

Environment of Tryptophan Side Chains in Proteins

Uttamkumar Samanta, Debnath Pal, and Pinak Chakrabarti*

Department of Biochemistry, Bose Institute, Calcutta, India

ABSTRACT Although relatively rare, the tryptophan residue (Trp), with its large hydrophobic surface, has a unique role in the folded structure and the binding site of many proteins, and its fluorescence properties make it very useful in studying the structures and dynamics of protein molecules in solution. An analysis has been made of its environment and the geometry of its interaction with neighbors using 719 Trp residues in 180 different protein structures. The distribution of the number of partners interacting with the Trp aromatic ring shows a peak at 6 (considering protein residues only) and 8 (including water and substrate molecules also). The means of the solvent-accessible surface areas of the ring show an exponential decrease with the increase in the number of partners; this relationship can be used to assess the efficiency of packing of residues around Trp. Various residues exhibit different propensities of binding the Trp side chain. The aromatic residues, Met and Pro have high values, whereas the smaller and polar-chain residues have weaker propensities. Most of the interactions are with residues far away in sequence, indicating the importance of Trp in stabilizing the tertiary structure. Of all the ring atoms NE1 shows the highest number of interactions, both along the edge (hydrogen bonding) as well as along the face. Various weak but specific interactions, engendering stability to the protein structure, have been identified. *Proteins* 2000;38:288–300. © 2000 Wiley-Liss, Inc.

Key words: tryptophan; hydrogen bond; weak interactions; protein stability; solvent accessibility

INTRODUCTION

In many aspects tryptophan (Trp) is a special amino acid. Of all the residues it has the largest surface area¹ and is a highly preferred component of residue-clusters in protein structures.² It is a part of a novel cofactor that has been found in bacterial methylamine dehydrogenase,^{3–4} the site of radical formation in the catalytic cycle of cytochrome c peroxidase⁵ and is also implicated in electron-transfer pathways.^{6–7} The fused heterocyclic ring system in the Trp side chain is similar to the purine bases of DNA, and the Trp binding sites in protein structures may resemble how DNA-binding proteins wrap around the bases while binding DNA. In the light of our observation⁸ that the adenine ring of different cofactors is embedded in proteins such that branched side-chain atoms sit on top of ring N atoms, it would be of interest to analyze the

chemical entities that interact with the face of the Trp residue and their positions relative to the ring.

Trp plays an important role in the structure of many proteins and how they interact with other molecules. For example, the *trp*-repressor of *Escherichia coli* is a dimeric DNA binding protein⁹ that represses transcription of a few operons in the presence of excess tryptophan, and the important residues in the binding pocket have been identified.¹⁰ Similarly, *trp* RNA-binding attenuation protein¹¹ regulates negatively the expression of the tryptophan operon of *Bacillus subtilis* in response to intracellular levels of L-tryptophan. The stacking interaction between an aromatic ring (usually Trp) and the carbohydrate residue is an important component of carbohydrate recognition by lectins,¹² many of whose family members have the characteristic sequence pattern, Gln–any residue–Trp (QXW).¹³ The stacking interaction between Trp and a cofactor is also found in methanol dehydrogenase whose α -subunit has a superbarrel structure composed of eight “propeller blades” the interaction between which is facilitated by a pair of Gly and Trp residues held close to each other.¹⁴ Another example of the conservation of Trp in the sequence is provided by the so called WSXWS motif of members of class I of the cytokine receptor superfamily.¹⁵ The presence of Trp in the binding site is exemplified by acetylcholinesterase,¹⁶ streptavidin,¹⁷ myosin,¹⁸ and chitin-binding protein.¹⁹ The recently described WW domains are used as modules for protein-protein interaction and their presence in specific proteins has been implicated in diseases such as hypertension or muscular dystrophy.²⁰ Trp can impart thermal stability,²¹ stabilize quaternary structure^{22–23} and help in the folding process.^{24–25} The importance of Trp extends to the folding and assembly of membrane proteins also.²⁶ It has a higher propensity for the extracellular face of membrane proteins,²⁷ and is found near the lipid-water interface of the ion channel formed by gramicidin.^{28–29}

Trp residues are often used as intrinsic fluorescence probes, since they are very sensitive to changes in their local environment.^{30–31} There have been attempts to understand fluorescence lifetime in terms of local structural features of the Trp environment.³² Because of the importance of Trp in protein structure and function and their spectroscopic investigations we have undertaken a detailed analysis of the surroundings around Trp residues in protein structures, concentrating mainly on the fluoro-

*Correspondence to: Dr. P. Chakrabarti, Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VII-M, Calcutta 700 054, India. E-mail: pinak@boseinst.ernet.in

Received 1 July 1999; Accepted 23 September 1999

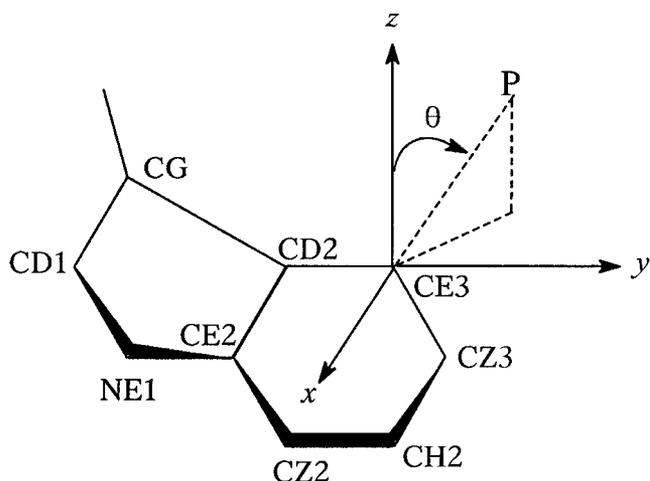


Fig. 1. Atom labels in the aromatic part of Trp and the definition of the angle θ used to determine if the partner atom, P, in contact with a ring atom (shown here as CE3) is on the face or the edge of the ring. The face shown is designated as β .

phore, i.e., the indole ring of the side chain (Fig. 1). Both intra- and intermolecular (by the application of symmetry in the crystal structure) contacts involving protein, cofactor/substrate/water molecules are considered. The pattern in the distribution of various residues and their constituent atoms around Trp should be useful in our understanding of protein stability, molecular recognition and binding.

MATERIALS AND METHODS

The protein structures were selected from the Brookhaven Protein Data Bank (PDB, 1996 version)³³ with constraints of 25% maximum sequence identity upon pairwise alignment³⁴ crystallographic resolution better than 2.0 Å and R-value of ≤ 0.20 . Out of 211 protein entries that qualify, 21 had no Trp. So as to consider only the well-ordered Trp residues those with more than two ring atoms having a thermal factor of $> 30 \text{ \AA}^2$ or with a partial occupancy factor were excluded. Any residue having any exceptional contact ($< 2 \text{ \AA}$, especially noted when intermolecular interactions were involved) were shunned out. With these screening there were 719 Trp residues from 180 structures.

To find out the partners for a Trp ring, all atoms (inclusive of those with water, cofactor, substrate etc. present in the structure, and all crystallographic symmetry applied) within 4 Å of any Trp atom were identified in a file which was sorted in the descending order of distance. This was then edited as follows. 1) Any interacting atom with a thermal factor of $> 30 \text{ \AA}^2$ or with a fractional occupancy was deleted. 2) The atom with the shortest contact distance was considered a partner. Because of the constraints of being covalently attached to such an atom other atoms may be brought into proximity of Trp without there being any preferential interaction; such atoms were excluded. However, atoms one position away were eligible for consideration. For example, if CA atom of a residue is having the shortest contact it is taken as the partner atom,

which precludes bonded neighbors N, C, and CB from consideration, but not CG, O, C of the previous residue and N of the next (if they satisfy the distance criterion), and so on. 3) If one partner atom is interacting with two or more Trp atoms, the shortest one is retained. 4) For the aromatic side chain if more than one atom is interacting (most likely a stacking arrangement) only the atom with the shortest contact is used. 5) The main-chain atoms (N, CA, C, and O) of the two adjacent residues are excluded. (While dealing with the surface area we needed to find out the partners in contact with the whole Trp residue. For this, the adjacent CB atoms were also omitted). The list obtained after the above filtering was again sorted on the atom-record-number (in the PDB file), so that the partner atoms originating from the same residue appear consecutively. Any residue that has one or more of its atoms in association with Trp was considered a partner residue.

The location of a partner atom in different regions relative to the Trp ring was next assigned. The angle, θ , between the normal to the plane at the Trp atom in contact and the direction from this atom to the partner atom, was found out (Fig. 1). Depending on the value of θ ($0 \leq \theta \leq 45^\circ$ or $45 < \theta \leq 90^\circ$) the partner was assumed to be on the face or the edge. A subset of the edge partners interacting with NE1 was considered to be hydrogen bonded. The two faces were distinguished following the recommendation of Rose et al.³⁵: the face in which a progression along CG \rightarrow CD1 \rightarrow NE1 traverses in an anticlockwise fashion is the β face. Another way of the face assignment is by expressing the partner coordinates in a molecular axial system shown in Fig. 1. Depending on the sign (+ or -) of the transformed z coordinate a partner is on the β or the α face.

The propensity, P_x , of a residue to be in the Trp environment was calculated as the proportion of the particular amino acid to be in contact with Trp divided by the proportion of all amino acids in contact, as shown below:

$$P_x = (N_x/T_x) / (N_p/T_p)$$

where N_x is the number of residues of a particular amino acid, X, in contact, T_x is the total number of residues of that amino acid in all the proteins used in the study, N_p is the total number of residues in contact and T_p is the total number of residues in the dataset. The standard deviation associated with P_x was also calculated.³⁶

For sp^2 nitrogen atom on the face we found out if the N-bearing group is stacked against Trp or the two planes are inclined to each other so that the N—H bond points towards the Trp face engendering an NH— π interaction. This was done by calculating the interplanar angle between the planes of the sp^2 N atom and the Trp ring. When this angle was less than 30° the interaction was categorized as stacked. For a larger value of the angle, we tested for the presence of NH— π interaction by finding out if the N—H vector (where the H position was obtained on stereochemical consideration) was pointing towards the Trp ring; if the deviation of H from the Trp plane was less than the deviation of N, the interaction was accepted as an NH— π interaction. The procedure followed here is slightly

different from that of Mitchell et al.³⁷ according to whom a contact on the face required the N atom to be within 20° (as against 45° here) of the perpendicular to the ring plane through its contact atom. Additionally, for the NH— π interaction any of the N—H...C angles for all possible aromatic acceptors (N...C distance < 3.8 Å) had to be 120° or more.

To find out how the solvent-accessible surface area (ASA) varies with the number of surrounding protein groups, Trp residues were segregated depending on the number of partners, which was redefined after excluding symmetry-related contacts. This was necessary to be consistent with the calculation of ASA which considered only the subunit Trp was located in (ignoring all intermolecular contacts and nonprotein molecules present in the structure). Likewise, for a comparative study the partner number was also calculated by considering all the atoms (not just the aromatic ones) in Trp residues. ASA was computed using the program ACCESS,³⁸ which is an implementation of the Lee and Richards³⁹ algorithm. We used the default van der Waals radii in the ACCESS program and the solvent probe size was 1.4 Å. The mean ASA values were plotted against the number of partners, and a curve was fitted in a weighted (based on the number of observations at each data point) least-squares manner.

The PDB codes of the proteins containing Trp, with the subunit identifier (if present) and the number of Trp residues in parenthesis, are given below:

131l(3), 153l(3), 193l(6), 1ade(A,4), 1amp(5), 1aoz(A,14), 1arb(7), 1asu(5), 1atl(A,2), 1bam(3), 1bbp(D,5), 1bdm(B,3), 1bec(6), 1bp2(1), 1bri(A,3), 1byb(11), 1ccr(1), 1cel(B,9), 1cew(I,1), 1cfb(6), 1chd(1), 1chm(B,5), 1clc(8), 1cmb(A,2), 1cns(A,6), 1cpc(A,1), 1cse(E,1), 1csh(8), 1cus(1), 1cyo(1), 1daa(A,3), 1ddt(5), 1dsb(A,2), 1dts(5), 1dyr(6), 1eca(1), 1ede(6), 1fba(B,3), 1fkj(1), 1fnc(6), 1gar(A,1), 1gky(1), 1gof(8), 1gox(1), 1gp1(A,2), 1gpb(8), 1gse(B,1), 1han(5), 1hbq(3), 1hle(A,2), 1hpm(1), 1hsl(A,1), 1huw(3), 1hxn(4), 1i1b(1), 1iae(3), 1iisc(A,7), 1knb(4), 1kpt(A,1), 1lcp(A,7), 1lct(3), 1lis(3), 1lts(D,1), 1mls(2), 1mml(3), 1mpp(2), 1mrj(1), 1msc(1), 1nar(4), 1nba(A,4), 1nfp(2), 1nif(3), 1onc(1), 1ora(4), 1oyc(7), 1pbe(6), 1pbn(3), 1pbp(8), 1pda(2), 1pgs(8), 1phg(5), 1pii(3), 1pne(2), 1poc(2), 1ppn(5), 1ptx(1), 1rcf(4), 1rci(2), 1rec(3), 1reg(Y,3), 1rsy(1), 1rva(A,4), 1sac(A,5), 1sat(7), 1sbp(7), 1scs(4), 1slt(A,1), 1sri(A,6), 1tca(5), 1tgx(A,1), 1tml(8), 1ton(4), 1tph(1,5), 1trb(1), 1trk(B,8), 1tsp(6), 1ttb(A,2), 1tys(7), 1udg(7), 1vca(A,1), 1vhh(3), 1wht(A,5; B,5), 1xnb(11), 1xyl(A,6), 1xyz(B,7), 2abk(2), 2acq(6), 2ak3(B,3), 2alp(1), 2ayh(8), 2aza(A,2), 2bbk(H,5; L,2), 2bop(A,1), 2cba(7), 2ccy(A,3), 2chs(A,1), 2cpl(1), 2ctc(7), 2cwg(A,1), 2cyp(7), 2dnj(A,3), 2end(1), 2er7(E,5), 2fal(2), 2gbp(5), 2gdm(2), 2gst(A,4), 2hbg(2), 2hft(4), 2hmz(A,3), 2hpd(A,5), 2hpe(A,1), 2hts(4), 2kau(C,7), 2mnr(6), 2nac(B,6), 2olb(A,13), 2pgd(6), 2phy(1), 2pia(4), 2prd(1), 2prk(2), 2rn2(6), 2scp(A,3), 2sil(6), 2tgi(2), 3bcl(7), 3chy(1), 3cla(3), 3cox(9), 3dfr(4), 3est(7), 3grs(3), 3pga(1,4), 3pte(2), 3rub(L,4), 3sdh(B,2), 3sic(E,3), 3tgl(4), 4blm(A,3), 4enl(5), 4fgf(1), 4fxn(2), 4gcr(4), 5rub(B,5), 8abp(5), 8acn(9), 8fab(C,3; D,5), 8tln(E,3), 9rnt(1).

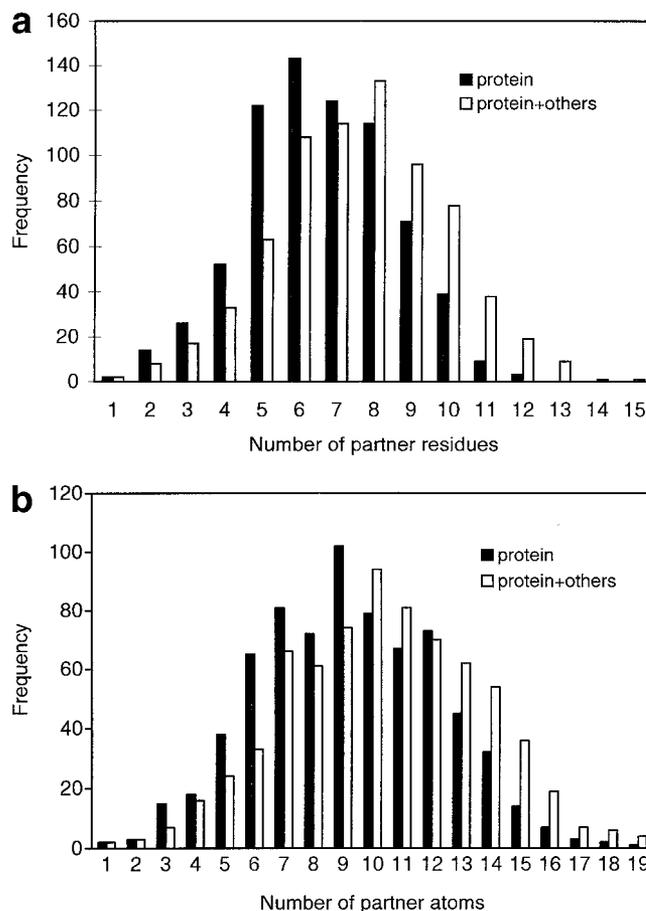


Fig. 2. Histogram of Trp residues having different number of partners, from a consideration of surrounding (a) residues and (b) atoms, from protein as well as non-proteinous molecules.

RESULTS

Number of Partner Residues/Atoms and Their Distribution Around Trp Ring

Considering all types of molecules (protein and non-protein) the maximum number of Trp rings have eight partners, the average of the distribution in Figure 2a being $7.6(\pm 2.3)$. Trp residues with more number of partners (> 10) are likely to have more water molecules (2.3 on average) around them. Restricting only to protein residues the most probable number of partners is six (the peak in the distribution) and the average value is $6.6(\pm 2.0)$. If partners are counted in terms of atoms (Fig. 2b) there are $9.2(\pm 3.0)$ protein atoms and $10.0(\pm 3.3)$ all types of atoms in direct contact with the Trp ring.

The location of partners in the various regions relative to Trp ring is provided in Figure 3. Of the 719 Trp residues, 676 have at least one partner on the α face, 686 on β face, 706 at the edge, and 524 are involved in hydrogen bonding interaction at NE1, thus showing that there are a few Trp rings with a face completely exposed and a significant number with no hydrogen-bonded partner at all. Most of the Trp rings have two protein residues on its face (the order on the basis of the number of partners is $2 > 1 > 3 > 4$)

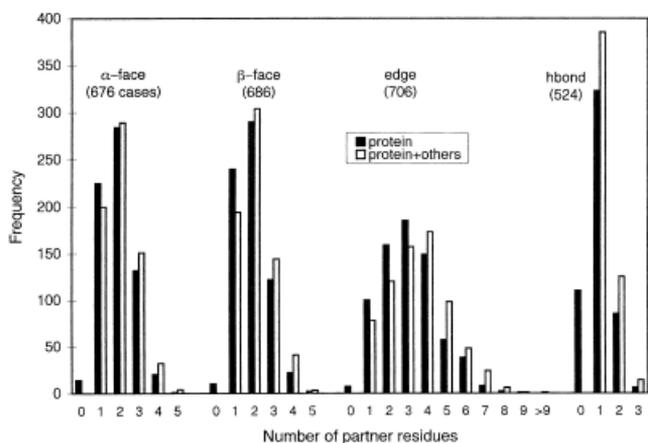


Fig. 3. Histogram of Trp residues having different number of partner residues (considering only proteins, and including other molecules also) on α and β faces, edge and interacting through hydrogen bonding at NE1. The total numbers of Trp residues with at least one partner in the four regions are 676, 686, 706, and 524, respectively.

and 2 to 4 at the edge. Although one hydrogen-bonded neighbor is the most common occurrence, a higher number is not unusual. Fifty-four percent (54%) of Trp residues have water in contact, one having as many as seven molecules (Fig. 4). Water prefers to interact at the edge or through hydrogen-bonding at NE1 rather than be on the face.

We also considered if the same residue can span different regions of the Trp environment (Table I), and about 10% of residues qualify this condition. Among these residues, 73% of the hydrophobic ones (including Ala, Arg, Lys, Asn, and Gln) interact with the edge and a face (α or β), whereas about 56% Asp, Glu, and Ser prefer to hydrogen-bond and bind to the edge or a face as well.

Propensity To Be Partners

The propensities of all amino acid residues to be in Trp environment, as well as bound to faces only, are shown in Figure 5, from which the following observations can be made. (1) Small (Gly, Ala), negatively-charged (Asp, Glu), and polar (Ser, Thr) residues avoid the Trp ring. (2) Between the two structurally similar residues, the more polar one tends to have a lower propensity value—for example, Tyr [1.67 (standard deviation, 9)] and Phe [1.93(9)], Asp [0.70(5)] and Asn [0.85(6)], and Glu [0.68(5)] and Gln [0.97(7)]. (3) Among the positively charged residues Arg is preferred to Lys. (4) Val is neutral to being in the Trp neighborhood, whereas the longer branched-chain residues, Leu and Ile, have higher propensities. (5) As a group the aromatic residues have high propensity values. (6) The overall value (1.07) of Pro increases to 1.36 when only the face-specific interactions are considered. (7) Separate calculations show that some residues have different propensities for the two faces; α and β face propensities for a few representative cases are: Met (2.49, 1.40), Phe (2.11, 1.75), Ile (1.51, 1.22), Arg (1.20, 1.44) and Trp (1.72, 2.23). (8) considering other region-specific interactions, Trp (2.28),

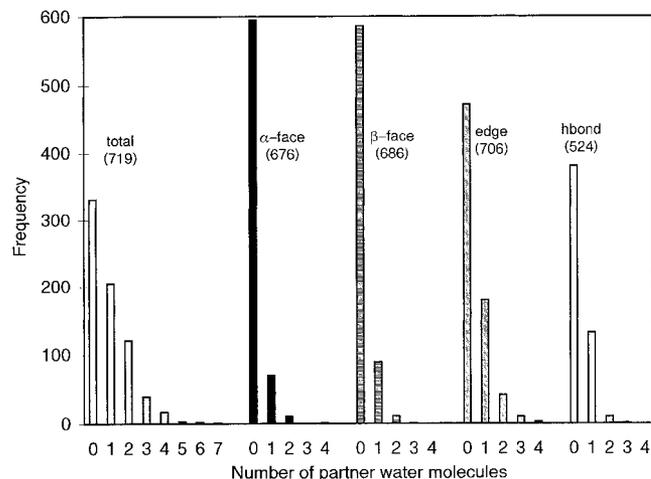


Fig. 4. Histogram indicating Trp residues with different number of water molecules in contact. The five groups correspond to the overall environment and its different regions (α or β face, edge or hydrogen-bonded contacts). In each case the total number of Trp residues are mentioned in the inset (for example, there are 676 Trp residues with at least one partner (protein or non-protein) on the α face).

Tyr (2.10), Phe (1.98), Met (1.74) have very high propensities for the edge, whereas residues with acceptor atoms (and surprisingly, Phe) are found to be engaged through hydrogen bonding [Tyr (2.06), Glu (1.95), Asp (1.73), Phe (1.37), Ser (1.29), Asn (1.28), Gln (1.26)].

Trp Atoms in Contact

The nine heteroaromatic-ring atoms have different number of contacts along the edge and the face (Fig. 6). With no attached proton and being sterically-hindered (Fig. 1), CD2, CE2, and CG rarely interact along the edge. Additionally, the 6-membered ring has more interactions on the face than the 5-membered ring. This may be because of the linking of the indole ring of Trp to the main-chain through CG, which makes the more remote 6-membered ring sterically more accessible. Dougherty's work⁴⁰ on cation— π interaction has indicated that the benzene ring of indole is the preferred cation— π binding site over the 5-membered pyrrole-type ring. Residues at the edge in contact with NE1 are hydrogen-bonded, and NE1, of all the atoms, has the maximum number of contacts in this direction. Interestingly enough however, of all the face-specific contacts it is again NE1 that exhibits the maximum number. This could be due to the electronegative nitrogen atom pulling the electron cloud of the ring towards itself making the nearby area electron-rich and thus providing more stable interaction on the face.⁴¹

Partner Atoms in Contact

A detailed count of all the partner atoms that interact with the face and the edge of Trp ring are provided in Table II, the main-chain contributing 35% of the atoms and side chain, 65%. The distribution of different atom types among the α and β faces and the edge (including hydrogen bonding interaction) is as follows. N: 31, 32, and 37%;

TABLE I. Statistics of the Same Residue Interacting With More Than One Region of Trp Ring[†]

	Combinations of regions						
	a + e	a + h	b + e	b + h	e + h	a + e + h	b + e + h
Number of cases	151	42	168	53	69	4	6

[†]a = α face, b = β face, e = edge, and h = hydrogen bonded. Symmetry-related residues are excluded. There is no residue interacting with both the faces simultaneously. The number of different residues contributing to the table: Gly (12), Ala (22), Pro (13), Ser (20), Cys (6), Met (19), Glu (37), Gln (24), Lys (26), Arg (27), Leu (47), Asp (40), Asn (34), His (8), Phe (31), Tyr (43), Trp (11), Val (22), Ile (25), Thr (26), total (493).

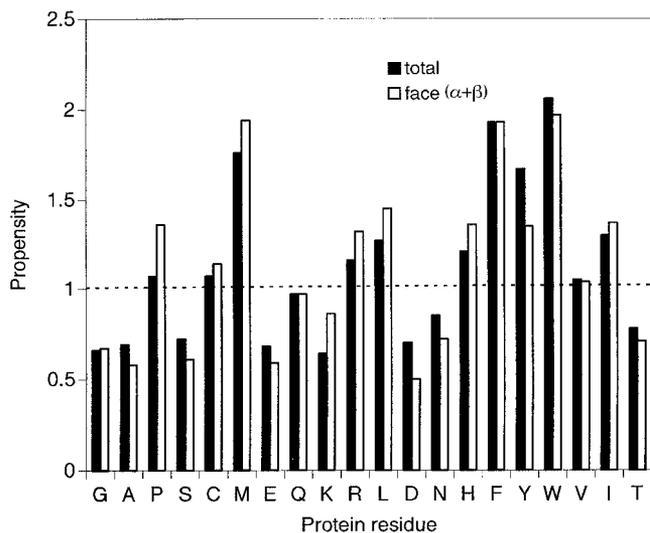


Fig. 5. Propensities of different residues to be in Trp environment and only on Trp face. The dotted horizontal line represents the average propensity (a value of 1). A value greater than (or less than) 1 suggests that the amino acid residue is favored (or disfavored) in the Trp environment.

C: 32, 32, and 35%; O: 14, 13, and 73%; water O: 14, 17, and 69%. Only the oxygen atoms have a distinct preference to occupy the edge region.

The following inferences can be drawn for the atoms interacting with the face. (1) For the branched aliphatic side chains (Val, Leu, and Ile) the terminal atoms have very large number of contacts. This is akin to what was observed for the packing of the adenine-containing cofactors in protein structures.⁸ (2) For most of the other residues, CB has the highest number of contacts. This is true for aromatic residues also, although if the whole aromatic moiety is taken as an entity the total number is greater than this. (3) The bridging atoms, CG (for all aromatics), CZ (Tyr), CD2 and CE2 (Trp), with no proton attached, have very few contacts. (4) Of the two atoms at the γ position, CG2 has more contacts than OG1 in Thr. (5) NZ of Lys has few contacts in comparison to side-chain N atoms of Arg, which may be one of the reasons why Lys is less frequent than Arg in the Trp-binding site (Fig. 5). (6) For S-containing residues the maximum number of contacts is with the atoms next to the S atom. This could be because we have looked for the shortest contact (see Methods), whereby S atoms with greater atomic radius got excluded in comparison with the neighboring C atoms. Consequently, for Met at least, the total number of atoms bound to the $-\text{CH}_2\text{-S-CH}_3$ group may provide a better

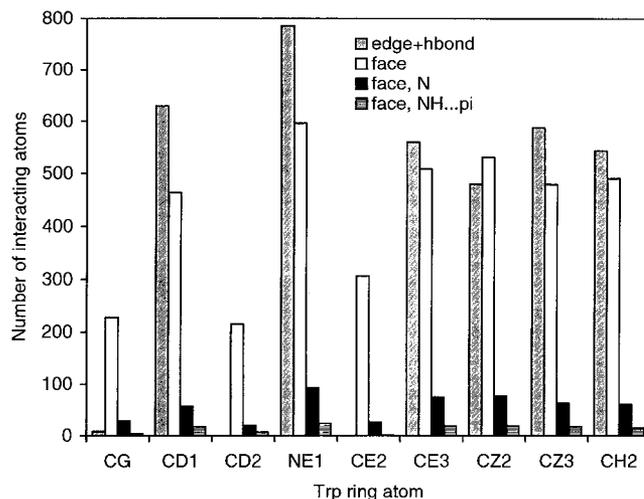


Fig. 6. Distribution showing the contact of partner atoms (protein and non-protein) with different Trp atoms when the interaction is on the face (α and β together) and edge (includes hydrogen-bonded cases also) (1st two bars). The 3rd bar on each atom corresponds to the number of partner N atoms (protein only) on the face (the portion of which exhibiting $\text{NH}\cdots\pi$ interaction is given in the 4th bar).

indication of the stereochemical preference of the S atom towards the aromatic face.

Turning our attention to the edge, (1) main-chain O atoms from various residues show, in general, the highest number of contacts. O atoms from the side chains of Ser, Thr, and Tyr show equal, if not higher number of contacts. For acidic residues the side chain O atoms show the highest number of contacts, whereas for the corresponding amide the side chain as well as the main-chain have quite comparable numbers. (2) For S-containing residues, S, in preference to the other atoms, has a significantly higher number of contacts.

The following statements can be made for the edge vs. face comparison. (1) Water molecules are more abundant near the edge. (2) For Cys and Met if we consider the contacts made by S as well as their bonded neighbors the interaction is more with the face. (3) Except Ala and O-containing side chains (Ser, Thr, Tyr, Asp, and Glu), residues have more interactions with the face than the edge.

Relative Position of the Partner in the Sequence

The sequential difference in the position between Trp and its partner (Fig. 7) suggests that most of the partners are beyond nine residues (69% α , 67% β , 72% edge and 79% hydrogen-bonded residues). In other

TABLE II. Statistics on the Interaction of the Various Atoms of the Partner Residues With the Face and the Edge of Trp Ring[†]

Residue	Total Number	N	CA	C	O	CB	CG		CD		CE			CZ	CZ3	CH2
							OG1	OG	SD	OD1	OD2	OE1	OE2			
(a) Face																
Gly	181	35	76	28	42											
Ala	154	23	12	4	20	95										
Val	237	10	7	7	18	11	84	100								
Leu	405	28	11	4	30	85	13		115	119						
Ile	241	9	4	3	15	16	31	77	86							
Ser	134	28	14	14	20	39	19									
Thr	139	12	9	6	21	8	22	61								
Cys	52	9	2	1	6	23	11									
Met	122	4	8	2	8	17	20		5		58					
Pro	221	25	11	3	18	52	46		66							
Phe	244	8	6	5	11	44	4		34	25	42	31		34		
Tyr	180	10	2	9	16	46	3		21	22	15	15		2	19	
Trp	99	4	3	1	7	22	1		8	1	9		12	5	14	12
His	116	14	6	5	10	26	2		9	22	14	8				
Arg	286	20	14	9	19	41	34		42		26			11	35	35
Lys	215	27	10	13	17	46	30		29		31			12		
Asp	107	13	11	5	21	29	2		10	16						
Asn	148	11	8	4	19	39	3		25	39						
Glu	118	17	7	6	14	18	26		4		12	14				
Gln	158	12	5	6	17	21	35		3		33	26				
HOH	210															
Others	54															
(b) Edge																
Gly	140 (39)	18 (5)	19	13	90 (28)											
Ala	205 (24)	21 (1)	10	12	82 (16)	80										
Val	204 (8)	13 (1)	3	5	64 (9)	6	59	54								
Leu	265 (41)	18 (2)	5	10	81 (27)	25	6		59	61						
Ile	187 (19)	15 (2)	1	6	47 (10)	8	23	39	48							
Ser	163 (54)	11 (4)	5	10	60 (18)	13	64 (26)									
Thr	170 (42)	7 (2)	4	10	42 (7)	8	61 (22)	38								
Cys	40 (6)	4	1	2	9 (2)	7	17 (3)									
Met	91 (9)	6	1	6	12 (5)	9	8		29 (4)		20					
Pro	101 (21)	11 (2)	1	2	45 (12)	14	15		13							
Phe	227 (31)	9 (1)	4	6	51 (14)	23	6		22	22	28	33		23		
Tyr	248 (45)	12	4	5	41 (8)	25	7		17	20	24	25		7	61 (20)	
Trp	94 (8)	3 (1)	1	2	17 (2)	6	3		5	6	6 (1)	1	8	10	10	16
His	64 (13)		1	1	13 (3)	12	2		5 (3)	15	8	7 (2)				
Arg	137 (17)	14 (1)	3	6	40 (11)	11	14		12		5			4	14	14 (1)
Lys	94 (12)	9	4	8	41 (7)	8	11		6		3			4 (1)		
Asp	198 (71)	10 (1)	5	6	43 (4)	20	8		48 (20)	56 (31)						
Asn	162 (38)	12 (1)	3	3	53 (15)	18	2		43 (11)	28 (7)						
Glu	167 (75)	4	4	3	36 (6)	16	15		5		34 (20)	50 (29)				
Gln	126 (31)	4	2	2	34 (5)	17	5		1		39 (18)	22 (2)				
HOH	459 (155)															
Others	54															

[†]Residues that have been named Asx and Glx in PDB files have been included under Asp and Glu respectively; the count of such residues is 5 and 2 on the face, and 2 and 0 at the edge. In only one case the carboxy-terminal atom, OXT (of a Lys residue), interacts; this has been entered under the atom O of Lys (Edge). For the edge residues the total number of atoms and only the electronegative atoms (that can form hydrogen bond) in contact with NE1 are given in parentheses. The total number of N atoms in the table is 837 (521 from the main chain and 316 from the side chain). The corresponding number for O atoms is 1900 (1251 + 649), and for C atoms, 3955 (563 + 3392).

words, the interactions are long range. Considering the relatively rare short range contacts, the edge residues have a peak at two which falls off progressively as one moves to nine. The residues on the β face have peaks at positions ± 2 and -4 . Overall there are 348 cases with difference of ± 2 , of which in 116 cases (33%) Trp and its

partner are on the same β -strand, whereas for 282 cases with ± 4 difference 128 (45%) have them on α -helices. This shows that while short-range interactions are not especially prominent, when present they can reflect the location of Trp and its partner in α -helices or β -strands.

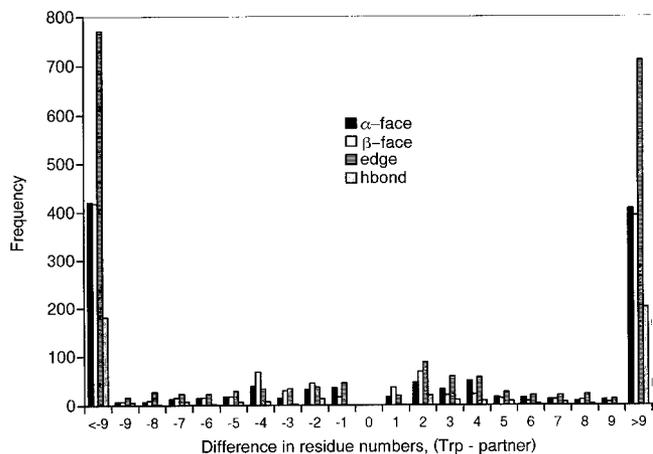


Fig. 7. Distribution of the relative position of Trp and its partner along the polypeptide chain (excluding symmetry-related cases).

Hydrogen Bonding

The hydrogen bond interaction involving the indole NH group of Trp and an acceptor (nitrogen, oxygen, or sulfur) atom was studied by Ippolito et al.,⁴² who found a distance of $3.0(\pm 0.2)$ Å. Here we have considered all edge-atoms that are in contact with NE1 as hydrogen bonded, and a few nonpolar atoms are also in the list. Whether these are due to the presence of some unconventional hydrogen bonding has not been examined. The average distances for 785 partners (protein + others) at NE1 along the hydrogen bond direction are: $3.1(\pm 0.3)$ Å for 572 O atoms (of which 155 are water), $3.6(\pm 0.4)$ Å for 41 N atoms, $3.7(\pm 0.2)$ Å for 165 C atoms and $3.6(\pm 0.1)$ Å for 7 S atoms. The distribution of hydrogen-bonded oxygen atoms are shown in Figure 8.

Stacking Versus NH— π Interaction

Although the methods employed to identify N atoms on the face to the category of stacked or NH— π interaction are to some extent different from those of Mitchell et al.³⁷ for the interaction of the amino group with the aromatic ring of Phe and Tyr, the results are qualitatively similar (Table III). While the earlier study found that in 10% of the interactions the sp^2 hybridized N atoms are positioned above the ring and of these instances those showing stacked interaction outnumber those with NH— π interaction by 2.5:1. In our study 514 cases of N atoms (including 12 NZ of Lys), out of a total of 815, are on the face (63%), and the stacked geometry is favored over the amino—aromatic hydrogen bond by around 2:1. Interestingly however, N atom on the face distinctly prefers to be positioned over NE1 (Fig. 6).

Variation of the Accessible Surface Area

We calculated the average (and the standard deviation) of the solvent-accessible surface area (ASA) for all Trp residues having a given number of protein residues around them. Although our thrust has been to identify the residues around the aromatic part of Trp, for this calculation

we also defined the number of partners considering the whole Trp residue. Results for the whole residue and aromatic part are presented in Figure 9 (a and b). The standard deviations of the mean values decrease with the increasing number of partners. This is because a partner residue can contribute one or more atoms for binding Trp. This causes a large disparity between ASA values when the number of partners is small. As the number increases the total number of contributed atoms reaches the saturation point and the ASA values of Trp residues with a given number of partners are nearly identical, resulting in smaller standard deviations.

ASAs decrease with the increase in the number of partners and one can adequately represent the variation in an exponential form (Fig. 9). Extrapolation to $x = 0$ (i.e., no partner) gives a value of 246.6 \AA^2 for the whole residue and 189.6 \AA^2 for the aromatic part, in excellent agreement with the values obtained for a Trp-containing tripeptide in an extended conformation (Table IV).

We wondered if we could have plotted the ASA values for the whole residue (“whole ASA”) against the number of partners calculated using the aromatic part (“aromatic partner”). This gives a curve (Fig. 9c, symbol \blacktriangle) which is closer to the “aromatic ASA” — “aromatic partner” plot (symbol \blacksquare), rather than the “whole ASA” — “whole partner” plot (symbol \bullet), and the limiting ASA value (when there is no partner) of 221.4 \AA^2 is considerably smaller than the value for the whole residue (Table IV). This suggests that a meaningful result is obtained by correlating the ASA and the number of partners both defined for the same fragment.

DISCUSSION

There have been attempts to analyze the microenvironment surrounding specific protein sites, like calcium-binding, or serine protease active site.⁴³ Similarly, studies have been conducted to measure and understand residue associations by identifying the surrounding residues around a given one.^{44–47} Here we have not only analyzed residues that are in contact with Trp rings (Fig. 1), but also their spatial relationship, as well as the atoms involved and the effect of the local neighborhood on the accessibility of Trp. Results that emanate are put into proper perspective in relation to other known stabilizing interactions.

General Features of the Environment and Trp-Binding Propensities

Trp prefers to have six protein residues around it, although the peak shifts to eight when non-protein residues are also included; the corresponding number for the atoms in contact are nine and ten, respectively (Fig. 2). The most likely number of protein residues to be found on a face is two, three at the edge (which increases to four on including non-protein residues) and one hydrogen-bonded (Fig. 3).

As expected different residues have different propensities to bind Trp (Fig. 5), but interestingly, there are some trends based on size and polarity. For example, among the aromatic residues, the more polar Tyr and His have lower

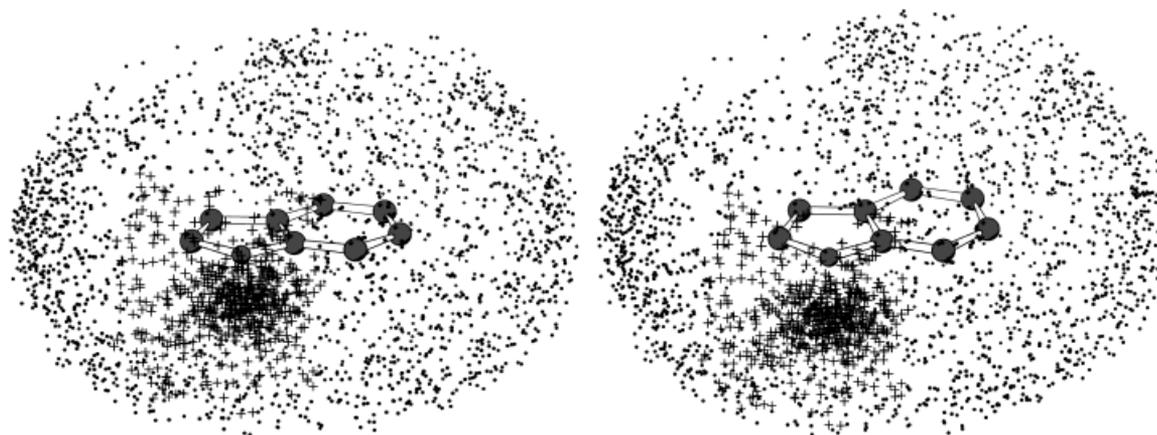


Fig. 8. Scatterplot (in stereo) of superimposed oxygen atoms (protein + others) bound to Trp; those hydrogen-bonded are indicated by the plus sign.

TABLE III. Interaction of the N-Containing Side Chain (by Residue) and Main Chain With Trp Face

Residue	On the face	NH- π	Stacked	Others ^a
Arg	96	21	52	23
Asn	38	11	18	9
Gln	26	8	14	4
His	16	0	12	4
Trp	9	4	2	3
All side-chain	185	44	98	43
All main-chain	317	81	151	85

^aThose not belonging to the categories in the previous two columns.

values. Among the isostructural pair of residues (Ser/Cys, Asp/Asn, Glu/Gln) the less polar residue has the higher value. Considering the branched aliphatic residues, Ile and Leu have high values, whereas the shorter Val is almost neutral towards Trp-binding. Small (Gly and Ala) and acidic residues avoid Trp, although, if one considers only the hydrogen-bonded interaction at NE1, the latter residues along with Tyr and Ser have high values. The dependence of the propensity value on the residue size prompted us to find out the correlation between them, and indeed if the four charged residues are excluded the correlation coefficient is very high (0.88) (Fig. 10). It is interesting that the largest of the amino acid residues, Trp, which usually have long range interactions (Fig. 7), has a tendency to contact residues commensurate to their size. Some residues have different propensities depending on the region of the Trp environment, a notable example being Pro, which is more likely to be found on the face than the edge.

Hydrogen Bonding and Other Contact Features

As in the case of adenine, where hydrogen-bonding with protein residues is not the dominant motif of binding,⁸ 27% of Trp residues (195 out of 719) are without any hydrogen-bond partner (Fig. 4). Most hydrogen bonds in globular proteins are local, i.e., between partners that are close in sequence.⁴⁸ However, NE1 of Trp forms bonds with groups

that are nine residues or more away from it (Fig. 7). Other types of contacts are also long-range, thus pointing to the importance of Trp in the stabilization of the tertiary structure. 3.2% contacts (150 out of 4,728 protein residues) involve symmetry-related residues indicating that Trp is a good interfacial residue that can mediate interaction between two molecules or protein subunits.

How Hydrophobic Is Tryptophan?

This was the question asked by Fauchere⁴⁹ because Trp ranks as one of the most hydrophobic amino acids on the basis of its partitioning into polar solvents such as octanol,⁵⁰ whereas scales based on partitioning into nonpolar solvents like cyclohexane^{51–52} rank it as only intermediate in hydrophobicity. As we have a count of different types of atoms in contact with Trp, the ratio of the number of electronegative atoms (O & N) and C atoms to the total number of atoms offers an estimate of the polar and nonpolar characteristics of Trp ring. Excluding S atoms (which can not conclusively be termed as polar or nonpolar) from the total number, the values are 0.46 and 0.54, respectively, which suggest an ambivalent nature of the Trp ring. It should, however, be mentioned that these values are obtained subject to the conditions of our methodology which, for example, does not include the whole Trp residue and considers only one atom of a stacked aromatic ring as the partner atom even though its adjacent atom may also be within four Å (although examples of aromatic rings stacked against Trp are very small⁵³). Consequently, the number of nonpolar atoms is underestimated to some extent in our analysis. Nevertheless, this offers a method, qualitatively similar to the one based on surrounding hydrophobicity,⁵⁴ for measuring the amphipathic characteristics of any amino acid residue.

Weak Interactions and Stability of Protein Structures

Although hydrogen-bonding is the most easily identifiable feature of all secondary structures, it may not be

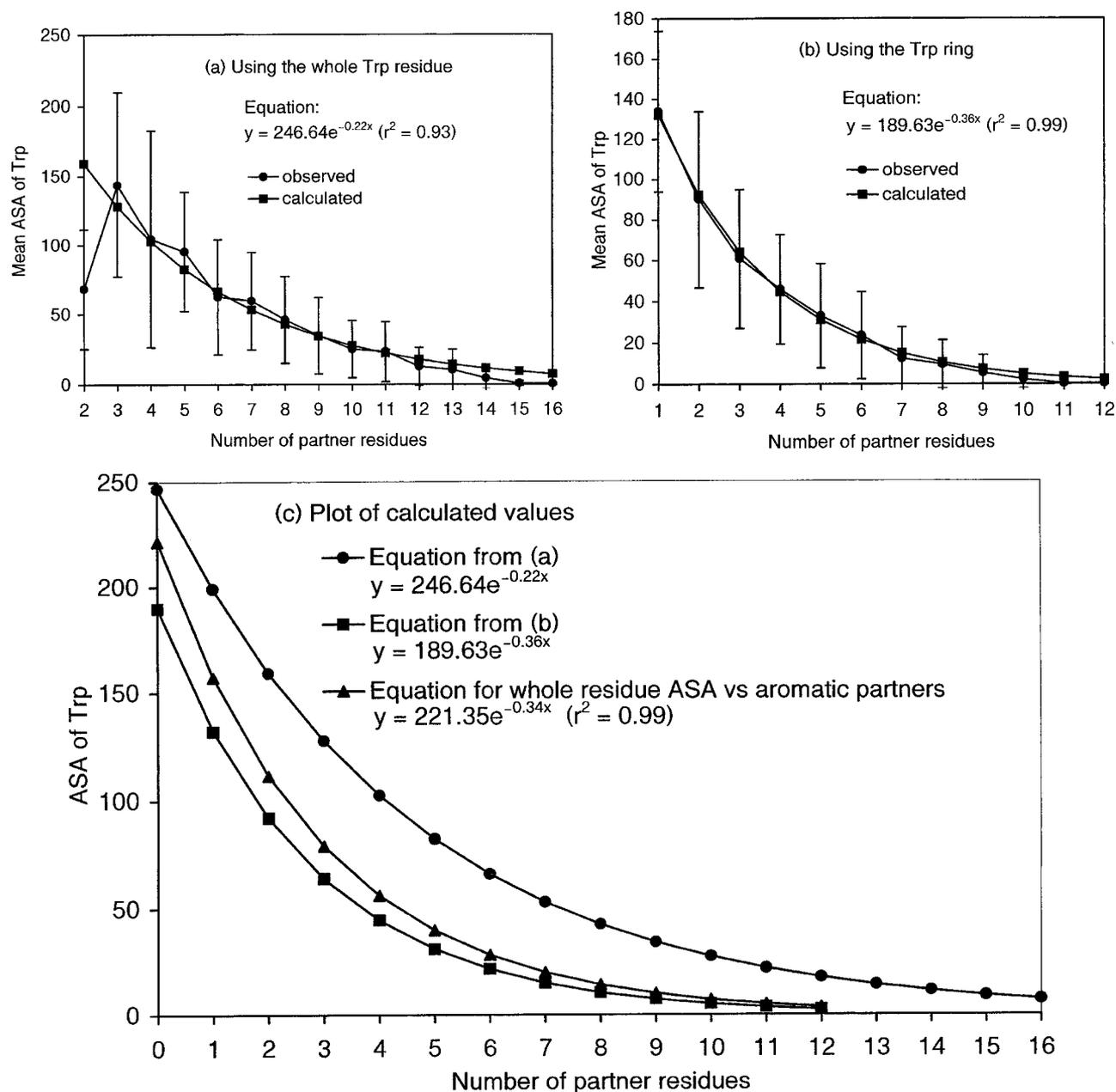


Fig. 9. Variation of the mean accessible surface areas (\AA^2), (ASA) (standard deviations given as bars) with the number of partner residues considering (a) the whole Trp residue, and (b) the aromatic Trp ring only. There are no cases with zero number of partners, and also with one in (a). The equations for the best least-squares fit are provided. In (c) are plotted

the calculated values, using equations in (a) and (b), of ASA corresponding to a given number of partners; the 3rd set of values are from the equation obtained if the ASA for the whole residue is plotted against the number of partners corresponding to the aromatic ring.

the dominant folding force.⁵⁵ Various other weak interactions have been identified, and those prevalent in the Trp environment are aromatic–aromatic,^{56–57} CH– π ,^{58,8} S–aromatic,^{59–61} OH– π and NH– π ,^{41,62–63,37} and CH–O.^{64–66}

Aromatic–aromatic interaction

As a group the aromatic residues provide the largest number of partners to the Trp environment. Trp can pack

another aromatic residue in a stacked face-to-face or perpendicular face-to-edge fashion, and other arrangements between these two limiting orientations. A detailed analysis of the packing geometry is published separately,⁵⁴ but two points are worth mentioning here. His has a higher propensity to interact with Trp face (Fig. 5), and indeed more than the expected number of His–Trp pairs are found in the face-to-face orientation; for the Phe–Trp pair on the other hand, the Phe-edge interacting

TABLE IV. Accessible Surface Area (ASA) for Trp in Model Peptides^a and Calculated^b in Absence of Any Partner

	ASA (Å ²)	
	Whole residue	Aromatic part
Gly-Trp-Gly	267.9	186.1
Ala-Trp-Ala	254.6	186.1
Trp, calculated	246.6	189.6

^aIn extended conformation.^bUsing the equations in Figure 9.

with the Trp-face is more favorable than the interaction between the Phe-face and Trp-edge.

CH— π interaction

In a survey of the interactions involving the adenine face it was found that the branched side chains have a preponderance and the contact usually involves a N atom on the ring.⁸ Ab initio calculations have shown that the binding energy is the highest when a C—H group is placed on an aromatic face close to (but not exactly on top of) a ring N atom,⁴¹ although there may be some argument as to the exact nature of the force involved.⁶⁷ For Trp if we consider the ring atoms involved in the face-specific interaction, NE1 dominates (Fig. 6). Moreover, as in the case of adenine the branched end of the side chains of Val, Leu, Ile, and Thr (CG2 atom only) have a large number of contacts (Table II). Additionally, the CB atom of most other residues has a value higher than any other atom in the residue. For Gly, CA has the maximum number, whereas for Pro, CG and CD also make large contributions in addition to CB. As to the reason why CB is so highly used, the CH— π interaction can contribute significantly to ΔG when there is minimum entropic loss.⁸ So CB, which gets fixed with the main-chain conformation and is unchanged by the rotation of the side chain, is preferred. By the same token CA of Gly and the ring atoms of the conformationally rigid pyrrolidine ring are also involved. As these atoms are already rather ordered, there is no significant loss in conformational entropy when they are engaged in CH— π interaction.

Interestingly, even among the aromatic residues CB has the highest number of contacts. Although the hydrophobic effect is believed to manifest itself by reducing the exposure of nonpolar surface area to the solvent, the fact that specific atom—atom contacts are maintained while doing so suggests that enthalpic factors also contribute in addition to entropy⁵⁵ to what constitute hydrophobic forces. Moreover the face of a Trp ring can even interact with a polar side chain by engaging the CB group of the latter through CH— π interaction (Fig. 11). Though weak,⁴¹ being numerous, these interactions can contribute significantly to the stability of the folded structure.

NH— π and OH— π interactions

Mitchell et al.³⁷ found that 26 out of 55 protein chains (47%) contain at least one NH— π interaction involving the aromatic ring of Phe and Tyr, although overall they are rare: stacked geometry is favored over the NH— π interac-

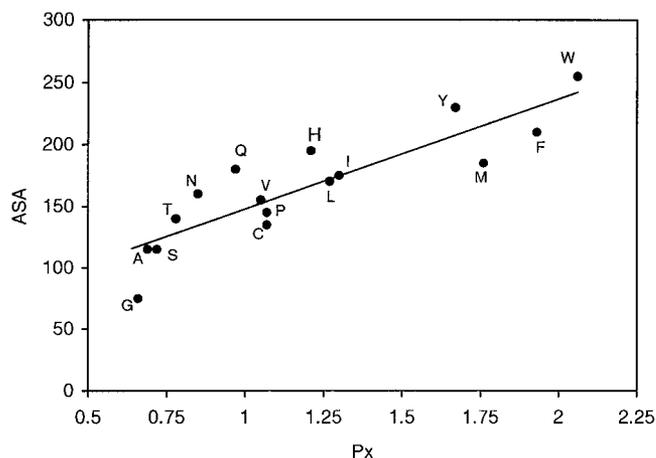


Fig. 10. Propensities, P_x , of different residues (excluding Asp, Glu, Arg, and Lys) to bind Trp plotted against their accessible surface areas (Å²), ASA, in a tripeptide Gly-X-Gly in an extended conformation.¹ The equation of the least-squares line is: $ASA = 89.2 P_x + 58.8$ (with $r^2 = 0.77$).

tion in the ratio 2.5:1. When the aromatic ring is from a Trp residue, the NH— π interaction is exhibited by 79 out of 180 protein chains (44%) (Table III), and the trend of parallel orientations of the planes being preferred over the NH— π bond is maintained, although to a lesser degree (2:1, the ratio is, however, 1.9:1 when the N atom is from the main chain). There is, however, one remarkable difference. A distinct preference of the N atom to be positioned over the N atom of the Trp ring is observed, suggesting that the NH— π interaction may be more stable when the NH group is directed towards NE1. It may be relevant here to point out that the NH— π interaction may lead to anomalous (upfield-shifted) NMR chemical shift of the partner NH proton,⁶⁸ and such an interaction in conjunction with other weak forces may very well be important in molecular recognition in specific cases (Fig. 11).

Table II shows that there are 270 examples of the side-chain hydroxyls of Ser, Thr, and Tyr, and water to be on the face (however, unlike the NH— π interaction no preference for the OH group to be positioned over NE1 is observed). These may constitute OH— π interaction (although in absence of any knowledge of the H position it can not definitely be said that the O—H group is directed towards the Trp π face). It is to be noted, however, that these groups are found more along the edge than the face (in the ratio 2.4:1). Again, of the two groups at the branched position of Thr side chain, the methyl group outnumbers the hydroxyl group by 2.8:1 in their interaction with the face. All these suggest that although energetically stable⁴¹ the use of the OH— π interaction is not quite striking in protein structures.

S—aromatic interaction

In a recent paper dealing with the stabilizing interactions involving Cys residues it was found that the free sulfhydryl group prefers to interact with the face rather than the edge of aromatic rings.⁶¹ From the data presented

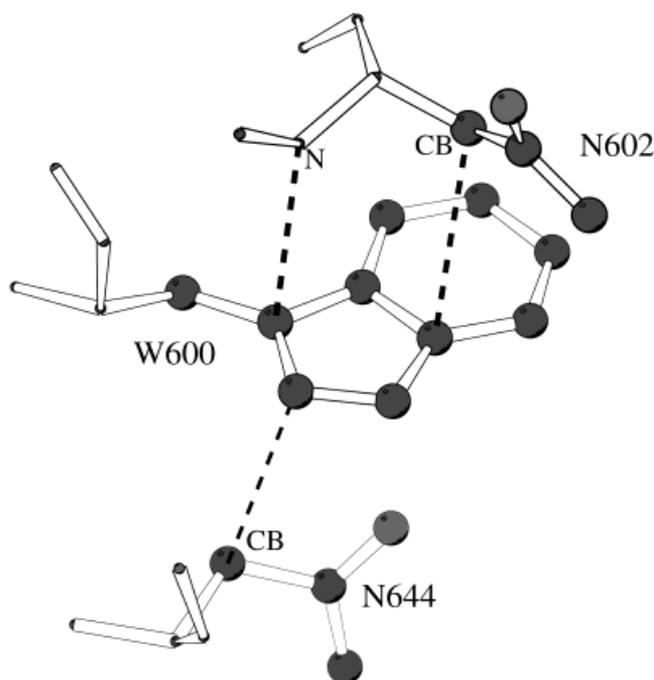


Fig. 11. Diagram showing the location of residues Asn602 on the β face, and Asn644 on the α face of Trp600 in the B-subunit of the structure, 1XYZ (the side-chain atoms are indicated by balls). The first residue has two interactions, CB...CE2 (3.69 Å) and N...CG (3.77 Å), and the second has one, CB...CD1 (3.83 Å). (Figures 8 and 11 were made using MOLSCRIPT⁷²).

in Table II only 26% of S contacts are with the face. However, as has been pointed out in Results, while considering S atoms it would be more appropriate if the count of the bonded neighbors (CB of Cys, and CG and CE of Met) are also included. When this is done, 59% of the contacts are face-specific.

CH—O interaction

The oxygen—aromatic interaction in proteins was first described by Thomas et al.,⁶⁹ who on analyzing the atomic environments of the Phe aromatic rings found that there is a statistically significant preference for the oxygen atoms to be found in the aromatic plane near the H atoms. An examination of the highly accurate small-molecule structures containing Phe led Gould et al.⁷⁰ to come to the same conclusion that the location of the oxygen atoms in the periphery of aromatic ring is stabilizing. Likewise, Flanagan et al.⁷¹ found that the dominant interaction of solvent molecules is with the edge and not with the face of the Phe ring. These interactions are the manifestation of what is now termed as CH—O hydrogen bond.⁶⁴ In the case of Trp ring also there is a preferential distribution of oxygen atoms all around the edge (Fig. 8). The possibility of the formation of CH—O hydrogen bonds along with the normal NH—O hydrogen bond at NE1 makes the edge of Trp suitable for interaction with solvents or other polar molecules.

Accessibility as a Function of the Number of Partners

We enquired if it is possible to have an estimate of the accessible surface area (ASA) of a Trp residue having a given number of protein residues (and vice versa) in its environment. With this aim we calculated the ASA of Trp residues and their mean after grouping them on the basis of the number of partners. One can nicely fit exponential curves passing through the mean values of ASAs calculated for the “whole” Trp residue or just the “aromatic” part (Fig 9a, b). Although the ASA values when there is no partner around were not included while deriving the exponential equations, these can be obtained (247 and 190 Å² for the whole and the aromatic, respectively) by putting $x = 0$ in the equations. One can also calculate the ASAs (Table IV) for Trp residues in peptide fragments, Gly—Trp—Gly and Ala—Trp—Ala modelled in an extended conformation so that the adjacent residues have the minimum contact with the Trp residue in the center. Although the ASA of the central residue in these peptides is, in general, assumed to represent the ASA of the residue in the unfolded state, it is to be noted that there is some variation depending on the type of the adjacent residue. The value, 247 Å², we obtain by extrapolation is the average over all types of flanking residues in the data base, and may be a better approximation of the unfolded state.

As the equations provide expected ASA values (whole/aromatic) for a Trp residue with a given number of partners, a comparison with the observed values will give an indication of the efficiency of packing of partner residues around Trp. If the observed ASA is below the curve in Figure 9a, it would mean that a larger surface area of the Trp residue has been covered by its partners and the packing/binding of Trp by the partners is likely to be stronger (and vice versa). As an extension of this work we are now studying if the exponential dependence of ASA on the number of partners is true for any residue type. In that case we will have 20 analytical expressions providing the expected ASA at any given number of partners for all the 20 amino acid residues. These can then be used to gauge the packing of various residues against each other in protein structures.

CONCLUSION

In conclusion, we have carried out a comprehensive analysis of the binding of Trp rings in protein structures, the types of residues and their constituent atoms interacting with different regions and atoms of the ring. The nature of the surrounding atoms provide an estimate of the amphipathic character of Trp. Because of the large number of partners involved, which are usually quite remote in the sequence, Trp has an important role in stabilizing the tertiary structure. Stereospecific interactions observed in Trp environment have relevance in the understanding of protein folding. The accessible surface area of Trp decreases exponentially with the number of residues around it, and this relationship provides a way to assess the efficiency of packing around any Trp residue.

ACKNOWLEDGMENTS

The authors would like to thankfully acknowledge the Department of Science and Technology for a grant, the Council of Scientific and Industrial Research for fellowships, and the Bioinformatics Center for the use of its facilities.

REFERENCES

- Chothia C. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* 1976;105:1–14.
- Heringa J, Argos P. Side-chain clusters in protein structures and their role in protein folding. *J Mol Biol* 1991;220:151–171.
- McIntire WS, Wemmer DE, Chistoserdov A, Lidstrom ME. A new cofactor in a prokaryotic enzyme: tryptophan tryptophylquinone as the redox prosthetic group in methylamine dehydrogenase. *Science* 1991;252:817–824.
- Chen L, Matthews FS, Davidson VL, Huizinga EG, Vellieux FMD, Hol WGJ. Three-dimensional structure of quinoprotein methylamine dehydrogenase from *P. denitrificans* determined by molecular replacement at 2.8 Å resolution. *Proteins* 1992;14:288–299.
- Prince RC, George GN. Tryptophan radicals. *Trends Biochem Sci* 1990;15:170–172.
- Stayton PS, Sligar SG. Structural microheterogeneity of a tryptophan residue required for efficient biological electron transfer between putidaredoxin and cytochrome P-450_{cam}. *Biochemistry* 1991;30:1845–1851.
- Witt H, Malatesta F, Nicoletti F, Brunori M, Ludwig B. Tryptophan 121 of subunit II is the electron entry site to cytochrome-c oxidase in *Paracoccus denitrificans*. Involvement of a hydrophobic patch in the docking reaction. *J Biol Chem* 1998;273:5132–5136.
- Chakrabarti P, Samanta U. CH/π interaction in the packing of adenine ring in protein structures. *J Mol Biol* 1995;251:9–14.
- Joachimiak A, Haran TE, Sigler PB. Mutagenesis supports water mediated recognition in the *trp* repressor-operator system. *EMBO J* 1994;13:367–372.
- Komeiji Y, Fujita I, Honda N, Tsutsui M, Tamura T, Yamato I. Glycine 85 of *trp*-repressor of *E. coli* is important in forming the hydrophobic tryptophan binding pocket: experimental and computational approaches. *Protein Eng* 1994;7:1239–1247.
- Antson AA, Otridge J, Brzozowski AM, et al. The structure of *trp* RNA-binding attenuation protein. *Nature* 1995;374:693–700.
- Elgavish S, Shaanan B. Lectin-carbohydrate interactions: different folds, common recognition principles. *Trends Biochem Sci* 1997;22:462–467.
- Hazes B. The (QxW)₃ domain: a flexible lectin scaffold. *Protein Sci* 1996;5:1490–1501.
- Anthony C, Ghosh M, Blake CCF. The structure and function of methanol dehydrogenase and related quinoproteins containing pyrrolo-quinoline quinone. *Biochem J* 1994;304:665–674.
- Bazan JF. Structural design and molecular evolution of cytokine receptor superfamily. *Proc Natl Acad Sci USA* 1990;87:6934–6938.
- Harel M, Kleywegt GJ, Ravelli RBG, Silman I, Sussman JL. Crystal structure of an acetylcholinesterase-fasciculin complex: interaction of a three-fingered toxin from snake venom with its target. *Structure* 1995;3:1355–1366.
- Katz BA. Binding to protein targets of peptidic leads discovered by phage display: crystal structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence. *Biochemistry* 1995;34:15421–15429.
- Yengo CM, Fagnant PM, Chrin L, Rovner AS, Berger CL. Smooth muscle myosin mutants containing a single tryptophan reveal molecular interactions at the actin-binding interface. *Proc Natl Acad Sci USA* 1998;95:12944–12949.
- Zeltins A, Schrepf H. Specific interaction of the *Streptomyces* chitin-binding protein CHB1 with alpha-chitin—the role of individual tryptophan residues. *Eur J Biochem* 1997;246:557–564.
- Staub O, Rotin D. WW domains. *Structure* 1996;4:495–499.
- Chang Y, Zajicek J, Castellino FJ. Role of tryptophan-63 of the kringle 2 domain of tissue-type plasminogen activator in its thermal stability, folding, and ligand binding properties. *Biochemistry* 1997;36:7652–7663.
- Perraut C, Clottes E, Leydier C, Vial C, Marcillat O. Role of quaternary structure in muscle creatine kinase stability: tryptophan 210 is important for dimer cohesion. *Proteins* 1998;32:43–51.
- Skoging U, Liljestrom P. Role of the C-terminal tryptophan residue for the structure-function of the alphavirus capsid protein. *J Mol Biol* 1998;279:865–872.
- Jonasson P, Aronsson G, Carlsson U, Jonsson BH. Tertiary structure formation at specific tryptophan side chains in the refolding of human carbonic anhydrase II. *Biochemistry* 1997;36:5142–5148.
- Matthews JM, Ward LD, Hammacher A, Norton RS, Simpson RJ. Roles of histidine 31 and tryptophan 34 in the structure, self-association, and folding of murine interleukin-6. *Biochemistry* 1997;36:6187–6196.
- Schiffer M, Chang C-H, Stevens FJ. The functions of tryptophan residues in membrane proteins. *Protein Eng* 1992;5:213–214.
- Landolt-Marticorena C, Williams KA, Deber CM, Reithmeier RAF. Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. *J Mol Biol* 1993;229:602–608.
- Doyle DA, Wallace BA. Crystal structure of the gramicidin/potassium thiocyanate complex. *J Mol Biol* 1997;266:963–977.
- Hu W, Cross TA. Tryptophan hydrogen bonding and electric dipole moments: functional roles in the gramicidin channel and implications for membrane proteins. *Biochemistry* 1995;34:14147–14155.
- Callis PR. 1La and 1Lb transitions of tryptophan: applications of theory and experimental observations to fluorescence of proteins. *Methods Enzymol* 1997;278:113–150.
- Chattopadhyay A, Mukherjee S, Rukmini R, Rawat SS, Sudha S. Ionization, partitioning, and dynamics of tryptophan octyl ester: implications for membrane-bound tryptophan residues. *Biophys J* 1997;73:839–849.
- Dahms TES, Willis KJ, Szabo AG. Conformational heterogeneity of tryptophan in a protein crystal. *J Am Chem Soc* 1995;117:2321–2326.
- Sussman JL, Lin D, Jiang J, Manning NO, Prilusky J, Ritter O, Abola EE. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr Sect D* 1998;54:1078–1084.
- Hobohm U, Sander C. Enlarged representative set of protein structures. *Protein Sci* 1994;3:522–524.
- Rose IA, Hanson KR, Wilkinson KD, Wimmer MJ. A suggestion for naming faces of ring compounds. *Proc Natl Acad Sci USA* 1980;77:2439–2441.
- Williams RW, Chang A, Juretic D, Loughran S. Secondary structure predictions and medium range interactions. *Biochim Biophys Acta* 1987;916:200–204.
- Mitchell JBO, Nandi CL, McDonald IK, Thornton JM. Amino/aromatic interactions in proteins: is the evidence stacked against hydrogen bonding? *J Mol Biol* 1994;239:315–331.
- Hubbard SJ. ACCESS, a program for calculating accessibilities. Department of Biochemistry and Molecular Biology, University College London; 1991.
- Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 1971;55:379–400.
- Ma JC, Dougherty DA. The cation—π interaction. *Chem Rev* 1997;77:1303–1324.
- Samanta U, Chakrabarti P, Chandrasekhar J. Ab initio study of energetics of X—H...π (X = N, O, and C) interactions involving a heteroaromatic ring. *J Phys Chem A* 1998;102:8964–8969.
- Ippolito JA, Alexander RS, Christianson DW. Hydrogen bond stereochemistry in protein structure and function. *J Mol Biol* 1990;215:457–471.
- Bagley SC, Altman RB. Characterizing the microenvironment surrounding protein sites. *Protein Sci* 1995;4:622–635.
- Karlin S, Zuker M, Brocchieri L. Measuring residue associations in protein structures: possible implications for protein folding. *J Mol Biol* 1994;239:227–248.
- Sippl MJ. Calculation of conformational ensembles from potentials of mean force: an approach to the knowledge-based prediction of local structures in globular proteins. *J Mol Biol* 1990;213:859–883.
- Kocher J-PA, Rooman MJ, Wodak SJ. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J Mol Biol* 1994;235:1598–1613.
- Bahar I, Jernigan RL. Inter-residue potentials in globular pro-

- teins and the dominance of highly specific hydrophilic interactions at close separation. *J Mol Biol* 1997;266:195–214.
48. Stickle DF, Presta LG, Dill KA, Rose GD. Hydrogen bonding in globular proteins. *J Mol Biol* 1992;226:1143–1159.
 49. Fauchere JL. How hydrophobic is tryptophan? *Trends Biochem Sci* 1985;10:268.
 50. Fauchere JL, Pliska V. Hydrophobic parameters π of amino-acid side chains from the partitioning of *N*-acetyl-amino-acid amides. *Eur J Med Chem* 1983;18:369–375.
 51. Wolfenden RV, Cullis PM, Southgate CCF. Water, protein folding, and the genetic code. *Science* 1979;206:575–577.
 52. Radzicka A, Wolfenden R. Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapour phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* 1988;27:1664–1670.
 53. Samanta U, Pal D, Chakrabarti P. Packing of aromatic rings against tryptophan residues in proteins. *Acta Crystallogr Sect D* 1999;55:1421–1427.
 54. Ponnuswamy PK. Hydrophobic characteristics of folded proteins. *Prog Biophys Mol Biol* 1993;59:57–103.
 55. Dill KA. Dominant forces in protein folding. *Biochemistry* 1990;29:7133–7155.
 56. Burley SK, Petsko GA. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* 1985;229:23–28.
 57. Singh J, Thornton JM. The interaction between phenylalanine rings in proteins. *FEBS Lett* 1985;191:1–6.
 58. Nishio M, Hirota M, Umezawa Y. The CH/ π interaction. Evidence, nature, and consequences. New York: Wiley-VCH; 1998.
 59. Morgan RS, McAdon JM. Predictor for sulphur-aromatic interactions in globular proteins. *Int J Pept Protein Res* 1980;15:177–180.
 60. Reid KSC, Lindley PF, Thornton JM. Sulphur-aromatic interactions in proteins. *FEBS Lett* 1985;190:209–213.
 61. Pal D, Chakrabarti P. Different types of interactions involving cysteine sulfhydryl group in proteins. *J Biomol Struct Dyn* 1998;15:1059–1072.
 62. Malone JF, Murray CM, Charlton MH, Docherty R, Lavery AJ. X—H... π (phenyl) interactions: theoretical and crystallographic observations. *J Chem Soc Faraday Trans* 1997;93:3429–3436.
 63. Burley SK, Petsko GA. Amino-aromatic interactions in proteins. *FEBS Lett* 1986;203:139–143.
 64. Desiraju GR. The C—H...O hydrogen bond: structural implications and supramolecular design. *Acc Chem Res* 1996;29:441–449.
 65. Derewenda ZS, Lee L, Derewenda U. The occurrence of C—H...O hydrogen bonds in proteins. *J Mol Biol* 1995;252:248–262.
 66. Chakrabarti P, Chakrabarti S. C—H...O hydrogen bond involving proline residues in α -helices. *J Mol Biol* 1998;284:867–873.
 67. Umezawa Y, Tsuboyama S, Honda K, Uzawa J, Nishio M. CH/ π interaction in the crystal structure of organic compounds: a database study. *Bull Chem Soc Jpn* 1998;71:1207–1213.
 68. Plesniak LA, Wakarchuk WW, McIntosh LP. Secondary structure and NMR assignments of *B. circulans* xylanase. *Protein Sci* 1996;5:1118–1135.
 69. Thomas KA, Smith GM, Thomas TB, Feldmann RJ. Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. *Proc Natl Acad Sci USA* 1982;79:4843–4847.
 70. Gould RO, Gray AM, Taylor P, Walkinshaw MD. 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* 1985;107:5921–5927.
 71. Flanagan K, Walshaw J, Price SL, Goodfellow JM. Solvent interactions with π ring systems in proteins. *Protein Eng* 1995;8:109–116.
 72. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 1991;24:946–950.