [25, 26]. We found that, all the aromatic residues such as Phe, Tyr, Trp and His residues were in the buried regions. This

Table 3. Frequency of Occurrence of CH/π Interaction Forming Residue in Different Secondary Structures

Residue	Strand	Coil	Helix
Phe	30.29 **(27.17)**	41.79 **(42.25)**	27.91 **(30.57)**
Tyr	50.00 **(52.97)**	18.75 **(14.17)**	31.25 **(32.86)**
Trp	20.00 **(17.32)**	46.00 **(47.22)**	34.00 **(35.44)**
His	35.71 **(36.36)**	21.43 **(18.34)**	42.85 **(45.31)**

Parenthesis value shows frequency of secondary structure of the residues in the whole data set.

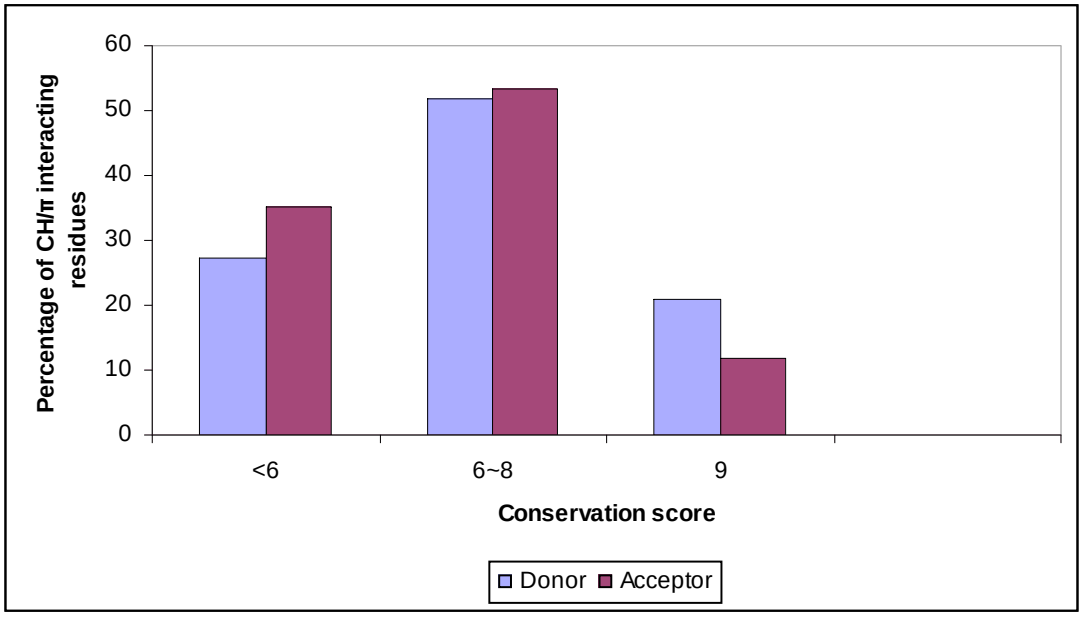

Fig. (5). Solvent accessibility of different ASA range in (%) *vs.* interacting residues.

observation is quite reasonable in the sense that, the aromatic residues are in principle, non polar residues, and tend to be buried. According to Manfred S. Weiss *et al.* [37], CH/π interactions involving aromatic residues either as donor or as acceptor groups are found mostly in the interior of the protein and tend to be buried in nature. These might be one of the reasons for their nature of solvent accessibility.

Sequential Separation

The contribution of CH/π interactions in therapeutic proteins could define either the local or the global stability of the proteins. Therefore, there is a need to evaluate the contribution of inter-residual CH/π interactions. The contribution from <±4 are treated as short-range contacts, >4 to <±20 as medium-range contacts and >20 are treated as long-range contacts [30]. This classification enables us to evaluate the contribution of short-range, medium-range and long-range contacts in the formation of CH/π interactions. The sequential distance between residues that contributed donor and acceptor atoms to inter-residue CH/π interactions were calculated. 50%, 29% and 21% of inter-residue CH/π interactions were found to be long-range, medium-range and short-range interactions respectively. Our results are consistent with RNA binding protein reported by our group earlier [38].

Stabilization Centers

We used the SCide server for computing the stabilization centers in the therapeutic proteins data ser. We found that 20% of the amino acid residues that contribute donor atoms to CH/π interactions had one or more stabilization centers in addition to their contribution to CH/π interactions and similarly 23% of the amino acid residues that contribute acceptor atoms to CH/π interactions had one or more stabilization centers in addition to their contribution to CH/π interactions. From this we infer that, these residues might contribute additional stability to the therapeutic proteins in addition to their participation in CH/π interactions.

Conservation Score

We used the ConSurf server to compute the conservation score of amino acid residues involved in CH/π interactions in therapeutic proteins, and the results are shown in Fig. (**6**). 21% of the amino acid residues that contributed donor atoms in CH/π interactions had the highest conservation score of 9, while 52% of the amino acid residues had a conservation score, in the range of 6-8. Thus, 73% of the donor amino acid residues had a high conservation score. In the case of amino acid residues, that contributed acceptor atoms in CH/π interactions, 12% of the acceptor amino acid residues had the highest conservation score of 9, while 53% of the amino acid residues had a conservation score, in the range of 6-8. Thus, 65% of the acceptor amino acid residues had a high conservation score. From these observations, we were able to infer that, most of the amino acid residues involved in CH/π interactions might be conserved in therapeutic proteins.

Stabilizing Residues

We thought it would be useful to identify any patterns of correlation between the CH/π interactions in a given therapeutic protein and the theoretically predicted stabilizing residues [35]. Stabilizing residues were computed using the parameters such as surrounding hydrophobicity, long-range order, stabilization center and conservation score. We used the server SRide for this purpose. None of the stabilizing residues are involved in the CH/π interactions. Hence apart from the stabilizing residues, the CH/π interactions may play a big role in the stability of the therapeutic proteins.

DISCUSSION

CH/π interactions have been identified to occur in 43% of therapeutic proteins in the data set. Unlike the involvement of main chain CH/π interactions in RNA binding protein, therapeutic protein CH/π interactions appear to occur relatively frequently between side-chains residues. The most prominent representatives are the interactions between

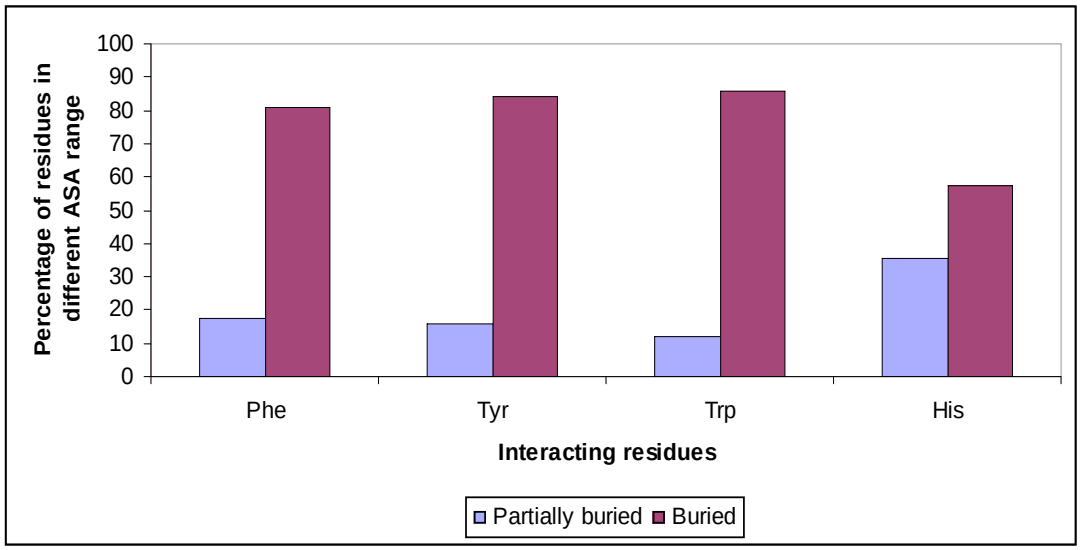

Fig. (6). Relation between conservation score and percentage of CH/π interacting residues in therapeutic protein.

aromatic C-H donor groups and aromatic π-acceptors. The geometric parameters calculated for these interactions suggest that CH/π interactions can be classified as weak H-bonds, and occur mainly in the distances greater than 4.00Å and 3.76Å from C atom and H atom respectively in the data set. The atom vice contribution to CH/π interactions shows that, Phe has the highest occurrence among the donor residues and Trp contribution higher than the other aromatic amino acid in the acceptor side. CH/π interactions involving aromatic π-systems as a donor or acceptor groups are generally found closer to the center of the protein and hence is buried in nature. The secondary structure preference analysis of CH/π interacting residues showed that Phe and Trp occurred most frequently in coil segments, while Tyr and His, respectively, occurred more in strand and helix segments. Thus, therapeutic proteins are therefore confronted with a very diverse array of surfaces resulting in the differences of amino acids to a particular secondary structure conformation. Long-range CH/π interactions are the predominant type of interactions in therapeutic proteins. The contribution of the medium-range and short-range interactions is comparatively less in the two types of inter-residue CH/π interactions studied. These results indicate that, long-range CH/π interactions contribute significantly to the global conformational stability of therapeutic proteins. Significant percentage of both donor and acceptor residues involved in CH/π interactions had one or more stabilization centers. From this we infer that, these residues might contribute additional stability to the therapeutic proteins in addition to their participation in CH/π interactions. None of the stabilizing residues are involved in the CH/π interactions. Hence apart from the stabilizing residues, the CH/π interactions may play a big role in the stability of the therapeutic proteins. 73% of the donor amino acid residues and 65% of the acceptor amino acid residues had a high conservation score. It might be due to their involvement in CH/π interactions and to the stability or the function of the protein. We hope this scrutiny will assist structural biologist and medicinal chemist to design better and safer drugs.

CONCLUSION

In conclusion, therapeutic proteins are stabilized by CH/π interactions. Aromatic amino acids contribute extensively to this CH/π interactions in therapeutic proteins data set. Majority of the interacting aromatic residues are located interior of the protein surface and tends to be buried in nature. Therapeutic proteins are confronted with a very diverse array of surfaces resulting in the differences of amino acids to a particular secondary structure conformation. The resulting high conservation score might be due to their involvement in the stability of therapeutic proteins. Thus from the cumulative analysis, we can infer that, the CH/π interactions do contribute significantly to the stability of therapeutic proteins based on our data.

ACKNOWLEDGEMENTS

The authors thank the management of Vellore Institute of Technology, for providing the facilities to carry out this work.

ABBREVIATIONS

TNF = Tumour necrosis factor

Tnf-r2 = Tumour necrosis factor receptor-2

G-csf = Granulocyte colony-stimulating factor

BMP-2 = Bone morphogenetic protein-2

Il-2 = Interleukin-2

IgG Fc = Immunoglobulin-G constant fragment

REFERENCES

[1] Bordo D, Argos P. The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins. J Mol Biol 1994; 243: 504-19.

[2] Dill KA. Dominant forces in protein folding. Biochemistry 1990; 29: 7133-55.

[3] Fersht AR, Serrano L. Principles of protein stability derived from protein engineering experiments. Curr Opin Struct Biol 1993; 3: 75-83.
[4] Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984; 44: 97-179.
[5] Jeffery GA, Saenger W. Hydrogen Bonding in Biological Systems. Springer-Verlag, New York 1991.
[6] McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994; 238: 777-93.
[7] Tamres M. Aromatic compounds as donor molecules in hydrogen bonding. J Am Chem Soc 1952; 74: 3375-8.
[8] Reeves LW, Schneider WG. Nuclear magnetic resonance measurements of complexes of chloroform with aromatic molecules and olefins. Can J Chem 1957; 35: 251-61.
[9] Nishio M, Hirota M, Umezawa Y, Ed. The CH/π Interaction. Wiley-VCH, New York 1998.
[10] Matsui I, Matsui E, Sakai Y, *et al.* The molecular structure of hyperthermostable aromatic aminotransferase with novel substrate specifiity from Pyrococcus horikoshii. J Biol Chem 2000; 275: 4871-9.
[11] Chakrabarti P, Samanta U. CH/π interaction in the packing of the adenine ring in protein structures. J Mol Biol 1995; 251: 9-14.
[12] Umezawa Y, Nishio M. CH/π interactions as demonstrated in the crystal structure of guaninenucleotide binding proteins, Src homology-2 domains and human growth hormone in complex with their specific ligands. Bioorg Med Chem 1998; 6: 493- 504.
[13] Muraki M, Harata K, Sugita N, Sato IK. Protein-carbohydrate interactions in human lysozyme probed by combining site-directed mutagenesis and affinity labeling. Biochemistry 2000; 39: 292-9.
[14] Umezawa Y, Nishio M. CH/π interactions in the crystal structure of class I MHC antigens and their complexes with peptides. Bioorg Med Chem 1998; 6: 2507-15.
[15] Jabs A, Weiss MS, Hilgenfeld R. CH/π interactions in the crystal structure of class I MHC antigens and their complexes with peptides. J Mol Biol 1999; 286: 291-304.
[16] Shimohigashi Y, Maeda I, Nose T, *et al.* Chymotrypsin inhibitory conformation induced by amino acid side-chain-side-chain intramolecular CH/π interaction. J Chem Soc [Perkin 1] 1996; 1: 2479-85.
[17] Shimohigashi Y, Nose T, Yamauchi Y, Maeda I. Design of serine protease inhibitors with conformation restricted by amino acid side-chain side-chain CH/π interaction. Biopolymers 1999; 51: 9-17.
[18] Steiner T. Angew. The hydrogen bond in the solid state. Chem Int Ed Engl 2002; 41: 48-76.
[19] Steiner T, Koellner GJ. Hydrogen bonds with π-acceptors in proteins: frequencies and role in stabilizing local 3D structures. J Mol Biol 2001; 305: 535-57.
[20] Meyer EA, Castellano RK, Diederich F. Interactions with aromatic rings in chemical and biological recognition. Angew Chem Int Ed Engl 2003; 42: 1210-50.
[21] Berman HM, Westbrook JZ, Feng G, *et al.* The protein data bank. Nucleic Acids Res 2000; 28: 235-42.
[22] Tiwari A, Panigrahi SK. *HBAT*: a complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. In Silico Biol 2007; 7: 651-61.
[23] Babu MM. NCI: a server to identify non-canonical interactions in protein Structures. Nucleic Acid Res 2003; 31: 3345-8.
[24] Kabsch W, Sander C. Dictionary of protein secondary structure:pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983; 22: 2577-637.
[25] Gilis D, Rooman M. Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials. J Mol Biol 1996; 257: 1112-26.
[26] 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: 276-90.
[27] Gromiha MM, Selvaraj S. Influence of medium and long range interactions in different structural classes of globular proteins. J Biol Phys 1997; 23: 151-62.
[28] Gromiha MM, Santhosh C, Shandar A. Structural analysis of cation-π interactions in DNA binding proteins. Int J Biol Macromol 2004; 34: 203- 211.
[29] Selvaraj S, Gromiha MM. Role of hydrophobic clusters and long-range contact networks in the folding of (alpha/beta)8 barrel proteins. Biophys J 2003; 84: 1919-25.
[30] Gromiha MM, Selvaraj S. Inter-residue interactions in protein folding and stability. Prog Biophys Mol Biol 2004; 86: 235-77.
[31] Dosztanyi ZS, Fiser A, Simon I. Stabilization centers in proteins: identification, characterization and predictions. J Mol Biol 1997; 272: 597-612.
[32] Dosztanyi ZS, Magyar CS, Tusnady E, Simon I. Scide: indentification of stabilization centers in proteins. Bioinformatics 2003; 19: 899-900.
[33] Glaser F, Pupko T, Paz I, *et al*. ConSurf: identification of functional regions in proteins by surfacemapping of phylogenetic information. Bioinformatics 2003; 19: 163-4.
[34] Boeckman B, Bairoch A, Apweiler R, *et al.* The swiss prot protein knoweldge base and its supplement TrEMBL in 2003. Nucleic Acids Res 2003; 31: 365-70.
[35] Gromiha MM, Pujadas G, Magyar C, Selvaraj S, Simon I. Locating the stabilizing residues in (alpha/beta)8 barrel proteins based on hydrophobicity, long-range interactions, and sequence conservation. Proteins 2004; 55: 316-29.
[36] Derewenda ZS, Lee L, Derewenda U. The occurrence of C-H….O hydrogen bonds in proteins. J Mol Biol 1995; 252: 248-62.
[37] Maria B, Weiss MS, Andreas J, JuE rgen SuE hnel, Rolf H. C-H….π interactions in proteins. J Mol Biol 2001; 307: 357-77
[38] Anand A, Sudha A, Babu MM, Sethumadhavan R. Investigations on C-H….π interactions in RNA binding proteins. Int J Biol Macromol 2007; 41: 251-9.

Received: June 23, 2008 Revised: September 19, 2008 Accepted: September 24, 2008