Available online at www.sciencedirect.com
ScienceDirect
Progress in Biophysics and Molecular Biology 95 (2007) 83–137
Progress in Biophysics & Molecular Biology
www.elsevier.com/locate/pbiomolbio

Review

Geometry of nonbonded interactions involving planar groups in proteins

Pinak Chakrabarti*, Rajasri Bhattacharyya

Department of Biochemistry and Bioinformatics Centre, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700 054, India

Available online 5 June 2007

Abstract

Although hydrophobic interaction is the main contributing factor to the stability of the protein fold, the specificity of the folding process depends on many directional interactions. An analysis has been carried out on the geometry of interaction between planar moieties of ten side chains (Phe, Tyr, Trp, His, Arg, Pro, Asp, Glu, Asn and Gln), the aromatic residues and the sulfide planes (of Met and cystine), and the aromatic residues and the peptide planes within the protein tertiary structures available in the Protein Data Bank. The occurrence of hydrogen bonds and other nonconventional interactions such as C–H$\cdots\pi$, C–H$\cdots$O, electrophile–nucleophile interactions involving the planar moieties has been elucidated. The specific nature of the interactions constraints many of the residue pairs to occur with a fixed sequence difference, maintaining a sequential order, when located in secondary structural elements, such as α-helices and β-turns. The importance of many of these interactions (for example, aromatic residues interacting with Pro or cystine sulfur atom) is revealed by the higher degree of conservation observed for them in protein structures and binding regions. The planar residues are well represented in the active sites, and the geometry of their interactions does not deviate from the general distribution. The geometrical relationship between interacting residues provides valuable insights into the process of protein folding and would be useful for the design of protein molecules and modulation of their binding properties.

Keywords: Packing; Relative orientations between planar groups; S$\cdots$aromatic interactions; Protein folding; Protein stability

Contents

*Corresponding author. Fax: +91 33 2355 3886.
E-mail addresses: pinak@boseinst.ernet.in, pinak_chak@yahoo.co.in (P. Chakrabarti).

0079-6107/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pbiomolbio.2007.03.016

1. Introduction

Deciphering the mechanism for the coding of protein structures by their amino acid sequence has been the Holy Grail for the biologists. Natural proteins fold into specific compact structures despite the huge number of possible configurations. Whereas the burial of hydrophobic groups serves as the primary source of stabilization energy in the folded structure (Kauzmann, 1959; Dill, 1990), it is important to understand the extent to which protein conformation is determined by packing interactions within the hydrophobic core (Behe et al., 1991). The packing density within a protein resembles a crystalline solid rather than oil (Richards, 1974), indicating that the stereospecific packing of amino acid side chains (Richards and Lim, 1994) and the secondary structures, such as α-helices (Crick, 1953; Harbury et al., 1993)—like pieces of a three-dimensional jigsaw puzzle—is the key determinant that links a sequence to a given fold. A contrary point of view suggests that the geometry of side-chain interactions is completely random and that the close packing arises simply on account of compaction within a constrained volume, like what happens to an ensemble of nuts and bolts in a jar (Bromberg and Dill, 1994; Liang and Dill, 2001). An analysis by Banerjee et al. (2003) tends to support the

jigsaw puzzle model as most interactions between buried hydrophobic residues are characterized by high surface complementarity and a heightened constraint in inter-residue geometry.

Singh and Thornton (1990 and 1992) published an atlas of the 400 (20×20) normalized contact propensities of side chain interaction in proteins. They calculated the propensity of contact between residues A and B as the ratio of the observed frequency (total number of atom–atom contacts between side chains of A and B) and the expected frequency of contacts (the product of the probabilities of getting any interaction between residue A and B and the total number of interactions between all 20 side chains). The distribution of the propensities of the residue pairs suggested that many of them resulted from directional, rather than random packing. Superposition of all the contacts for a pair in a common frame of reference and calculating the directionality of the interaction using polar coordinates enumerated the mode of interaction for the pairs. A χ^2 analysis was used to indicate any significant preference in the directionality of interaction. However, as results were reported for individual residue pairs, it was difficult to compare the results for two or more pairs.

Restricting to nine side chains with planar regions, and grouping the residue pairs into categories based on chemical similarity, Mitchell et al. (1997) found that for some categories the distributions of interplanar angles are significantly different from the sinusoidal one expected for random packing. In addition to the interplanar angle, Brocchieri and Karlin (1994) also used two additional angles (θ_1, θ_2) determined by the line connecting the centroids of the two planar groups and each of the two planes, and studied residue pairs in 15 chemical groups. It is possible to use one θ angle when analyzing the orientations of various groups relative to the residue of a given type taken as the 'central' residue (McGaughey et al., 1998). Though the two angular parameters are insufficient to describe the intermolecular geometry completely (Brocchieri and Karlin, 1994; Mitchell et al., 1997), they provide an intuitively easy way to compare different pairs of interactions and have been used to identify structurally meaningful features involving aromatic residues, proline, the sulfide group, etc. (Samanta et al., 1999; Bhattacharyya and Chakrabarti, 2003; Bhattacharyya et al., 2002, 2003, 2004). In this article, we carry out an analysis of the geometry of interaction between planar moieties of ten side chains (Phe, Tyr, Trp, His, Arg, Pro, Asp, Glu, Asn and Gln), the aromatic residues and the sulfide planes (of Met and cystine), and the aromatic residues and the peptide planes within the protein tertiary structures in an enlarged data set. As the percentage compositions of Trp, Cys and Met in proteins are on the lower side (Chakrabarti and Pal, 2001), the use of larger number of structures can allow a meaningful statistical summary. Moreover, aromatic-backbone interactions have been observed in helical structures in proteins (Tóth et al., 2004) and the geometry of this interaction is also studied. Most of the residues considered here are found in side-chain clusters in protein structures (Heringa and Argos, 1991; Karlin and Zhu, 1996; Kannan and Vishveshwara, 1999). The geometrical relationship and the nature of interactions between side chains help us to refine our understanding on inter-residue contacts (Gromiha and Selvaraj, 2004) in providing stability to the native fold.

1.1. Interactions favoring specific orientations

While hydrogen bonding imposes a rather tight geometry between the interacting groups (Baker and Hubbard 1984; Jeffrey and Saenger, 1991; McDonald and Thornton, 1994), there exist other noncovalent interactions that also have geometric constraints (Burley and Petsko, 1988) and are thus specific.

1.1.1. Aromatic–aromatic interactions

Since the seminal work of Burley and Petsko (1985), which was put in the right statistical context by Blundell et al. (1986), a large number of studies have dealt with the subject of aromatic–aromatic interaction in proteins (Singh and Thornton, 1985; Hunter et al., 1991; McGaughey et al., 1998; Samanta et al., 1999, 2000; Bhattacharyya et al., 2002, 2003; Thomas et al., 2002a), as well as the interaction of amide and guanidinium groups with aromatic residues (Burley and Petsko, 1986; Flocco and Mowbray, 1994; Mitchell et al., 1994). The vast majority of medicinal agents contain aromatic substituents and their differential recognition by proteins is likely to be facilitated by the noncovalent interactions involving aromatic residues (Gilman et al., 1993) (Fig. 1), which are thus pivotal to the process of drug design. Interactions between aromatic residues and nucleic acid bases can play crucial role in protein–DNA recognition (Rooman et al., 2002; Mazza et al., 2002; Sathyapriya and Vishveshwara, 2004), an example of which is shown in Fig. 2. In addition to interacting with

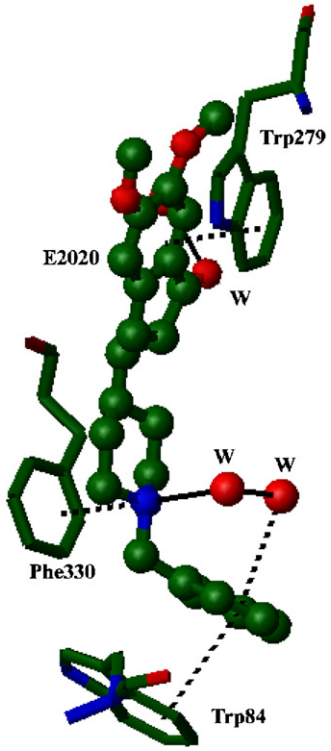

Fig. 1. Interactions involving aromatic residues in the binding of the anti-Alzheimer drug E2020 (in ball and stick) with acetylcholinesterase from *Torpedo californica* (PDB code: 1eve). (Kryger et al., 1999). Water molecules have been labeled as W; thin, solid lines indicate conventional hydrogen bonds and dashed lines represent O–H$\cdots\pi$ (involving E2020), cation$\cdots\pi$ (Phe330) and $\pi\cdots\pi$ (Trp84 and Trp279) interactions.

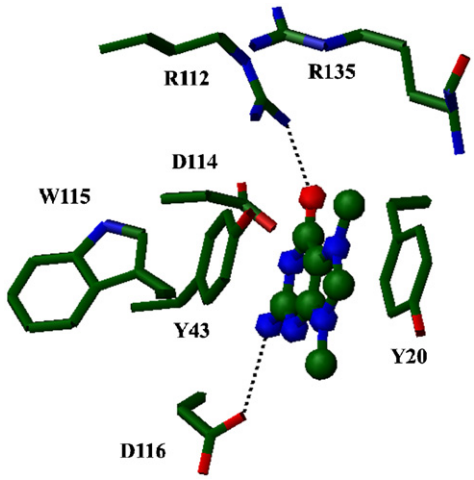

Fig. 2. The specificity for the methylated guanosine (in 5′ cap structure of RNA) is achieved by sandwiching the base between the aromatic residues, Tyr20 and Tyr43, in heterodimeric nuclear cap-binding complex (PDB code: 1h2t) (Mazza et al., 2002).

another aromatic ring, an aromatic residue can engage other moieties through X–H$\cdots\pi$ interactions (Section 1.1.2) or cation$\cdots\pi$ interactions (Dougherty, 1996; Ma and Dougherty, 1997) (Fig. 1). The diversity of these interactions can be exploited to design peptide motifs (Aravinda et al., 2003; Waters, 2004) and synthetic receptors (Hunter, 1994; Waters, 2002; Meyer et al., 2003). Aromatic clusters have been suggested to

be a determinant of thermal stability of thermophilic proteins (Kannan and Vishveshwara, 2000) and the introduction of even a single aromatic–aromatic interaction has been found to increase the apparent melting temperature (T_m) by 9 °C in the thermal unfolding of a xylanase (Georis et al., 2000). A cluster of aromatic residues can also be seen near the binding sites of carbohydrates in proteins (Fig. 3).

Because aromatic–aromatic interactions are so prevalent across chemistry and biology, a large body of work has focused on the preferred mode of binding of the prototype, namely the benzene dimer (Hobza et al., 1994; Jaffe and Smith, 1996; Chipot et al., 1996; Chelli et al., 2002). The calculated interaction energies of the parallel, edge–face (T-shaped) and offset stacked are −1.48, −2.46 and −2.48 kcal/mol, respectively (Tsuzuki et al., 2002), and the major source of attraction is not short-range (such as charge-transfer), but long-range interactions (quadrupole–quadrupole electrostatic and dispersion) (Hunter and Sanders, 1990; Morozov et al., 2004). The latter two structures (Fig. 4) are nearly isoenergetic and empirical force field studies indicate that depending on the magnitude of the partial charges ($q_H = -q_C$) one may be favored over the other—small charges (<0.153) favor offset stacked geometries, whereas large partial charges (>0.3) favor edge–face

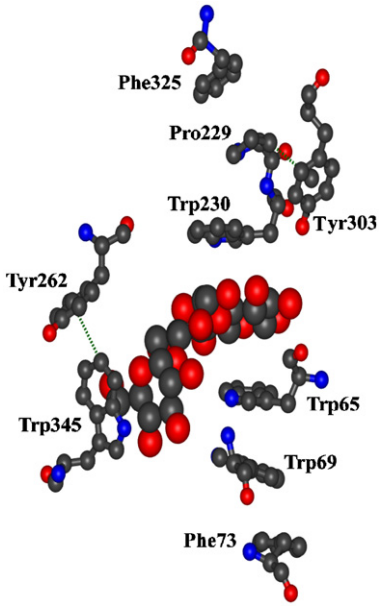

Fig. 3. Clustering of aromatic residues and a proline in the binding site of maltodextrin-binding protein (PDB code: 1elj) (Evdokimov et al., 2001). The oligosaccharide is shown in the CPK mode and the C–H···π interactions are shown in dotted lines.

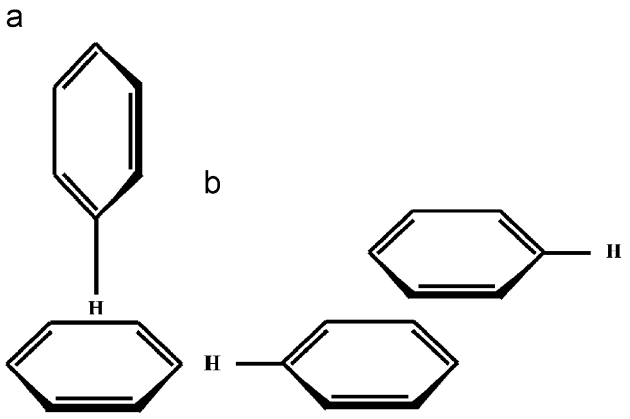

Fig. 4. Two favored geometries for aromatic–aromatic interactions: (a) edge face and (b) offset stacked.

configuration (Sun and Bernstein, 1996). For flexible organic molecules, edge–face interactions are energetically favorable in the solid state (where conformational entropy effects are negligible) (Jennings et al., 2001). Mutating an aromatic pair (the edge of Tyr17 interacting with the face of Tyr13) on the solvent exposed face in an α-helix of barnase to alanine provided an aromatic–aromatic interaction energy of −1.3 kcal/mol in the folded protein relative to solvation by water in the unfolded protein (Serrano et al., 1991). It has been suggested that the perpendicular and the parallel-displaced configurations are more common than the sandwich geometry as these, especially the former one exposes three aromatic faces to the outside, offering greater possibility for additional interactions with other groups (Chipot et al., 1996). Thus there may be differences in the preferential geometry depending on whether the aromatic pairs are isolated or part of higher-order clusters of aromatic residues. Considering only the former category of interactions in high-resolution X-ray structures, the off-centered parallel orientation was found to be preferred with 0.5–0.75 kcal/mol higher stability than the T-shaped structure (McGaughey et al., 1998).

1.1.2. X–H···π interactions

The conventional definition of hydrogen bond, X–H···A, in which both X and A are electronegative atoms and the distance H···A or even X···A being shorter than the sum of the van der Waals radii, has been expanded in recent years to include weaker donors, such as the C–H group and acceptors like the π electron cloud of an aromatic ring (Jeffrey and Saenger, 1991; Desiraju and Steiner, 1999; Nishio et al., 1998). These nonconventional hydrogen bonds, with the aromatic ring as the acceptor, are observed in benzene–water and benzene–ammonia dimers (Atwood et al., 1991; Suzuki et al., 1992; Rodham et al., 1993), and are quite numerous in proteins (Levitt and Perutz, 1988; Pal and Chakrabarti, 1999; Samanta et al., 2000; Steiner and Koellner, 2001; Brandl et al., 2001; Weiss et al., 2001; Tóth et al., 2001; Bhattacharyya and Chakrabarti, 2003), and taken together may contribute substantially to the overall stability of the structures. About three-quarters of the Trp rings, half of all Phe and Tyr rings and a quarter of all His rings are involved as acceptors in C–H···π interactions (Brandl et al., 2001); apart from the larger aromatic surface, the conjugate nature of the two rings might increase the π-acceptor strength of Trp compared to Tyr and Phe (Steiner and Koellner, 2001). Though N–H···π interaction has been reported with His acting as an acceptor (Vásquez et al., 1998), the frequency of occurrence of such bonds is low owing to the unsuitability of imidazole ring in this role when charged. The interactions are also conspicuous at the binding sites of substrates, cofactors (Waksman et al., 1992; Liu et al., 1993; Chakrabarti and Samanta, 1995; Umezawa and Nishio, 1998, 2005; Matsui et al., 2000; Sampson and Vrielink, 2003), carbohydrate (Muraki et al., 2000; Jeyaprakash et al., 2002) and DNA (Umezawa and Nishio, 2002). Though Brandl et al. (2001) did not find any clear trend relating thermostability of a protein to its overall C–H···π content, Ibrahim and Pattabhi (2004) reported a higher occurrence of C^{α}–H···π and C^{α}–H···O interactions in thermophilic proteins than in the mesophilic counterpart. C–H···π interaction has been utilized for the design of conformationally restricted peptides for use as inhibitors to serine proteases (Shimohigashi et al., 1999).

Ab initio calculations indicate that the strengths of O–H···π interactions follow the trend Trp > His > Tyr∼Phe (Scheiner et al., 2002). Computational and NMR studies attribute an energy of 0.88 kcal/mol to C–H···π hydrogen bond (Takagi et al., 1987; Dasgupta et al., 2007). Electron-rich heteroaromatic rings are good acceptors and hence X–H···π interactions may play crucial role in the binding of cofactors (Chakrabarti and Samanta, 1995). Model calculations involving water/ammonia/methane with pyridine (Samanta et al., 1998) indicate that the maximum stabilization of 2.9, 1.8 and 0.8 kcal/mol, respectively, is obtained when the X–H group is directly above (shifted slightly towards the ring center when X = C) the ring N; when the aromatic protons of benzene are involved, a herringbone structure two C–H bonds facing the pyridine ring can have a stabilization energy of 2.7 kcal/mol.

1.1.3. C–H···O interactions

Another nonconventional hydrogen bond, C–H···O interaction (Taylor and Kennard, 1982; Desiraju, 2002) is weaker as compared to N–H···O and O–H···O hydrogen bonds, but it constitutes 20–25% of the total number of hydrogen bonds in proteins (Weiss et al., 2001). Though these can provide overall stability to the structure (Derewenda et al., 1995; Bella and Berman, 1996), these are of particular importance in the context of the secondary structures, such as β-sheets (Fabiola et al., 1997; Ho and Curmi, 2002), along

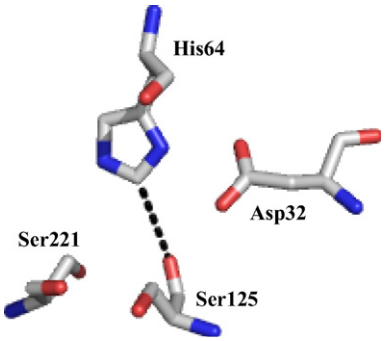

Fig. 5. C–H···O interaction involving the proton at CE1 of histidine (belonging to the catalytic triad, Ser221-His64-Asp32) and the main-chain carbonyl oxygen atom of Ser125, in subtilisin (PDB code: 2st1) (Bott et al., 1988).

the main body and the termination of α-helices (Chakrabarti and Chakrabarti, 1998; Babu et al., 2002; Bhattacharyya and Chakrabarti, 2003; Manikandan and Ramakumar, 2004). These are observed at protein–protein interfaces (Jiang and Lai, 2002; Saha et al., 2007) and could be relevant in the ligand-binding sites (Pierce et al., 2002), where the ligands are observed to use their stronger hydrogen bond capabilities for use with the protein residues, leaving the weaker C–H···O interactions to bind water (Sarkhel and Desiraju, 2004). C–H···O hydrogen bonds involving the C^{α} protons, especially of Gly, Ser and Thr, are found at the packing interfaces between transmembrane helices, providing an explanation for the involvement of the GxxxG motif in the organization of these helices (Senes et al., 2001, 2004). Likewise, C^{α}–H···O hydrogen bonds have been found to stabilize the structural segments that are involved in binding cholorophyll *a* molecules in photosystem I (Loll et al., 2003). These are also observed in nucleic acid and carbohydrate structures (Wahl and Sundaralingam, 1997; Ghosh and Bansal, 1999) and are involved in protein–nucleic acid interactions (Mandel-Gutfreund et al., 1998).

Whereas the acceptor group in X–H···π interactions is an aromatic side chain in proteins, for C–H···O hydrogen bonds it is usually the peptide carbonyl oxygen atom; the donor could be from the side chain, or the main chain C^{α}–H group (Fig. 5). *Ab initio* calculations indicate that the C–H···O hydrogen bond between C^{α}–H groups of amino acids and water molecules is about half as strong as a classical one between two water molecules (Scheiner et al., 2001). Likewise, a C^{α}–H···O=C interaction (of energy −2.1 kcal/mol) has been found to be one-half the strength of the N–H···O=C hydrogen bond (Vargas et al., 2000). A protonated residue such as $HisH^{+}$ makes for a very powerful proton donor, such that even its C–H···O hydrogen bonds are stronger than the conventional bonds formed by neutral groups (Scheiner et al., 2002), and such interactions are observed in the active sites of serine hydrolases (Derewenda et al., 1994).

1.1.4. Electrophile–nucleophile and S···aromatic interactions

When two nonbonded chemical groups come within the sum of the van der Waals radii of contacting atoms, certain directional preferences exist depending on the chemical nature and stereochemistry of the contacting atoms and groups (Bent, 1968; Rosenfield et al., 1977; Dunitz, 1979; Bürgi and Dunitz, 1983; Ramasubbu et al., 1986). Such intermolecular associations usually result from the interactions of highest occupied molecular orbital with the lowest unoccupied molecular orbital (HOMO–LUMO) (Fukui et al., 1952). These have been termed as incipient electrophilic and nucleophilic attack, as the structural parameters relating the molecular fragments, observed in a large variety of crystalline frameworks, vary in a systematic manner tracing paths that roughly follow potential energy valleys in the appropriate parameter space (Bürgi and Dunitz, 1983). The patterns of molecular deformations found in crystal environments mirror the distortion that the fragments would undergo along a reaction coordinate and can be assumed to represent chemical reaction paths. For example, in structures having a close contact between the amino group (nucleophile) and the carbonyl carbon atom (electrophile), the nitrogen atom is always approximately in the bisecting plane of the >C=O group, making an angle of about 107° with the C=O bond (Bürgi et al., 1973). In the realm of protein structures, such an electrophile–nucleophile pairing has been observed involving the peptide carbonyl

group and the thiolate anion of the metal-bound cysteine residues (Chakrabarti and Pal, 1997). Free sulfhydryl groups are also inclined to participate in S···C=O interactions (Pal and Chakrabarti, 1998).

Unlike other protein atoms, it is not easy to characterize S atom as hydrophobic or hydrophilic (Eisenberg et al., 1989; Pal and Chakrabarti, 2001). If involved in a hydrogen bond, the cysteine –SH group would rather act as a proton donor (usually to a carbonyl group) than an acceptor (Pal and Chakrabarti, 1998). On the other hand, Met residues are not adept in forming hydrogen bonds (Gregoret et al., 1991); very few that may exist do not reveal any particular stereochemical preference, hallmark of all oxygen-containing residues (Ippolito et al., 1990). Even small molecule structures containing the divalent sulfur show little inclination to act as hydrogen bond acceptors (Allen et al., 1997), though they are not averse to form direct S···O contacts (Burling and Goldstein, 1993; Nagao et al., 1998; Iwaoka et al., 2002b). The short nonbonded contacts of the Y–S–Z (Y, Z=C, N, O or S) group with an atom X fall into two groups: (a) nucleophilic X approaching S in the Y–S–Z plane along the posterior extensions (Y→S or Z→S) of the two bonds, interpreted as an X interaction with the LUMO (which is the antibonding orbital of the S–Y or S–Z bond), and (b) electrophilic X approaching S at angles $<40^{\circ}$ from the perpendicular to the Y–S–Z plane, interpreted as an X···HOMO (which is the lone pair orbital on S) interaction (Rosenfield et al., 1977). As the protein structure results from the optimization of myriads of weak nonbonded interactions, the stereoelectronic requirements around the S atom may be swamped by the demands of other stronger interactions (Carugo, 1999), but the geometry of interaction of metal ions (electrophiles) with the Met S (Fig. 6a) (Chakrabarti, 1989) and oxygen atoms (nucleophile) with the S atom in Met or disulfide groups (Fig. 6b) (Pal and Chakrabarti, 2001; Iwaoka et al., 2002a; Bhattacharyya et al., 2004) essentially follows the trend observed in small molecule structures.

The stereospecificity of interactions of divalent sulfur (–CH_2–S–CH_3 and –CH_2–S–S–CH_2–) in proteins with the aromatic side chains can be formulated in simple terms—interaction along the edge or the face of the aromatic ring. Whereas the work by Reid et al. (1985), based on 36 proteins, indicated the former mode of binding, other studies (Klingler and Brutlag, 1994; Pal and Chakrabarti, 2001) starting with that by Morgan and McAdon (1980), based on 22 proteins, indicated the latter mode, with S interacting with the π electron cloud, a remarkable example of which can be seen in lysozyme (Fig. 7). Morgan et al. (1978) provided some other early examples of chains of S···aromatic interactions. Questions were asked if the aromatic ring could be considered a nucleophile in its interaction with the S atom, akin to the S···O interaction (Pal and Chakrabarti, 2001; Bhattacharyya et al., 2004). In an aromatic ring, the π electrons are spread out and not localized as in the

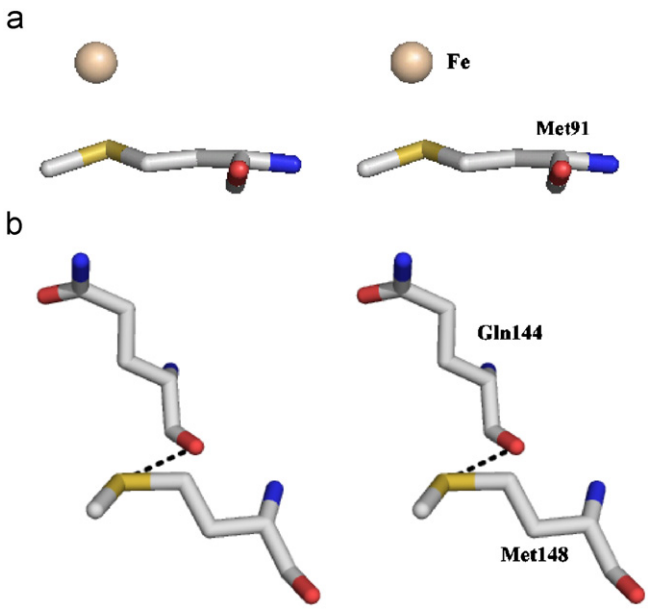

Fig. 6. Stereoviews of the interaction of Met sulfur atom with (a) Fe in cytochrome *c*2 (reduced) (PDB code: 3c2c) (Salemme et al., 1973), and (b) the main-chain oxygen of arginine kinase (PDB code: 1bg0) (Zhou et al., 1998).

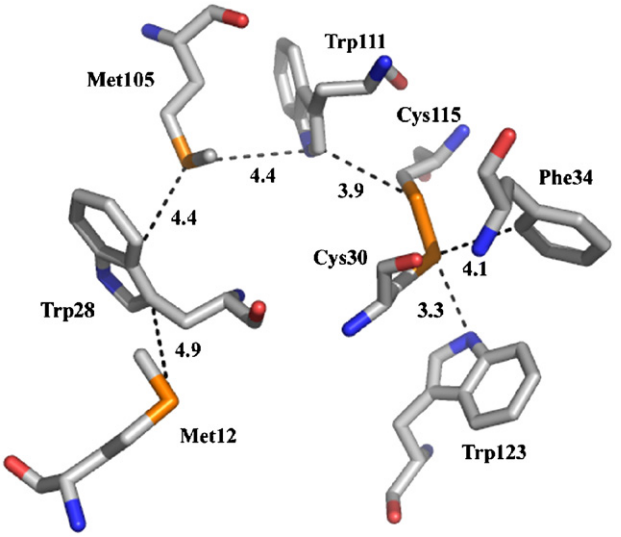

Fig. 7. A series of S···aromatic interactions in hen egg-white lysozyme (PDB code: 6lyz) (Diamond, 1974). The shortest contact distances involving S (in Met and cystine) and the aromatic atoms are shown.

lone pair orbital and hence the directionality of the interaction would be less pronounced. Nevertheless, the avoidance of the geometries that place the S lone pair orbital on top and facing the π electron cloud is quite pronounced (Bhattacharyya et al., 2004). The preferred mode of interaction of the –SH group of a free Cys residue is also along the aromatic face (Pal and Chakrabarti, 1998).

1.2. Planar groups and propensities of residues to interact with each other

It is useful to put the propensities of planar groups to interact with each other in the context of all possible types of interactions in protein structures (Narayana and Argos, 1984; Karlin et al., 1994; Russell and Barton, 1994; Keskin et al., 1998; Saha et al., 2005). The propensity, P_{XY} of a residue X to be in the environment (defined to be within a distance of 4.5 Å of any atom) of Y (the central residue) can easily be calculated (Saha and Chakrabarti, 2006). $P_{XY} \approx 1.0$ indicates a neutral preference of X to be in the environment of Y; a value >1.0 implies preference to associate; <1.0, to avoid. The propensity values are depicted in Fig. 8; in (a) all the residue atoms were included in the calculation, whereas in (b) it was restricted to the atoms in the side chains only. As the features of the side chain distinguish one residue from another, the preferential association (or avoidance) between residues is much more prominent in Fig. 8b (which has more of blue or red color, from the two ends of the spectrum, compared to Fig. 8a). The expected associations of side-chain interactions (Karlin et al., 1994) between oppositely charged residues, two Cys residues and among hydrophobic residues can be seen.

Pro has high propensities to have aromatic side chains in its environment, but avoids or has neutral preferences to other six planar residues. Association between the same type of residues is strong among aromatics. Asn has a high propensity to have another Asn or Gln in its environment, but Gln shows preference only for Asn. Arg, Asn and Gln have high preference to have aromatic side chains, especially Tyr and Trp, in the neighborhood. It may be mentioned that although in general charged residues avoid like charges, there are instances when two acidic groups in the protonated state form hydrogen bond to each other and thereby be in close proximity (Sawyer and James, 1982; Flocco and Mowbray, 1995; Torshin et al., 2003).

Among the aromatic residues, Phe prefers to pair up with another Phe, Tyr is more frequently involved with a Phe, Trp with Phe, and His with Tyr (the numbers of interactions are available as supplementary material, Fig. S1), as noted by Thomas et al. (2002a). Indeed, the propensity of His to interact with a planar group is the highest (1.74) with Tyr, much more than that with a negatively charged residue, Asp (1.10) (Bhattacharyya et al., 2003). His can be found in varied chemical environment in protein structures, sometimes behaving as an aromatic residue, or as a metal ligand, and at other times forming salt bridges with acidic groups. As a result,

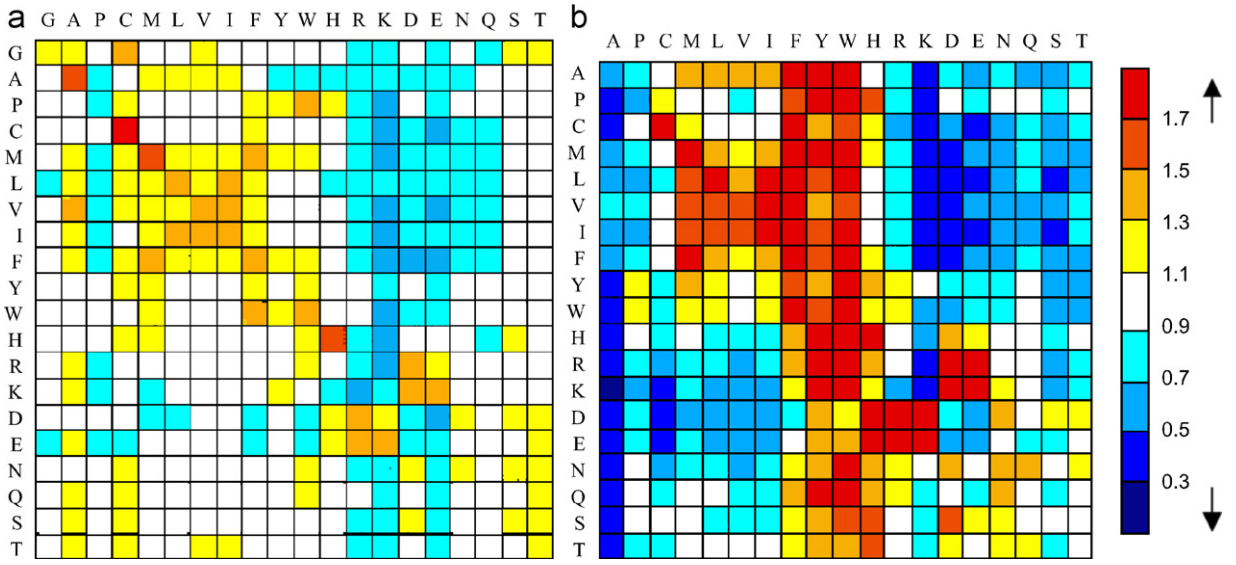

Fig. 8. Color-coded representation of propensities of different residues (in columns) to be in the environment of any given residue (row) in protein structures. In (a) all the atoms in the residue are considered during the calculation of contacts, whereas in (b) it is restricted to the side-chain atoms only. The equivalence between the propensity values and the color is shown on the right (taken from Saha et al., 2005).

if one performs an environment-based classification of amino acid residues, His forms a distinct class separate from all other residue types (Saha et al., 2005).

2. Dataset, methodology and convention

To analyze nonbonded interactions protein structures were taken from the Protein Data Bank (PDB) at the Research Collaboratory for Structural Bioinformatics (http://www.rcsb.org/pdb/) (Berman et al., 2000). In total, 1608 protein chains were selected using PISCES (Wang and Dunbrack, 2003) from the PDB files (as of April 2005) with an R-factor $\leqslant 20\%$, resolution $\leqslant 2.0$ Å and the sequence identity between any pair of polypeptide chains $\leqslant 25\%$. Table S1 (supplementary material) lists the four-lettered PDB codes (in lowercase, with the chain identifier, if present, appended in uppercase). A pair of residues, X and Y, having a nonbonded interaction is indicated as X–Y; if any one or both of X and Y stand for an atom then the interaction is represented by X$\cdots$Y.

The secondary structural elements were determined using the algorithm (DSSP) (Kabsch and Sander, 1983). Unless mentioned, all helix types (H, G or I in DSSP notation) were grouped as H, β-strands (E or B) as E and turns (S or T) as T. Hydrogen bond interactions were identified using the program HBPLUS (McDonald and Thornton, 1994). Molecular diagrams were generated using MOLMOL (Koradi et al., 1996) and PYMOL (DeLano, 2002).

To analyze the nonbonded interactions we grouped the planar moieties in two classes. The first group contains the planar side chains of ten residues: Phe, Tyr, Trp, His, Arg, Pro, Asp, Glu, Asn and Gln, while the second group contains the sulfur-containing moieties (Met and cystine) along with the aromatic rings. We also considered the interaction of the peptide moiety with the aromatic rings. It may be mentioned that the pyrrolidine ring is not strictly planar and has a puckering (Ramachandran et al., 1970; Ashida and Kakudo, 1974; Pal and Chakrabarti, 1999; Chakrabarti and Pal, 2001); but we considered the least-squares plane.

Only those residues were considered for which the fractional occupancies of all the atoms defining the planar moieties—CG onwards for Phe, Tyr, Trp, His, Glu and Gln, CD onwards for Arg, CB onwards for Asp and Asn, and all the pyrrolidine ring atoms for Pro—were 1.00. IUPAC-IUB Commission on Biochemical Nomenclature (1970) has described the convention for atom labels.

2.1. Selection of the interatomic distance for the identification of the interacting residues

A large number of cut-off values have been used by different workers, usually in the range 5–14 Å, to describe the environment of different amino acid residues in proteins and in the construction of potentials of mean force for inter-residue interactions (Narayana and Argos, 1984; Nishikawa and Ooi, 1986; Bahar and Jernigan, 1997; Baud and Karlin, 1999; Zhang and Kim, 2000). However, as we are interested in identifying the specific interactions, we have restricted to the limiting values of 4.3 Å for the S···C distance in S···aromatic interactions and 4.5 Å for the C···C distance for the other interactions for reasons given below.

2.1.1. Aromatic–aromatic or aromatic–aliphatic interactions

McGaughey et al. (1998) analysed R_{clo}, the closest contact distance between any two aromatic amino acids with centroid–centroid separation, R_{cen} less than 12.0 Å in proteins. The distribution (Fig. 9) was found to be bimodal with a prominent minimum between 4.5 and 5 Å (the distribution of R_{cen} showed a similar behavior with a minimum at ~7.5 Å). This was interpreted to represent the distance at which the interaction between the aromatic rings drops below the Boltzman temperature factor (~0.6 kcal/mol at 300 K). Inside the minimum in the distribution there is a binding interaction between the rings, with a sharp peak occurring at 3.8 Å (Thomas et al., 2002a); outside the minimum any direct ring–ring interaction is lost because of random thermal motion. Thus using residue pairs with $R_{clo} < 4.5$ Å would reveal any preferential orientations that they may have in their interactions.

The cutoff distance is just above twice the sum of the lengths of a C–H bond (1.1 Å) and the van der Waals radius of H (1.1 Å) to take into account possible coordinate errors of protein atoms (Brandl et al., 2001). That 4.5 Å is the optimum distance to be used also comes out from another, rather unrelated study. Samanta et al. (2002) defined the partner number (PN) of each residue as the number of atoms that are in contact (distance less than a cut-off distance) with it and found that this was related exponentially to the solvent accessible surface area (ASA) of the residue. One can then theoretically compute the ASA based on PN and compare the value with the observed value of ASA. When averaged over the whole protein structure, one finds that the match between the observed and the calculated values is the best when a cut-off value of 4.5 Å was used to define PN.

2.1.2. S···aromatic interactions

Bhattacharyya et al. (2004) found out that the threshold distance for the interaction of cystine sulfur atom with aromatic residues was 4.3 Å, and the method can be discussed in relation to methionine sulfur using the

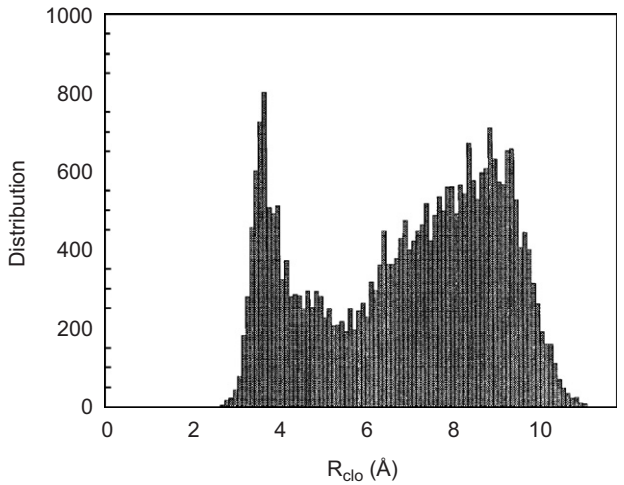

Fig. 9. R_{clo} distribution for 30,444 pairs of aromatic–aromatic amino acid side chains when $R_{cen} < 12.0$ Å, found in 505 proteins; R_{clo} is the closest distance between any two atoms in a pair of aromatic residues with a centroid-to-centroid separation, R_{cen} (taken from McGaughey et al., 1998).

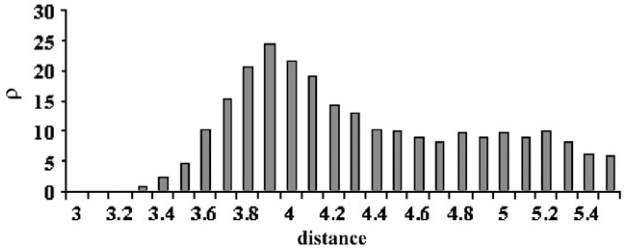

Fig. 10. The distribution of the density function, ρ, involving the interaction between methionine S atom and aromatic ring.

present database. A density function, ρ can be determined for aromatic atoms found at distances, $r = 3.0$, $3.1, \ldots, 5.5$ Å from the S atom. A shell of width 0.1 Å was assumed at each distance and the number of atoms (C or N from aromatic side chains) in it was found out (for a residue in contact, even if more than one atoms satisfied the criterion, only one was accepted). If the outer radius r_2 had N_2 occurrences and the inner radius r_1 had N_1, then

$$\rho = (N_2 - N_1)/(4/3\pi(r_2^3 - r_1^3)).$$

Met–aromatic interaction shows a sharp peak at about 3.9 Å that reaches a plateau beyond 4.3 Å (Fig. 10), indicating the specific nature of the interaction within this range. In an analogous manner, though using the ring-centroid-to-sulfur separation (instead of the closest atom–atom contact), Zauhar et al. (2000) found a local maximum at about 5.0 Å in the distribution of the distance that was absent in a control distribution.

2.2. Relative orientations involving ten planar groups

For the calculation of the geometry between the planar groups, their centroids were first determined (the center of mass of the five- or six-membered ring for Pro, His, Phe and Tyr, the mid-point of CD2 and CE2 atoms for Trp, the position of CZ for Arg, CG for Asp and Asn, and CD for Glu and Gln). A molecular axial system was defined with the origin at the centroid and the z-axis along the normal to that plane. The interplanar angle, P and the angle θ, made between the z-axis and the line joining the centroid of the second group to the origin of the first were computed (Fig. 11a). For each interaction, the geometry was placed in one of the nine grid elements (each spanning a range of 30° along P and θ). We used the same schematic representation as given in Bhattacharyya et al. (2002) to indicate the relative orientation of the second planar group with respect to the first at the nine grid elements (Fig. 11b). Each relative orientation is designated by a two-letter code (*ff*, *ot*, *ef*, etc.). The first letter indicates if the planar group of the first (or the so-called central) residue in a pair is interacting with its face (*f*), edge (*e*) or has the centroid of the second group (or the partner) in an intermediate (offset or *o*) position. The second letter indicates if the planar group of the second residue is tilted (*t*) with respect to the first or has its face (*f*) or edge (*e*) pointing towards the first. If O_{ij} is the observed frequency of occurrence in the grid element corresponding to the ith row and jth column (i and j varying from 1 to 3), the corresponding expected value (E_{ij}) can be calculated as the product of the sum of the observed numbers in the elements in the ith row and that in the jth column, divided by the total number of observations in all the nine grid elements (Samanta et al., 1999).

The above classification of geometries also corresponds to other nomenclature of aromatic–aromatic interactions (Singh and Thornton, 1985). The *ff*, *of* and *ee* orientations are equivalent to fully stacked, staggered stacking and parallel in-plane interactions, respectively; *ft*, *ot* and *et* orientations to the tilted interactions; *fe* and *ef* to edge-to-face or T-shaped geometry, and *oe* to the cogwheel or L-shaped geometry.

2.2.1. Identification of $X–H\cdots\pi$ interactions

For an aromatic–Pro or aromatic–aromatic contact, the existence of C/N–H$\cdots\pi$ interaction (with the C/N–H donor located on the central residue and the π electron system on the partner residue) was found out. First, hydrogen atoms are added to the residues using REDUCE (Word et al., 1999). Next we calculated an angle, ϕ, made between the line joining the centroid (Ar_{cen}) of the aromatic ring (partner) to the atom, P, of

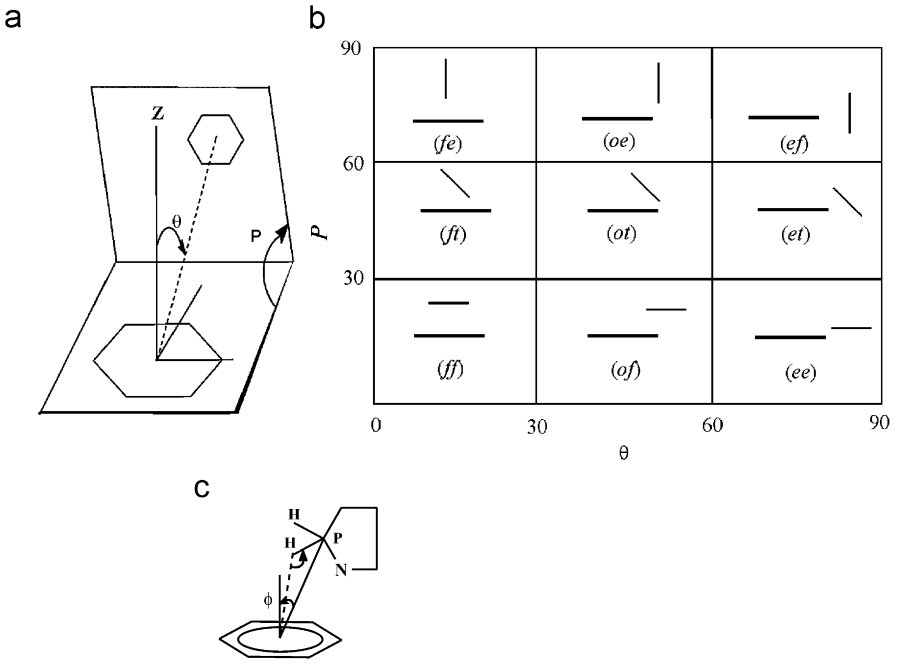

Fig. 11. (a) Definitions of P, the angle between the planes (represented by hexagons), and θ, the angle between the normal to the plane of the central residue and the direction linking its centre to that of the partner residue. (b) Schematic representation and their nomenclature for the relative orientations of the planes corresponding to various combinations of P and θ values (in degrees). Lines signify the planes (the thicker one for the central residue and the thinner one for the partner) edge-on. (c) Parameters (angles ϕ and P–H$\cdots$Ar$_{cen}$) for the definition of X–H$\cdots\pi$ interaction; P is the C/N atom in an aromatic or Pro ring having the shortest contact with the partner aromatic ring, whose centroid is Ar$_{cen.}$

the central residue with the shortest contact and the z-axis along the normal to the plane of the partner (Fig. 11c) (Steiner and Koellner, 2001; Bhattacharyya and Chakrabarti, 2003). If this angle is $\leqslant 30^\circ$ and the angle centered at the hydrogen atom, $\angle$P–H$\cdots$Ar$_{cen}$ is $\geqslant 120^\circ$, we assumed it to be a C–H$\cdots\pi$ interaction. If either ND1 or NE1 or NE2 is in shortest contact with aromatic ring, the interaction is assumed to a N–H$\cdots\pi$ interaction. As they are not bonded to any hydrogen atom, if N of Pro, CG of Phe, CG, CZ of Tyr, CG, CD2, CE2 of Trp and CG of His are in shortest contact, the atom with the next shortest contact is used to check for C/N–H$\cdots\pi$ interaction.

2.3. Orientation of the sulfide plane (of Met or cystine) relative to an aromatic ring

Disulfide bonds were identified using a cut-off distance of 2.3 Å between the SG atoms of two Cys residues. Only those Cys, Met and aromatic residues were considered for which the fractional occupancies of all atoms of Cys and Met and all ring atoms of aromatic residues were 1.00 and temperature (or B) factors of SG (Cys) and SD (Met) were $\leqslant 30\,\text{Å}^2$.

For the calculation of the geometry, the centroid (defined in Section 2.2) of the aromatic residue was first determined. A molecular axial system was defined with the origin at the centroid and the z-axis along the normal to the aromatic plane. The interplanar angle, P and the angle θ, made between the z-axis and the line joining the centroid of the aromatic residue to SG (for half cystine) or SD (for Met) were computed (Fig. 12a). For each interaction, the geometry was placed in one of the elements in a 3×3 grid (each element spanning a range of 30° along P and θ) (Fig. 12b). Each relative orientation is designated by a two-letter code (*fp*, *ot*, *en*, etc.). The first letter indicates if the aromatic residue is interacting with its face (*f*), or the SG of half cystine/SD of Met is near its edge (*e*) or located in an intermediate (offset or *o*) position. The second letter denotes if the sulfide plane (passing through CB–SG–SG′ of cystine and CG–SD–CE of Met) is normal (*n*) or parallel (*p*) to the aromatic ring or has a tilted (*t*) orientation. These designations (the second letter, in particular) are slightly

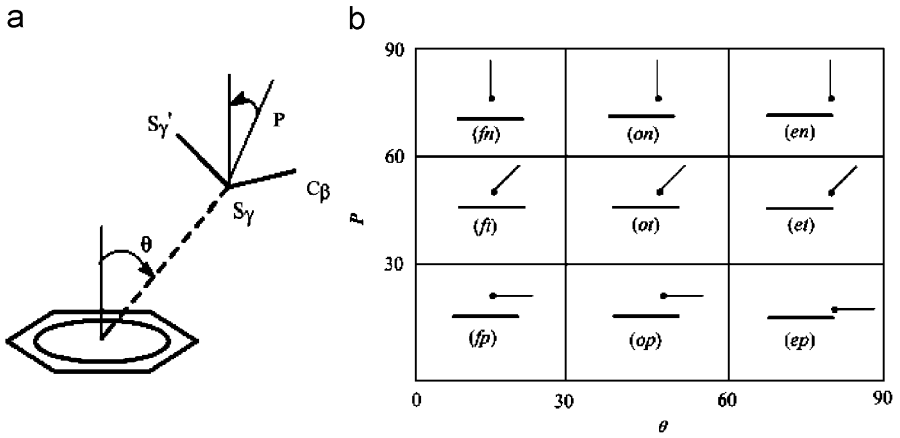

Fig. 12. (a) Parameters (P and θ) describing the orientation of the disulfide plane (defined by the SG atom having the contact and its two bonded neighbors) relative to the aromatic ring. The two normals passing through the aromatic centroid and the SG atom of the disulfide plane are shown. The interplanar angle, P is the angle between the two vectors; θ is the angle between the normal to the aromatic ring and its center to SG direction. For the calculation involving a Met residue, the atoms CG, SD and CE would define the plane. (b) Schematic representations for orientations corresponding to various combinations of P and θ values (in degrees). Lines signify planes (the longer one for the aromatic ring and the shorter one for the sulfide plane and the dot represents the S atom in contact) perpendicular to the paper.

different from those given in Fig. 11b, as here the focal point is the sulfur atom, rather than the whole plane. The expected values in grid elements were calculated as given in Section 2.2.

2.4. Aromatic–peptide interaction

For each aromatic residue, the contact with the main- and the side-chain atoms of all residues (taken consecutively) was found out. Any residue having a contact with the side chain was excluded. If the aromatic ring is within 4.5 Å of any main chain atom of a qualifying residue (at position i), then depending on the atom involved (N or CA in one case, and CA, C or O in another), the peptide group was constructed with the atoms in the preceding or the following residue. As CA (used synonymously as C^{α}) is at the intersection of two peptide groups, it is important to identify which of the two groups should be considered as interacting with the aromatic ring. To resolve this, the numbers of atoms of both the peptide groups within 4.5 Å of the aromatic ring were found out. The peptide plane with the greater number of interactions was accepted. However, if the numbers turned out to be equal (which happened in only a few cases) this interaction was excluded. The atom of the peptide group having the shortest contact distance was used for finding out the presence of any specific type of interaction. Once the (i, $i+1$) peptide group had been identified as interacting, the calculation was repeated starting with the residue $i+2$.

To analyze the geometry of interaction between peptide and aromatic planes, the centroid of peptide plane was taken as the mid point of C and N atoms. The interplanar angle, P and the angle, θ were determined for the orientation of the peptide group relative to the aromatic ring, as given in Section 2.2. We have checked whether the interaction can be categorized into N/C^{α}–H$\cdots\pi$ and C–H$\cdots$O based on the type of atom of the peptide plane that has the shortest contact with the aromatic ring. For the former, N or CA of the peptide group is the proton donor, while it is the aromatic group in the latter. For the C–H$\cdots$O interaction, the distance between any carbon atom of the aromatic ring and the peptide oxygen atom is <4.0 Å and the $\angle$C–H$\cdots$O centered at the hydrogen atom $\geqslant 120^{\circ}$.

2.5. Limitations of defining the relative orientations using only two angular parameters

When the two rings are at 90°, a rotation about the vertical axis passing through the partner ring does not change the values of P and θ, although the relative orientations may be altered. As a result, two further limiting orientations are shown in two extra boxes at the top of Fig. 13a. These are designated as *OE* and *EE*,

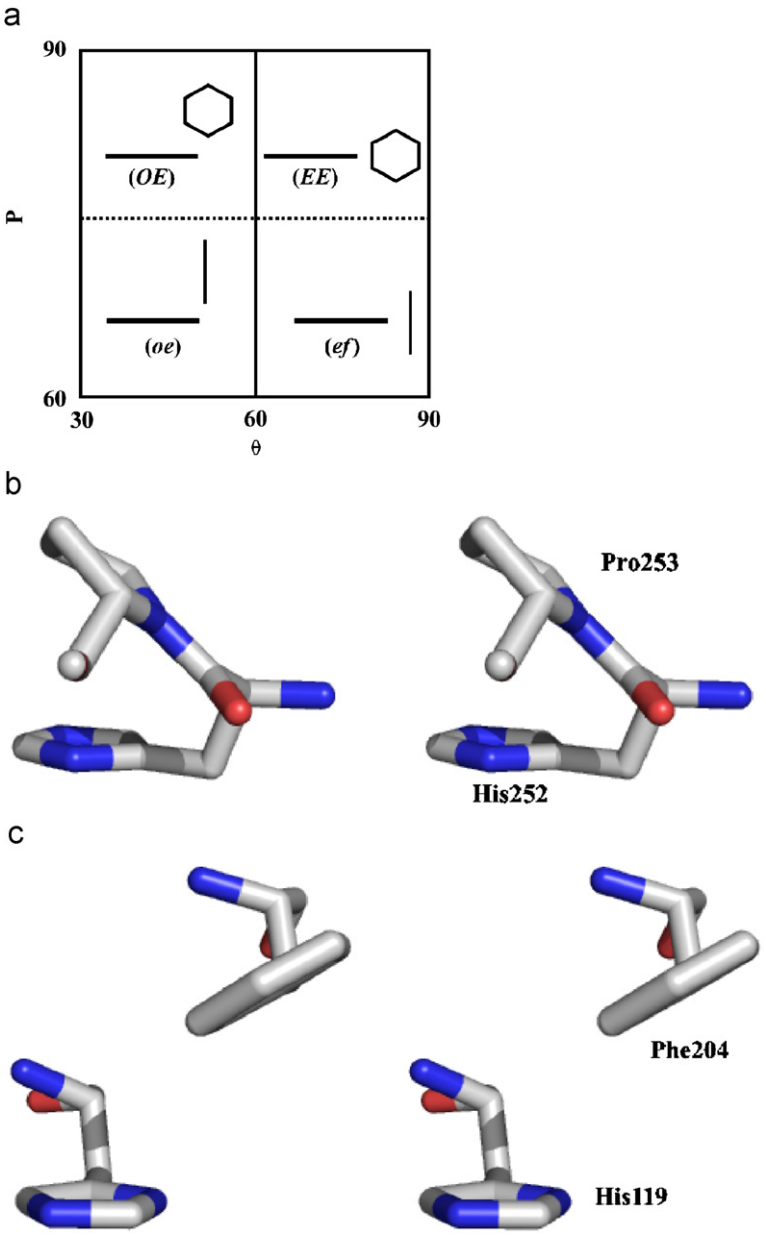

Fig. 13. (a) Two distinct alternatives of the orientations of the partner group (shown as a hexagon) when P is in the range 60–90°, combined with $\theta = 30$–$60°$ or 60–90°. The relative orientation of *ot* as observed in (b) 1a4y and (c) 1a8e, shown in stereo.

to distinguish them from *oe* and *ee*, which correspond to other geometries in Fig. 11b. It is not possible to uniquely represent the orientation using the two parameters, though it may be possible to resolve the ambiguity in an indirect way (Bhattacharyya et al., 2003), which however has not been attempted here.

Even in other grid elements, a particular interplanar angle does not fix the geometry precisely. For example, the relative orientation of *ot* is shown in two cases in Fig. 13b and c. In the first case, the partner ring points towards the extension of the central plane and the projection of the former falling on top of the latter. However, in the second case, the partner is pointing towards the face, but the projection lies beyond the central residue. It may be possible to distinguish between the two possibilities depending on the

occurrence of X–H$\cdots\pi$ interaction. However, this distinction is not done and *ot* is used to indicate both the configurations.

3. Pairwise geometries and salient features

The percentage of hydrogen bond interactions among contacts between residues that are capable of forming hydrogen bonds is provided in Table 1, and is quite similar to the one reported by Mitchell et al. (1997). A protonated His can only act as hydrogen bond donor, as are Arg and Trp. The fact that there are examples of Trp–His, Arg–His and His–His hydrogen bonds (3–11%) indicates that all His residues are not protonated. Hydrogen bond between larger residues is quite low in occurrence (4% for Tyr–Tyr/Trp). Salt bridges involving Arg-Asp/Glu make these pairs exhibit the highest percentage of hydrogen bonds (57%); when the acid is replaced by the amide group the percentage drops down to 23%. About 34% of the contacts involving His or Tyr and an acidic group are through hydrogen bonds.

The distribution of relative orientations in different residue pairs, along with the percentage of hydrogen bonds and C/N–H$\cdots\pi$ interactions (involving only Pro/aromatic–aromatic pairs) in different orientations are given in Fig. S1 as supplementary material and a selected few are presented in Fig. 14 for the discussion of the general features below. It may be mentioned that a small sequence difference between the interacting residues may influence the relative geometry (Bhattacharyya et al., 2002; Bhattacharyya and Chakrabarti, 2003; Thomas et al., 2002b; Meurisse et al., 2003, 2004). However, most of the interactions, especially involving aromatic residues are long-range (Section 4) and any effect of local steric constraints should not mask the overall trend that is under consideration here.

Table 1
Percentage of contacts that are hydrogen bonded

Interacting pair	No. of interactions	Fraction hydrogen-bonded
Tyr–Tyr	3516	0.04
Tyr–Trp	2084	0.04
Tyr–His	1893	0.16
Tyr–Arg	3566	0.11
Tyr–Asp	3347	0.34
Tyr–Glu	3270	0.33
Tyr–Asn	2669	0.14
Tyr–Gln	2180	0.15
Trp–His	952	0.03
Trp–Asp	1410	0.19
Trp–Glu	1570	0.21
Trp–Asn	1325	0.07
Trp–Gln	1291	0.08
His–His	1668	0.11
His–Arg	1709	0.07
His–Asp	2610	0.36
His–Glu	2269	0.34
His–Asn	1416	0.13
His–Gln	1083	0.14
Arg–Asp	7204	0.58
Arg–Glu	7868	0.56
Arg–Asn	2874	0.22
Arg–Gln	2480	0.24
Asp–Asn	3996	0.21
Asp–Gln	2449	0.27
Glu–Asn	2897	0.20
Glu–Gln	2223	0.25
Asn–Asn	3646	0.21
Asn–Gln	2213	0.23
Gln–Gln	1684	0.26

3.1. Aromatic–aromatic (Fig. 14a–f)

A wide variety of conformations spanning over all the 9 grid elements are observed, though the regions that are electrostatically unfavorable are sparsely populated (Hunter et al., 1991). Broadly, *of* and *ef* (Fig. 24a) are preferred, as has been indicated by most of the earlier studies (Section 1.1.1). However, there are subtle differences between individual residues. For His-His, it is *ff* instead of *of*, and metal binding (Section 6) can also change the geometry (Bhattacharyya et al. 2003). Neither *ff* nor *of* is favored for Trp-Trp pairs (Samanta et al., 1999). For the Phe-Tyr pair, the preferential interaction is between the edge of Phe with the face of Tyr. About 30% of all unique aromatic–aromatic contacts also form C/N–H$\cdots\pi$ interactions.

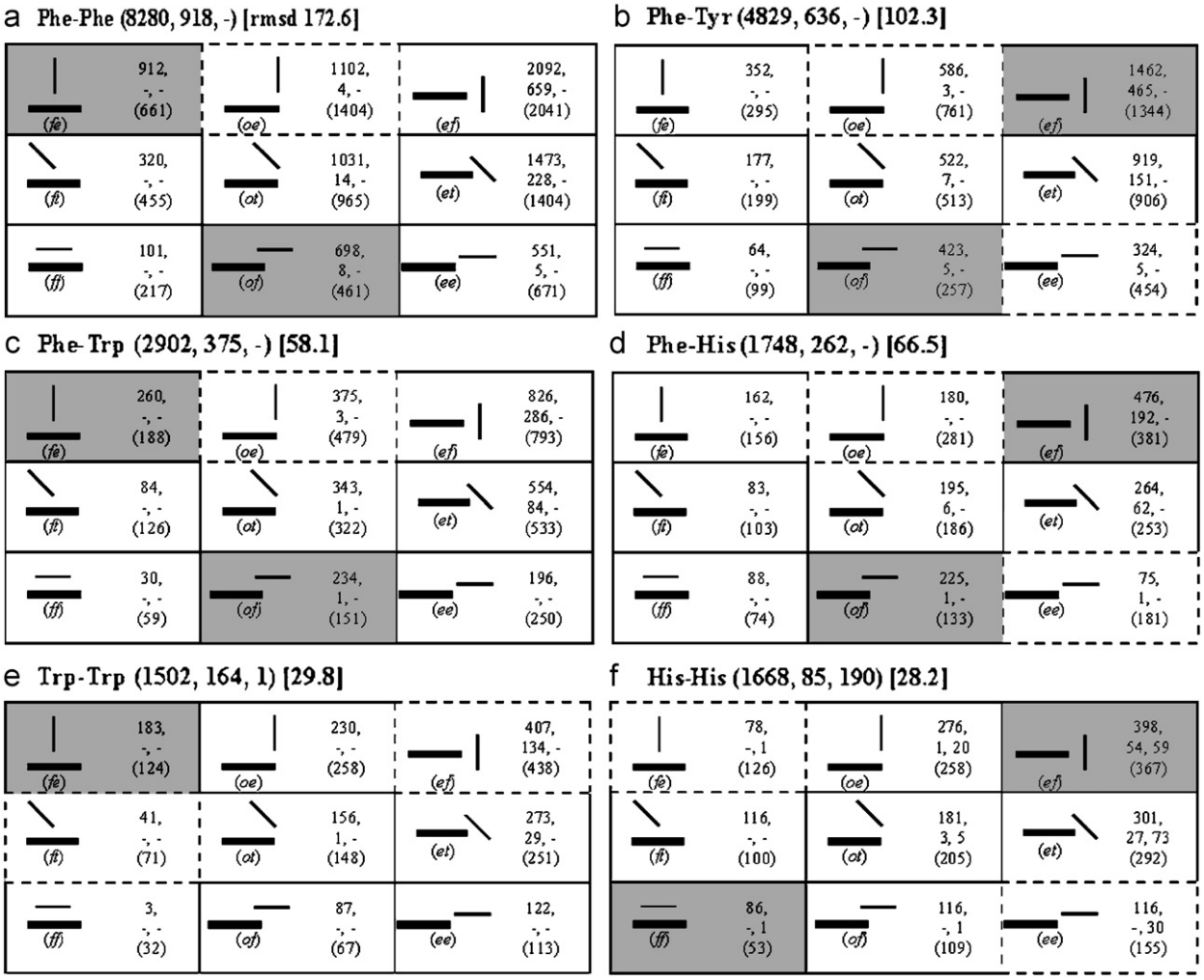

Fig. 14. The distribution of the geometry of a few selected pairs of residues, extracted from the complete set provided as supplementary material (Fig. S1). The name of the central residue, followed by that of the partner is given on top. The numbers in each grid element are the observed number, C/N–H$\cdots\pi$ interactions (calculated only for aromatic–aromatic and Pro–aromatic pairs, and with the central residue acting as the donor), hydrogen bonds and the expected number (in parentheses). On top of each grid, the total numbers of the first three quantities are provided (in parentheses), along with the rmsd (σ) value [in square bracket] for the difference between the observed and calculated values over the nine grid elements. The boxes in which the observed number exceeds the expected number by 1σ are shaded and where it is less is indicated with perforated border. If a particular interaction is absent (or is not calculated, as for example, C/N–H$\cdots\pi$ interactions between residues other than Pro and aromatic), a hyphen is put in place of the corresponding number.

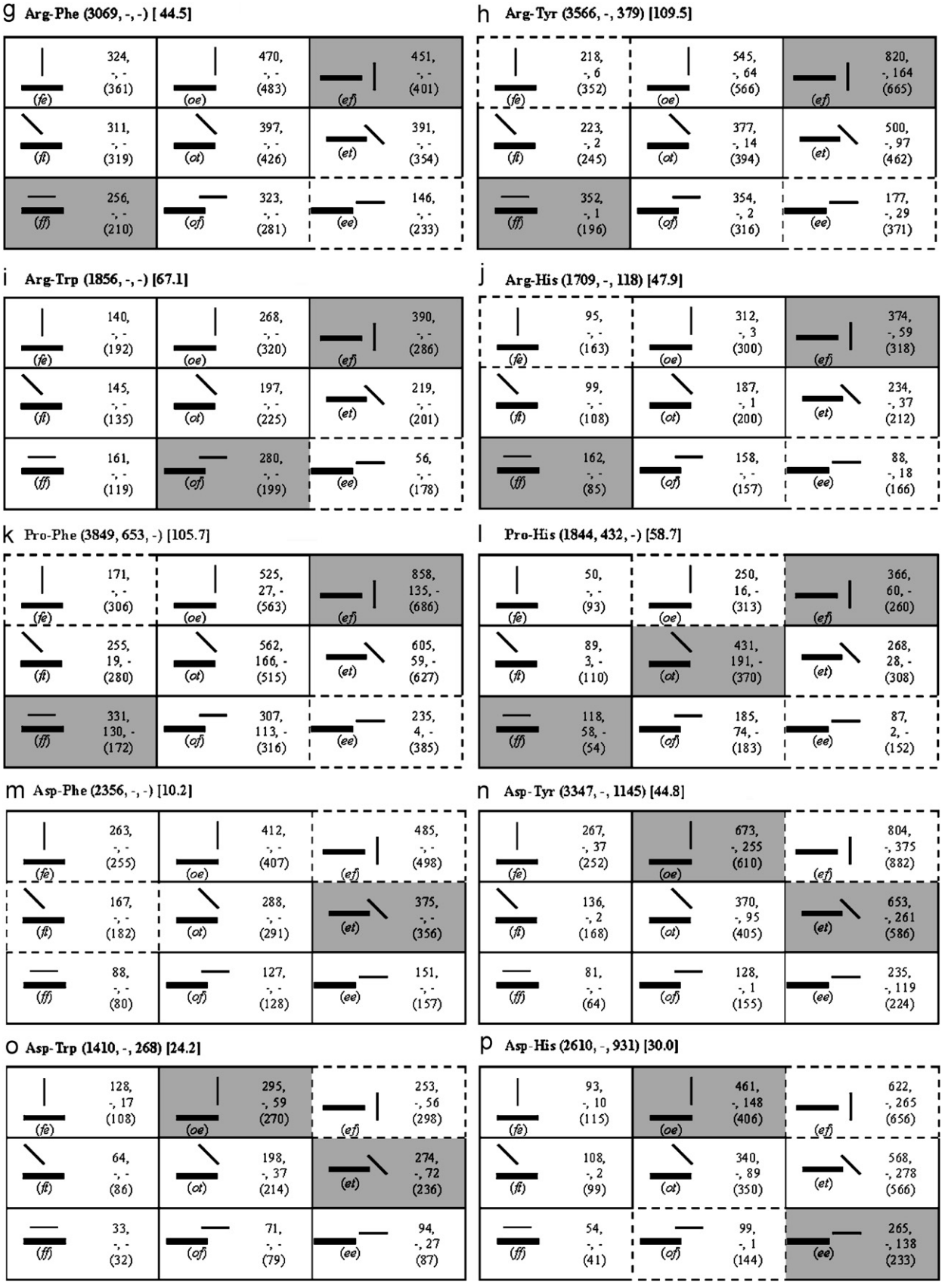

Fig. 14. (*Continued*)

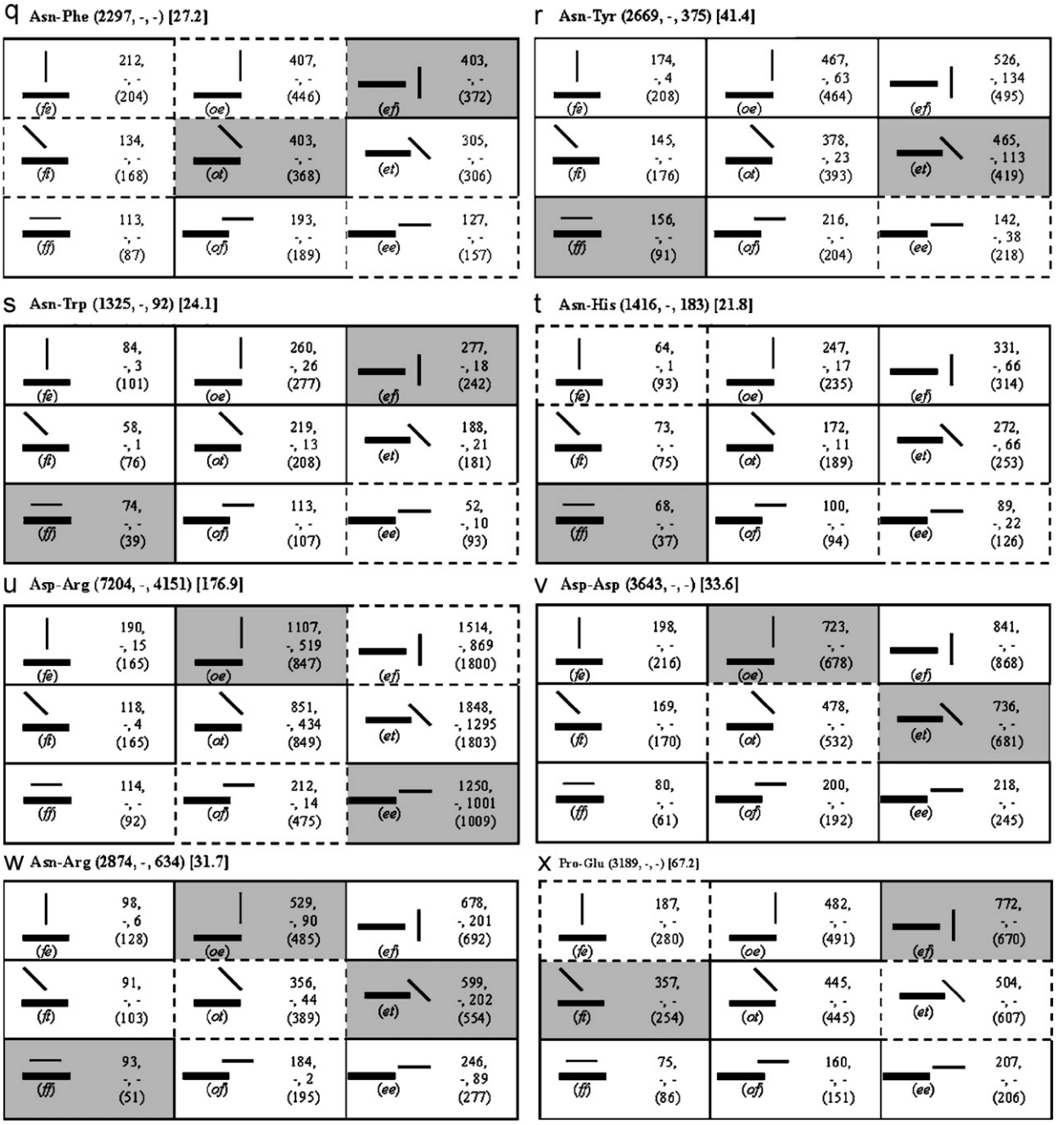

Fig. 14. (*Continued*)

3.2. Arg–aromatic (Fig. 14g–j)

Preferred orientations are *ff* (*of* with Trp) and *ef*. With Tyr and His, 81% of the hydrogen-bonded cases are observed with grid elements in the right column. It has been suggested that a parallel stacking arrangement of the guanidinium group directly over the center of the rings of aromatic side chains serves to orient the Arg side chain without interfering with its ability to form hydrogen bonds elsewhere; in a survey, 34% of proteins were found to contain at least one example of this type of interaction (Flocco and Mowbray, 1994). However, in the

ef orientation there is the possibility of the formation of N–H$\cdots\pi$ interactions (Mitchell et al., 1994; Steiner and Koellner, 2001) and some of these are important functionally (Waksman et al., 1992).

3.3. Pro–aromatic (Fig. 14k and l)

As with Arg–aromatic, the preferred geometries are *ff* (Fig. 20j) and *ef*; 40% in the former and 15% in the latter participate in C–H$\cdots\pi$ interactions, thus providing an explanation for the higher preference for these two geometries (Bhattacharyya and Chakrabarti, 2003). Neighboring grid elements can also support C–H$\cdots\pi$ interactions (Fig. 23b), and for His *ot* is also a preferred geometry.

3.4. Acid–aromatic (Fig. 14m–p)

Relative to the acidic group *et* is the most common pattern of interaction for the aromatic residue. With Tyr, Trp and His, 67% of hydrogen bond interactions occur with geometries in a grid element in the right column. It may be noted that being linked by hydrogen bond does not necessarily constraint the two groups (for example, Asp and His) to be coplanar (geometry *ee*) (Bhattacharyya et al., 2003).

3.5. Amide–aromatic (Fig. 14q–t)

ff is the most common geometry, with *ef* being the next propitious. The stacked geometry has been found to outnumber orientations leading to N–H$\cdots\pi$ interactions by around 2.5:1 (Mitchell et al., 1994; Samanta et al., 2000—these publications, however, consider the amino groups of Asn, Gln, His, Trp and Arg side chains together). *Ab initio* calculations (Duan et al., 2000), performed on the benzene–formamide complex to model the aromatic–amide interaction, revealed a significant stabilization energy of −4.0 kcal/mol, with two energetically similar conformations—the parallel stacking and the perpendicular, T-shaped geometry with the N–H$\cdots\pi$ interaction (Fig. 20a). Molecular mechanics calculations have shown that when all the hydrogen bonding potential of an amide group is satisfied by surrounding solvents, the stacked geometry is preferred; however, in the absence of the solvent, the amide group being perpendicular to the aromatic ring is favored (Worth and Wade, 1995). As in acid–aromatic category, 72% of hydrogen bond interactions occur in geometries belonging to the right column; however, *ee* geometry is observed in less than expected numbers.

3.6. Acid–acid (Fig. 14v)

oe and *et* seem to have the preference (Fig. 25m, Table 3). It is to be noted that while Flocco and Mowbray (1995) considered the acidic pairs in which two oxygen atoms are in close contact, here the contact is not limited to oxygen atoms only, and interactions mediated by metal binding are also included.

3.7. Acid–Arg (Fig. 14u)

ee is preferred and 80% of contacts with this geometry are hydrogen bonded. The *oe* geometry with the two planes nearly perpendicular to each other (Fig. 20b) is also observed more than expected, but here only 48% of the contacts are hydrogen bonded. The in-plane *ee* geometry is expected when two atoms are involved from each side in the formation of the salt bridge; however, even when the interaction is of the type single oxygen–single nitrogen, which is of the most frequent occurrence, there is a preference to lie in the plane (Mitchell et al., 1992; Singh et al., 1987).

3.8. Amide–Arg (Fig. 14w)

ff is now preferred. Unlike 80% of contacts occurring in *ee* grid being hydrogen bonded for acid–Arg, it is only 36% for amide–Arg (Fig. 20c) (in general, a lower percentage of amides, as compared to acids are hydrogen bonded to Arg—Table 1). This makes the pair neutral towards this geometry.

With reference to Fig. S1, it can be seen that the preference is *ff* for Arg–Arg, Arg–Pro and amide–amide, *ef* for Pro–Pro, and *ff* and *ef* for Pro–Asp/Asn. The four residues belonging to acid–amide pair show a diverse range of geometries when interacting among themselves (Fig. 20d and e).

3.9. Pro–Glu/Gln (Fig. 14x)

The preference for *ff* that was observed with Pro–Asp/Asn is shifted to a more tilted geometry *ft* (Fig. 20g) with Glu/Gln. The other preference for *ef* remains the same.

3.10. Aromatic–Met (Fig. 15a)

There appears to be two preferred orientations of the CG–SD–CE plane in Met relative to an aromatic ring, one being *op*, the other being *et* (Fig. 21a) or *en*. The real distribution of the sulfide planes in Met relative to the His ring is shown in Fig. 16 (about one-third of the contacts involve ring N atoms), so that one can have an idea how this translates into the distribution of geometries in Fig. 15. There have been conflicting results on the preference of interaction of S with the face (Morgan et al., 1978; Pal and Chakrabarti, 2001) or the edge (Reid et al., 1985; Iwaoka et al., 2002a). Yet another study suggested that the preference of S to interact at a significant angle of elevation above the plane of the ring in proteins gets changed to a position in the plane in small molecule structures (though most of the latter structures have S as part of a ring and is thus not directly comparable to the situation in proteins) (Zauhar et al., 2000). Usually, these observations were based on the analysis of the polar angle θ only. Use of two parameters provides a better representation of the orientation: *et* (or *en*) indicates an interaction closer to the ring edge, and *op* indicates an interaction with the face, but displaced from the center. The preferred orientations seem to have the sulfur lone pairs directed away from the ring centroid (Bhattacharyya et al., 2004).

3.11. Aromatic–cystine (Fig. 15b)

The interaction between the disulfide plane CB–SG–SG′ and the aromatic ring essentially follows the pattern of aromatic–Met interactions (Fig. 21b and c). The earlier analysis (Bhattacharyya et al., 2004) reported a similar finding and indicated that of all the aromatic residues Trp has the highest propensity of interactions. 94% of the disulfide bonds were found to have direct S⋯O contacts with carbonyl oxygens, an interaction that would be competing to retain its own geometry in the environment of the divalent sulfur. Despite this, the consistency of the geometry observed for S⋯aromatic interactions is quite remarkable.

Expanding on the theme of sulfur interactions, we found out if S⋯S contacts are important when two Met residues interact, and found that out of 530 cases of Met–Met interactions in 32 cases the two sulfur atoms are within 4.5 Å. Such interactions may also follow well-defined geometry (Iwaoka et al., 2002a).

3.12. Aromatic–peptide interactions

The preferred orientations (Fig. 17a–d) are quite similar (*of* and *ef*) to those found with aromatic–aromatic interactions—however, for both His–peptide and His–His *ff*, rather than *of*, is observed. A large percentage of the contacts are associated with N–H⋯π, C–H⋯π interactions, C–H⋯O and conventional hydrogen bonds (Fig. 22). Of the total contacts, 2.4% are of the type X–H⋯π (X–H located in the peptide group), 32.7% are C–H⋯O (C–H located in the aromatic ring) and 6.8% are hydrogen bonds. Steiner and Koellner (2001) observed that the peptide N–H and C^{α}–H groups are involved in X–H⋯π interactions in frequencies 0.17 and 1%, respectively. Thus the C–H⋯O interaction is the most favored interaction and indeed in general the oxygen atoms, including those of the solvent molecules, are found in the periphery of aromatic rings (Thomas et al., 1982; Gould et al., 1985; Singh and Thornton, 1990; Flanagan et al., 1995; Samanta et al., 2000).

Of all the residues, Gly is the one having the closest contact of a main-chain atom with an aromatic ring (Fig. 17e). Gly has been found to be a predominant contributor to both C–H⋯O and C–H⋯π interactions, brought about by the presence of two hydrogens at the C^{α} atom (Brandl et al., 2001; Ibrahim and Pattabhi,

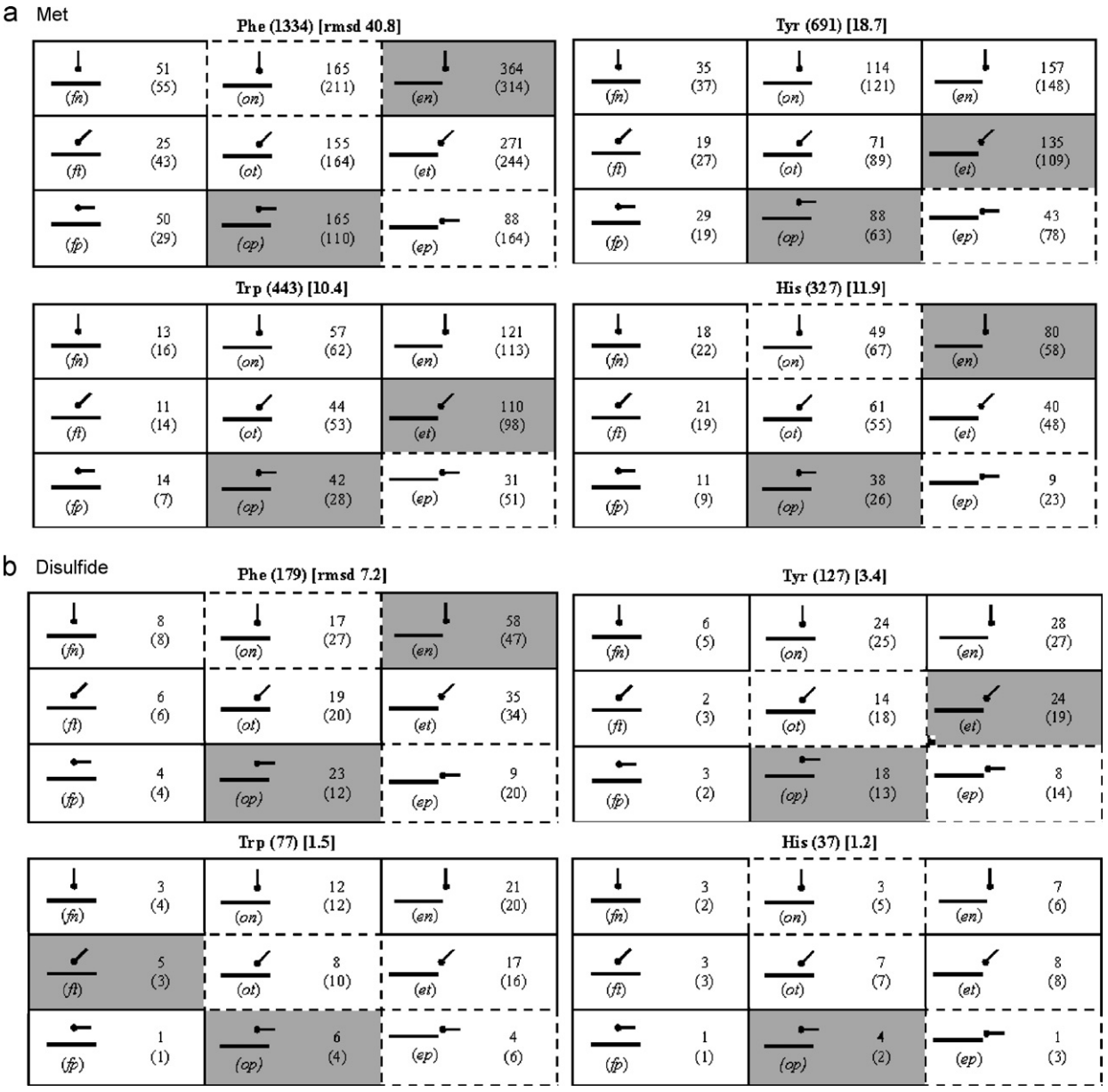

Fig. 15. (a) The observed and the expected (in parentheses) numbers in grid elements corresponding to different geometrical orientations of the CG–SD–CE planar moiety of Met relative to the aromatic rings. (b) Same as in (a), but for the CB–SG–SG′ plane of the disulfide group. For other details see the legend to Fig. 14.

2004) and the closer approach that is possible between peptide and aromatic planes in the absence of a side chain.

4. Sequence difference between interacting residues

The short- and long-range natures of aromatic–aromatic interactions have been studied (Thomas et al., 2002a, b; Bhattacharyya et al., 2003; Meurisse et al., 2003, 2004)—the former generally contributing to the stability of the secondary structure and the latter holding the tertiary structure together. Fig. 18 provides the distribution of the sequence differences of the interacting pairs; cases with $|\Delta| \leqslant 5$ can be considered to

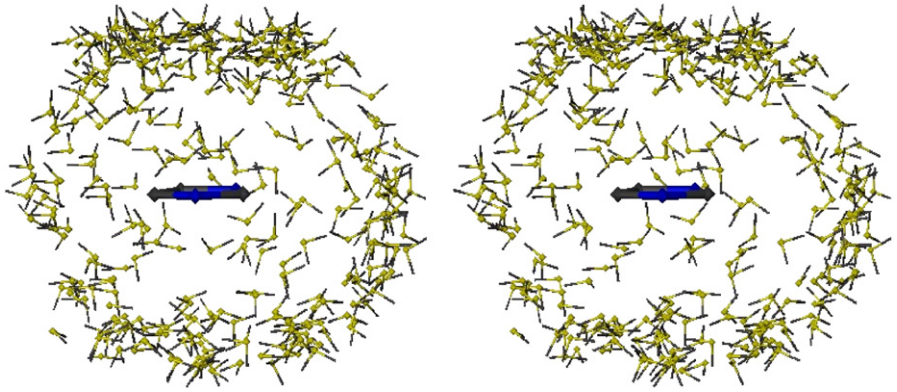

Fig. 16. Stereoview of the scatter plot showing the disposition of different Met thioether groups around the His ring.

represent short-range interactions. Combining both $\pm\Delta$ regions, 60–80% of all interactions involving the aromatic residues are long range. This number is much smaller (30–45%) for the interactions of Asp with acid/amide side chains and Pro, and increases somewhat (40–50%) for the longer residue, Glu. Asn and Gln exhibit patterns similar to Asp and Glu. Thus Pro, Asp and Asn contribute relatively more towards local structures. The long-range interactions involving aromatic pairs mainly bridge two β-strands, followed in preference by a bridge between a β-strand and a helix, and then between a β-strand and a non-regular region (Thomas et al., 2002a).

For each pair of residues interacting in the short range, the preferred secondary structure at a given Δ value is provided in Table 2, which can be used to understand the location in secondary structures that causes the occurrence of the shaded bars (when a bar on one side is distinctly longer than the equivalent bar on the other side) in Fig. 18. Location in an α-helix is the most dominant secondary structural motif (43% of the cases) for Phe/Tyr/Trp–His pair with $\Delta = 4$. Residues in i, $i+4$ positions in a helix may interact, but the specific interactions can take place only when a given pair of residues are in a particular order and would be lost if the residues are interchanged (Fig. 24c). The order seen in two other pairs is aromatic–Arg (63% of the cases of $\Delta = 4$ are in helices), aromatic–amide (58%, especially involving Tyr/Trp and Gln) (Fig. 20a). The contact between aromatic and Arg side chains has been attributed to cation–π interaction, for which there is a strong bias toward Trp among the aromatics and for Arg (rather than Lys) as a cationic side chain (Gallivan and Dougherty, 1999). However, the stabilizing effect, leading to increased helicity of model peptide with the (i, $i+4$) placement of aromatic–polar pair (Tsou et al., 2002), can also result from the hydrophobic contacts between the aromatic ring and the CH_2 groups of the cationic side chain (Andrew et al., 2002).

Instead of four, a sequence difference of three is also observed in helices and two sequential pairs are acid–Arg (68%) (Fig. 20b) and amide–Arg (63%) (Fig. 20c). A salt bridge can be formed when an acidic and a basic group are on the same face of the helix, three or four residues away (Barlow and Thornton, 1983; Kumar and Nussinov, 1999; Musafia et al., 1995). Interestingly however, with $\Delta = 4$ the order in which the two residues occur is immaterial. But with $\Delta = 3$, the acid/amide group is located first. Being in adjacent positions, the Arg–aromatic sequence is found in helices 46% of the time.

Two acidic groups can occur in helices (amphipathic ones, in particular) with $\Delta = 1$ and 3. Whereas Glu–Glu is found with both the Δ values, when Asp is associated with Glu and $\Delta = 3$, the order is Asp–Glu. Likewise, when there is a Gln in place of Glu, the same order is observed (Fig. 20d). Pro–Arg/amide with $\Delta = 3$ can occur near the N-terminal end of helices. Pro–acid with $\Delta = 2$ or 3 can connect a Pro occurring before the start of a helix to an acid group in the helix.

Acid and amide side chains can be in α-helices separated by three or four residues and there can be hydrogen bond linking them (Fig. 20e and f); however, the existence of any specific order of the residue type is not apparent. This is because with these residue pairs, a residue at position i can act as both proton donor, as well as acceptor (and likewise at position $i+1$). A free energy of interaction of -1.0 kcal/mol is observed for a hydrogen bond in the Gln–aspartate (i, $i+4$) pair in peptide helices (Huyghues-Despointes et al., 1995). For

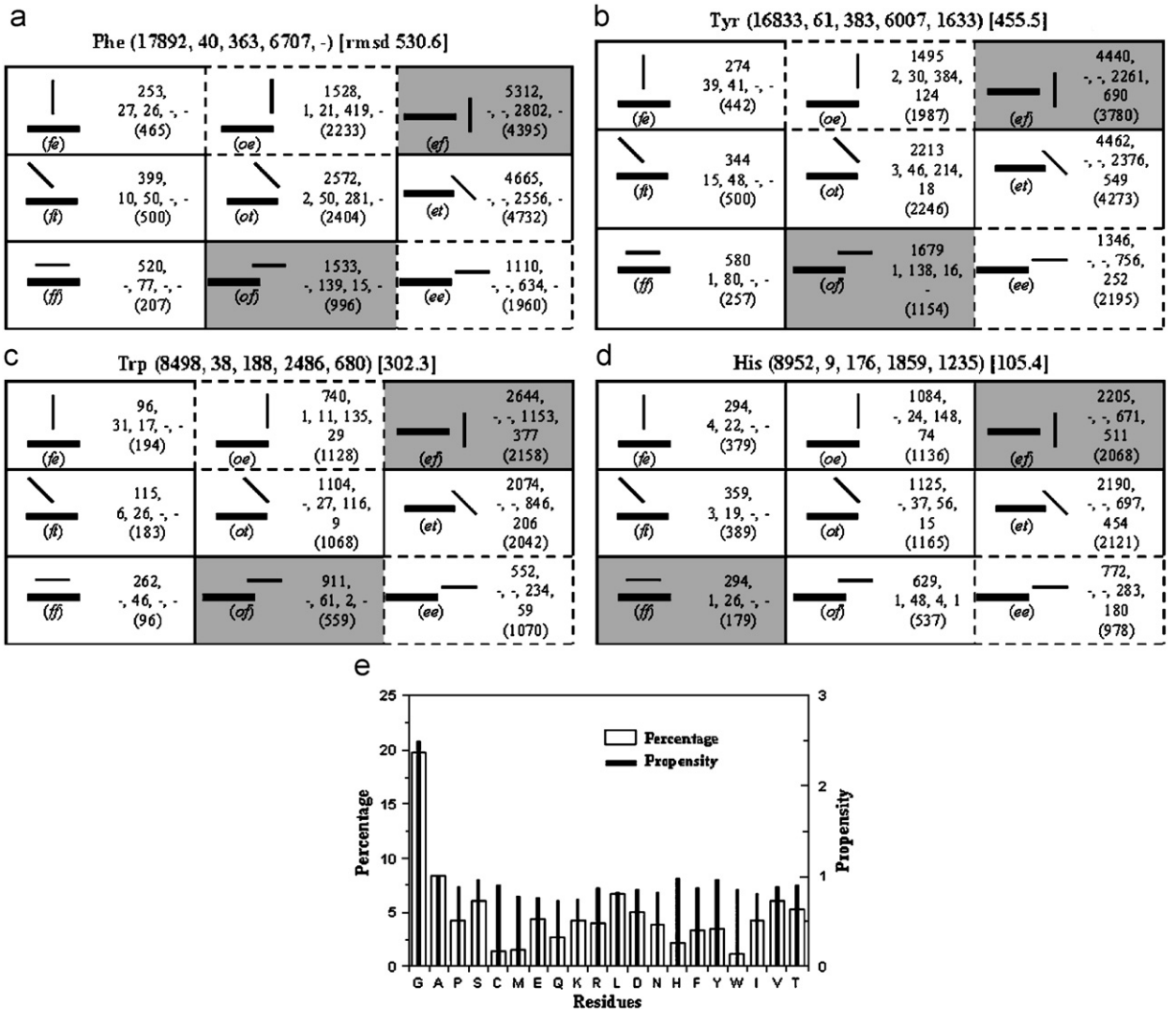

Fig. 17. (a)–(d) The distribution of the peptide groups in different orientations relative to the four aromatic residues. Various features are explained in the legend to Fig. 14. Following the name of the central residue, the numbers in parentheses are: the observed number, and the interactions of types N–H$\cdots\pi$, C–H$\cdots\pi$, C–H$\cdots$O and N–H$\cdots$O hydrogen bond. (The X–H group in the first two interactions is located on the peptide group, and on the aromatic ring for the third.) (e) Amino acid distribution—for the residues in a peptide group having the shortest contact (involving a main-chain atom) with an aromatic ring—shown as the percentage composition (left vertical axis) and propensity values (right axis); the latter values were obtained by dividing the observed percentage composition by the overall percentage composition of residues in a database (Chakrabarti and Pal, 2001).

a Gln–Asn pair at the same spacing the experimental value is between -0.4 and -0.7 kcal/mol (Stapley and Doig, 1997).

Compared to helices, the β-strands do not exhibit much asymmetry in the sequence difference of the interacting residues. Amide–aromatic (especially, Asn–Phe) with $\Delta = 2$ (Fig. 20h) is found in β-strands 32% cases. Amide–Arg with $\Delta = 2$ has similar features. As β-turns involve only short stretches of residues, only pairs with $\Delta = 1$ can show any asymmetry, which can be explained by the residue preference at the four positions i to $i+3$ in different turn types (Hutchinson and Thornton, 1994). The sequence His–Pro is favored at the first two positions of type I β-turn as this allows the two rings to stack with the possibility of formation of C–H$\cdots\pi$ interactions, with the His simultaneously acting as a hydrogen bond acceptor from the backbone NH group at position $i+2$ in the turn (Fig. 20i) (Bhattacharyya and Chakrabarti, 2003). Pro is the most

favored residue at the $i+1$ position of type I β-turn; an Asp or Asn can be at position i and an aromatic residue, such as His can occupy $i+2$ (Fig. 20j). While the sequence Asp–Pro is very conspicuous at the first two positions of type I β-turns, Pro–Asp is also found in such turns (68%) with Asp occurring at $i+2$

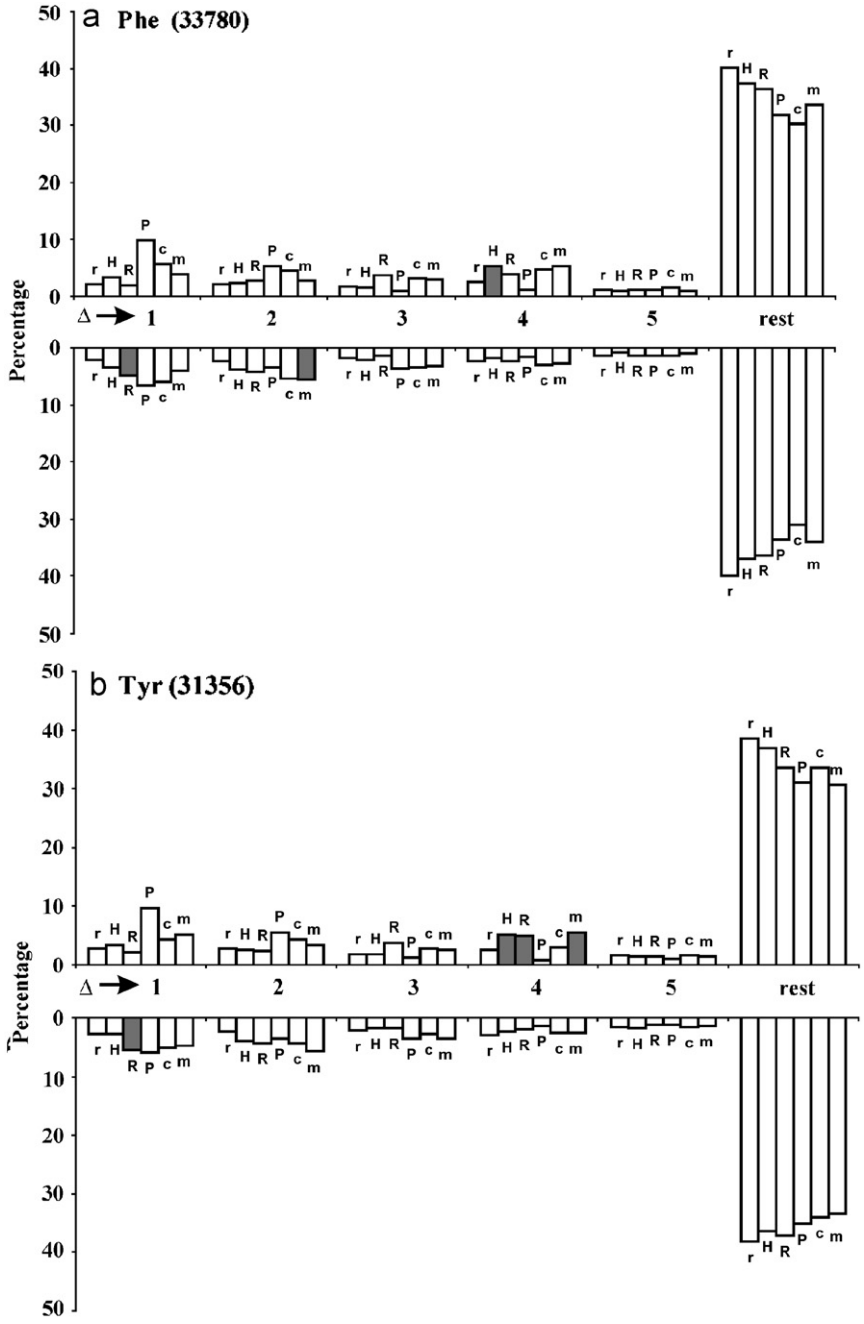

Fig. 18. (a)–(j) Histogram of sequence differences, Δ (planar-X, where X is the residue under consideration, indicated on top of each plot, interacting with another planar residue). The total number of interactions for X is given in parentheses. The capital letters on top of the bars correspond to the amino acid code of the interacting residue, whereas the small letters represent a group of similar residues: r: Phe, Tyr and Trp; c: Asp and Glu; m: Asn and Gln. The percentage is calculated with respect to the total number of each interacting pair. $\Delta>0$ is shown on the top panel and $\Delta<0$ in the bottom. A comparison has been made for the bars in the upper and lower panels for a given $|\Delta|$ value for each residue type. If any (or both) of the bars has a height of 5 and one is at least double the other, then the longer bar is shaded.

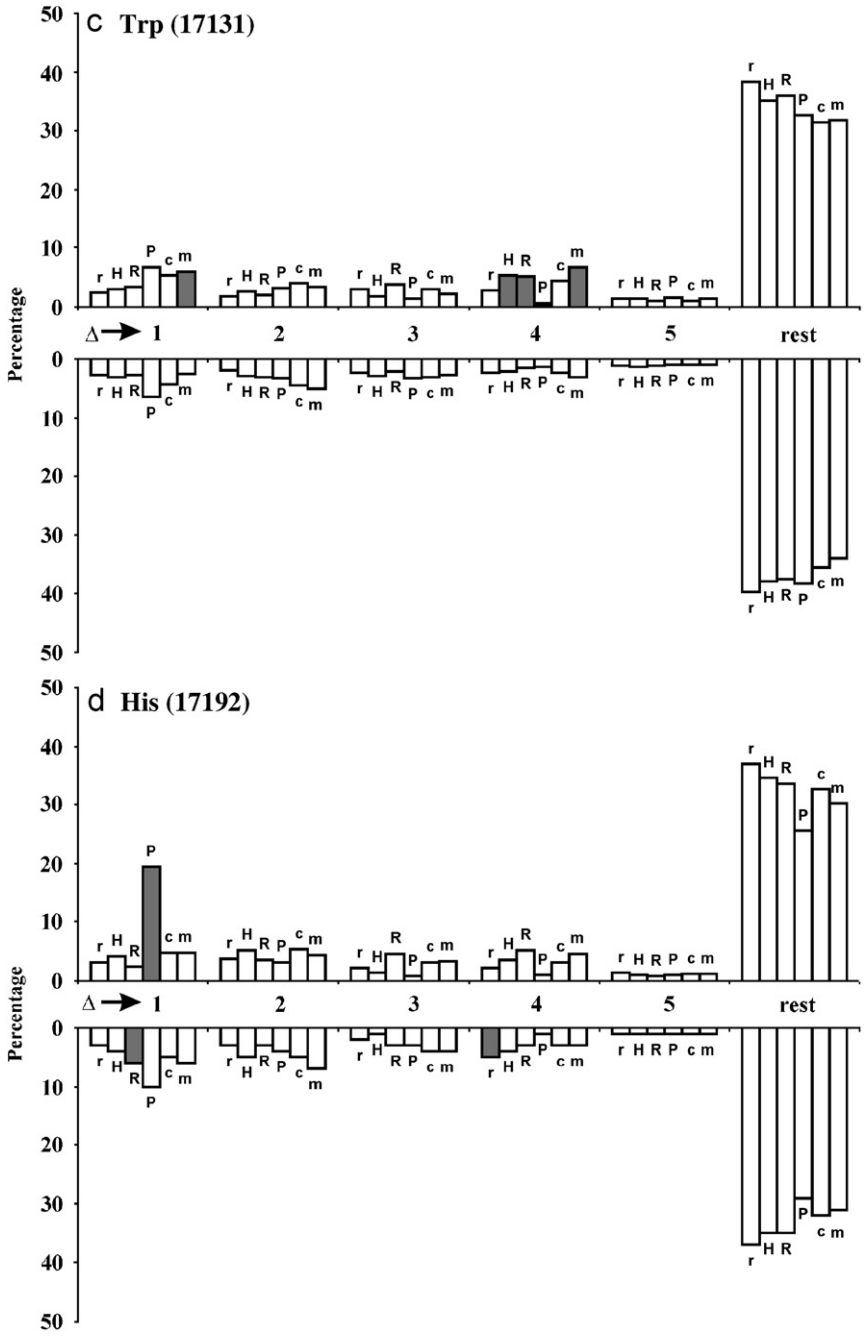

Fig. 18. (*Continued*)

(Fig. 20i). Interestingly however, the sequence Pro–Glu is found in helices (60%) (Fig. 20g) rather than in turns; 58% of 303 occurrences of the pairs in α-helices are involved in C–H···O interactions, as shown.

The distribution of the sequence differences of aromatic···S (S belonging to Met or cystine) interactions, shown in Fig. 19 indicates that about 40% of the cases are long-range. There is an asymmetry between the upper and lower panels when $|\Delta| = 4$ (for both Met and half cystine) and one (only for half cystine), such that the sulfur-containing residue has a higher sequence number than the aromatic. This is due to the usual

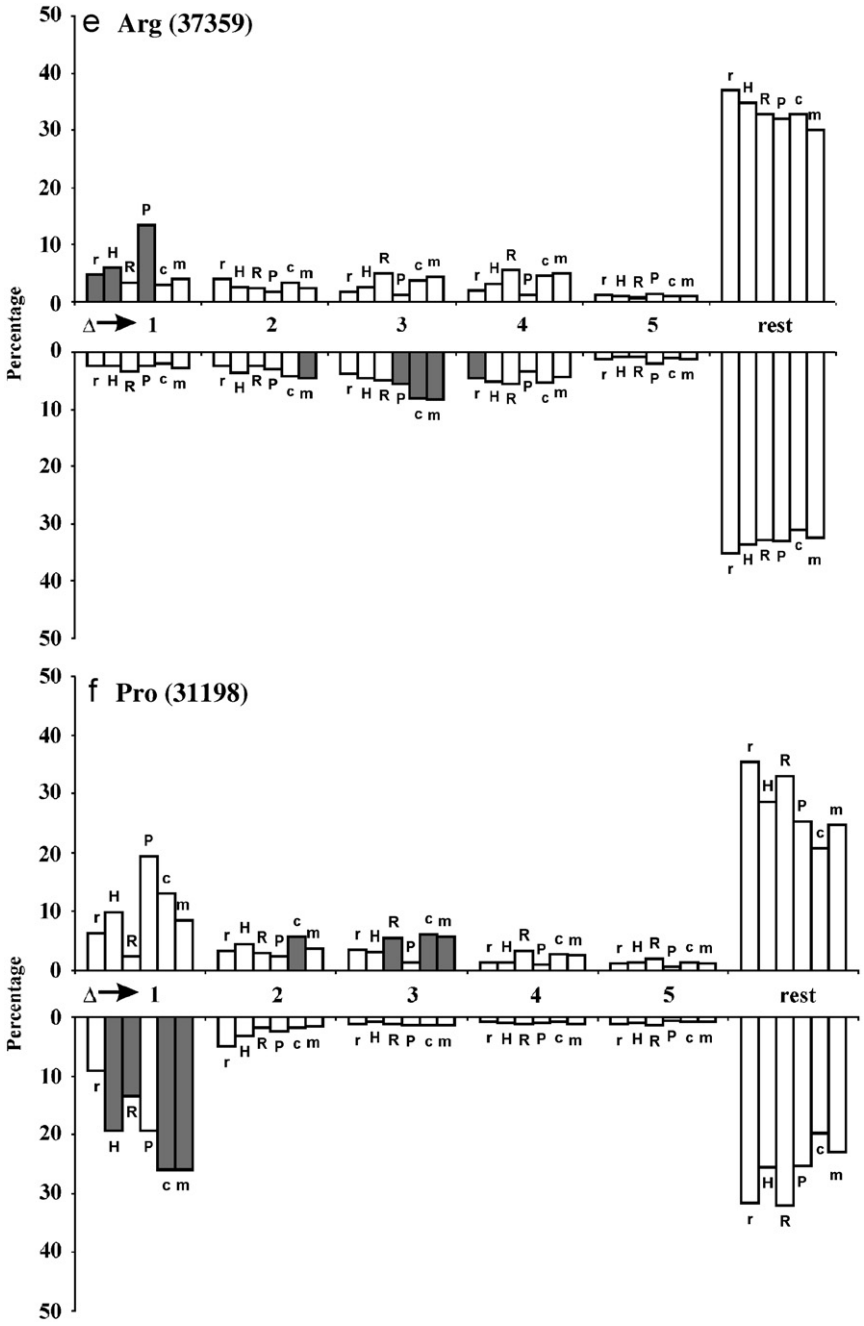

Fig. 18. (*Continued*)

occurrence of the aromatic ring and the S-containing residue at i, $i+4$ or i, $i+1$ positions in α-helices (Fig. 21) (Klingler and Brutlag, 1994; Pal and Chakrabarti, 2001; Bhattacharyya et al., 2004).

4.1. Sequence difference in aromatic–peptide interactions

In 11% of all aromatic–peptide interactions the residues involved (the aromatic and the two constituent residues of the peptide group) are in helices. When in α-helices the local interactions with $\Delta = -4$ and -3

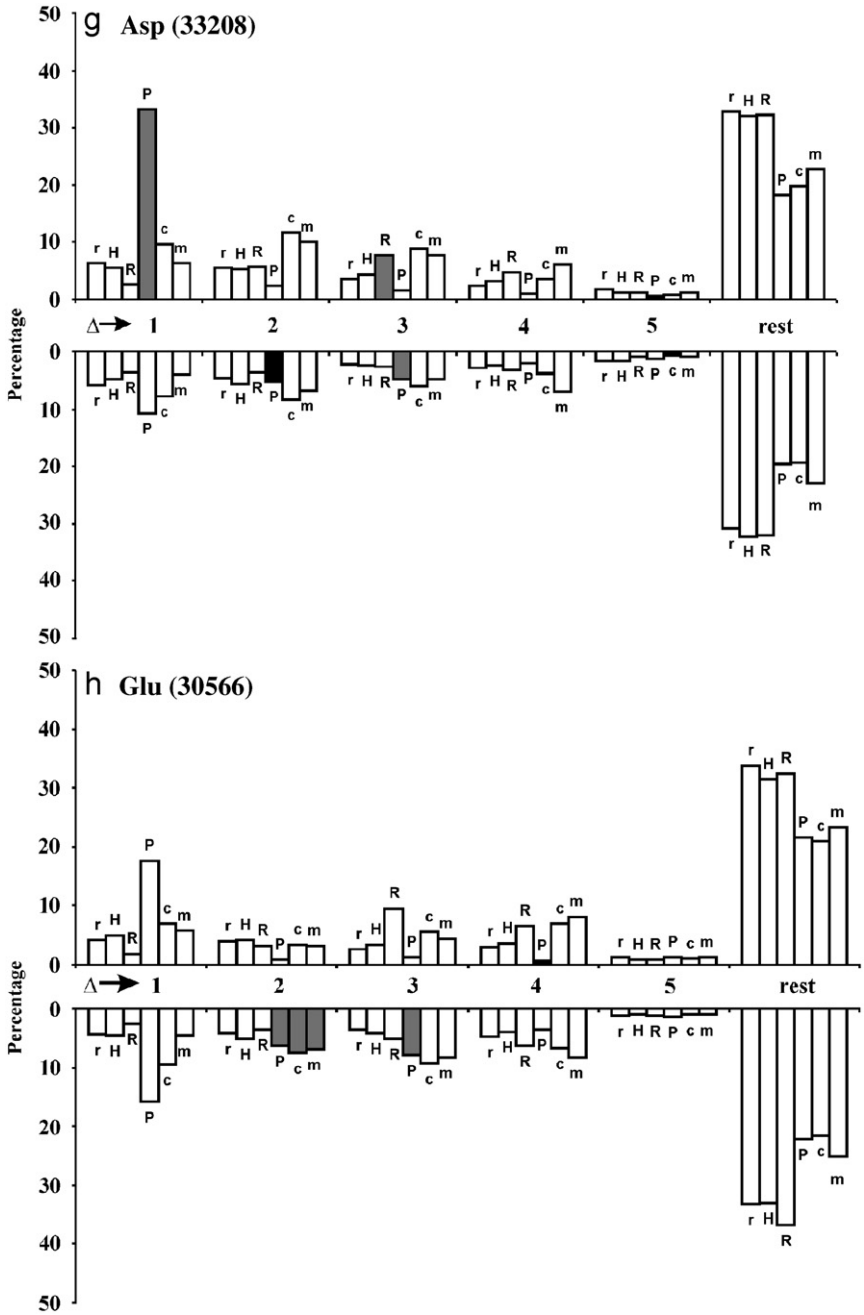

Fig. 18. (*Continued*)

(Δ = the number of the first residue in the peptide group—aromatic residue number) are observed in higher number compared to nonlocal interactions ($|\Delta| > 5$) (Tóth et al., 2004; Manikandan and Ramakumar, 2004). Thus for Phe–peptide, 25% and 33% of the interactions are with Δ values of -4 and -3, respectively, and the residues belong to the same helix. (The corresponding values for the other aromatic residues are, Tyr: 27%, 32%; Trp: 25%, 27% and His: 38% and 34%). When $\Delta = -4$, about 44% interactions involving Phe, Tyr and Trp have C–H···O interactions, and 26% with $\Delta = -3$. An example of Trp–peptide contact with C–H···O interaction is shown in (Fig. 22a). Compared to the other aromatic residues, His–peptide contacts with $\Delta = -4$ have a smaller percentage (29%) with C–H···O interactions; however, 14% are associated with

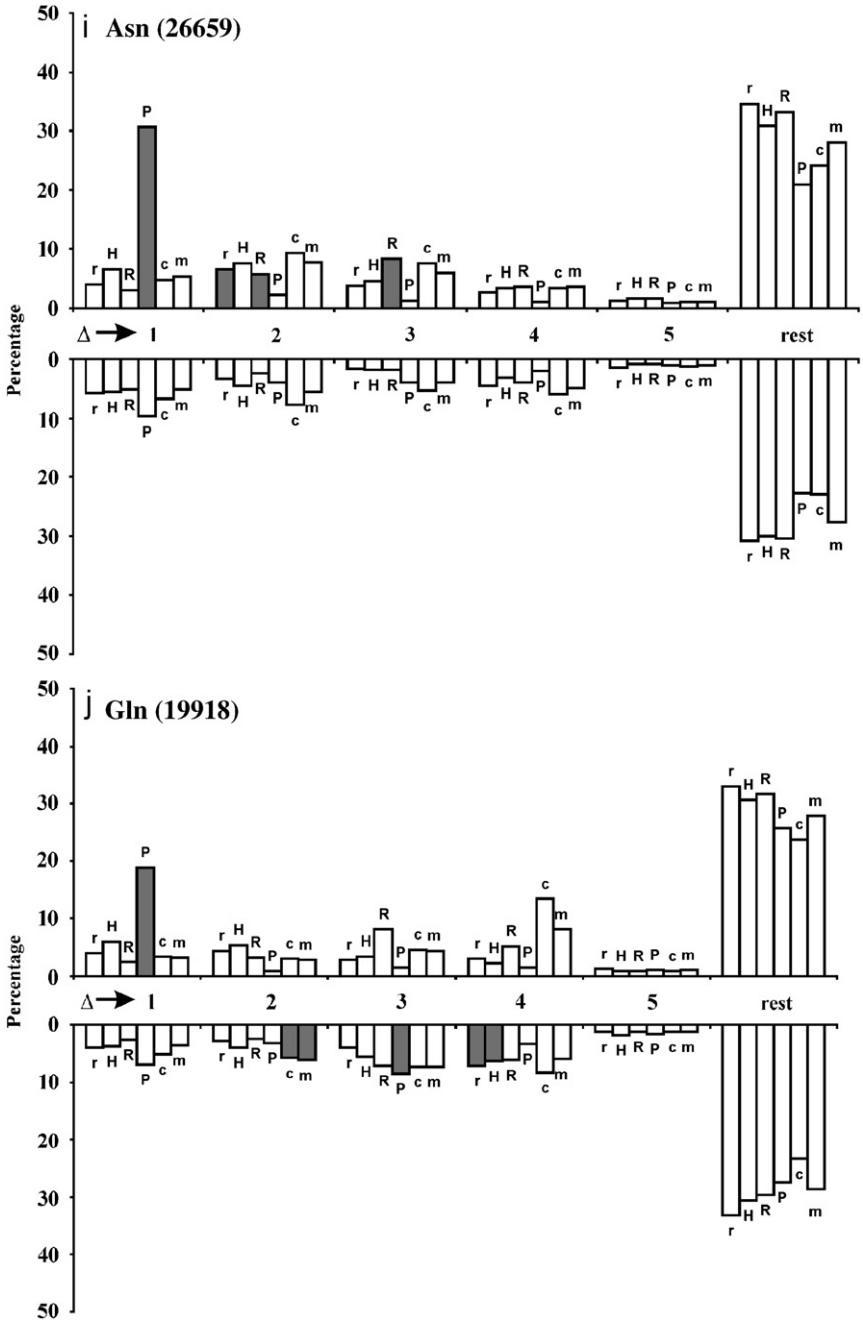

Fig. 18. (*Continued*)

hydrogen bonding (Fig. 22b). Very little interactions of these two types are observed when $\Delta = -3$. Additionally, Brandl et al. (2001) noted that 7.1% of all C^{α}–H$\cdots\pi$ interactions in proteins have a sequence difference (donor–accptor) of −3, most of which occur in helices or near their termini; in helical conformation, the hydrogen atoms on the C^{α} atoms point upward almost along the direction of the helix—side chains with an acceptor π-system following along the chain after three residues can therefore easily saturate the interaction potential of the C^{α}–H.

About 18% of all the aromatic–peptide interactions (individually, the percentage is the lowest, 11%, involving His) have the 3 involved residues located in β-sheets. With $\Delta = -2$, peptide and aromatic planes in

Table 2
Preferred secondary structure of interacting residues with sequence difference, Δ, −1 to 4

Interaction type	Pair[a]	Δ[b]	Percentage composition of the top ranking combination of secondary structures[c,d]
Aromatic–histidine	r–H	**4**	CC = 3; HH = 43 (24c)
Aromatic(r and H)–arginine	Ar–R	**−1**	CC = 2; HH = 46
	Ar–R	**4**	CC = 4; HH = 63
Aromatic–proline	r–P	1	CC = 28; TT = 15
	Y–P	1	EE = 12
	W–P	1	HH = 13
	H–P	**1**	CC = 16; CT = 26 (20i)
	H–P	−1	CC = 9; TT = 48 (20j)
Aromatic–acid	F/Y/H–c	−1	HH = 33
	F/W/H–c	1	HH = 58
	F/H–c	−2	EE = 27
	F/H–c	2	EE = 24
	F–c	4	HH = 42
Aromatic–amide	r–m	**−2**	CC = 8; EE = 32 (20h)
	r–m	4	CC = 2; HH = 58 (20a)
	W–m	1	HH = 60
	H–m	**−2**	CC = 5; EE = 28
	H–m	−1	HH = 34
	H–m	1	HH = 40
	H–m	**4**	HH = 49
Proline–arginine	P–R	**−1**	CC = 30; CT = 15
	P–R	**3**	CC = 6; HH = 40
Proline–proline	P–P[e]	1	CC = 43; TC = 22
Proline–acid	P–D	1	CC = 4; TT = 68
	P–E	1	CC = 2; HH = 60 (20g)
	P–D	**−1**	CC = 14; TC = 32; HC = 26
	P–E	−1	CC = 28; TC = 15
	P–c	**2**	CC = 5; CH = 44
	P–c	**3**	CC = 9; HH = 32; CH = 31
Proline–amide	P–N	1	CC = 7; TT = 67
	P–N	**−1**	CC = 15; TC = 26
	P–Q	**−1**	CC = 29; CT = 15
	P–m	**3**	CC = 9; HH = 36
Arginine–acid	R–c	**−3**	CC = 2; HH = 68 (20b)
	R–c	−4	HH = 69
	R–c	4	HH = 67
Arginine–amide	R–m	**−3**	CC = 3; HH = 63 (20c)
	R–m	**−2**	CC = 8; EE = 21
Acid–Acid	D–c	1	CC = 5; HH = 27
	D–D[e]	1	TT = 28
	D–c	2	CC = 4; CT = 32; CH = 31
	D–c	3	CC = 9; CH = 30
	D–E	3	HH = 28
	E–E[e]	1	CC = 4; HH = 62
	E–E[e]	3	CC = 9; HH = 58
Acid–amide	D–m	2	CC = 5; CT = 36
	D–m	3	CC = 9; CH = 25
	D–Q	3	HH = 31 (20d)

Table 2 (*continued*)

Interaction type	Pair[a]	Δ[b]	Percentage composition of the top ranking combination of secondary structures[c,d]
	E–N	−3	CC = 5; HC = 38
	E–Q	−3	CC = 10; HH = 60
	E–m	−4	CC = 6; HH = 65
	E–m	4	CC = 2; HH = 72 (20e)
Amide–amide	Q–m	4	CC = 3; HH = 75 (20f)
	Q–m	**−2**	CC = 8; TC = 18

[a]If all residues within a group show similar trend, then the group name (as given in the legend to Fig. 18) is mentioned. However, if a particular residue within the group shows any other feature this is also indicated.

[b]For a pair X–Y in the previous column, Δ = Y–X. The value is in bold if it corresponds to the shaded bar in Fig. 18.

[c]Considering the cases when at least one of the residues has a regular secondary structure. However, the value for the combination 'CC' is provided for comparison. The order of the secondary structure is the same as the order of residues given in column 2. The convention of secondary structure, defined by DSSP, is $3_{10}/\alpha$-helix = 'H', β-strand/B = 'E', hydrogen/non-hydrogen bonded turn = 'T' and non-regular structure = 'C'. Molecular diagrams of some of the entries are provided; for these the figure numbers are given in parentheses.

[d]The percentage is calculated with respect to the total number of each interacting pair given in column 2 corresponding to that Δ value. Only those interacting types are chosen for which the percentage of interaction at a given Δ value is $\geqslant 5$ in Fig. 18.

[e]When the pair involves residues of the same type, the sign of Δ has no significance. Both the residues were used as the central residue in turn, and consequently the height of the bars in the upper and lower panels (for example, Pro-Pro, with $\Delta = 1$) in Fig. 18 is the same.

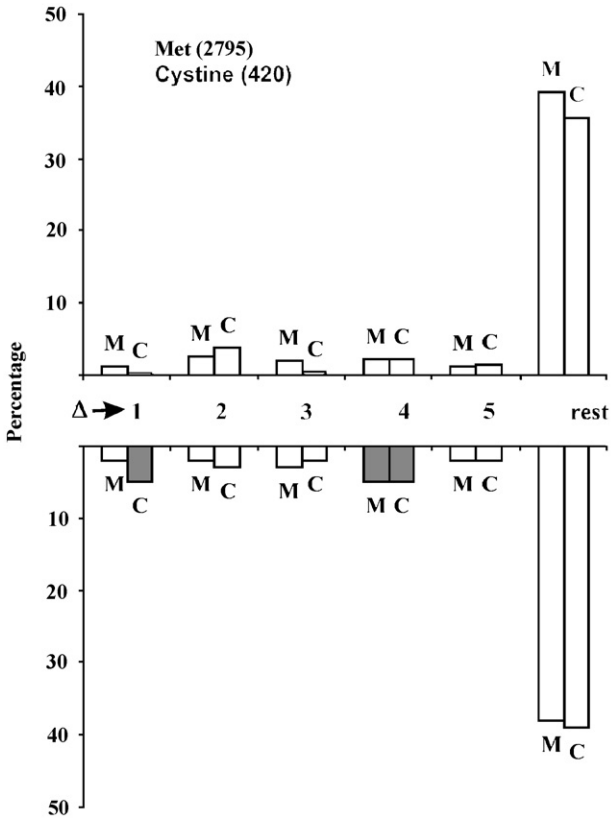

Fig. 19. Histogram of sequence differences, Δ (aromatic—Met/half cystine) for aromatic···S interactions. For the explanation of shading, see legend to Fig. 18.

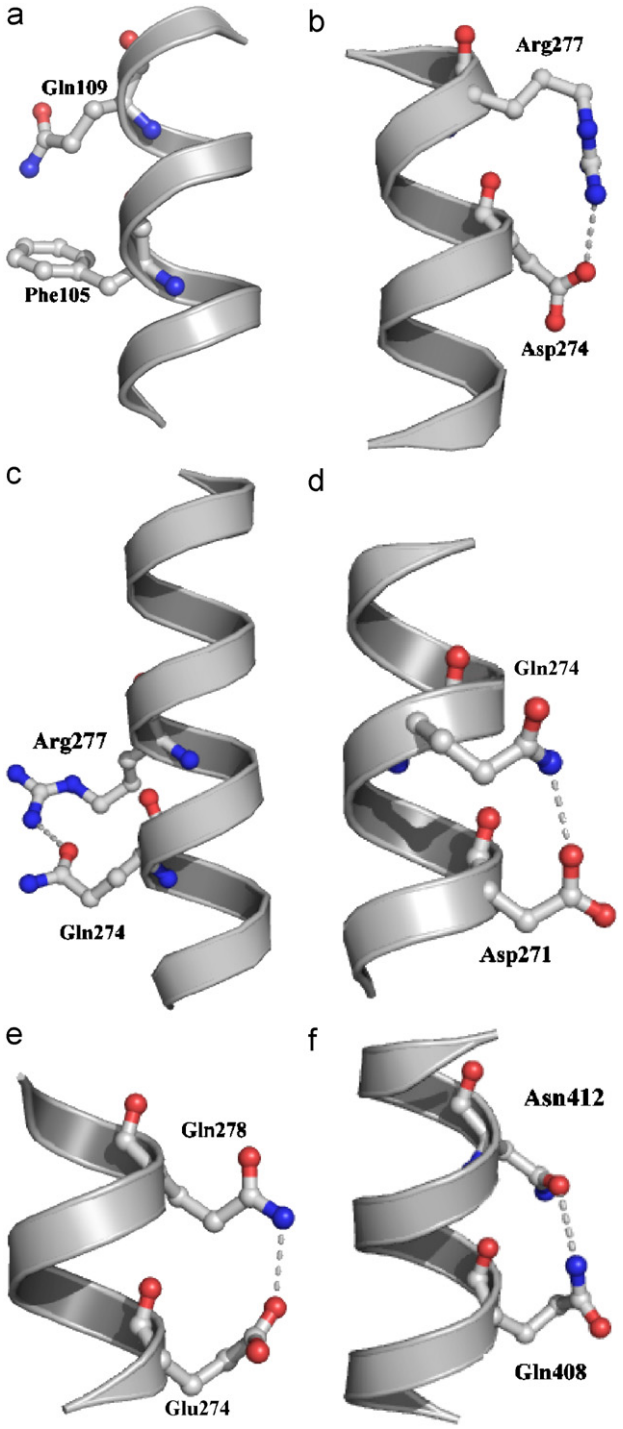

Fig. 20. Examples showing the correlation of the sequence difference in residue pairs with their location in secondary structural elements. Only the interacting side chains are shown. The residue pair, Δ, secondary structure (as given in Table 2), PDB code and (geometry, for the second residue relative to the first) are: (a) r-m 4 HH 1pa7 (*fe*); (b) R-c –3 HH 1a4y (*ef*); (c) R-m –3 HH 1bs0 (*ee*); (d) D-Q 3 HH 1gpu (*et*); (e) E-m 4 HH 1cxp (*oe*); (f) Q-m 4 HH 1a8i (*et*); (g) P-E 1 HH 1a81 (*ft*); (h) r-m –2 EE 1jnd (*oe*) (i) H-P-D 1 β-turn (type I) 1rq2 (P-H, *ot*; P-D, *ot*); (j) P-H 1 β-turn (type I) 1a8e (*ff*). The C–H$\cdots$O and C–H$\cdots\pi$ interactions are shown as black dotted lines and hydrogen bond is shown in gray dotted lines. In (i) the four positions of the β-turn are indicated.

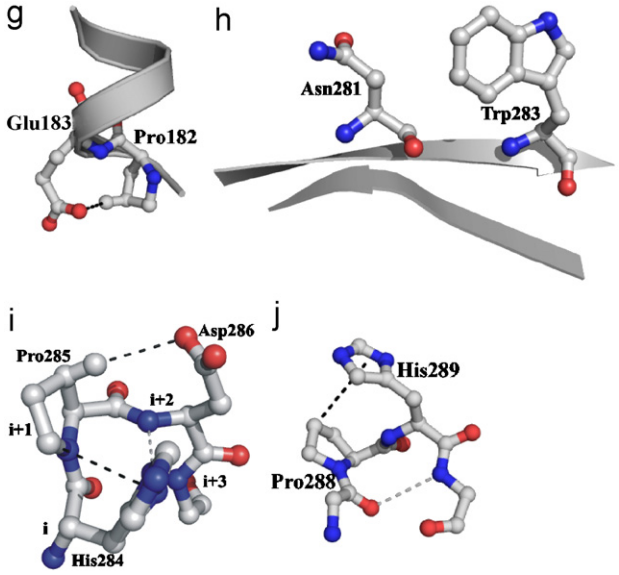

Fig. 20. (*Continued*)

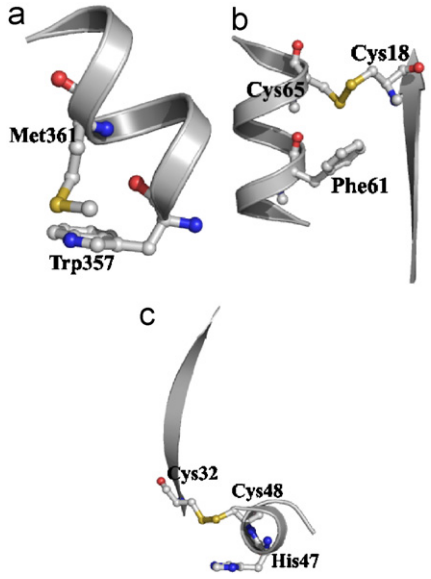

Fig. 21. Typical examples of aromatic···S interactions in α-helices, involving (a) Met in 1ogs (geometry, *et*); (b,c) half cystine in 1kp6 (*en*) and 1p3c (*op*). The relative positions are i, $i+4$ in (a) and (b), and i, $i+1$ in (c).

the same β-strand can interact (8% of the cases when the secondary structural designation of all the three residues is 'E', according to DSSP). When $\Delta = -2$, ~44% of contacts involving Phe, Tyr and Trp are through C–H···O interactions. As has been observed in helices, even in β-strands and with $\Delta = -2$, a lower percentage (21%) of His–peptide contacts have C–H···O interactions and 12% are hydrogen bonded. Two examples of Phe–peptide interactions involving C–H···π and C–H···O interactions in β-strands are shown in Fig. 22c and d. For a residue in β-sheet, the C^{α}–H group is oriented in-plane that makes it possible for it to be engaged in interstrand C–H···O interactions (Derewenda et al., 1995; Fabiola et al., 1997). However, when the residue is Gly, the second C^{α}–H bond is oriented out of the sheet plane and is free to form nonconventional hydrogen

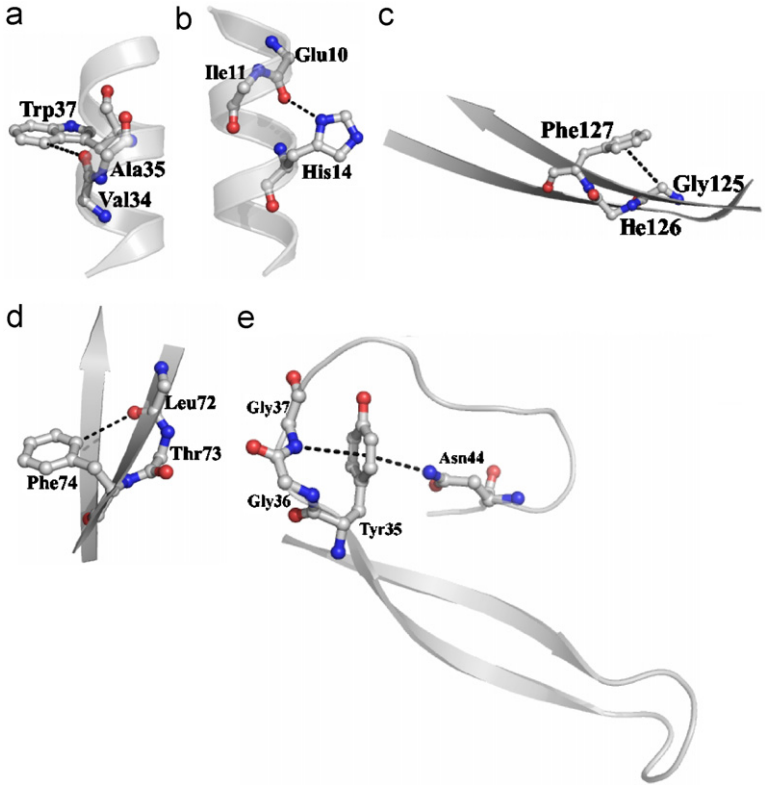

Fig. 22. Examples of aromatic–peptide interactions. (a) C–H···O interaction in 1ji7 (geometry: *ef*); (b) hydrogen bonding in 1b5e (*et*); (c) C–H···π interaction in 1m0w (*ff*); (d) C–H···O interaction in 1c5e (*ef*) and (e) N–H···π interaction in 5pti (*fe*).

bonds. With Gly there is also the possibility of both N–H and C^{α}–H groups forming a pair of π-hydrogen bonds with the large face of a single Trp (Steiner and Koellner, 2001).

The most frequent peptide N–H to π hydrogen bond has been reported to be the motif $i \rightarrow (i-2)$ (representing the two residue numbers) (Worth and Wade, 1995; Steiner and Koellner, 2001). Next to the highest number of occurrences in loop regions, the incorporation of the motif in secondary structure elements is observed at the carbonyl end of β-strands—$(i-2)$ being the last residue of the strand and the donor residue (strongly preferred to be Gly) being located on the loop beyond. An example from bovine pancreatic trypsin inhibitor (Wlodawer et al., 1984) is shown in Fig. 22e, in which the aromatic Tyr ring has a second N–H···π interaction involving a side-chain amide group. The involvement of a peptide N–H group means that such an interaction is not feasible in central parts of helices and sheets, and is usually operative for edge and terminus stabilization (Steiner and Koellner, 2001).

5. Stereospecific interactions and protein folding

Proteins in general are quite tolerant to mutations in their interiors provided the hydrophobic composition of the core is not altered (Lim and Sauer, 1989; Axe et al., 1996). X-ray crystallographic studies on phage T4 lysozyme showed that its fold could accommodate the mutation of seven residues to methionine, albeit with a reduction in the thermal stability and activity of the protein (Gassner et al., 1996). Other mutational studies have also indicated that the fold is generally robust, though there are structural adjustments of both the side- and main-chain atoms (Eriksson et al., 1992; Buckle et al., 1996). An analysis of distantly related globins has shown that changes in volumes of buried residues are accompanied by changes in the geometry of the helix packing, but multiple shifts are coupled so that the heme pocket remains unaltered and thus the functional activity is retained (Lesk and Chothia, 1980). However, when dealing with a particular structure, the

introduction of even small changes to a hydrophobic core typically diminishes the packing quality, the stability and often the activity of the protein (Kellis et al., 1988; Karpusas et al., 1989; Eriksson et al., 1992; Sandberg and Terwilliger, 1989; Shortle et al., 1990; Lim and Sauer, 1991). Though sequences that simply maintain hydrophobicity at core positions are able to adopt the overall fold, the specific stereoelectronic features of these residues and their packing interactions clearly act as important determinants in attaining a fully functional, thermally stable protein. Thus even the substitution of buried Cys residues, not part of any disulfide bridge, can have deleterious effect on stability (Hiraga and Yutani, 1996)—presumably the different types of interactions that a Cys sulfhydryl group can participate in (Pal and Chakrabarti, 1998) cannot be sustained in the mutants. A residue, such as Pro, which is notionally assumed to engage aromatic side chains through hydrophobic forces, can indeed form C–H$\cdots\pi$ interactions, and it is found that the relative orientations of the rings that favor these interactions outnumber those that cannot sustain these stereospecific interactions (Section 3; Bhattacharyya and Chakrabarti, 2003). These interactions may thus be important for the structure and binding, and the residues would be under evolutionary pressure to be conserved so that these are retained.

The hydrogen bonds, both strong and weak, are thought to play an important role in providing kinetic pathways during the folding process (Myers and Oas, 1999; Pal et al., 2003; Powers et al., 2005). An important issue concerns the kinetic accessibility of the native structures, and whether there are selection principles imposed by kinetics (Baker and Agard, 1994). Both local and nonlocal interactions can influence the thermodynamics and kinetics of protein folding (Govindarajan and Goldstein, 1995; Abkevich et al., 1995).

5.1. Stabilization of local structures and peptide design

Some examples of the interactions between planar groups with bearing on stability, local conformation and possibly folding kinetics are enumerated below. Others have been discussed in Section 4. These interactions are useful in the design of peptides, especially the ones with helical conformation (Chakrabartty and Baldwin, 1995; Serrano, 2000; Shi et al., 2002).

5.1.1. Phe–His at helix end

The interaction of Phe8 with His12 has been shown to account for the unexpected stability of the C-peptide helix (the N-terminal 13 residues of RNase A) (Dadlez et al., 1988; Shoemaker et al., 1990; Armstrong et al., 1993). Likewise, Trp and His at positions i, $i+4$, respectively, can stabilize the helix by 1 kcal/mol (and thereby increase the helical content of model peptides) when the histidine is protonated (Fernández-Recio et al., 1997). The results can be explained by the X–H$\cdots\pi$ interaction, involving the His X–H group and the aromatic π electrons, whose strength is increased as His is protonated (with the resultant increase in the partial positive charge on its protons); such a pair of residues usually occur at the C-terminal end of a helix (Fig. 24c) (Bhattacharyya et al., 2002). The stabilizing nature of the edge of His interacting with the face of an aromatic ring is also exemplified by a tertiary interaction between His18 and Trp94 in barnase, where the protonation of His increases the stability of the protein by 1 kcal/mol (Loewenthal et al., 1992).

5.1.2. Pro–aromatic and Gly–aromatic in β-sheet

Among the nonlocal interactions ($|\Delta|>5$), the one between Pro and an aromatic residue at the nonhydrogen-bonded site of antiparallal β-strands is easily identifiable (Fig. 23a) (Bhattacharyya and Chakrabarti, 2003). As Pro does not have a main-chain NH group, it cannot form hydrogen bond with a neighboring strand and is located at the edge strand with its CO group pointing outside. An aromatic residue placed at the non–hydrogen-bonded site can stack against the pyrrolidine ring. Tyr and Trp are particularly favored, and in a large number of cases the aromatic residue and Pro occur at the first and the last positions of their respective strands, respectively.

Aromatic–peptide interaction can be exhibited by the aromatic–X–Gly sequence, (i, $i+2$) motif, at the carbonyl end of β-strands, in which the N–H group of Gly points to the aromatic face (Fig. 22e) (Worth and Wade, 1995; Steiner and Koellner, 2001). These sequences form local structure early in the protein folding pathway and mutation studies indicate that they influence the rate of folding (Kemmink and Creighton, 1993). The tetrapeptide, YTGP, containing the motif, is structured in solution, retaining the N–H$\cdots\pi$ interaction and

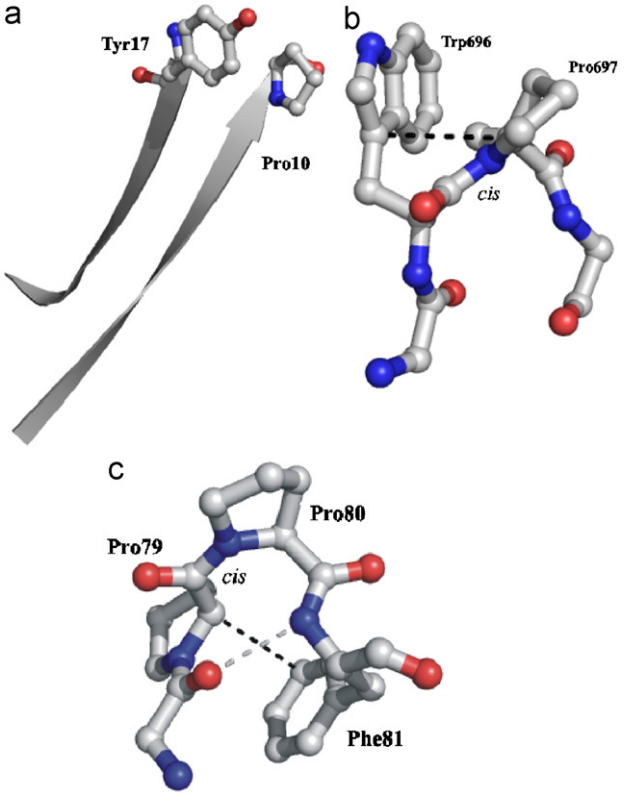

Fig. 23. Examples of Pro–aromatic interactions in (a) the non-H-bonded sites of antiparallel β-strands (in 1bxo, geometry *of*); (b) Pro–aromatic *cis* peptide bond (in 1oac, *et*); and (c) Pro–Pro–Phe moiety (in 1aoc, *of* for the Pro79–Phe81 pair).

this can also be reproduced by molecular dynamics simulations (van der Spoel et al., 1996). However, a peptide YTAP cannot make such an interaction because of steric hindrance of the Ala side chain and hence does not show a well-defined conformation. The presence of two C^{α}–H groups and the lack of any side chain make it rather facile for a Gly to interact with an aromatic residue.

5.1.3. *Pro– aromatic in cis peptide*

In protein structures the peptide bond connecting an aromatic residue with Pro occurs in the *cis* conformation 7–10% of the time (Stewart et al., 1990; Frömmel and Preissner, 1990; MacArthur and Thornton, 1991; Reimer et al., 1998; Pal and Chakrabarti, 1999). In the r-Pro cases with $\Delta = 1$ in Table 2, 10% have *cis* peptides linking them; however, when His is the aromatic residue the percentage is lower (2%), suggesting that the direct contact across the *cis* peptide bond may be lower when His is involved. In the majority of these cases, the CA atom of Pro has a close contact of $3.6(\pm 1)$ Å with the CG atom of the aromatic residue and interacting with its face (Fig. 23b) (Pal and Chakrabarti, 1999). It should be noted that the C–H group of CA (and CD) atom in Pro is made more acidic by an adjacent electron-withdrawing N atom (Pedireddi and Desiraju, 1992; Bhattacharyya and Chakrabarti, 2003), which would make the C–H$\cdots\pi$ interaction confer stability to the *cis* peptide unit. The occurrence of a *cis* peptide bond introduces a type VI turn and isolated peptides containing aromatic–Pro sequence show high population of the *cis* isomeric form in solution (Yao et al., 1994; Wu and Raleigh, 1998).

About 4.1% of Pro-r and 6.3% of Pro-His interactions ($\Delta = 1$) occur after a prolyl *cis* peptide bond. It is not just the aromatic–Pro in *cis* peptide bond that can have specific interactions; with an aromatic residue following a prolyl *cis* peptide, especially in Pro–Pro, there is an interaction between the CA of the first Pro and the π face of the aromatic residue (Fig. 23c) (Pal and Chakrabarti, 1999). Thus a peptide Phe–Pro–Pro–Phe has the Pro–Pro bond in the *cis* form ~50% of the time in solution, and when the Pro$\cdots$Phe interaction is

abolished in Ala–Pro–Pro–Ala peptide, the *cis* content is reduced to ~10% (Dasgupta et al., 2007). Thus one can take recourse to the aromatic–Pro interaction in stabilizing the *cis* form and thereby inducing a turn conformation in the design of compact peptide structures.

5.1.4. Pro–···aromatic in β-turn

In addition to the *cis*-peptide induced turn conformation (Section 5.1.3), the aromatic–Pro interaction is also an important component in other types of β-turn, notably type I (Hutchinson and Thornton, 1994). Among the four positions, i to $i+3$ defining a β-turn, Pro has a high propensity to occur at $i+1$. When an aromatic residue occupies an adjacent position, there could be stacking interactions of the type C–H···π (Bhattacharyya and Chakrabarti, 2003). Thus His is particularly favored at i, in which position it can additionally form a hydrogen bond to the backbone NH group at position $i+2$, thereby enhancing the stability of the local structure (Fig. 20i).

5.1.5. S···aromatic in α-helix

Like the Phe/Tyr/Trp–His pair found at i, $i+4$ positions in helices, aromatic···S interaction is also found at the same spacing (Section 4, Figs. 19 and 21). Circular dichroism studies on polyalanine-based peptides indicated that the Phe–Cys pair could contribute up to 2 kcal/mol to the α-helix stability (Viguera and Serrano, 1995). The value obtained for the Phe–Met pair was 0.65 kcal/mol, while another study reported 0.75 kcal/mol (Stapley et al., 1995).

5.1.6. Directionality in the sequence in the context of a secondary structure

Some of the sequence information may get converted in a sequential way into the three-dimensional structure during folding. For example, the comparison of sequences of the shortest possible helices with the longer ones has led to the suggestion that the α-helix gets nucleated near its N-terminal end and then it propagates towards the C-terminal end (Pal et al., 2003; Dasgupta et al., 2004), which has been validated by molecular dynamics simulations (Monticelli, et al., 2005). Likewise, when two residues are interacting, the order in which the residues occur in the sequence may be important, and usually one is preferred over the other.

Figs. 14 and 15 provide the general distribution of relative orientations without any considerations on the effect of sequence difference on the observed geometry. However, smaller sequence differences do impose some restrictions (Bhattacharyya and Chakrabarti, 2003), especially when they occur in a secondary structure, such as helix (Bhattacharyya et al., 2002). Thus in an α-helix when $\Delta = 1$, the predominant orientation of the second residue relative to the first (in the sequence) is *ef*, whereas with $\Delta = 4$, *ot* is a preferred configuration (Fig. 24a and b). Thus if a pair is such that there is an extra stability resulting from the edge of the one at $i+4$ position pointed in a tilted fashion towards the face of the residue at position i, then this pair would be favored compared to other combinations of residues. This is what happens with nonhistidine aromatic residue and His pair at i, $i+4$ positions in an α-helix (Fig. 24c); the X–H protons of a protonated His ring would carry more partial positive charge, resulting in stronger X–H···π interactions with the π electrons of the aromatic ring in the previous turn of the helix. If the order of the residues is reversed, the X–H group, now located on the aromatic ring (with lesser partial positive charge on the proton) would interact with the π electron cloud of a His ring carrying a formal positive charge, resulting in a relatively weaker interaction. Thus the Phe–His pair, rather than His–Phe, is preferred in the C-terminal end of helix.

There is also directionality in the preference for the aromatic–Met/Cys pair at i, $i+4$ positions in helices (Figs. 19 and 21a). It has been suggested (Viguera and Serrano, 1995) that in the Phe–Cys pair, the rotamers of the Phe and Cys side chains are very favorable in α-helices, and consequently the energy cost of making the interactions is low. When it is the opposite order, the Phe residue needs to adopt a high-energy rotamer to contact the Cys or Met residue and the helix needs to be slightly distorted at the Phe position, with the corresponding energy cost. The difference in the relative orientations of the sulfur plane relative to the aromatic ring in the two categories may also provide an explanation. There are 127 cases of aromatic–Met (i, $i+4$) pairs in helices. Of these 65% have geometry *op*, which is the favored pattern for aromatic–Met/half cystine interactions (Fig. 15). However, for the 28 cases with the helix having the residues in the reverse order, only three are found with this geometry.

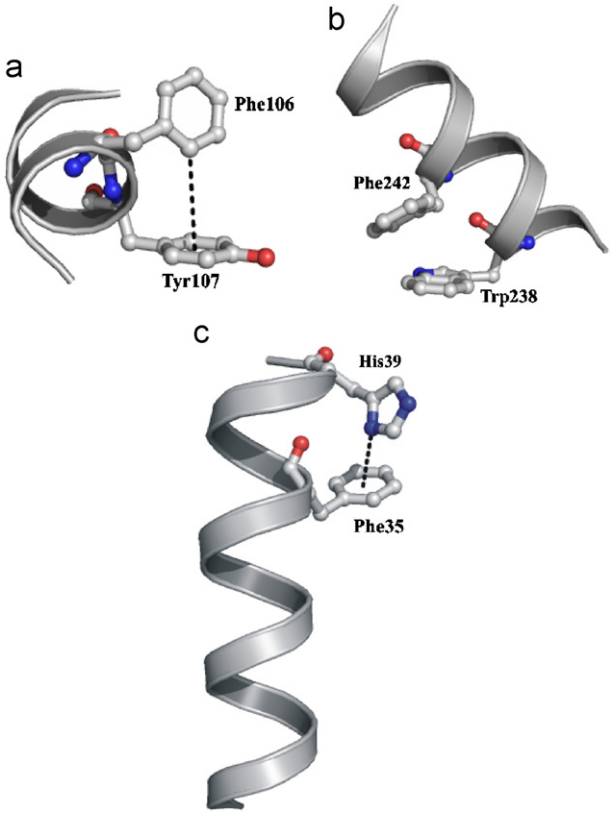

Fig. 24. (a, b) Examples of interacting aromatic pairs, with sequence difference, $\Delta = 1$ (PDB code: 1jdw, geometry of the residue with higher sequence number relative to the lower, *ef*) and 4 (1cxl, *ot*), in α-helices. (c) Aromatic–His interaction in α-helix with $\Delta = 4$ in 1mty, geometry *ft*.

Acid–Arg pair with $\Delta = 3$ (Table 2) provides yet another example of preferential order in the sequence. Though not commented upon, the data provided by Sundaralingam et al. (1987) show that acid–Arg (i, $i+3$ in α-helices) ion pairs are more in number when the acidic group is N-terminal to Arg, which is not seen when the spacing is i, $i+4$. Klingler and Brutlag (1994) found significant (i, $i+3$) sequence correlation for the pair Asp–Arg. In helices only 17% of the i, $i+3/4$ ion pairs are involved in direct hydrogen bonds, with further 43% bridged through water molecules, and the remaining 40% are separated by more than 7 Å. It is likely that the i, $i+3$ spacing for the acid–Arg pair offers the right sequential and spatial separation for forming hydrogen bond (Fig. 20b).

The above observations taken together with the results in Table 2 can be generalized as follows. If there is a recurrent pattern (i, $i+3/4$) of side chain interactions in a helix usually there exists conventional or nonconventional hydrogen bond interactions, such that the proton acceptor (an oxygen-containing group or an aromatic face) is the first in the sequence, followed by the proton donor (C/N–H that can also belong to His). The position of the proton donor is taken up by sulfur in S⋯aromatic interaction.

6. Changes in the geometrical preferences in the metal-binding site

Several structural motifs, involving two or more His residues, binding the same metal ion have been identified (Bhattacharyya et al., 2003). Such His residues are usually in contact (based on the threshold distance that has been used), but their geometry is likely to be modulated by the requirement of the metal site. For such His–His pairs, the geometry is found mostly in the three grid elements at the top-right-hand corner (Fig. 14f) (Bhattacharyya et al., 2003). Comparison of the metal-bound form with the apo structure indicates that the metal-binding site is essentially preformed and can accommodate the cation with only minor

structural adjustments. Most metal-binding sites with a pair of His residues have a relative orientation that generates C–H$\cdots\pi$ interactions, which may provide extra stability and affect the redox properties of the metal centers. In an analogous situation, by virtue of their common role as metal ligands two carboxylate groups (Flocco and Mowbray, 1995), or an acidic side chain and His can also come within a distance of 3.0 Å and may have restricted geometries (Fig. 25m).

7. Interactions in the active site

Planar residues, such as His, Tyr, Asp, Arg, Pro, etc. are important at the catalytic and binding sites (Meurisse et al., 2004). The PROSITE database enumerates the functional motifs in proteins (Hulo et al., 2006). The key residues involved in functional sites as well as in protein–protein interaction have been identified using an automated graph theoretic approach for the detection of recurring structural patterns of 3–6 amino acids (Wangikar et al., 2003; Tendulkar et al., 2003). We observe that the planar residues in the active sites interact among themselves with geometries that are similar to those observed in higher numbers in tertiary structures. Some examples of interactions at the active sites are discussed (Table 3).

A battery of biochemical and biophysical studies on the high-affinity maltose transport system in *E. coli* indicated that the maltodextrin-binding protein is extremely resistant to heat and chemical denaturation; it has been proposed that the abundance of salt bridges, hydrophobic interactions and proline residues make it

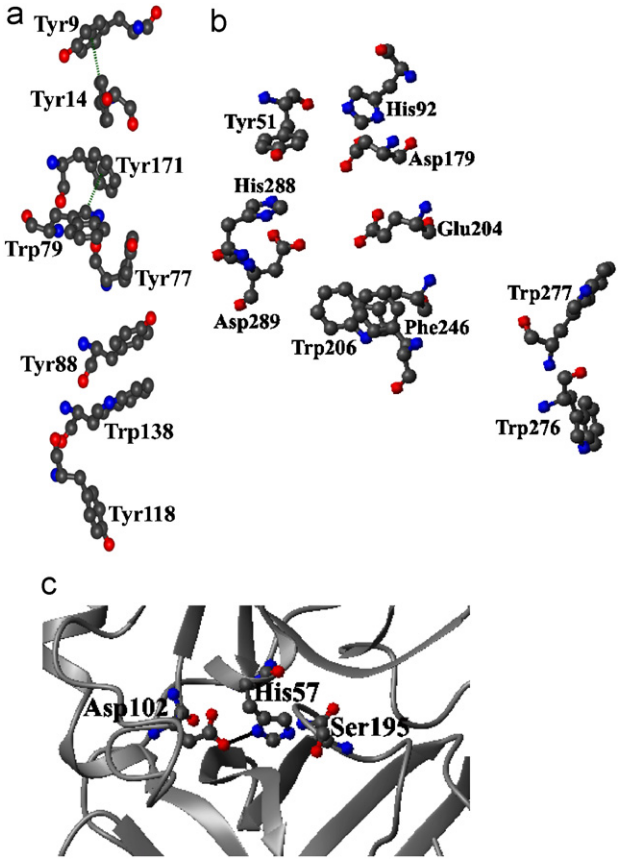

Fig. 25. Some typical examples of planar residues interacting at the active site. The C–H$\cdots\pi$, cation$\cdots\pi$ and C–H$\cdots$O interactions are indicated as dotted lines, and hydrogen bonds as solid lines. Protein names, PDB codes, residues involved and relative orientations are provided in Table 3. The PDB codes with references are (a) 1xnd (Campbell et al., 1993); (b) 1amy (Kadziola et al., 1994); (c) 1mctA (Huang et al., 1993); (d) 1ppn (Pickersgill et al., 1992); (e) 1glqA (García- Sáez et al., 1994); (f) 1vip (Carredano et al., 1998); (g) 1bbzA (Pisabarro et al., 1998); (h) 1ir3A (Hubbard, 1997); (i) 1amuA (Conti et al., 1997); (j) 1pg2A (Crepin et al., 2003); (k) 1i6mA (Doublié et al., 1995); (l) 1cxvA (Botos et al., 1999) and (m) 1cdmA (Meador et al., 1993).

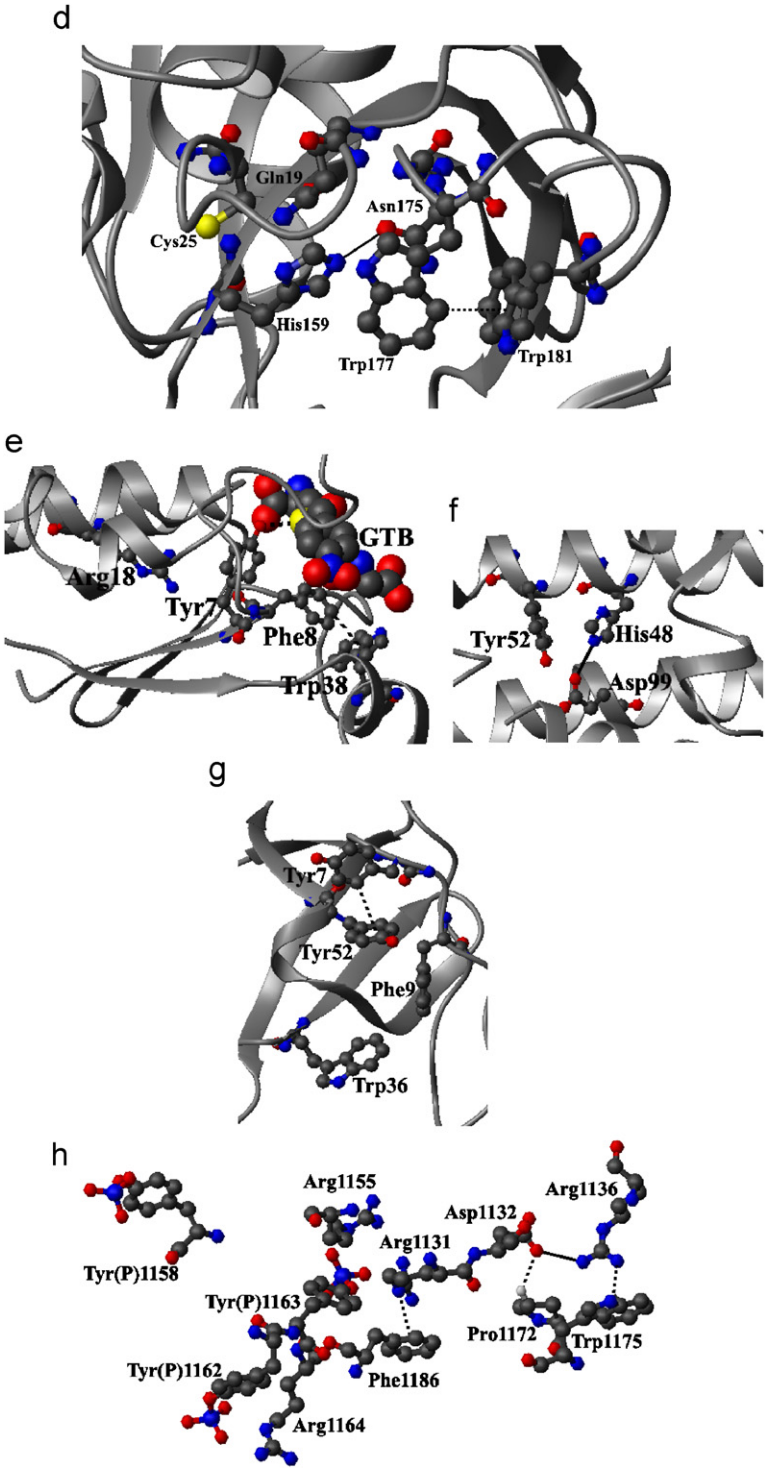

Fig. 25. (*Continued*)

highly thermostable (Evdokimov et al., 2001). At the maltotriose-binding site (Fig. 3), aromatic residues and a proline form a cluster. The carbohydrate ligand-binding residue, Trp345 is involved in a C–H$\cdots\pi$ interaction with Tyr262 in a perpendicular fashion. Additionally, Pro229 interacts in a tilted orientation with Tyr303 through a C–H$\cdots\pi$ interaction.

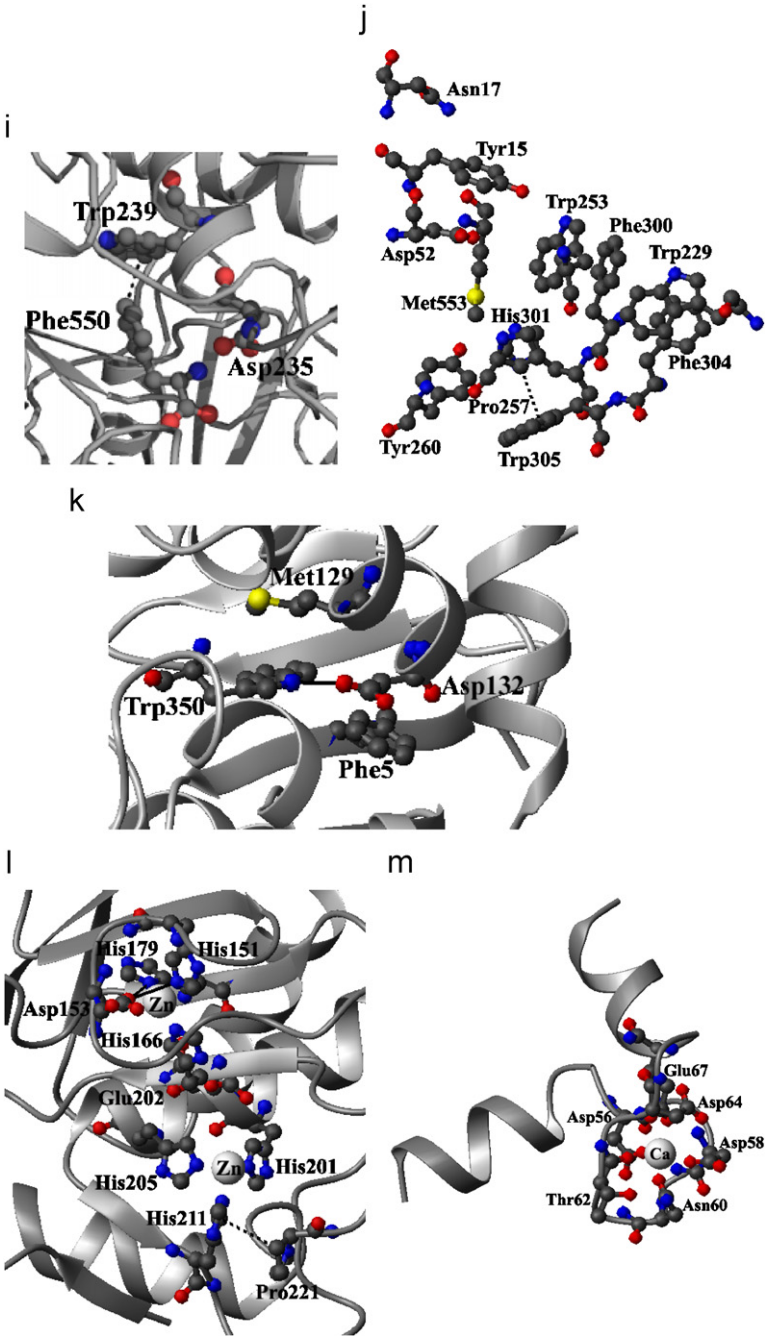

Fig. 25. (*Continued*)

Endo-β-1,4-xylanases catalyze the depolymerization of xylans and these can be categorized into two families (Henrissat and Bairoch, 1993). The overall three-dimensional structures of the xylanases of family 11 are similar and the xylan-binding site is made of aromatic residues. The active site of a xylanase of family 11 from *Trichoderma harzianum* is shown in Fig. 25a. Among the aromatic residues, Tyr9 and Tyr14 interact through a C–H$\cdots\pi$ interaction. A similar interaction also connects Tyr171 with Trp79, though the latter is not a primary component of the active site.

Table 3
Geometrical orientation of the interacting residue pairs at the active sites

Protein	PDB code[a]	Interacting residue pair[b]	Geometry[c]	HB/C–H$\cdots\pi$[d]
Maltodextrin-binding protein	1eljA (3)	W65–W69	*of of*	
		W69–F73	*of of*	
		Y174–W345	*ef ef*	
		P229–W230	*of of*	
		P229–Y303	*ot ft*	P229–Y303
		P229–F325	*ff ff*	
		Y262–W345	*ef fe*	W345–Y262
Xylanase	1xnd (25a)	Y9–Y14	*ef fe*	Y14–Y9
		Y14–Y171	*ot et*	
		Y77–Y88	*et ft*	
		Y77–Y171	*et ot*	
		W79–Y171	*ef oe*	W79–Y171
		Y88–W138	*ff of*	
Amylase	1amy (25b)	Y51–D179	*et ft*	
		Y51–H288	*of of*	
		H92–D179	*oe ef*	
		D179–E204	*of ee*	
		E204–W206	*ot et*	
		E204–F246	*et ot*	
		E204–D289	*ee ee*	
		W206–F246	*of of*	
		W206–D289	*ee of*	
		H288–D289	*ft ft*	
Trypsin	1mctA (25c)	H57–D102	*et et*	HB
Cysteine proteinase	1ppn (25d)	H159–Q19	*et et*	
		H159–N175	*ef ef*	HB
		H159–W177	*ot ot*	
		N175–W177	*ot ft*	N–H$\cdots\pi$
		N175–W181	*ef oe*	N–H$\cdots\pi$
		W177–W181	*ef fe*	W177–W181
Glutathione S transferase	1glqA (25e)	Y7–R18	*et ot*	
		Y7–F8	*ot et*	
		F8–W38	*ef fe*	F8–W38
Phospholipase A2	1vip (25f)	H48–Y52	*ot ft*	
		H48–D99	*ee ee*	HB
Abl tyrosine kinase (SH3 domain)	1bbzA (25g)	Y7–Y52	*ef fe*	Y7– Y52
		F9–W36	*oe ef*	
Tyrosine kinase	1ir3A (25h)	R1131–R1155	*et ft*	
		R1131–Y1163	*ft ot*	
		R1131–F1186	*ef fe*	Cation$\cdots\pi$
		D1132– R1136	*et et*	HB
		R1136–P1172	*ef oe*	
		R1136–W1175	*ef fe*	Cation$\cdots\pi$
		R1155–Y1163	*ef oe*	HB with pTyr[e]
		Y1162– R1164	*oe oe*	HB with pTyr[e]
		Y1163–F1186	*et ot*	
		P1172–W1175	*et et*	
		P1172–D1132	*oe oe*	C–H$\cdots$O
Gramicidine synthetase 1	1amuA (25i)	F550–W239	*ef fe*	F550–W239
		F550–D235	*oe oe*	

Table 3 (*continued*)

Protein	PDB code[a]	Interacting residue pair[b]	Geometry[c]	HB/C–H···π[d]
Methionyl-tRNA synthetase	1pg2A (25j)	Y15–N17	*ot et*	
		Y15–D52	*of ff*	
		W229–F300	*oe ef*	
		W229–F304	*ff ff*	
		W253–F300	*ff ff*	
		W253–H301	*ef oe*	
		W253–P257	*ef oe*	
		P257–W305	*ef fe*	P257–W305
		P257–Y260	*ef oe*	
		Y260–H301	*et et*	
		Y260–W305	*et ot*	
		F300–F304	*ot et*	
		H301–W305	*oe oe*	
		W253–M553	*ep*	
		Y260– M553	*on*	
		H301– M553	*en*	
Tryptophanyl– tRNA synthetase	1i6mA (25k)	F5–D132	*of of*	
		W350–F5	*ot et*	
		W350–D132	*ee ee*	HB
		W350–M129	*fp*	
Matrix metalloproteinase	1cxvA (25l)	H151–H166	*ee ee*	
		H151–H179	*ef oe*	
		H151–D153	*oe oe*	HB
		D153–H166	*oe ef*	
		D153–H179	*ee ee*	HB
		H166–H179	*ef ef*	
		H201–H205	*oe oe*	
		H201–H211	*ee ee*	
		H201–E202	*et ot*	
		H205–H211	*ef oe*	
		H205–E202	*et ot*	
		H211–P221	*ff ff*	P221–H211
Calmodulin	1cdmA (25m)	D56–D58	*ef ef*	
		D56–N60	*et et*	
		D56–E67	*ef oe*	
		D58–N60	*ef oe*	
		N60–E67	*et et*	
		D64–E67	*ot ot*	

[a]Molecular diagrams of some of the entries are provided; for these the figure numbers are given in parentheses.
[b]One letter amino acid code followed by the residue number is given.
[c]The geometrical orientations from both the perspectives (the second residue relative to the first, and vice versa) are given.
[d]Hydrogen bonding and C–H···π interaction (the first residue is the proton donor and the second residue is the π acceptor), if present, are indicated. One C–H···O interaction is marked. N–H···π and cation···π interactions, identified in the original publications, are shown.
[e]The hydrogen bond is with the phosphate group of a phosphorylated Tyr.

All the α-amylases have a considerable 3D structural similarity though they have differences in the amino acid sequence (Kadziola et al., 1994). Three acidic groups (two Asp and one Glu) act as catalytic residues, which in barley malt α-amylase (isoform AMY-2) are Asp179, Asp289 and Glu204. However, a few aromatic residues are in close contact and thereby providing the right scaffold to the active site (Fig. 25b). Among these, Tyr51 and His92 interact with Asp179, Trp206 and His288 with Asp289, while Trp206 and Phe246 are in contact with Glu204. Also, there are stacking interactions, Tyr51 with His288, and Phe246 with Trp206. Two residues, Trp276 and Trp277, are present at the starch granule-binding site.

The catalytic triad of serine proteases consists of Ser–His–Asp residues. In trypsin the planar carboxylate moiety of Asp is in close contact with His in a tilted geometry (Fig. 25c). As mentioned in Section 3, when

there is a hydrogen bond between His and an acidic side chain the in-plane *ee* geometry is not the most common. The same feature is retained in the active site, where the geometry is *et*.

In the papain family of cysteine proteinase, the catalytic triad is constituted by Cys, His and an Asn residue. Apart from the catalytic residues, two nearby aromatic residues (Trp177 and Trp181) are conserved in the papain family (Brömme et al., 1996). The interactions among the planar groups at the active site are shown in Fig. 25d. Trp177 interacts with catalytic His159 in a tilted geometry. All known structures of this class of enzymes contain a conserved C–H$\cdots\pi$ interaction between Trp177 and Trp181. At the same time, both the indole rings of Trp177 and Trp181 act as hydrogen bond acceptors from the side chain amide group of catalytic Asn175 (not shown in the figure). Thus the network of interactions suggests an extended catalytic site of cysteine proteinases.

Glutathione-*S*-transferase (GST) helps to metabolize a variety of compounds in glutathione (GSH) conjugation pathway (Allocati et al., 2000) and can be classified into several groups. An example of a GST from π-class with its active site is shown in Fig. 25e. The active site is composed of two subsites. In the hydrophobic-binding site, the nitrobenzyl group of the inhibitor molecule is sandwiched between two aromatic residues Phe8 and Tyr108, and Phe8 is involved in C–H$\cdots\pi$ interaction with Trp38. At the GSH-binding site, Tyr7, which is the active site residue, forms hydrogen bond with the sulfur atom of GSH.

Phospholipase A2 hydrolyzes phospholipids into fatty acids and lysophospholipids. These are abundant in snake venom. The catalytic triad of phospholipases is made up of three planar residues—His, Tyr and Asp. His–Tyr pair has a tilted orientation, while an edge-to-edge geometry is observed for His–Asp (Fig. 25f). Interestingly, acutohaemolysin is another phospholipase A2 lacking catalytic and hemolytic activity. It has a Phe residue (Phe1102) that stacks against the catalytic His1048 blocking the catalytic triad (Liu et al., 2003). Thus aromatic–aromatic interaction can be important both for the function, as well as malfunction of a protein.

SH3 domains containing about 60 residues are found in proteins that are needed to assemble signaling complexes (Pawson, 1994). These domains recognize target ligands with Pro-rich motifs. The peptide-binding site of SH3 domain is composed of aromatic residues. In abl tyrosine kinase, these aromatic residues interact among themselves (Fig. 25g), giving rise to one C–H$\cdots\pi$ interaction. Specific interactions are also involved in the binding of peptides containing the PxxP (x being any amino acid) motif with the SH3 domain and this has been discussed in Bhattacharyya and Chakrabarti (2003).

Phosphorylation of cellular proteins by protein kinases provides control mechanisms for various signaling processes (Schenk and Snaar-Jagalska, 1999; Cohen, 2000; Hunter, 2000; Huse and Kuriyan, 2002). The catalytic site of the kinase consists of a number of functional elements, one of which is the activation segment containing a conserved Ser/Thr/Tyr, the phosphorylation of which leads to the formation of a network of residue contacts at the active site that are well conserved (Krupa et al., 2004). Ser/Thr and Tyr kinase families can be distinguished on the basis of the variations in the residue contacts and the consequent fall-out on residue conservation. Many of these protein kinases are also termed as RD kinases as they contain an Arg preceding the catalytic aspartate in the catalytic loop. As an example, insulin receptor tyrosine kinase can be considered, in which Tyr1158, Tyr1162 and Tyr1163 are the three phosphorylation sites (indicated by the letter 'P' against the residue names in Fig. 25h). In the inactive form, Tyr1162 is hydrogen bonded to the catalytic residue, Asp1132. In the phosphorylated state the conformation of the activation loop is changed and Tyr1162 is hydrogen bonded to Arg1164 (Table 3). The most striking feature of the active site is the presence of two cation–π interactions, the first involving Arg1131 in the RD motif and the conserved Phe1186—the interaction purported to channel conformational changes occurring at the activation segment to the catalytic loop—and the other involving Arg1136 in the catalytic loop and Trp1175 in the activation segment. Among other interactions, Arg1155 is in an edge-to-face geometry with Tyr1163 facilitating the hydrogen bond between the guanidinium and the phosphate groups. Phosphorylated Tyr1163 is in an offset-tilted orientation relative to the conserved Phe1186. Pro1172, which is also conserved along the kinase family, is in close contact with Trp1175 and is involved in C–H$\cdots$O interaction with catalytic Asp1132.

The phenylalanine substrate (Phe550) binds at the hydrophobic pocket of phenylalanine-activating subunit of gramicidine synthetase 1 in a ternary complex with AMP (Conti et al., 1997) (Fig. 25i). In this hydrophobic region, Phe550 is involved in a C–H$\cdots\pi$ interaction with the pocket lining aromatic residue Trp239. The conserved Asp235, in an almost perpendicular orientation with Phe550, has a hydrogen bond with the α-amino group of the latter as the primary interaction.

Aminoacyl-tRNA synthetases can be divided into two classes (Eriani et al., 1990; Cusack et al., 1991). In class-1 aminoacyl tRNA synthetase, the catalytic domain contains a Rossmann fold with the signature sequence "HIGH" and "KMSKS" (Webster et al., 1984; Hountondji et al., 1986), while class-2 contains antiparallel β-sheet with three conserved sequence motifs (Cusack et al., 1990; Ruff et al., 1991). Both methionyl and tryptophanyl tRNA synthetases belong to class 1. In methionyl tRNA synthetase (Fig. 25j), the Met-binding pocket is constituted by the amino acids—Ala12, Leu13, Tyr15, Trp253, Ala256, Pro257, Tyr260, Ile297, His301 and Trp305—located in two helices and two loops. Tyr15 and Trp253 play key roles in the strength of the binding of Met and its analogues (Crepin et al., 2003). Tyr15 also controls the size of the hydrophobic pocket. Met binding induces conformational changes of the aromatic residues in active site and stacking is observed in Phe300–Trp253 and Phe304–Trp229 pairs. His301, Trp253 and Tyr260 participate in the recognition of the sulfur atom of the incoming Met (Met553). Pro257, located at the border of the cavity is involved in C–H$\cdots\pi$ interaction with Trp305.

At the binding pocket of tryptophanyl tRNA synthetase, the incoming Trp (Trp350 in Fig. 25k) has Met129 on its face in one side, and has Phe5 in an offset-tilted orientation on the other. Asp132 located at the edge of Trp, makes a hydrogen bond with the nitrogen atom of the indole ring. It is quite interesting that in the active sites of both methionyl and tryptophanyl tRNA synthetases, Met–Trp interaction plays an important role. Further, Doublié et al. (1995) have pointed out that the interactions of substrate Trp with Met129 and Asp132 bear a striking resemblance to functionally significant clusters of the same three amino acids in another enzyme, cytochrome *c* peroxidase, for which mutagenesis has shown that Met230 and Asp235 are essential for Trp191 to function as a free-radical intermediate in the electron transfer process (Fishel et al., 1991). Thus the Met–Trp–Asp triad may be an important motif in the enzyme active sites.

Matrix metalloproteinases (MMPs), also named matrixins, cleave proteins of the extracellular matrix. Collagenase-3 (MMP-13) is a member of this family, which cleaves type II collagen, cartilage, fibronectin and aggrecan (Botos et al., 1999). The structures of MMPs are characterized by two tetrahedrally coordinated zinc ions, one catalytic and the other structural. The catalytic zinc ion is coordinated by the imidazole nitrogen atoms of His201, His205, His211 and a carboxyl oxygen of Glu202, while the structural zinc ion is coordinated by His151, His166, His179 and Asp153 (Fig. 25l). Pro221 is involved in C–H$\cdots\pi$ interaction with His211 in a face-to-face orientation and Asp153 makes a hydrogen bond with His179.

The close contact between residues mediated by metal ion can be illustrated using the structure of calmodulin, a calcium sensor having two domains, which binds and regulates the activity of many protein kinases along with other proteins in a calcium dependent manner (Crivici and Ikura, 1995; Nelson and Chazin, 1998). Each domain consists of two EF hand motifs containing helix-turn-helix supersecondary structure (Fig. 25m). Residues number 1, 3, 5, 7, 9 and 12 of the middle turn conformation bind calcium in a pentagonal bipyramidal configuration. The ligand groups (mainly side-chain carboxylates/amides and hydroxyl) are in close contact. The planar moieties of Asp/Asn/Glu are positioned almost at perpendicular orientations (Table 3, from which Thr62, not a residue under consideration, has been excluded).

7.1. Conservation of residues

Residue positions that are well conserved and clustered together in space usually correspond to the active site and/or folding nuclei, which are requisite for the fast and correct folding of proteins into their stable three-dimensional structures (Mirny and Shakhnovich, 1999; Ptitsyn and Ting, 1999; Kisters-Woike et al., 2000; Panchenko et al., 2004). It has been posited that the evolutionary conservation of amino acid residues in a structure correlates with the degree of their packing (Liao et al., 2005). Structurally conserved residues can also distinguish between binding sites and exposed protein surfaces on one hand (Valdar and Thornton, 2001; Ma et al., 2003; Caffrey et al., 2004), and different types of protein–protein interfaces on the other (Elcock and McCammon, 2001; Mintseris and Weng, 2005; Guharoy and Chakrabarti, 2005). Methodologies, based on evolutionary relations among proteins as revealed by inferred phylogenetic trees, have been devised that delineate the functional epitope and identify residues critical to binding specificity (Lichtarge et al., 1996; Armon et al., 2001). In general, the above studies look into the conservation of a group of residues, such as the ones located in the interface, rather than pairs of planar residues, the subject under consideration here.

While the conservation of catalytic residues cannot be disputed (Rennell et al., 1991), some residues in close proximity, but with no direct role in the catalytic mechanism, are also conserved. Structure-based sequence alignment of 728 sequences of different globin subfamilies shows that in each subfamily there is a functional cluster that not only includes six heme-binding residues, but seven others that do not bind the heme but belong to its immediate neighborhood (Ptitsyn and Ting, 1999). Likewise, the extended active site of cysteine proteinase, discussed in Section 7, encompasses two Trp residues (Fig. 25d), which are conserved (Brömme et al., 1996).

Salt bridges constitute the prime example of pair-wise interactions considered here and their conservation has attracted considerable attention (Barlow and Thornton, 1983; Sundaralingam et al., 1987; Kumar and Nussinov, 1999). Barlow and Thornton (1983) reported that highly conserved salt bridges occur either in catalytic sites, or as connecting units between different protein subunits (Gibbs et al., 1990). Similar to the observation of Thornton (1981) on disulfide bridges, residues forming an isolated salt-bridge have a tendency to disappear in pairs and sometimes acid-base pairs can be changed to base-acid pairs. However, complex salt bridges (those joining more than two charged residues) are found to be more conserved than the isolated ones (Musafia et al., 1995). In general, conserved salt bridges are less exposed than the non-conserved ones and a consideration of 3 modes of conservation—preservation (same charges at the same residue positions), compensation (reversal of charges) and complementation (maintenance of a salt bridge between two segments of secondary structures)—indicated that preservation was the predominant mode in the conserved interactions, complementation was of secondary importance and compensation occurred rarely and mostly in exposed salt bridges (Schueler and Margalit, 1995).

Free cysteines may mutate rapidly, but cystines are generally better conserved (Thornton, 1981). In 50% of the cases when the disulfide bond is fully conserved, it is in contact with at least one aromatic residue that is also conserved and the geometry is *en* or *et* (Bhattacharyya et al., 2004). In the majority of these cases of conserved pairs the relative accessibilities of both the half cystines is $<7\%$. When the conservation of the disulfide bridge is partial, the members retaining the bridge also retain an S$\cdots$aromatic contact. Conservation of the triad involving Cys–Cys and Trp was reported in the Fab molecule (Lesk and Chothia, 1982), and was then found to be a common feature in almost all immunoglobulin domains (Ioerger et al., 1999). Besides immunoglobulin superfamily the C–CW triad is found, though not always with the same configuration or close packing, in many other proteins, which are usually secreted into the extracellular space and the protection of disulfide bonds from hostile, reducing environments may be a role of the Trp residue. However, when the contact is typical face-on (geometry *en*, involving Trp40 in the PDB file 1qfo, as given in Table 5 of Bhattacharyya et al. (2004)), the specific interaction may provide extra stability to the structure. That a specific interaction can contribute to the stability and specificity of binding can be seen in many of the binding sites involving Pro, for which the interacting molecule also harbors a conserved aromatic residue such that a Pro–aromatic interaction can be sustained in the interface (Bhattacharyya and Chakrabarti, 2003).

8. Conclusions

The native structure is the compromise of a large number of noncovalent interactions that exist in proteins and the geometrical features relating two residue-types are expected to be rather broad. However, based on the distribution of interplanar angles it was suggested that there is non-randomness in the packing of side chains (Mitchell et al., 1997). Using two angular parameters to define the relative orientations, each pair of planar residues is indeed found to exhibit preference for one or two geometries, and similarly avoidance for some. The origin of this selectivity may be due to the avoidance of unfavorable electrostatic interactions, for example, in aromatic–Met interaction, or the occurrence of favorable nonconventional, but directional interactions (such as C–H$\cdots\pi$, in Pro-aromatic interaction) and stronger hydrogen bonds (for example, in Asp–His, Asp–Arg). The geometry of X–H$\cdots\pi$ interaction is much softer than that of conventional hydrogen bonds, allowing large lateral displacements of the donor and strong bending of the hydrogen bond angle without much of a change in energy (Steiner and Koellner, 2001)—thus even if a particular grid element is optimum according to the geometric criteria of an interaction, the neighboring element is also likely to sustain a similar interaction. The geometries that are observed in abundance are not necessarily the ones that have the highest interaction energy between the two moieties in a pair, but the ones that can provide the maximum

overall stability to the protein structure by the optimum use of all hydrogen bonding sites. For example, for the Phe–peptide interaction (Fig. 17a), the *fe* orientation that can sustain the stable N–H$\cdots\pi$ interaction is observed less than the *of* geometry having no such interaction, but which allows the formation of conventional N–H$\cdots$O interactions with other groups in the structure.

The uneven distribution of sequence differences indicates strong structural specificities for a few residue pairs. Thus Phe–His is found at i, $i+4$ positions in an α-helix (Fig. 24c). The optimum spacing in the sequence and the order of the residues may facilitate interaction between them when located in a secondary structure. Such motifs should be important in the design of peptides. Though we were concerned with planar groups only, the importance of Gly residue (in a peptide moiety) stood out as an important donor group participating in X–H$\cdots\pi$ interaction and stabilizing structural motifs. The general pair-wise interaction geometries are maintained in protein active sites also. In an interesting example of the reciprocity of the use of side chains in a pair, in methionyl tRNA synthetase a Trp residue is employed in the active site to bind Met, and the role is reversed in tryptophanyl tRNA synthetase, where a Met in the active site engages the incoming Trp (Fig. 25j and k). In addition to their catalytic role, many of the planar groups can also be part of extended active sites or binding regions and thus be conserved in homologous proteins. The general trend in the geometry of interaction observed within protein tertiary structures is maintained when the groups are located at the binding sites and the interaction is across the protein–protein interface—thus pointing to the similarity of chemical interactions that occur during protein folding and binding (Saha et al., 2007).

Acknowledgments

We are grateful to many present and past members of the lab whose contributions have made this article possible. The work was supported by the Department of Biotechnology and the Council of Scientific and Industrial Research, India. P.C. is indebted to Prof. Y.T. Thathachari Award for Science 2005.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pbiomolbio.2007.03.016

References

Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I., 1995. Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. J. Mol. Biol. 252, 460–471.

Allen, F.H., Bird, C.M., Rowland, R.S., Raithby, P.R., 1997. Hydrogen-bond acceptor and donor properties of divalent sulfur (Y–S–Z and R–S–H). Acta Crystallog. B 53, 696–701.

Allocati, N., Casalone, E., Masulli, M., Polekhina, G., Rossjohn, J., Parker, M.W., Di Ilio, C., 2000. Evaluation of the role of two conserved active-site residues in beta class glutathione S-transferases. Biochem. J. 351, 341–346.

Andrew, C.D., Bhattacharjee, S., Kokkoni, N., Hirst, J.D., Jones, G.R., Doig, A.J., 2002. Stablizing interactions between aromatic and basic side chains in α-helical peptides and proteins. Tyrosine effects on helix circular dichroism. J. Am. Chem. Soc. 124, 12706–12714.

Aravinda, S., Shamala, N., Das, C., Sriranjini, A., Karle, I.L., Balaram, P., 2003. Aromatic–aromatic interactions in crystal structures of helical peptide scaffolds containing projecting phenylalanine residues. J. Am. Chem. Soc. 125, 5308–5315.

Armon, A., Graur, D., Ben-Tal, N., 2001. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. J. Mol. Biol. 307, 447–463.

Armstrong, K.M., Fairman, R., Baldwin, R.L., 1993. The (i, $i+4$) Phe–His interaction studied in an alanine-based α-helix. J. Mol. Biol. 230, 284–291.

Ashida, T., Kakudo, M., 1974. Conformations of prolyl residues in oligopeptides. Bull. Chem. Soc. Japan 47, 1129–1133.

Atwood, J.L., Hamada, F., Robinson, K.D., Orr, G.W., Vincent, R.L., 1991. X-ray diffraction evidence for aromatic π hydrogen bonding to water. Nature 349, 683–684.

Axe, D.D., Foster, N.W., Fersht, A.R., 1996. Active barnase variants with completely random hydrophobic cores. Proc. Natl. Acad. Sci. USA 93, 5590–5594.

Babu, M.M., Singh, S.K., Balaram, P., 2002. A C–H$\cdots$O hydrogen bond stabilized polypeptide chain reversal motif at the C terminus of helices in proteins. J. Mol. Biol. 322, 871–880.

Bahar, I., Jernigan, R.L., 1997. Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. J. Mol. Biol. 266, 195–214.

Baker, D., Agard, D.A., 1994. Kinetics versus thermodynamics in protein folding. Biochemistry 33, 7505–7509.
Baker, E.N., Hubbard, R.E., 1984. Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44, 97–179.
Banerjee, R., Sen, M., Bhattacharya, D., Saha, P., 2003. The jigsaw puzzle model: search for conformational specificity in protein interiors. J. Mol. Biol. 333, 211–226.
Barlow, D.J., Thornton, J.M., 1983. Ion-pairs in proteins. J. Mol. Biol. 168, 867–885.
Baud, F., Karlin, S., 1999. Measures of residue density in protein structures. Proc. Natl. Acad. Sci. USA 96, 12494–12499.
Behe, M.J., Lattman, E.E., Rose, G.D., 1991. The protein-folding problem: the native fold determines packing, but does packing determine the native fold? Proc. Natl. Acad. Sci. USA 88, 4195–4199.
Bella, J., Berman, H.M., 1996. Crystallographic evidence for C^{α}–H···O=C hydrogen bonds in a collagen triple helix. J. Mol. Biol. 264, 734–742.
Bent, H.A., 1968. Structural chemistry of donor–acceptor interactions. Chem. Rev. 68, 587–648.
Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242.
Bhattacharyya, R., Chakrabarti, P., 2003. Stereospecific interactions of proline residues in protein structures and complexes. J. Mol. Biol. 331, 925–940.
Bhattacharyya, R., Samanta, U., Chakrabarti, P., 2002. Aromatic–aromatic interactions in and around α-helices. Protein Eng. 15, 91–100.
Bhattacharyya, R., Saha, R.P., Samanta, U., Chakrabarti, P., 2003. Geometry of interaction of the histidine ring with other planar and basic residues. J. Proteome Res. 2, 255–263.
Bhattacharyya, R., Pal, D., Chakrabarti, P., 2004. Disulfide bonds, their stereospecific environment and conservation in protein structures. Protein Eng. 17, 795–808.
Blundell, T., Singh, J., Thornton, J., Burley, S.K., Petsko, G.A., 1986. Aromatic interactions. Science 234, 1005.
Botos, I., Meyer, E., Swanson, S.M., Lemaître, V., Eeckhout, Y., Meyer, E.F., 1999. Structure of recombinant mouse collagenase-3 (MMP-13). J. Mol. Biol. 292, 837–844.
Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., Power, S., 1988. The three-dimensional structure of *Bacillus amyloliquifaciens* subtilisin at 1.8 Å resolution and an analysis of the structural consequences of peroxide inactivation. J. Biol. Chem. 263, 7895–7906.
Brandl, M., Weisss, M.S., Jabs, A., Sühnel, J., Hilgenfeld, R., 2001. C–H···π interactions in proteins. J. Mol. Biol. 307, 357–377.
Brocchieri, L., Karlin, S., 1994. Geometry of interplanar residue contacts in protein structures. Proc. Natl. Acad. Sci. USA 91, 9297–9301.
Bromberg, S., Dill, K.A., 1994. Side-chain entropy and packing in proteins. Protein Sci. 3, 997–1009.
Brömme, D., Bonneau, P.R., Purisima, E., Lachance, P., Hajnik, S., Thomas, D.Y., Storer, A.C., 1996. Contribution to activity of histidine–aromatic, amide–aromatic and aromatic–aromatic interactions in the extended catalytic site of cysteine proteinases. Biochemistry 35, 3970–3979.
Buckle, A.M., Cramer, P., Fersht, A.R., 1996. Structural and energetic responses to cavity-creating mutations in hydrophobic cores: observation of a buried water molecule and the hydrophilic nature of such hydrophobic cavities. Biochemistry 35, 4298–4305.
Bürgi, H.B., Dunitz, J.D., 1983. From crystal statics to chemical dynamics. Acc. Chem. Res. 16, 153–161.
Bürgi, H.B., Dunitz, J.D., Shefter, E., 1973. Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. J. Am. Chem. Soc. 95, 5065–5067.
Burley, S.K., Petsko, G.A., 1985. Aromatic–aromatic interaction: a mechanism of protein structure stabilization. Science 229, 23–28.
Burley, S.K., Petsko, G.A., 1986. Amino–aromatic interactions in proteins. FEBS Lett. 203, 139–143.
Burley, S.K., Petsko, G.A., 1988. Weakly polar interactions in proteins. Adv. Prot. Chem. 39, 125–189.
Burling, F.T., Goldstein, B.M., 1993. A database study of nonbonded intramolecular sulfur-nucleophile contacts. Acta Crystallogs. B 49, 738–744.
Caffrey, D.R., Somaroo, S., Hughes, J.D., Mintseris, J., Huang, E.S., 2004. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? Protein Sci. 13, 190–202.
Campbell, R.L., Rose, D.R., Wakarchuk, W.W., To, R., Sung, W., Yaguchi, M., 1993. A comparison of the structures of the 20 kD xylanases from *Trichoderma harzianum* and *Bacillus circulans*. In: Suominen, P., Reinikainen, T. (Eds.), Proceedings of the Second TRICEL Symposium on *Trichoderma reesei* Cellulases and Other Hydrolases, Espoo, Finland. Foundation for Biotechnical and Industrial Fermentation Research, Helsinki, pp. 63–72.
Carredano, E., Westerlund, B., Persson, B., Saarinen, M., Ramaswamy, S., Eaker, D., Eklund, H., 1998. The three-dimensional structures of two toxins from snake venom throw light on the anticoagulant and neurotoxic sites of phospholipase A2. Toxicon 36, 75–92.
Carugo, O., 1999. Stereochemistry of the interaction between methionine sulfur and the protein core. Biol. Chem. 380, 495–498.
Chakrabarti, P., 1989. Geometry of interaction of metal ions with sulfur-containing ligands in protein strcutures. Biochemistry 28, 6081–6085.
Chakrabarti, P., Chakrabarti, S., 1998. C–H···O hydrogen bond involving proline residues in α-helices. J. Mol. Biol. 284, 867–873.
Chakrabarti, P., Pal, D., 1997. An electrophile–nucleophile interaction in metalloprotein structures. Protein Sci. 6, 851–859.
Chakrabarti, P., Pal, D., 2001. The interrelationships of side-chain and main-chain conformations in proteins. Prog. Biophys. Mol. Biol. 76, 1–102.
Chakrabarti, P., Samanta, U., 1995. CH/π interaction in the packing of the adenine ring in protein structures. J. Mol. Biol. 251, 9–14.
Chakrabartty, A., Baldwin, R.L., 1995. Stability of α-helices. Adv. Protein Chem. 46, 141–176.
Chelli, R., Gervasio, F.L., Procacci, P., Schettino, V., 2002. Stacking and T-shape competition in aromatic–aromatic amino acid interactions. J. Am. Chem. Soc. 124, 6133–6143.

Chipot, C., Jaffe, R., Maigret, B., Pearlman, D.A., Kollman, P.A., 1996. Benzene dimer: a good model for π–π interactions in proteins? A comparison between the benzene and the toluene dimers in the gas phase and in an aqueous solution. J. Am. Chem. Soc. 118, 11217–11224.

Cohen, P., 2000. The regulation of protein function by multisite phosphorylation—a 25 year update. Trends Biochem. Sci. 25, 596–601.

Conti, E., Stachelhaus, T., Marahiel, M.A., Brick, P., 1997. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidine *S*. The EMBO J. 16, 4174–4183.

Crepin, T., Schmitt, E., Mechulam, Y., Sampson, P.B., Vaughan, M.D., Honek, J.F., Blanquet, S., 2003. Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase. J. Mol. Biol. 332, 59–72.

Crick, F.H.C., 1953. The packing of α-helices: simple coiled coils. Acta Crystallogr. D 6, 689–697.

Crivici, A., Ikura, M., 1995. Molecular and structural basis of target recognition by calmodulin. Annu. Rev. Biophys. Biomol. Struct. 24, 85–116.

Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., Leberman, R., 1990. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA sythetase at 2.5 Å. Nature 347, 249–255.

Cusack, S., Hartlein, M., Leberman, R., 1991. Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. Nucleic Acids Res. 19, 3489–3498.

Dadlez, M., Bierzynski, A., Godzik, A., Sobocinska, M., Kupryszewski, G., 1988. Conformational role of His-12 in C-peptide of ribonuclease A. Biophys. Chem. 31, 175–181.

Dasgupta, B., Pal, L., Basu, G., Chakrabarti, P., 2004. Expanded turn conformations: characterization and sequence-structure correspondence in α-turns with implications in helix folding. Proteins Struct. Funct. Bioinform. 55, 305–315.

Dasgupta, B., Chakrabarti, P., Basu, G., 2007. A sequence motif stabilizing Pro–Pro *cis* peptide bond, under submission.

DeLano, W.L., 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA 〈http://www.pymol.org〉.

Derewenda, Z.S., Derewenda, U., Kobos, P.M., 1994. (His)C^{ε}–H$\cdots$O$=$C$<$ hydrogen bond in the active sites of serine hydrolases. J. Mol. Biol. 241, 83–93.

Derewenda, Z.S., Lee, L., Derewenda, U., 1995. The occurrence of C–H$\cdots$O hydrogen bonds in proteins. J. Mol. Biol. 252, 248–262.

Desiraju, G.R., 2002. Hydrogen bridges in crystal engineering: interactions without borders. Acc. Chem. Res. 35, 565–573.

Desiraju, G.R., Steiner, T., 1999. The Weak Hydrogen Bond in Structural Chemistry and Biology. Oxford University Press, Oxford.

Diamond, R., 1974. Real-space refinement of the structure of hen egg-white lysozyme. J. Mol. Biol. 82, 371–391.

Dill, K.A., 1990. Dominant forces in protein folding. Biochemistry 29, 7133–7155.

Doublié, S., Bricogne, G., Gilmore, C., Carter Jr., C.W., 1995. Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. Structure 3, 17–31.

Dougherty, D.A., 1996. Cation-π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science 271, 163–168.

Duan, G., Smith Jr., V.H., Weaver, D.F., 2000. Characterization of aromatic-amide (side-chain) interactions in proteins through systematic ab initio calculations and data mining analyses. J. Phys. Chem. A 104, 4521–4532.

Dunitz, J.D., 1979. X-ray analysis and the structure of organic molecules. Cornell University Press, Ithaca, NY.

Eisenberg, D., Wesson, M., Yamashita, M., 1989. Interpretation of protein folding and biding with atomic solvation parameters. Chem. Scr. 29A, 217–221.

Elcock, A.H., McCammon, J.A., 2001. Identification of protein oligomerization states by analysis of interface conservation. Proc. Natl. Acad. Sci. USA 98, 2990–2994.

Eriani, G., Delarue, M., Poch, O., Gangloff, J., Moras, D., 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347, 203–206.

Eriksson, A.E., Baase, W.A., Zhang, X.-J., Hienz, D.W., Blaber, M., Baldwin, E.P., Matthews, B.W., 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255, 178–183.

Evdokimov, A.G., Anderson, D.E., Routzahn, K.M., Waugh, D.S., 2001. Structural basis for oligosaccharide recognition by *Pyrococcus furiosus* maltodextrin-binding protein. J. Mol. Biol. 305, 891–904.

Fabiola, G.F., Krishnaswamy, S., Nagarajan, V., Pattabhi, V., 1997. C–H$\cdots$O hydrogen bonds in β-sheets. Acta Crystallogr. D 53, 316–320.

Fernández-Recio, J., Vázquez, A., Civera, C., Sevilla, P., Sancho, J., 1997. The tryptophan/histidine interaction in α-helices. J. Mol. Biol. 267, 184–197.

Fishel, L.A., Farnum, M.F., Mauro, J.M., Miller, M.A., Kraut, J., Liu, Y., Tan, X., Scholes, C.P., 1991. Compound I radical in site-directed mutants of cytochrome c peroxidase as probed by electron paramagnetic resonance and electron-nuclear double resonance. Biochemistry 30, 1986–1996.

Flanagan, K., Walshaw, J., Price, S.L., Goodfellow, J.M., 1995. Solvent interactions with π ring systems in proteins. Protein Eng. 8, 109–116.

Flocco, M.M., Mowbray, S.L., 1994. Planar stacking interactions of arginine and aromatic side-chains in proteins. J. Mol. Biol. 235, 709–717.

Flocco, M.M., Mowbray, S.L., 1995. Strange bedfellows: interactions between acidic side-chains in proteins. J. Mol. Biol. 254, 96–105.

Frömmel, C., Preissner, R., 1990. Prediction of prolyl residues in *cis*-conformation in protein structures on the basis of the amino acid sequence. FEBS Lett. 277, 159–163.

Fukui, K., Yonezawa, T., Shinga, H., 1952. A molecular orbital theory of reactivity in aromatic hydrocarbons. J. Chem. Phys. 20, 722–772.

Gallivan, J.P., Dougherty, D.A., 1999. Cation–π interactions in structural biology. Proc. Natl. Acad. Sci. USA 96, 9459–9464.
García- Sáez, I., Párraga, A., Phillips, M.F., Mantle, T.J., Coll, M., 1994. Molecular structure at 1.8 Å of mouse liver class pi glutathione *S*-transferase complex with S-(*p*-nitrobenzyl) glutathione and other inhibitors. J. Mol. Biol. 237, 298–314.
Gassner, N.C., Baase, W.A., Matthews, B.W., 1996. A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proc. Natl. Acad. Sci. USA 93, 12155–12158.
Georis, J., De Lemos Esteves, F., Lamotte-Brasseur, J., Bougnet, V., Devreese, B., Giannotta, F., Granier, B., Frère, J.-M., 2000. An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: structural basis and molecular study. Protein Sci. 9, 466–475.
Ghosh, A., Bansal, M., 1999. C–H···O hydrogen bonds in minor groove of A-tracts in DNA double helices. J. Mol. Biol. 294, 1149–1158.
Gibbs, M.R., Moody, P.C.E., Leslie, G.W., 1990. Crystal structure of aspartic acid-199> asparagine mutant of chloarmphenicol acetyltransferase to 2.35 Å resolution: structural consequences of disruption of a buried salt bridge. Biochemistry 29, 11261–11265.
Gilman, A.G., Rall, T.W., Miles, A.S., Taylor, P., 1993. The Pharmaceutical Basis of Therapeutics, 8th ed. McGraw Hill, Inc., New York.
Gould, R.O., Gray, A.M., Taylor, P., Walkinshaw, M.D., 1985. Crystal environments and geometries of leucine, isoleucine, valine, and phenylalanine provide estimates of minimum nonbonded contact and preferred van der Waals interaction distances. J. Am. Chem. Soc. 107, 5921–5927.
Govindarajan, S., Goldstein, R.A., 1995. Searching for foldable protein structures using optimized energy functions. Biopolymers 36, 43–51.
Gregoret, L.M., Rader, S.D., Fletterick, R.J., Cohen, F.E., 1991. Hydrogen bonds involving sulfur atoms in proteins. Proteins Struct. Funct. Genet. 9, 99–107.
Gromiha, H.M., Selvaraj, S., 2004. Inter-residue interactions in protein folding and stability. Prog. Biophys. Mol. Biol. 86, 235–277.
Guharoy, M., Chakrabarti, P., 2005. Conservation and relative importance of residues across protein–protein interfaces. Proc. Natl. Acad. Sci. USA 102, 15447–15452.
Harbury, P.B., Zhang, T., Kim, P.S., Alber, T., 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262, 1401–1407.
Henrissat, B., Bairoch, A., 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 293, 781–788.
Heringa, J., Argos, P., 1991. Side-chain clusters in protein structures and their role in protein folding. J. Mol. Biol. 220, 151–171.
Hiraga, K., Yutani, K., 1996. Study of cysteine residues in the α subunit of *Escherichia coli* tryptophan synthase. 1. Role in conformational stability. Protein Eng. 9, 425–431.
Ho, B.K., Curmi, P.M.G., 2002. Twist and shear in β-sheets and β-ribbons. J. Mol. Biol. 317, 291–308.
Hobza, P., Selzle, H.L., Schlag, E.W., 1994. Structure and properties of benzene-containing molecular clusters; nonempirical *ab initio* calculations and experiments. Chem. Rev. 94, 1767–1785.
Hountondji, C., Lederer, F., Dessen, P., Blanquet, S., 1986. *Escherichia coli* tyrosyl- and methionyl-tRNA synthatases display sequence similarity at the binding site for the 3′-end of tRNA. Biochemistry 25, 16–21.
Huang, Q., Liu, S., Tang, Y., 1993. Refined 1.6 Å resolution crystal structure of the complex formed between porcine β-trypsin and MCTI-A, a trypsin inhibitor of the squash family: detailed comparison with bovine β-trypsin and its complex. J. Mol. Biol. 229, 1022–1036.
Hubbard, S.R., 1997. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. EMBO J. 16, 5573–5581.
Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., Sigrist, C.J., 2006. The PROSITE database. Nucl. Acids Res. 34, D227–D230.
Hunter, C.A., 1994. The role of aromatic interactions in molecular recognition. Chem. Soc. Rev. 23, 101–109.
Hunter, C.A., Sanders, J.K.M., 1990. The nature of π–π interactions. J. Am. Chem. Soc. 112, 5525–5534.
Hunter, C.A., Singh, J., Thornton, J.M., 1991. π–π interactions: the geometry and energetics of phenylalanine–phenylalanine interactions in proteins. J. Mol. Biol. 218, 837–846.
Hunter, T., 2000. Signaling—2000 and beyond. Cell 100, 113–127.
Huse, M., Kuriyan, J., 2002. The conformational plasticity of protein kinases. Cell 109, 275–282.
Hutchinson, E.G., Thornton, J.M., 1994. A revised set of potentials for β-turn formation in proteins. Protein Sci. 3, 2207–2216.
Huyghues-Despointes, B.M.P., Klingler, T.M., Baldwin, R.L., 1995. Measuring the strength of side-chain hydrogen bonds in peptide helices: the Gln.Asp (i, $i+4$) interaction. Biochemistry 34, 13267–13271.
Ibrahim, B.S., Pattabhi, V., 2004. Role of weak interactions in thermal stability of proteins. Biochem. Biophys. Res. Commun. 325, 1082–1089.
Ioerger, T.R., Du, C., Linthicum, D.S., 1999. Conservation of cys–cys trp structural triads and their geometry in the protein domains of immunoglobulin superfamily members. Mol. Immunol. 36, 373–386.
Ippolito, J.A., Alexander, R.S., Christianson, D.W., 1990. Hydrogen bond stereochemistry in protein structure and function. J. Mol. Biol. 215, 457–471.
IUPAC-IUB Commission on Biochemical Nomenclature, 1970. Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969). Biochemistry 9, 3471–3479.
Iwaoka, M., Takemoto, S., Okada, M., Tomoda, S., 2002a. Weak nonbonded S···X (X = O, N, and S) interactions in proteins. Statistical and theoretical studies. Bull. Chem. Soc. Japan. 75, 1611–1625.
Iwaoka, M., Takemoto, S., Okada, M., Tomoda, S., 2002b. Statistical and theoretical investigations on the directionality of nonbonded S···O interactions. Implications for molecular design and protein enginerring. J. Am. Chem. Soc. 124, 10613–10620.

Jaffe, R.L., Smith, G.D., 1996. A quantum chemistry study of benzene dimer. J. Chem. Phys. 105, 2780–2788.
Jeffrey, G.A., Saenger, W., 1991. Hydrogen Bonding in Biological Structures. Springer, New York.
Jennings, W.B., Farrell, B.M., Malone, J.F., 2001. Attractive intramolecular edge-to-face aromatic interactions in flexible organic molecules. Acc. Chem. Res. 34, 885–894.
Jeyaprakash, A.A., Rani, P.G., Reddy, G.B., Banumathi, S., Betzel, C., Sekar, K., Surolia, A., Vijayan, M., 2002. Crystal structure of the jacalin-T-antigen complex and a comparative study of lectin-T-antigen complexes. J. Mol. Biol. 321, 637–645.
Jiang, L., Lai, L., 2002. C–H···O hydrogen bonds at protein–protein interfaces. J. Biol. Chem. 277, 37732–37740.
Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure. Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.
Kadziola, A., Abe, J.-I., Svensson, B., Haser, R., 1994. Crystal and molecular structure of barley α-amylase. J. Mol. Biol. 239, 104–121.
Kannan, N., Vishveshwara, S., 1999. Identification of side-chain clusters in protein structures by a graph sepectral method. J. Mol. Biol. 292, 441–464.
Kannan, N., Vishveshwara, S., 2000. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. Protein Eng. 13, 753–761.
Karlin, S., Zhu, Z.-Y., 1996. Characterizations of diverse residue clusters in protein three-dimensional structures. Proc. Natl. Acad. Sci. USA 93, 8344–8349.
Karlin, S., Zuker, M., Brocchieri, L., 1994. Measuring residue associations in protein structures. J. Mol. Biol. 239, 227–248.
Karpusas, M., Baase, W.A., Matsumura, M., Matthews, B.W., 1989. Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. Proc. Natl. Acad. Sci. USA 86, 8237–8241.
Kauzmann, W., 1959. Some factors in the interpretation of protein denaturation. Adv. Protein Chem. 14, 1–63.
Kellis Jr., J.T., Nyberg, K., Sali, D., Fersht, A.R., 1988. Contribution of hydrophobic interactions to protein stability. Nature 333, 784–786.
Kemmink, J., Creighton, T.E., 1993. Local conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor and their roles in folding. J. Mol. Biol. 234, 861–878.
Keskin, O., Bahar, I., Badretdinov, A.Y., Ptitsyn, O.B., Jernigan, R.L., 1998. Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. Protein Sci. 7, 2578–2586.
Kisters-Woike, B., Vangierdegom, C., Müller-Hill, B., 2000. On the conservation of protein sequences in evolution. Trends Biochem. Sci. 25, 419–421.
Klingler, T.M., Brutlag, D.L., 1994. Discovering structural correlations in α-helices. Protein Sci. 3, 1847–1857.
Koradi, R., Billeter, M., Wüthrich, K., 1996. MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. 14, 51–55.
Krupa, A., Preethi, G., Srinivasan, N., 2004. Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases. J. Mol. Biol. 339, 1025–1039.
Kryger, G., Silman, I., Sussman, J.L., 1999. Structure of acetylcholinesterase complexed with E2020 (Aricept): implications for the design of new anti-Alzheimer drugs. Structure 7, 297–307.
Kumar, S., Nussinov, R., 1999. Salt bridge stability in monomeric proteins. J. Mol. Biol. 293, 1241–1255.
Lesk, A.M., Chothia, C., 1980. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J. Mol. Biol. 136, 225–270.
Lesk, A.M., Chothia, C., 1982. Evolution of proteins formed by β-sheet. II. Core of the immunoglobulin domains. J. Mol. Biol. 160, 325–345.
Levitt, M., Perutz, M.F., 1988. Aromatic rings act as hydrogen bond acceptors. J. Mol. Biol. 201, 751–754.
Liang, J., Dill, K.A., 2001. Are proteins well-packed? Biophys. J. 81, 751–766.
Liao, H., Yeh, W., Chiang, D., Jernigan, R.L., Lustig, B., 2005. Protein sequence entropy is closely related to packing density and hydrophobicity. Protein Eng. Des. Sel. 18, 59–64.
Lichtarge, O., Bourne, H.R., Cohen, F.E., 1996. An evolutionary trace method defines binding surfaces common to protein families. J. Mol. Biol. 257, 342–358.
Lim, W.A., Sauer, R.T., 1989. Alternative packing arrangements in the hydrophobic core of λ repressor. Nature 339, 31–36.
Lim, W.A., Sauer, R.T., 1991. The role of internal packing interactions in determining the structure and stability of a protein. J. Mol. Biol. 219, 359–376.
Liu, Q., Huang, Q., Teng, M., Weeks, C.M., Jelsch, C., Zhang, R., Niu, L., 2003. The crystal structure of a novel, inactive, lysine 49 PLA2 from *Agkistrodon acutus* venom. J. Biol. Chem. 278, 41400–41408.
Liu, S., Ji, X., Gilliland, G.L., Stevens, W.J., Armstrong, R.N., 1993. Second-sphere electrostatic effects in the active site of glutathione S-trnasferase. Observation of an on-face hydrogen bond between the side chain of threonine 13 and the π-cloud of tyrosine 6 and its influence on catalysis. J. Am. Chem. Soc. 115, 7910–7911.
Loewenthal, R., Sancho, J., Fersht, A.R., 1992. Histidine–aromatic interactions in barnase. Elevation of histidine pKa and contribution to protein stability. J. Mol. Biol. 224, 759–770.
Loll, B., Raszewski, G., Saenger, W., Biesiadka, J., 2003. Functional role of C^{α}–H···O hydrogen bonds between transmembrane α-helices in photosystem I. J. Mol. Biol. 328, 737–747.
Ma, B., Elkayam, T., Wolfson, H., Nussinov, R., 2003. Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proc. Natl. Acad. Sci. USA 100, 5772–5777.
Ma, J.C., Dougherty, D.A., 1997. The cation–π interaction. Chem. Rev. 97, 1303–1324.
MacArthur, M.W., Thornton, J.M., 1991. Influence of proline residues on protein conformation. J. Mol. Biol. 218, 397–412.

Mandel-Gutfreund, Y., Margalit, H., Jernigan, R.L., Zhurkin, V.B., 1998. A role for C–H···O interactions in protein–DNA recognition. J. Mol. Biol. 277, 1129–1140.

Manikandan, K., Ramakumar, S., 2004. The occurrence of C–H···O hydrogen bonds in α-helices and helix termini in globular proteins. Proteins Struct. Funct. Bioinform. 56, 768–781.

Matsui, I., Matsui, E., Sakai, Y., Kikuchi, H., Kawarabayasi, Y., Ura, H., Kawaguchi, S., Kuramitsu, S., Harata, K., 2000. The molecular structure of hyperthermostable aromatic aminotransferase with novel substrate specificity from *Pyrococcus horikoshii*. J. Biol. Chem. 275, 4871–4879.

Mazza, C., Segreff, A., Mattaj, I.W., Cusack, S., 2002. Large-scale induced fit recognition of an m^7GpppG cap analogue by the human nuclear cap-binding complex. EMBO J. 21, 5548–5557.

McDonald, I.K., Thornton, J.M., 1994. Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. 238, 777–793.

McGaughey, G.B., Gagné, M., Rappé, A.K., 1998. π-stacking interactions: alive and well in proteins. J. Biol. Chem. 273, 15458–15463.

Meador, W.E., Means, A.R., Quiocho, F.A., 1993. Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. Science 262, 1718–1721.

Meurisse, R., Brasseur, R., Thomas, A., 2003. Aromatic side-chain interactions in proteins. Near- and far-sequence His-X pairs. Biochim. Biophys. Acta 1649, 85–96.

Meurisse, R., Brasseur, R., Thomas, A., 2004. Aromatic side-chain interactions in proteins: near- and far-sequence Tyr-X pairs. Proteins Struct. Funct. Bioinform. 54, 478–490.

Meyer, E.A., Castellano, R.K., Diederich, F., 2003. Interactions with aromatic rings in chemical and biological recognition. Angew. Chem. Int. Ed. 42, 1210–1250.

Mintseris, J., Weng, Z., 2005. Structure, function, and evolution of transient and obligate protein–protein interactions. Proc. Natl. Acad. Sci. USA 102, 10930–10935.

Mirny, L.A., Shakhnovich, E.I., 1999. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. J. Mol. Biol. 291, 177–196.

Mitchell, J.B.O., Thornton, J.M., Singh, J., Price, S.L., 1992. Towards an understanding of the arginine–aspartate interaction. J. Mol. Biol. 226, 251–262.

Mitchell, J.B.O., Nandi, C.L., McDonald, I.K., Thornton, J.M., Price, S.L., 1994. Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? J. Mol. Biol. 239, 315–331.

Mitchell, J.B.O., Laskowski, R.A., Thornton, J.M., 1997. Non-randomness in side-chain packing: the distribution of interplanar angles. Proteins Struct. Funct. Genet. 29, 370–380.

Monticelli, L., Tieleman, D.P., Colombo, G., 2005. Mechanism of helix nucleation and propagation: microscopic view from microsecond time scale MD simulations. J. Phys. Chem. B 109, 20064–20067.

Morgan, R.S., McAdon, J.M., 1980. Predictor for sulfur–aromatic interactions in globular proteins. Int. J. Peptide Protein Res. 15, 177–180.

Morgan, R.S., Tatsch, C.E., Gushard, R.H., McAdon, J.M., Warme, P.K., 1978. Chains of alternating sulfur and π-bonded atoms in eight small proteins. Int. J. Peptide Protein Res. 11, 209–217.

Morozov, A.V., Misura, K.M.S., Tsemekhman, K., Baker, D., 2004. Comparison of quantum mechanics and molecular mechanics dimerization energy landscapes for pairs of ring-containing amino acids in proteins. J. Phys. Chem. B 108, 8489–8496.

Muraki, M., Harata, K., Sugita, N., Sato, K., 2000. Protein–carbohydrate interactions in human lysozyme probed by combining site-directed mutagenesis and affinity labeling. Biochemistry 39, 292–299.

Musafia, B., Buchner, V., Arad, D., 1995. Complex salt bridges in proteins: statistical analysis of structure and function. J. Mol. Biol. 254, 761–770.

Myers, J.K., Oas, T.G., 1999. Contribution of a buried hydrogen bond to lambda repressor folding kinetics. Biochemistry 38, 6761–6768.

Nagao, Y., Hirata, T., Goto, S., Sano, S., Kakehi, A., Iizuka, K., Shiro, M., 1998. Intramolecular nonbonded S···O interaction recognized in (acylimino)thiadiazoline derivatives as angiotensin II receptor antagonists and related compounds. J. Am. Chem. Soc. 120, 3104–3110.

Narayana, S.V.L., Argos, P., 1984. Residue contacts in protein structures and implications for protein folding. Int. J. Peptide Protein Res. 24, 25–39.

Nelson, M.R., Chazin, W.J., 1998. Calmodulin as a calcium sensor. In: Eldik, L., Watterson, D. (Eds.), Calmodulin and Signal Transduction. Academic Press, New York, pp. 17–64.

Nishikawa, K., Ooi, T., 1986. Radial locations of amino acid residues in a globular protein: correlation with the sequence. J. Biochem. 100, 1043–1047.

Nishio, M., Hirota, M., Umezawa, Y., 1998. The CH/π Interaction. Evidence, Nature, and Consequences. Wiley-VCH, New York.

Pal, D., Chakrabarti, P., 1998. Different types of interactions involving cysteine sulfhydryl group in proteins. J. Biomol. Struct. Dyn. 15, 1059–1072.

Pal, D., Chakrabarti, P., 1999. *Cis* peptide bonds in proteins: residues involved, their conformations, interactions and locations. J. Mol. Biol. 294, 271–288.

Pal, D., Chakrabarti, P., 2001. Non-hydrogen bond interactions involving the methionine sulfur atom. J. Biomol. Struct. Dyn. 19, 115–128.

Pal, L., Chakrabarti, P., Basu, G., 2003. Sequence and structure patterns in proteins from an analysis of the shortest helices: implications for helix nucleation. J. Mol. Biol. 326, 273–291.

Panchenko, A.R., Kondrashov, F., Bryant, S., 2004. Prediction of functional sites by analysis of sequence and structure conservation. Protein Sci. 13, 884–892.

Pawson, T., 1994. SH2 and SH3 domains in signal transduction. Adv. Cancer Res. 64, 87–110.
Pedireddi, V.R., Desiraju, G.R., 1992. A crystallographic scale of carbon acidity. J. Chem. Soc.Chem. Commun., 988–990.
Pickersgill, R.W., Harris, G.W., Garman, E., 1992. Structure of monoclinic papain at 1.60 Å resolution. Acta Crystallogr. B, 48, 59–67.
Pierce, A.C., Sandretto, K.L., Bemis, G.W., 2002. Kinase inhibitors and the case for C–H···O hydrogen bonds in protein–ligand binding. Proteins Struct. Funct. Genet. 49, 567–576.
Pisabarro, M.T., Serrano, L., Wilmanns, M., 1998. Crystal structure of the Ab1-SH3 domain complexed with a designed high affinity peptide ligand: implications for SH3–ligand interactions. J. Mol. Biol. 281, 513–521.
Powers, E.T., Deechongkit, S., Kelly, J.W., 2005. Backbone–backbone H-bonds make context-dependent contributions to protein folding kinetics and thermodynamics: lessons from amide-to-ester mutations. Adv. Prot. Chem. 72, 39–78.
Ptitsyn, O.B., Ting, K.-L.H., 1999. Non-functional conserved residues in globins and their possible role as a folding nucleus. J. Mol. Biol. 291, 671–682.
Ramachandran, G.N., Lakshminarayanan, A.V., Balasubramanian, R., Tegoni, G., 1970. Energy calculations on proline residues. Biochim. Biophys. Acta 221, 165–181.
Ramasubbu, N., Parthasarathy, R., Murray-Rust, P., 1986. Angular preferences of intermolecular forces around halogen centers: preferred directions of approach of electophiles and nucleophiles around the carbon-halogen bond. J. Am. Chem. Soc. 108, 4308–4314.
Reid, K.S.C., Lindley, P.F., Thornton, J.M., 1985. Sulfur-aromatic interactions in proteins. FEBS Lett. 190, 209–213.
Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., Fischer, G., 1998. Side-chain effects on peptidyl–prolyl *cis*/*trans* isomerisation. J. Mol. Biol. 279, 449–460.
Rennell, D., Bouvier, S.E., Hardy, L.W., Poteete, A.R., 1991. Systematic mutation of bacteriophage T4 lysozyme. J. Mol. Biol. 222, 67–88.
Richards, F.M., 1974. The interpretation of protein structures total volume, group volume distributions and packing density. J. Mol. Biol. 82, 1–14.
Richards, F.M., Lim, W.A., 1994. An analysis of packing in the protein folding problem. Quart. Rev. Biophys. 26, 423–498.
Rodham, D.A., Suzuki, S., Suenram, R.D., Lovas, F.J., Dasgupta, S., Goddard, W.A., Blake, G.A., 1993. Hydrogen bonding in the benzene–ammonia dimer. Nature 362, 735–737.
Rooman, M., Liévin, J., Buisine, E., Wintjens, R., 2002. Cation-π/H-bond stair motifs at protein–DNA interfaces. J. Mol. Biol. 319, 67–76.
Rosenfield Jr., R.E., Parthasarathy, R., Dunitz, J.D., 1977. Directional preferences of nonbonded atomic contacts with divalent sulfur. 1. Electrophiles and nucleophiles. J. Am. Chem. Soc. 99, 4860–4862.
Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., Moras, D., 1991. Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA ^{Asp}. Science 252, 1682–1689.
Russell, R.B., Barton, G.J., 1994. Structural features can be unconserved in proteins with similar folds: an analysis of side-chain to side-chain contacts, secondary structure and accessibility. J. Mol. Biol. 244, 332–350.
Saha, R.P., Chakrabarti, P., 2006. Parity in the number of atoms in residue composition in proteins and contact preferences. Curr. Sci. 90, 558–561.
Saha, R.P., Bahadur, R.P., Chakrabarti, P., 2005. Interresidue contacts in proteins and protein-protein interfaces and their use in characterizing the homodimeric interface. J. Proteome Res. 4, 1600–1609.
Saha, R.P., Bhattacharyya, R., Chakrabarti, P., 2007. Interaction geometry involving planar groups in protein-protein interfaces. Proteins Struct. Funct. Bioinform. 67, 84–97.
Salemme, F.R., Freer, S.T., Xuong, N.H., Alden, R.A., Kraut, J., 1973. The structure of oxidized cytochrome c2 of *Rhodospirillum rubrum*. J. Biol. Chem. 248, 3910–3921.
Samanta, U., Chakrabarti, P., Chandrasekhar, J., 1998. Ab initio study of energetics of X–H···π (X = N, O, and C) interactions involving a heteroaromatic ring. J. Phys. Chem. A 102, 8964–8969.
Samanta, U., Pal, D., Chakrabarti, P., 1999. Packing of aromatic rings against tryptophan residues in proteins. Acta Crystallogr. D 55, 1421–1427.
Samanta, U., Pal, D., Chakrabarti, P., 2000. Environment of tryptophan side chains in proteins. Proteins Struct. Funct. Genet. 38, 288–300.
Samanta, U., Bahadur, R.P., Chakrabarti, P., 2002. Quantifying the accessible surface area of protein residues in their local environment. Protein Eng. 15, 659–667.
Sampson, N.S., Vrielink, A., 2003. Cholesterol oxidases: a study of nature's approach to protein design. Acc. Chem. Res. 36, 713–722.
Sandberg, W.S., Terwilliger, T.C., 1989. Influence of interior packing and hydrophobicity on the stability of a protein. Science 245, 54–60.
Sarkhel, S., Desiraju, G.R., 2004. N–H···O, O–H···O and C–H···O hydrogen bonds in protein-ligand complexes: strong and weak interactions in molecular recognition. Proteins Struct. Funct. Bioinform. 54, 247–259.
Sathyapriya, R., Vishveshwara, S., 2004. Interaction of DNA with clusters of amino acids in proteins. Nucleic Acids Res. 32, 4109–4118.
Sawyer, L., James, M.N.G., 1982. Carboxyl–carboxylate interactions in proteins. Nature 295, 79–80.
Scheiner, S., Kar, T., Gu, Y., 2001. Strength of the C^{α}H···O hydrogen bond of amino acid residues. J. Biol. Chem. 276, 9832–9837.
Scheiner, S., Kar, T., Pattanayak, J., 2002. Comparison of various types of hydrogen bonds involving aromatic amino acids. J. Am. Chem. Soc. 124, 13257–13264.
Schenk, P.W., Snaar-Jagalska, B.E., 1999. Signal perception and transduction: the role of protein kinases. Biochem. Biophys. Acta 1449, 1–24.
Schueler, O., Margalit, H., 1995. Conservation of salt bridges in protein families. J. Mol. Biol. 248, 125–135.

Senes, A., Ubarretxena-Belandia, I., Engelman, D.M., 2001. The C_α–H···O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. Proc. Natl. Acad. Sci. USA 98, 9056–9061.

Senes, A., Engel, D.E., DeGrado, W.F., 2004. Folding of helical membrane proteins: the role of polar GxxxG-like proline motifs. Curr. Opin. Struct. Biol. 14, 465–479.

Serrano, L., 2000. The relationship between sequence and structure in elementary folding units. Adv. Protein Chem. 53, 49–85.

Serrano, L., Bycroft, M., Fersht, A.R., 1991. Aromatic–aromatic interactions and protein stability: investigation by double-mutant cycles. J. Mol. Biol. 218, 465–475.

Shi, Z., Olson, C.A., Kallenbach, N.R., 2002. Stabilization of α-helix structure by polar side-chain interactions: complex salt bridges, cation–π interactions, and C–H···O H-bonds. Biopolymers 60, 366–380.

Shimohigashi, Y., Nose, T., Yamauchi, Y., Maeda, I., 1999. Design of serine protease inhibitors with conformation restricted by amino acid side-chain–side-chain CH/π interaction. Biopolymers 51, 9–17.

Shoemaker, K.R., Fairman, R., Schultz, D.A., Robertson, A.D., York, E.J., Stewart, J.M., Baldwin, R.L., 1990. Side-chain interactions in the C-peptide helix: Phe8···$His12^+$. Biopolymers 29, 1–11.

Shortle, D.A., Stites, W.E., Meeker, A.K., 1990. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. Biochemistry 29, 8033–8041.

Singh, J., Thornton, J.M., 1985. The interaction between phenylalanine rings in proteins. FEBS Lett. 191, 1–6.

Singh, J., Thornton, J.M., 1990. SIRIUS: an automated method for the analysis of the preferred packing arrangements between protein groups. J. Mol. Biol. 211, 595–615.

Singh, J., Thornton, J.M., 1992. Atlas of Protein Side-Chain Interactions, vols. I and II. IRL Press, Oxford, UK.

Singh, J., Thornton, J.M., Snarey, M., Campbell, S.F., 1987. The geometries of interacting arginine-carboxyls in proteins. FEBS Lett. 224, 161–171.

Stapley, B.J., Doig, A.J., 1997. Hydrogen bonding interactions between glutamine and asparagine in α-helical peptides. J. Mol. Biol. 272, 465–473.

Stapley, B.J., Rohl, C.A., Doig, A.J., 1995. Addition of side chain interactions to modified Lifson–Roig helix-coil theory: application to energetics of phenylalanine-methionine interactions. Protein Sci., 2383–2391.

Steiner, T., Koellner, G., 2001. Hydrogen bonds with π-acceptors in proteins: frequencies and role in stabilizing local 3D structures. J. Mol. Biol. 305, 535–557.

Stewart, D.E., Sarkar, A., Wampler, J.E., 1990. Occurrence and role of *cis* peptide bonds in protein structures. J. Mol. Biol. 214, 253–260.

Sun, S., Bernstein, E.R., 1996. Aromatic van der Waals clusters: structure and nonrigidity. J. Phys. Chem. 100, 13348–13366.

Sundaralingam, M., Sekharudu, Y.C., Yathindra, N., Ravichandran, V., 1987. Ion pairs in alpha helices. Proteins Struct. Funct. Genet. 2, 64–71.

Suzuki, S., Green, P.G., Bumgarner, R.E., Dasgupta, S., Goddard, W.A., Blake, G.A., 1992. Benzene forms hydrogen bonds with water. Science 257, 942–945.

Takagi, T., Tanaka, A., Matsuo, S., Maezaki, H., Tani, M., Fujiwara, H., Sasaki, Y., 1987. Computational studies on C–H···π interactions. J. Chem. Soc., Perkin Trans. 2, 1015–1018.

Taylor, R., Kennard, O., 1982. Crystallographic evidence for the existence of C–H···O, C–H···N, and C–H···Cl hydrogen bonds. J. Am. Chem. Soc. 104, 5063–5070.

Tendulkar, A.V., Wangikar, P.P., Sohoni, M.A., Samant, V.V., Mone, C.Y., 2003. Parameterization and classification of the protein universe via geometric techniques. J. Mol. Biol. 334, 157–172.

Thomas, A., Meurisse, R., Charloteaux, B., Brasseur, R., 2002a. Aromatic side-chain interactions in proteins. I. Main structural features. Proteins Struct. Funct. Genet. 48, 628–634.

Thomas, A., Meurisse, R., Brasseur, R., 2002b. Aromatic side-chain interactions in proteins. II. Near- and far-sequence Phe-X pairs. Proteins Struct. Funct. Genet. 48, 635–644.

Thomas, K.A., Smith, G.M., Thomas, T.B., Feldmann, R.J., 1982. Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. Proc. Natl. Acad. Sci. USA 79, 4843–4847.

Thornton, J.M., 1981. Disulphide bridges in globular proteins. J. Mol. Biol. 151, 261–287.

Torshin, I.Y., Harrison, R.W., Weber, I.T., 2003. Close pairs of carboxylates: a possibility of muticenter hydrogen bonds in proteins. Protein Eng. 16, 201–207.

Tóth, G., Murphy, R.F., Lovas, S., 2001. Stabilization of local structures by π–CH and aromatic-backbone amide interactions involving prolyl and aromatic residues. Protein Eng. 14, 543–547.

Tóth, G., Kövér, K.E., Murphy, R.F., Lovas, S., 2004. Aromatic-backbone interactions in α-helices. J. Phys. Chem. B 108, 9287–9296.

Tsou, L.K., Tatko, C.D., Waters, M.L., 2002. Simple cation–π interaction between a phynyl ring and a protonated amine stabilizes an α-helix in water. J. Am. Chem. Soc. 124, 14917–14921.

Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., Tanabe, K., 2002. Origin of attraction and directionality of the π/π interaction: model chemistry calculations of benzene dimer interaction. J. Am. Chem. Soc. 124, 104–112.

Umezawa, Y., Nishio, M., 1998. CH/π interactions in the crystal structure of class I MHC antigens and their complexes with peptides. Bioorg. Med. Chem. 6, 2507–2515.

Umezawa, Y., Nishio, M., 2002. Thymine–methyl/π interaction implicated in the sequence-dependent deformability of DNA. Nucleic Acids Res. 30, 2183–2192.

Umezawa, Y., Nishio, M., 2005. CH/π hydrogen bonds as evidenced in the substrate specificity of acetylcholine esterase. Biopolymers 79, 248–258.

Valdar, W.S.J., Thornton, J.M., 2001. Protein–protein interfaces: analysis of amino acid conservation in homodimers. Proteins Struct. Funct. Genet. 42, 108–124.

van der Spoel, D., van Buuren, A.R., Tieleman, D.P., Berendsen, H.J., 1996. Molecular dynamics simulations of peptides from BPTI: a closer look at amide-aromatic interactions. J. Biomol. NMR 8, 229–238.

Vargas, R., Garza, J., Dixon, D.A., Hay, B.P., 2000. How strong is the C^{α}–H···O═C hydrogen bond? J. Am. Chem. Soc. 122, 4750–4755.

Vásquez, G.B., Ji, X., Fronticelli, C., Gilliland, G.L., 1998. Human carboxyhemoglobin at 2.2 Å resolution: structure and solvent comparisons of R-state, R2-state and T-state hemoglobins. Acta Crystallogr. D 54, 355–366.

Viguera, A.R., Serrano, L., 1995. Side-chain interactions between sulfur-containing amino acids and phenylalanine in α-helices. Biochemistry 34, 8771–8779.

Wahl, M.C., Sundaralingam, M., 1997. C–H···O hydrogen bonding in biology. Trends Biochem. Sci. 22, 97–102.

Waksman, G., Kominos, D., Robertson, S.C., Pant, N., Baltimore, D., Birge, R.B., Cowburn, D., Hanafusa, H., Mayer, B.J., Overduin, M., Resh, M.D., Rios, C.B., Silverman, L., Kuriyan, J., 1992. Crystal structure of the phosphotyrosine recognition domain SH2 of v-*src* complexed with tyrosine-phosphorylated peptides. Nature 358, 646–653.

Wang, G., Dunbrack Jr., R.L., 2003. PISCES: a protein sequence culling server. Bioinformatics 19, 1589–1591.

Wangikar, P.P., Tendulkar, A.V., Ramya, S., Mali, D.N., Sarawagi, S., 2003. Functional sites in protein families uncovered via an objective and automated graph theoretic approach. J. Mol. Biol. 326, 955–978.

Waters, M.L., 2002. Aromatic interactions in model systems. Curr. Opin. Chem. Biol. 6, 736–741.

Waters, M.L., 2004. Aromatic interactions in peptides: impact on structure and function. Biopolymers 76, 435–445.

Webster, T., Tsai, H., Kula, M., Mackie, G.A., Schimmel, P., 1984. Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. Science 226, 1315–1317.

Weiss, M.S., Brandl, M., Sühnel, J., Pal, D., Hilgenfeld, R., 2001. More hydrogen bonds for the (structural) biologist. Trends Biochem. Sci. 26, 521–523.

Wlodawer, A., Walter, J., Huber, R., Sjölin, L., 1984. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. J. Mol. Biol. 180, 301–329.

Word, J.M., Lovell, S.C., Richardson, J.S., Richardson, D.C., 1999. Asparagine and glutamine: using hydrogen atom contacts in the choice side-chain amide orientation. J. Mol. Biol. 285, 1735–1747.

Worth, G.A., Wade, R.C., 1995. The aromatic-($i+2$) amine interaction in peptides. J. Phys. Chem. 99, 17473–17482.

Wu, W.-J., Raleigh, D.P., 1998. Local control of peptide conformation: stabilization of *cis* proline peptide bonds by aromatic proline interactions. Biopolymers 45, 381–394.

Yao, J., Feher, V.A., Espejo, B.F., Reymond, M.T., Wright, P.E., Dyson, H.J., 1994. Stabilization of a type VI turn in a family of linear peptides in water solution. J. Mol. Biol. 243, 736–753.

Zauhar, R.J., Colbert, C.L., Morgan, R.S., Welsh, W.J., 2000. Evidence for a strong sulfur–aromatic interaction derived from crystallographic data. Biopolymers 53, 233–248.

Zhang, C., Kim, S.-H., 2000. Environment-dependent residue contact energies for proteins. Proc. Natl. Acad. Sci. USA 97, 2550–2555.

Zhou, G., Somasundaram, T., Blanc, E., Parthasarathy, G., Ellington, W.R., Chapman, M.S., 1998. Transition state structure of arginine kinase: implications for catalysis of bimolecular reactions. Proc. Natl. Acad. Sci. USA, 95, 8449–8454.