Protein Engineering vol.14 no.1 pp.7–15, 2001

Assessing the role of tryptophan residues in the binding site

Uttamkumar Samanta and Pinak Chakrabarti[1]

Department of Biochemistry and Bioinformatics Centre, Bose Institute, P-1/12, CIT Scheme VIIM, Calcutta 700 054, India
[1]To whom correspondence should be addressed.
E-mail: pinak@boseinst.ernet.in

Instead of looking at the interfacial area as a measure of the extent of a protein–protein recognition site, a new procedure has been developed to identify the importance of a specific residue, namely tryptophan, in the binding process. Trp residues which contribute more towards the free energy of binding have their accessible surface area reduced, on complex formation, for both the main-chain and side-chain atoms, whereas for the less important residues the reduction is restricted only to the aromatic ring of the side chain. The two categories of residues are also distinguished by the presence or absence of hydrogen bonds involving the Trp residue in the complex. A comparison of the observed change in the accessible surface area with the value calculated using an analytical expression provides another way of characterizing the Trp residues critical for binding and this has been used to identify such residues involved in binding non-proteinaceous molecules in protein structures.

Keywords: accessible surface area/molecular recognition/protein–protein complexes/substrate binding/tryptophan

Introduction

The recognition and association between macromolecules are fundamental to the functioning of biological systems. The affinity between two molecules for the formation of non-covalent complexes can be quantified on a structural basis (Janin, 1995a,b). For example, in protein–protein complexes, such as those between protease and inhibitor, antibody and antigen, etc., the interface covers an area of ~1500 Å^2 and contains ~10 hydrogen bonds (Chothia and Janin, 1975; Janin and Chothia, 1990). The two surfaces have good shape and electrostatic complementarity (Norel *et al.*, 1994; Jones and Thornton, 1996; McCoy *et al.*, 1997). Although analyzing the whole surface, these studies do not provide much insight into the details of the contributions of individual residues to binding. To probe experimentally the energetic contributions of individual side chains to protein binding, alanine scanning mutagenesis (Wells, 1991) has been used to remove selectively individual side chains from an interface. Recently, Bogan and Thorn (1998) compiled a database of 2325 alanine mutants for which the change in free energy of binding upon mutation to alanine ($\Delta\Delta G$) has been measured. They found that the free energy of binding is not evenly distributed across interfaces; instead, there are hotspots ($\Delta\Delta G$ = 2 kcal/mol) of binding energy made up of a small subset of residues in the dimer interface. Of all the amino acid residues found in the interface, the likelihood of being in hotspots is the maximum for tryptophan (Trp).

In this context, it would be of interest to see if there is any structural or binding feature in the three-dimensional structure of a complex that one can use to distinguish a Trp residue in the hotspot from another which is energetically less important. Such characteristics can then be used to assess the importance of Trp in a protein in the binding of other non-proteinaceous molecules, such as carbohydrate, cofactor, substrate or drug.

We have recently analyzed the environment of Trp residues (the aromatic part of the side chain, in particular) in protein structures, the nature of the interacting residues (partners) and the exponential dependence of the accessible surface area of the Trp residue on its number of partners (other protein residues in contact with Trp) (Samanta *et al.*, 2000). As atoms buried at protein–protein interfaces are close-packed like the protein interior (Lo Conte *et al.*, 1999), the aforementioned features of Trp residues in proteins should also be transferable to the residues in the interface region. Consequently, one should be able to assess the role of Trp in the binding by finding the change in the number of its partner residues on complex formation and the associated loss in its accessible surface area and by looking at other elements of its environment and comparing the results with those found within protein structures. This paper is an anatomy of Trp residues in energetically hotspots and other less important regions in protein–protein interfaces, as well as those involved in the binding of other small molecules.

Materials and methods

Information on the Trp residues which are at the protein–protein interface, as revealed by the crystallographic analysis of the heterodimeric complex, was obtained from the file interface.xls in http://motorhead.ucsf.edu/~thorn/hotspot (Bogan and Thorn, 1998). Depending on their contribution towards the free energy of binding, these were classified as being or not being in hotspots. Only the complexes for which both thermodynamic and crystallographic data are available could be used and are given in Table I. As outlined by Samanta *et al.* (2000), any residue with an atom within 4 Å of any Trp atom was considered a partner. In protein–protein complexes the partner residues are provided by both molecules, whereas consideration of only the parent molecule (containing the Trp residue) gave the partners before complexation.

For the analysis of the role of Trp in binding small molecules (termed substrates in this paper), all non-proteinaceous molecules (excluding water) in contact with Trp residues were identified for a selected set of 180 protein structures from the Protein Data Bank (PDB) (Sussman *et al.*, 1998); the methodology and the files used are supplied in Samanta *et al.* (2000). The solvent-accessible surface area (ASA) was computed using the program ACCESS (Hubbard, 1991), which is an implementation of the Lee and Richards (1971) algorithm. The solvent probe size was 1.4 Å and the default van der Waals radii in the program were used in all calculations. Any hydrogen bond involving donor (>NH groups in the

Table I. Trp residues in the interface and their partner residues

PDB file[a]	Trp residue[b]	Change in partner no. on complexation[c]		Partners[b, d]	
		Whole	Aromatic	Before complexation	Additional ones after complexation
(i) *In hotspot*					
1dan	45T	6/8	5/7	*Q37T*, *I38T*, *S39T*, G43T, *L72T*, *R74T*	*F129H*, *R134H*
3hhr	104B	3/8	2/6	*I103B*, I105B, *D126B*	*P61A*, K168A, *D171A*, *K172A*, *T175A*
	169B	7/12	4/8	*E42B*, *R43B*, *E44B*, *I103B*, D164B, K167B, M170B	*N63A*, *R64A*, *T67A*, *T175A*, I179A
(ii) *Not in hotspot*					
3hhr	80B	6/6	4/4	*F67B*, *Y68B*, *T69B*, *T77B*, Q78B, K81B	–
1dvf	52B	8/13	3/8	Y32B, *M50B*, I51B, D54B, G55B, *N56B*, *D58B*, K71B	*R96A*, *Y98D*, *Y99D*, *Q100D*, *G100(A)D*
1vfb	92A	6/8	4/6	*N28A*, *I29A*, *H30A*, *Y32A*, H90A, T94A	*Q121C*, *R125C*
	52B	8/12	4/8	*M50B*, I51B, D54B, G55B, *N56B*, *D58B*, K71B, *Y101B*	*R96A*, *G117C*, *T118C*, *D119C*

[a]The proteins involved in the complexes (Bogan and Thorn, 1998) are 1dan, complex of human blood coagulation factor VIIa with human recombinant soluble tissue factor; 3hhr, human growth hormone complexed with the extracellular domain of its receptor; 1dvf, idiotopic antibody D1.3 Fv fragment–antiidiotopic antibody E5.2 Fv fragment complex; 1vfb, Fv fragment of mouse monoclonal antibody D1.3 complexed with hen egg lysozyme.
[b]The one-letter amino acid code followed by the residue number and the subunit identifier are given.
[c]Considering the whole Trp residue and only its aromatic part; only protein residues are considered as partners.
[d]Those interacting with the aromatic part of Trp are in italics, those interacting with the main-chain atoms or C^{β} are given in normal font and those having contacts with both the regions of Trp are underlined.

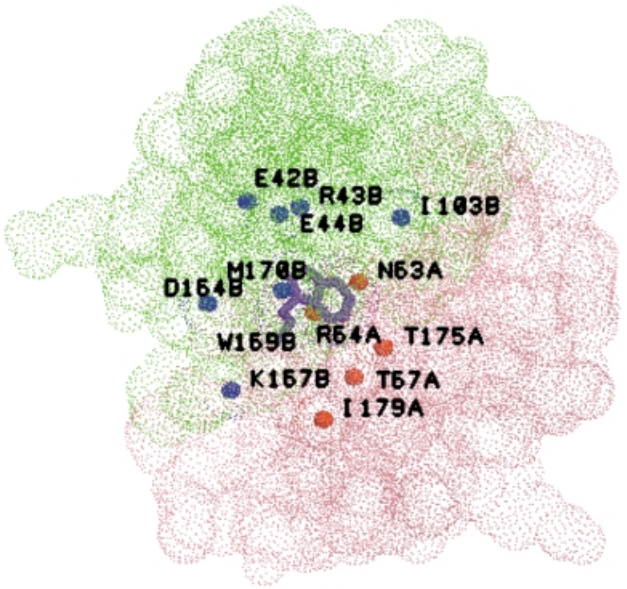


Fig. 1. Residues (C^{α} positions indicated) in contact with Trp169B in the PDB file 3hhr (details are available in Table I). The two subunits (and their residues) in the complex are drawn in different colours and the residues upto a distance of 15 Å from the Trp residue are used to draw the surface plot. The diagram was made using RASMOL (Sayle and Milner-White, 1995).

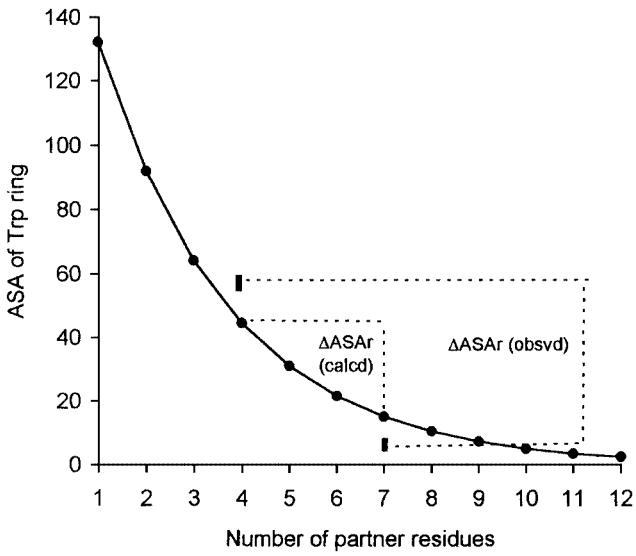


Fig. 2. A plot of the equation ASA = $189.63e^{-0.36x}$, relating the variation of the accessible surface area, ASA (Å^2) of the aromatic ring of a Trp residue and its number (x) of partners. If the number of partners increases from 4 to 7 on complex formation, the change in ASA (ΔASAr) can be calculated from the above equation and is shown. Depending on the initial and final values of ASA, the observed value of ΔASAr could be different. Both the observed and calculated values are provided Table II. Instead of the aromatic ring only, the whole Trp residue can also be considered for the ASA calculation and for finding out the number of partners, the corresponding equation is ASA = $246.64e^{-0.22x}$.

main-chain and the side-chain NE1 position) and acceptor (main-chain O atom) sites in Trp residues and complementary sites in other protein residues or substrates was identified first by noting such groups within a cutoff distance of 3.5 Å and then visually checking them on a graphics terminal.

Results and discussion

Trp residues in protein interface

Table I lists Trp residues which are/are not in hotspots, as elucidated by Bogan and Thorn (1998). The number of partner residues in contact with the Trp residue, considering either the whole residue or just the aromatic ring, before and after complex formation and the names of the partner residues are also provided. Figure 1 depicts a Trp in the interface and how its partners are disposed in the two subunits.

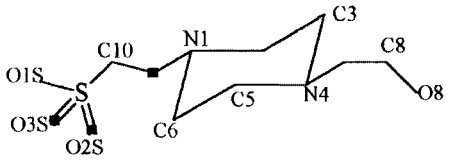
An analysis of the environment of the aromatic ring of Trp showed that the peak in the distribution of the number of protein residues in contact with the ring (the so-called partners) occurs at six (Samanta *et al.*, 2000). The number of partners of Trp residues coming from the same polypeptide chain is in the range 2–5, suggesting that the binding potential of these

Table II. Change in the accessible surface area of Trp residues in the interface on complex formation

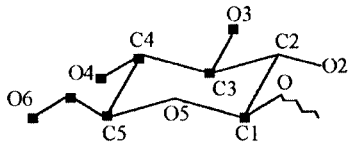
PDB file	Trp[a]	No. of partners		Considering the whole Trp residue[b]				Considering only the Trp aromatic ring[b]			
		Whole	Arom.	Before complexation	After complexation	ΔASAw (obs.)	ΔASAw (calc.)	Before complexation	After complexation	ΔASAr (obs.)	ΔASAr (calc.)
				ASAw	ASAw_s			ASAr	ASAr_s		
(i) *In hotspot*											
1dan	45T*†	6/8	5/7	98.2	36.9	61.3	23.5	65.2	23.2	42.0	16.1
3hhr	104B*†	3/8	2/6	163.2	7.5	155.7	85.1	117.8	0.4	117.4	70.4
	169B	7/12	4/8	89.7	1.3	88.4	35.3	77.3	0.4	76.9	34.3
(ii) *Not in hotspot*											
3hhr	80B	6/6	4/4	91.0	90.8	0.3	0.0	42.1	41.8	0.3	0.0
1dvf	52B	8/13	3/8	88.8	9.2	79.6	28.3	85.7	9.1	76.6	53.8
1vfb	92A	6/8	4/6	97.7	36.9	60.8	23.5	94.6	33.8	60.8	23.1
	52B	8/12	4/8	81.3	17.6	63.7	24.8	78.5	16.8	61.7	34.3

[a]On complex formation, if there is a hydrogen bond involving the side-chain (NE1) or main-chain atoms of Trp, such residues are tagged by * and †, respectively. In 1dan, the hydrogen bond marked by * is water mediated.

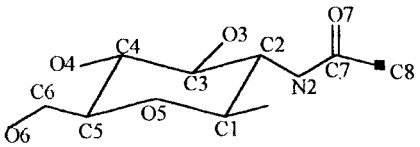
[b]ASAw and ASAr are the accessible surface areas (in $Å^2$) of Trp residues in the polypeptide chain in which they are located, considering the whole residue and only the aromatic ring, respectively. ASAw_s and ASAr_s are the equivalent values in the complex, i.e. when both the subunits are included in the surface area calculation. ΔASAw and ΔASAr are then the observed change in the accessible surface areas of Trp residues ('whole' and 'aromatic ring' only) on complex formation. The expected value of this change can also be calculated from the analytical expression that relates the accessible surface area of Trp residues on its number of partners; the numbers of partners before and after the complex is formed are given in the table.



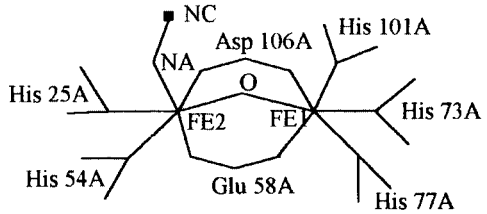
Hepes (EPE), $C_8H_{18}N_2O_4S$
1asu: EPE 252; 76 (O2S,O3S,C9).



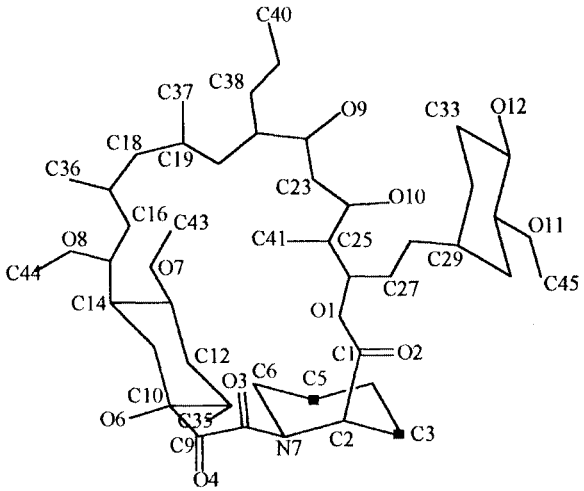
Glucose (GLC) and Galactose (GAL), $C_6H_{12}O_6$
1byb: GLC 496; 55 (O3), 198 (O6), 301 (C4,O3,C1).
1cel: GLC 437B; 367B (C6), 376B (C1,C3,C5,O4).
2gbp: GLC 310; 183 (O4,C5,C3).
1slt: GAL 402; 68A (C4).



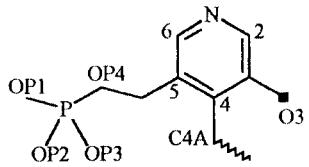
N-acetyl-D-glucosamine (NAG), $C_8H_{15}NO_6$
1cfb: NAG 3; 681 (C8).



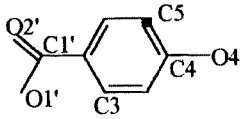
Monoazido-mu-oxo-diiron (FEA), N_3OFe_2
2hmz: FEA 101A; 97A (NC).



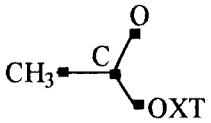
Ascomycin (FK5), $C_{44}H_{69}NO_{12}$
1fkj: FK5 108; 59 (C3,C5).



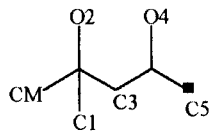
Pyridoxal 5'-phosphate (PLP), $C_8H_{10}NO_6P$
1gpb: PLP 999; 491 (O3).



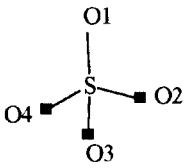
p-Hydroxybenzoic acid (PHB), $C_7H_6O_3$
1pbe: PHB 396; 185(C5).



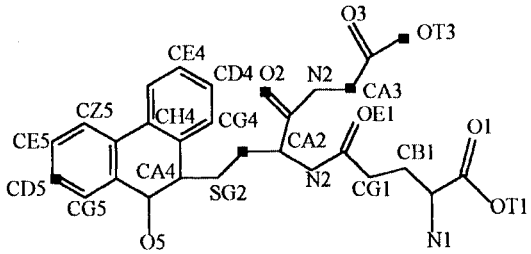
Acetic acid (ACY), $C_2H_4O_2$
Acetate ion (ACY or ACT), $C_2H_3O_2^-$
1gof: ACY 703; 290 (C). ACY 701; 336 (CH_3).
2hts: ACY 337; 203 (O), 217 (O), 258 (C, OXT).
2olb: ACT 1; 397A (CH3).



2-Methyl-2,4-pentanediol (MPD), $C_6H_{14}O_2$
1lcp: MPD 2; 82A (C5).

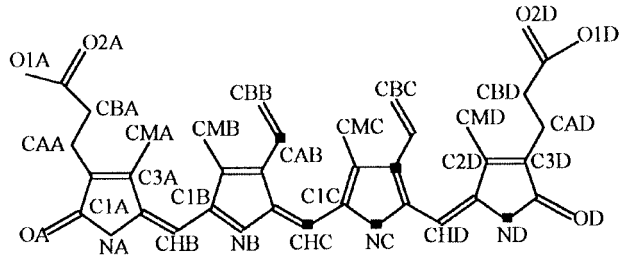


Sulphate, SO_4^{--}
1nba: SO_4 400A; 56A (O2), 111 A (O4).
1sbp: SO_4 310; 192 (O3).

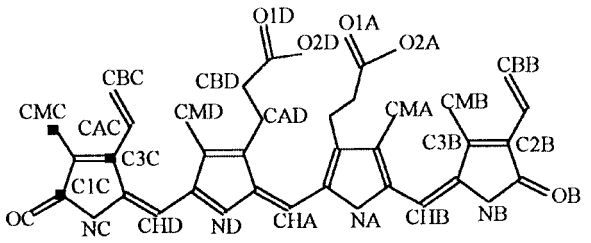


(9s,10s)-9-(S-glutathioyl)-10-hydroxy-9,10-dihydrophenanthrene (GPS), $C_{24}H_{29}N_3O_7S$
2gst: GPS 218I; 7A (O2,CB2,CD5), 45A (CA3,OT3).

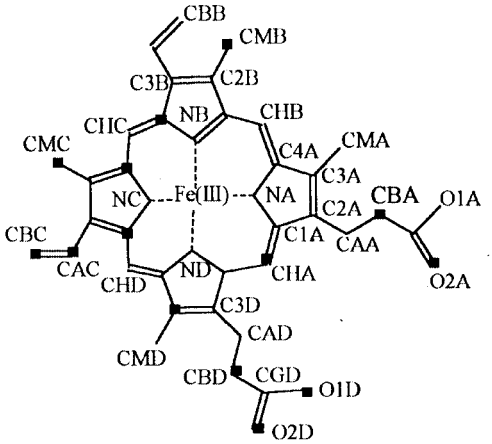
Fig. 3a



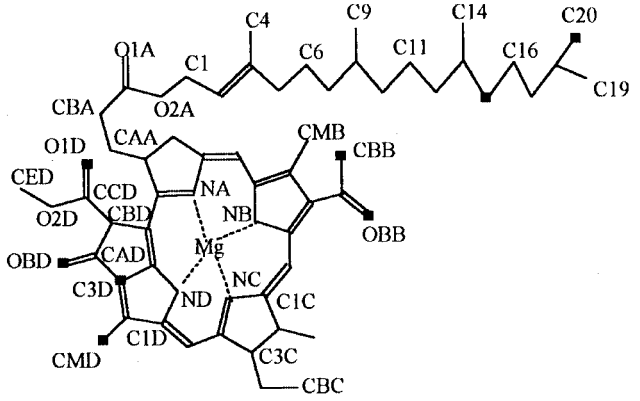
Biliverdin IX gamma chromophore (BLV), $C_{33}H_{35}N_4O_6$
1bbp: BLV 500 D; 133 D (CAB,CHC,C3C,NC,ND).



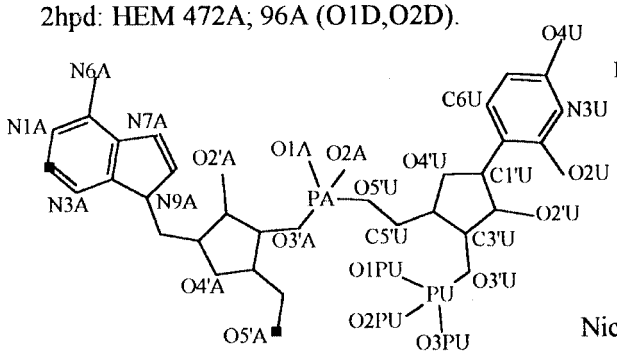
Phycocyanobilin chromophore (CYC), $C_{33}H_{37}N_4O_6$
1cpc: CYC 84A; 128A (C3C,CMC,C1C).



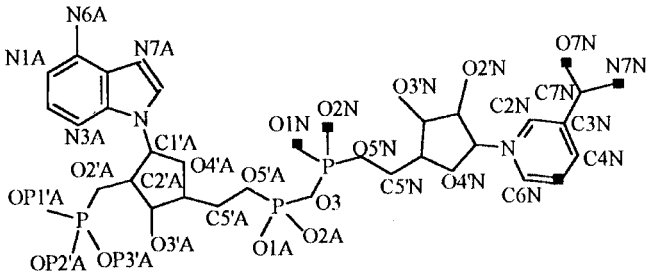
Protoporphyrin IX (HEM),
$C_{34}H_{34}N_4O_4Fe^{+++}$
1ccr: HEM 1; 67 (O2A,CBA).
2ccy: HEM 1A; 23A (CMB), 58A (C2D,CBD),
86A (CBC,CMC,C1C,C4C).
2cyp: HEM 1; 51 (C4B,CAC,C4C,CMC,C1C), 191 (CHA).
2hpd: HEM 472A; 96A (O1D,O2D).



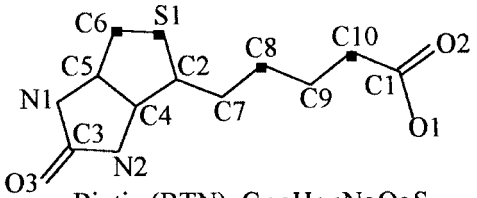
Bacteriochlorophyll A (BCL), $C_{55}H_{74}N_4O_6Mg$
3bcl: BCL 1; 178 (C20), 215 (C15).
BCL 4; 340 (CMD,OBD,C3D,O1D).
BCL 6; 178 (OBB, CBB), 231 (CMD,OBD).



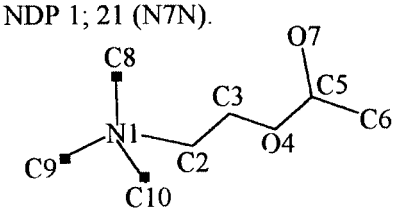
Adenylyl 3',5'-uridine-3'-monophosphate,
(APU), $C_{19}H_{23}N_7O_{15}P_2^-$
1ddt: APU 950; 153 (C2A,O5'A).



Nicotinamide adenine dinucleotide phosphate (NADP+) (NAP),
$C_{21}H_{26}N_7O_{17}P_3^+$; NADPH (NDP), $C_{21}H_{27}N_7O_{17}P_3$
2acq: NAP 316; 20 (O1N,O2N,C5N), 111 (O7N).
3dfr: NDP 1; 21 (N7N).



Biotin (BTN), $C_{10}H_{16}N_2O_3S$
1stp: BTN 300; 79 (S1,C8,C10),
92 (S1), 108 (C6).



Acetylcholine (ACH), $C_7H_{16}NO_2$
1ace: ACH 998; 84 (C8,C9,C10).

Fig. 3b

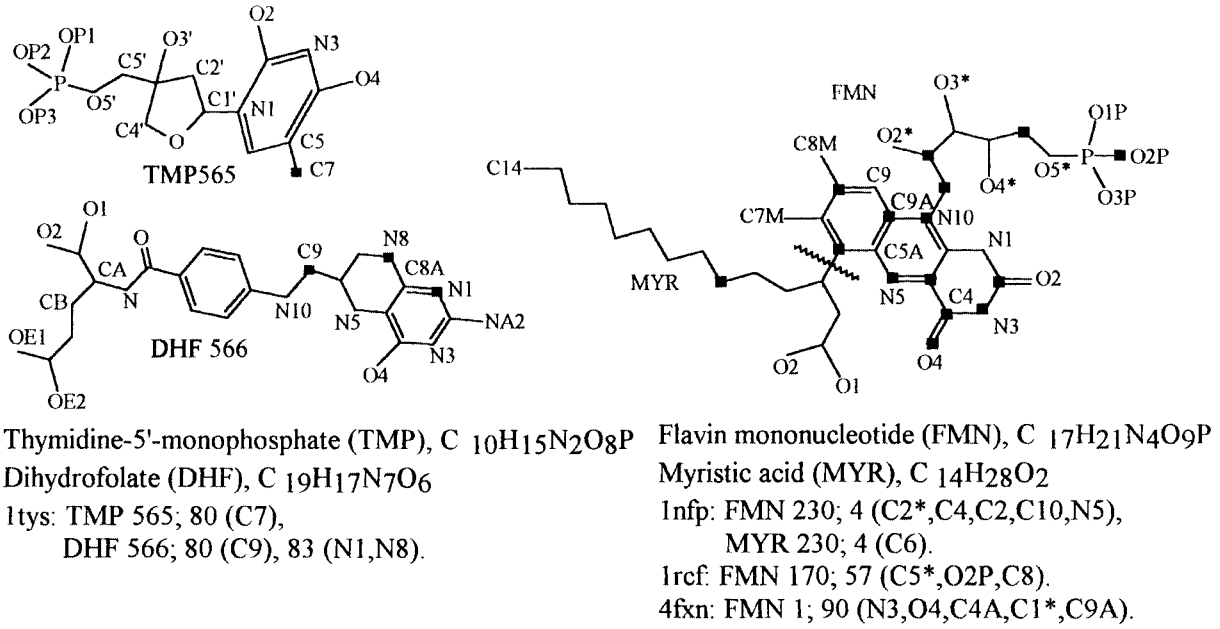


Fig. 3c

Fig. 3. Various types of substrates interacting with Trp residues. Atoms in contact with the indole ring of Trp (in all the different structures containing the substrate) are highlighted using squares. Each diagram is labelled by the name of the compound, its formula, the name of the PDB file containing it, its name and number (including the subunit identifier) in the file, the interacting Trp residue number and the atoms which are interacting with Trp. All the PDB files containing the substrate are shown.

Trp residues is not completely satisfied in their parent molecule. On complex formation, this number increases to 6–8. In one case of a Trp residue not in the hotspot, the number of partners does not change at all. However, there is an important feature which distinguishes Trp residues in hotspots from those which are not. If one compares the observed values of the change in accessible surface area of Trp residues on complex formation, ΔASAw and ΔASAr in Table II, the former considering the whole Trp residue and the latter only the aromatic ring, it is found that the two sets of values are significantly different for Trp residues in a hotspot, but are nearly the same when Trp residues are not in a hotspot. This suggests that for non-hotspot Trp residues, the change in the accessible surface area on complex formation is essentially restricted to the indole part of the side chain, but for residues in a hotspot the change in accessibility is spread over the whole residue (including the main chain and the CB atom). Moreover, when not in a hotspot, Trp residues are not found to form any extra hydrogen bonds in the complex. However, in two out of three hotspot residues there are hydrogen bonds engaging the Trp residues with the physiological partner molecule.

Using the analytical expression relating the accessible surface area of a Trp residue and its number of partners (Samanta *et al.*, 2000) (Figure 2), the expected values of the change in ASA of Trp residues on complex formation, ΔASAw and ΔASAr, can be calculated. These values are, in general, smaller than the observed values irrespective of whether or not the Trp residue is in a hotspot (Table II). This suggests that a Trp residue at the interface is less buried than an average Trp residue in the protein structure and/or on complex formation, the residue gets more buried than suggested by the increase in its number of partners. As discussed above, the observed value of ΔASAw is greater than that of ΔASAr for Trp residues in a hotspot and the calculated values also reflect the same trend. Additionally, for residues not in a hotspot, the trend is just the opposite (with one exception), i.e. ΔASAw (calc.) $<$ ΔASAr (calc.); in one case both the values are 0.0, as there is no change in the number of partners on complex formation.

Trp residues in substrate-binding site

Based on the above observations on Trp residues in the protein interface, we wanted to see if it is possible to assess the importance of Trp residues in the binding site of non-proteinaceous molecules in protein structures. For a residue to be important the following two conditions have to be satisfied: ΔASAw (obs.) $\geqslant$ 2ΔASAw (calc.) and ΔASAr (obs.) $\geqslant$ 2ΔASAr (calc.). These conditions are only approximate, as when applied to residues in Table II, these would have missed out one hotspot residue and also would have identified one non-hotspot residue as important. However, in the case of substrate binding these conditions should be more appropriate. As the substrates are usually much larger than the average size of an amino acid residue and ΔASA values are calculated assuming an increase in the number of partners by just one owing to the substrate binding, these values are expected to be smaller than the values actually observed if the Trp residue is crucial for the binding of the substrate. The other criterion for an important residue is the existence of a hydrogen bond between Trp and the substrate molecule.

The formulae of all the substrate molecules used in our analysis and their atoms which are found in contact with the indole ring of Trp residues in different PDB files are shown in Figure 3. Information on Trp residues, their partners, accessible surface areas and how these change on substrate binding is provided in Table III. In one respect these Trp residues are different from those in the protein interface. Whereas the latter residues have 2–5 partners (around the aromatic ring) in the parent molecule (Table II), the majority of the former residues have a value of $\geqslant$6. The substrate molecules are of different shapes and sizes. Trp residues which

Table III. Accessibility of Trp residues in presence of various substrates

PDB	Substrate	ASA_s	Trp[a]	No. of partners		Considering the whole Trp residue[b]				Considering only the Trp aromatic ring[b]				Remark[c]
				Whole	Arom.	Before substrate binding	After substrate binding	ΔASAw (obs.)	ΔASAw (calc.)	Before substrate binding	After substrate binding	ΔASAr (obs.)	ΔASAr (calc.)	
						ASAw	ASAw_s			ASAr	ASAr_s			
1ace	ACH	344.33	84	7	4	56.9	29.9	27.1	10.4	26.6	5.4	21.2	13.6	
1asu	EPE 252	408.9	76*	11	4	65.9	41.6	24.3	4.3	65.9	41.6	24.3	13.6	
1bbp	BLV 500D	919.0	133D	13	6	48.8	0.5	48.3	2.8	47.8	0.5	47.3	6.6	o
1byb	GLC 496	292.8	55	10	7	7.8	3.9	3.9	5.4	7.8	3.9	3.9	4.6	
			198	10	6	18.9	5.5	13.4	5.4	15.4	2.0	13.4	6.6	o
			301	10	7	24.9	4.5	20.4	5.4	22.2	1.9	20.3	4.6	o
1ccr	HEM 1	851.6	67*	11	7	17.5	0.0	17.5	4.3	17.5	0.0	17.5	4.6	o
1cel	GLC 437B	293.8	367B	7	5	85.6	69.2	16.4	10.4	63.1	46.6	16.5	9.5	
			376B	8	6	67.8	13.6	54.2	8.4	53.0	9.8	43.2	6.6	o
1cfb	NAG 3	360.0	681	7	4	76.5	61.8	14.7	10.4	75.4	61.8	13.6	13.6	
1cpc	CYC 84A	907.7	128A	13	7	25.0	4.5	20.5	2.8	19.6	4.5	15.1	4.6	o
1ddt	APU 950	752.7	153	13	8	35.9	24.9	11.0	2.8	16.4	5.4	11.0	3.2	o
1fkj	FK5 108	966.3	59	12	7	9.8	0.0	9.8	3.5	9.8	0.0	9.8	4.6	o
1gof	ACY 703	181.4	290*	10	8	42.6	33.5	9.1	5.4	28.5	19.3	9.2	3.2	
	ACY 701	183.3	336	13	9	16.1	14.1	2.0	2.8	9.5	7.6	1.9	2.3	
1gpb	PLP 999	388.1	491	10	7	0.8	0.3	0.5	5.4	0.8	0.3	0.5	4.6	
1lcp	MPD 2	280.5	82A	8	5	53.0	34.2	18.8	8.4	53.0	34.2	18.8	9.5	o
1nba	SO_4 400A	181.6	56A*	10	7	6.2	1.4	4.8	5.4	6.2	1.3	4.9	4.6	
			111A	14	9	8.1	3.4	4.7	2.2	8.1	3.4	4.8	2.3	o
1nfp	FMN 230	956.8	4	10	5	65.4	2.5	62.9	5.4	58.0	0.6	57.4	9.5	o
	MYR 230													
1pbe	PHB 396	278.6	185	12	7	7.2	1.3	5.9	3.5	5.7	0.0	5.7	4.6	
1rcf	FMN 170	637.3	57*	7	3	84.5	25.4	59.1	10.4	75.0	18.7	56.3	19.5	o
1sbp	SO_4 310	181.6	192*	11	8	6.0	5.7	0.3	4.3	3.1	2.8	0.3	3.2	
1slt	GAL 402	286.6	68A	5	4	123.4	88.0	35.4	16.2	83.5	49.2	34.4	13.6	o
1stp	BTN 300	414.52	79	11	8	21.0	1.3	19.7	4.3	19.7	0.0	19.7	3.2	o
			92	11	8	10.2	1.8	8.4	4.3	8.4	0.0	8.4	3.2	o
			108	10	8	46.8	30.4	16.4	5.4	44.1	27.7	16.4	3.2	o
1tys	TMP 565	467.6	80	11	6	24.7	2.4	22.3	4.3	24.2	1.9	22.3	6.6	o
	DHF 566	648.1	83	9	6	44.8	27.5	17.3	6.7	27.8	10.6	17.3	6.6	o
2acq	NAP 316	894.8	20†	9	5	69.1	14.0	55.1	6.7	60.0	14.0	46.0	9.5	o
			111	11	6	19.2	0.2	19.0	4.3	19.2	0.2	19.0	6.6	o
2ccy	HEM 1A	839.2	23A	11	6	40.7	28.0	12.7	4.3	39.8	28.0	11.9	6.6	
			58A	10	7	32.3	18.1	14.2	5.4	14.3	0.0	14.3	4.6	o
			86A	10	5	21.8	2.4	19.4	5.4	19.2	0.0	19.2	9.5	o
2cyp	HEM 1	831.2	51	9	5	49.5	0.2	49.3	6.7	42.9	0.2	42.7	9.5	o
			191	11	8	13.0	1.3	11.7	4.3	11.8	0.1	11.7	3.2	o
2gbp	GLC 310	305.3	183	11	9	19.4	6.2	13.2	4.3	13.1	0.0	13.1	2.3	o
2gst	GPS 218I	723.5	7A*	9	6	38.2	2.4	35.8	6.7	34.6	0.7	33.9	6.6	o
			45A*	12	7	6.8	0.0	6.8	3.5	6.8	0.0	6.8	4.6	
2hmz	FEA 101A	241.0	97A	9	5	42.6	42.5	0.1	6.7	34.6	34.5	0.1	9.5	
2hpd	HEM 472A	851.0	96A*	11	7	12.9	8.6	4.3	4.3	12.6	8.2	4.4	4.6	
2hts	ACY 337	182.8	203	4	1	139.6	121.3	18.3	20.2	134.5	120.3	14.2	40.0	
			217	8	6	76.5	64.6	11.9	8.4	52.5	40.6	11.9	6.6	
			258	10	8	34.6	16.3	18.3	5.4	34.6	16.3	18.3	3.2	o

Table III continued

PDB	Substrate	ASA_s	Trp[a]	No. of partners		Considering the whole Trp residue[b]				Considering only the Trp aromatic ring[b]				Remark[c]
				Whole	Arom.	Before substrate binding	After substrate binding	ΔASAw (obs.)	ΔASAw (calc.)	Before substrate binding	After substrate binding	ΔASAr (obs.)	ΔASAr (calc.)	
						ASAw	ASAw_s			ASAr	ASAr_s			
2olb	ACT 1	183.0	397	11	7	9.7	3.3	6.4	4.3	6.2	0.0	6.2	4.6	
3bcl	BCL 1	1162.4	178	8	4	79.8	53.1	26.7	8.4	57.9	35.4	22.5	13.6	
	BCL 6	1162.4	178*	8	4	79.8	46.9	32.9	8.4	57.9	25.0	32.9	13.6	o
	BCL 1	1162.4	215	8	5	41.9	6.3	35.6	8.4	41.1	6.3	34.8	9.5	o
	BCL 6	1162.4	231	10	6	16.5	4.8	11.7	5.4	16.0	4.3	11.7	6.6	
	BCL 4	1162.4	340	7	4	43.1	21.1	22.0	10.4	34.9	12.9	22.0	13.6	
3dfr	NDP 1	929.2	21	10	8	24.8	24.3	0.5	5.4	2.5	2.0	0.5	3.2	
4fxn	FMN 1	635.0	90†	0	0	238.4	186.0	52.4	48.7	169.2	117.4	51.8	57.3	

Any non-proteinaceous (and non-water) molecule in contact with a Trp residue is designated substrate. Trp residues which do not change the accessibility value on substrate binding are not included in the table. ASA_s stands for the solvent-accessible surface area (in $Å^2$) of the isolated substrate molecule. Other such values are explained in Table II footnotes, but with the following modification. ASAw and ASAr are the values not considering the substrate molecule, whereas ASAw_s and ASAr_s are the values in the presence of the substrate. While calculating ΔASAw and ΔASAr values it is assumed that the number of partners of the Trp residue is increased by 1 on binding the substrate molecule.

[a]On complex formation, if there is a hydrogen bond involving the side-chain (NE1) or main-chain atoms of Trp, such residues are tagged by * and †, respectively.

[b]Considering the whole Trp residue and only its aromatic part; only protein residues are considered as partners (if the substrate molecule is considered as a partner, this number would increase by 1).

[c]Trp residues which are likely to be important in substrate binding are marked by o. In these cases, ΔASAw (obs.) $\geq$ 2ΔASAw (calc.) and ΔASAr (obs.) $\geq$ 2ΔASAr (calc.).

are deemed to be important in substrate binding using the conditions on ASA are marked with dots in the last column in Table III. If in addition there is a hydrogen bond between the Trp residue and the substrate, the residue is likely to be important in substrate binding. One example is the binding of FMN by Trp57 in the structure, 1rcf. 1stp corresponds to the structure of streptavidin which binds biotin with exceptionally high affinity ($K_d = 10^{-15}$ M) (Green, 1975). There are three Trp residues in the binding site (Weber *et al.*, 1989) and all are shown to be important, thus lending credence to the predictive power of our methodology. Moreover, aromatic-sugar stacking is a typical feature of protein–carbohydrate interactions (Vyas, 1991; Kadziola *et al.*, 1998). In all the structures (1byb, 1cel, 1slt and 2gbp) where a carbohydrate molecule is bound, there is at least one Trp residue which is shown to be important. However, in all the cases the decrease in the accessible surface area on substrate binding may not be the best criterion to judge the role of a residue. For example, in the binding of the small sulfate ion (structure, 1sbp), there is hardly any change in ASA and the formation of the hydrogen bond could be the deciding factor in this case. Another situation where the comparison of the observed and calculated values of ΔASA may not yield the right result is when the number of partners is atypically small, e.g. 0. In 4fxn, although the observed value is one of the highest in the table, the calculated value is also large and their difference is very small. Nevertheless, this procedure provides some guidelines as to the importance of a Trp residue in binding, which can then be corroborated by protein engineering experiments.

Conclusion

Depending on the magnitude of contribution towards the binding energy, interface residues have been classified as being or not being in a hotspot (Bogan and Thorn, 1998). In this paper we analyzed whether it is possible to identify Trp residues in hotspots from those which are not, on the basis of crystal structure data. We find that for Trp residues not in hotspots, the change in accessible surface area of the Trp residue on complex formation is restricted to only the indole ring, whereas for hotspot residues the change involves the whole residue. Although the former residues do not form hydrogen bonds with the physiological partner molecule, a hydrogen bond is usually formed for the latter residues. Depending on the change in the number of partner residues, it is possible to calculate the expected change in the accessible surface area of a Trp residue due to complex formation. The observed values are always found to be greater than the calculated values. Similar comparisons between the observed and calculated values and the identification of any hydrogen bond linking Trp to the substrate molecule provides a way to assess the importance of Trp residues in the substrate-binding sites. Based on these encouraging results involving Trp, we are now in the process of extending the methodology to other residues.

Acknowledgements

The authors are grateful to the Department of Biotechnology and the Council of Scientific and Industrial Research for financial support.

References

Bogan,A.A. and Thorn,K.S. (1998) *J. Mol. Biol.*, **280**, 1–9.
Chothia,C. and Janin,J. (1975) *Nature*, **256**, 705–708.
Green,N.M. (1975) *Adv. Protein Chem.*, **29**, 85–133.
Hubbard,S.J. (1991) *ACCESS, a Program for Calculating Accessibilities*. Department of Biochemistry and Molecular Biology, University College London, London.
Janin,J. (1995a) *Biochimie*, **77**, 497–505.
Janin,J. (1995b) *Proteins: Struct. Funct. Genet.*, **21**, 30–39.
Janin,J. and Chothia,C. (1990) *J. Biol. Chem.*, **265**, 16027–16030.
Jones,S. and Thornton,J.M. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
Kadziola,A., Sogaard,M., Svensson,B. and Haser,R. (1998) *J. Mol. Biol.*, **278**, 205–217.
Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.
Lo Conte,L., Chothia,C. and Janin,J. (1999) *J. Mol. Biol.*, **285**, 2177–2198.
McCoy,A.J., Epa,V.A. and Colman,P.M. (1997) *J. Mol. Biol.*, **268**, 570–584.
Norel,R., Lin,S.L., Wolfson,H.J. and Nussinov,R. (1994) *Biopolymers*, **34**, 933–940.
Samanta,U., Pal,D. and Chakrabarti,P. (2000) *Proteins: Struct. Funct. Genet.*, **38**, 288–300.
Sayle,R.A. and Milner-White,E.J. (1995) *Trends Biochem. Sci.*, **20**, 374.
Sussman,J.L., Lin,D., Jiang,J., Manning,N.O., Prilusky,J., Ritter,O. and Abola,E.E. (1998) *Acta Crystallogr.*, **D54**, 1078–1084.
Vyas,N.K. (1991) *Curr. Opin. Struct. Biol.*, **1**, 732–740.
Weber,P.C., Ohlendorf, D.H., Wendoloski,J.J. and Salemme,F.R. (1989) *Science*, **243**, 85–88.
Wells,J.A. (1991) *Methods Enzymol.*, **202**, 390–411.

Received May 9, 2000; revised October 31, 2000; accepted November 9, 2000