J Mol Model (2011) 17:401–413
DOI 10.1007/s00894-010-0736-2

ORIGINAL PAPER

Analysis of structural water and CH···π interactions in HIV-1 protease and PTP1B complexes using a hydrogen bond prediction tool, HBPredicT

Joshy P. Yesudas · Fareed Bhasha Sayyed · Cherumuttathu H. Suresh

Received: 26 February 2010 / Accepted: 27 April 2010 / Published online: 19 May 2010

Abstract A hydrogen bond prediction tool HBPredicT is developed for detecting structural water molecules and CH···π interactions in PDB files of protein-ligand complexes. The program adds the missing hydrogen atoms to the protein, ligands, and oxygen atoms of water molecules and subsequently all the hydrogen bonds in the complex are located using specific geometrical criteria. Hydrogen bonds are classified into various types based on (i) donor and acceptor atoms, and interactions such as (ii) protein-protein, (iii) protein-ligand, (iv) protein-water, (v) ligand-water, (vi) water-water, and (vii) protein-water-ligand. Using the information in category (vii), the water molecules which form hydrogen bonds with the ligand and the protein simultaneously–the structural water–is identified and retrieved along with the associated ligand and protein residues. For CH···π interactions, the relevant portions of the corresponding structures are also extracted in the output. The application potential of this program is tested using 19 HIV-1 protease and 11 PTP1B inhibitor complexes. All the systems showed presence of structural water molecules and in several cases, the CH···π interaction between ligand and protein are detected. A rare occurrence of CH···π interactions emanating from both faces of a phenyl ring of the inhibitor is identified in HIV-1 protease 1D4L.

Keywords CH···π interactions · HIV protease · Hydrogen bond · Protein-ligand interactions · PTP1B · Structural water

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0736-2) contains supplementary material, which is available to authorized users.

J. P. Yesudas · F. B. Sayyed · C. H. Suresh (✉)
Computational Modeling and Simulation Section,
National Institute for Interdisciplinary Science and Technology (CSIR),
Trivandrum 695 019, India
e-mail: sureshch@gmail.com
URL: http://www.niist.res.in/english/scientists/suresh-c-h/personal.html

Introduction

In recent years, structure based drug design (SBDD) methods have played a vital role in the design of several drug candidates that were later used in clinical trials [1–4]. The successful application of SBDD methods depends on the availability of accurate three dimensional structures of bio-macromolecular complexes [5] as well as the identification and quantification (on the basis of geometric criteria) of non-covalent interactions in them [6]. The critical role of hydrogen bonding and CH···π interactions in stabilizing the protein structures has been well established [7, 8]. Hydrogen bond analysis provides valuable clues of the structural information of protein-ligand complexes. The interaction between complementary functional groups and the release of water molecules from ligand and protein to bulk drives the process of ligand binding in a protein active site [9]. However, active sites may contain other compounds, such as cofactors or coenzymes as well as structural water molecules that can simultaneously interact with protein residues and ligand functional groups to provide a bridging function between them [9, 10]. Of particular interest to us are structural water molecules, which play a dual role by acting as both donor and acceptor for hydrogen bonds at the interface of target protein and the ligand molecule, and thus contribute significantly to improve the binding ability of the ligand [11–13]. These structural water molecules play a vital role in the SBDD whose ignorance reduces the success of the drug-design strategy [11]. There are reported cases which show how water increases the range of specificity of

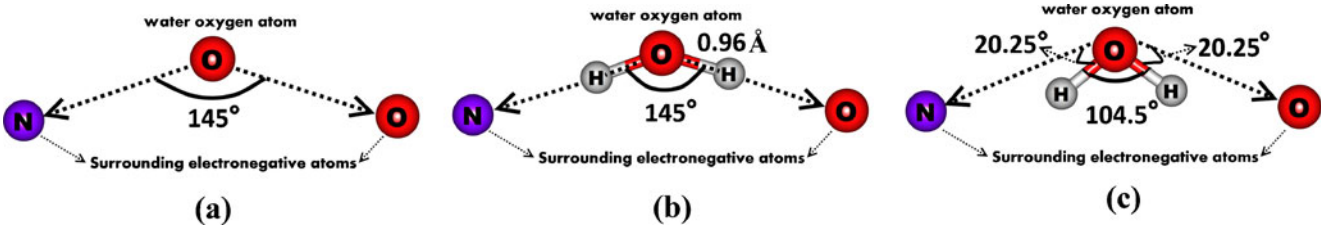


Fig. 1 Criteria used to add missing hydrogen atoms of the water molecules. An example is given wherein the proton acceptors are making an angle 145° with the oxygen of water. (**a**) An oxygen atom of the water molecule is surrounded by two electronegative (nitrogen and oxygen) atoms, ∠N-O-O = 145°. (**b**) Hydrogen atoms are first placed at a distance of 0.96 Å along the vector directions of the electronegative atoms. (**c**) ∠H-O-H is refined to 104.5° by repositioning the hydrogen atoms

a binding site, when additional water molecules are included [10, 11]. Recently, Suresh *et al*. [13] have studied the role of structural water in HIV-1 protease inhibitor complex and showed that the binding energy of the structural water molecule compensate for the strain energy that the ligand experiences in the protein bound structure. It is also understood that the structural water molecules show a strong influence on the flexibility of the proteins [14].

CH···π interactions, one of the weak hydrogen bonds which occur between soft acid (C-H) and soft base (π-electron system) is gradually gaining substantial importance in chemical and biological studies [15–17]. These attractive noncovalent interactions play an important role in stabilizing the structure of proteins [7], in protein–ligand interaction [18–20] as well as in crystal packing of organic compounds [21, 22]. These interactions occurs when partial charge transferred from the HOMO π-orbital to the LUMO σ*-orbital of C-H bond [7]. Role of CH···π interactions, in substrate binding by *E. coli* β-galactoside [23] and in the packing of adenine ring in protein structures involving aromatic residues [24] are extensively studied. More interestingly, it has been reported that CH···π interactions play an important role in binding the inhibitors to torpedo californica acetylcholine esterase (TcAchE) [25], which suggests the importance of considering CH···π interactions in designing the ligand.

Several tools are available for adding the missing hydrogen atoms and finding hydrogen bonds [26–28], but there are only a few which do the hydrogen bond analysis [29–31]. To the best of our knowledge, there is no tool to identify and visualize the structural water as well as to view the hydrogen bonds based on different classification (protein-protein, protein-water, protein-ligand *etc*.). Kaur and Raghava developed a program to determine the CH···π interactions in a protein using amino acid sequence based on recurrent neural network but it does not use the three dimensional structural details and the user has no option to visualize the CH···π interactions [32]. Moreover, there are no tools available to identify the CH···π interactions between ligand and protein in a complex system. In this article, we describe a newly developed computational tool HBPredicT for the identification of structural water, the CH···π interactions, and other typical hydrogen bond interactions in a protein-ligand complex. Further, using this tool, a detailed analysis of the nonbonding interactions in the crystal structures of human immunodeficiency virus type 1 (HIV-1) protease and protein tyrosine phosphatase non-receptor type-1B (PTP1B) complex with various inhibitors is made. Protein tyrosine phosphatases (PTPs) are responsible for the selective dephosphorylation of tyrosine residues [33]. Several PTPs which includes PTP1B and transmembrane receptor-like protein LAR are capable of dephosphorylating the insulin receptor [34]. The development of PTP1B inhibitors for the treatment of type-2 diabetes had more attention in recent years.

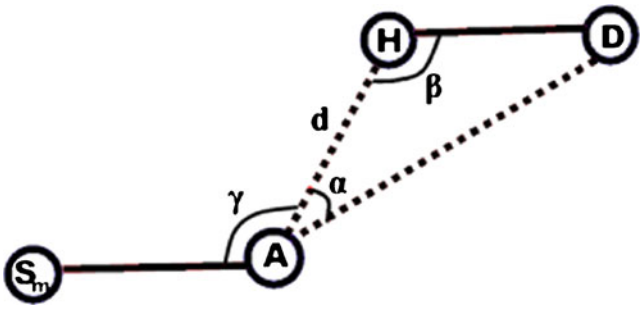
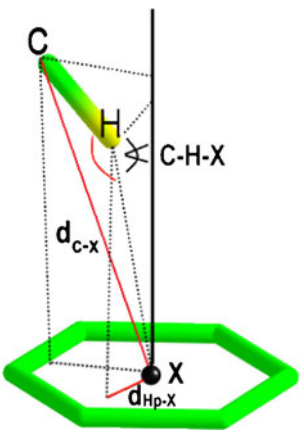


Fig. 2 Criteria used for identifying hydrogen bonds. (1) The distance between the acceptor atom (A) and the proton (H), d ≤ 2.5 Å, and the conditions for the angles α, β and γ are (2) 0° ≤ α ≤ 90° (3) 90° ≤ β ≤ 180° (4) 90° ≤ γ ≤ 180°

Fig. 3 Criteria used to identify CH···π interactions in the case of a six-membered π-ring is shown, where X is the center of mass. The maximum value of dC-X is 4.5 Å and ∠C-H-X, is in the range of 120°–180°. The distance (dHp-X) of the hydrogen positioned on to the π-plane from X, is less than 1.4 Å. The definition is taken from reference [41]



Methodology

HBPredicT is coded in PERL using a Pentium 4 computer (256 Mb RAM) running under Windows XP operating system. The program reads the input structure file in Protein Data Bank (PDB) format [35] and generates the output file "outfile.hbp" after computing the hydrogen bonds and another output file "outfile.chp" after computing the CH···π interactions. The "outfile.hbp" also contains the information about the structural water molecules. HBPredicT enables the users to view the identified structural water and the CH···π interactions, with the hydrogen bonds labeled with distance between proton and acceptor atom as broken lines, through the interface of RasMol [36]. In addition, the program shows histograms which give the composition of amino acids and nucleic acid bases contributing to hydrogen bonds with various types of classifications, which are discussed below based on the HIV-1 protease and PTP1B complexes. Furthermore, the detailed output files can be directly viewed in Microsoft Excel, a versatile computer package available for a wide range of applications, enabling the user to perform further statistical analysis on various hydrogen bond interactions in the system.

Adding missing hydrogen atoms

The three dimensional structural information of a protein is mainly derived by X-ray crystallography, but for mapping hydrogen atoms, rarely achieved ultrahigh-resolution (≤1.0 Å) is required [37]. Therefore, most of the X-ray structures will be devoid of hydrogen atoms and hence HBPredicT is designed to add the missing hydrogen atoms of amino acid residues and ligand systems based on the hybridization properties of atoms, directionality of hydrogen bonds, and standard bond length data [38]. On the basis of the information of those atoms surrounding the oxygen atom of water molecules, the missing hydrogen atoms to the water molecules can be added. The program locates the atoms of C, N, and O surrounded by the oxygen atom of water molecules in 3.0 Å distance. It will also identify the hydrogen atoms in the 2.5 Å vicinity of the oxygen atom of the water molecule. If there is only one proton acceptor, HBPredicT will place the first hydrogen along the axis of water oxygen and the proton acceptor, then it will add the next hydrogen at an angle of 104.5°. If there are no proton acceptors or hydrogen atoms of other molecules within 2.5 Å, hydrogen atoms will be added at random. For example, if an oxygen atom of a water molecule is surrounded by two proton acceptors (oxygen and nitrogen) as shown in Fig. 1a, hydrogen atoms are added at a distance of 0.96 Å along the vector direction of the surrounding atoms (Fig. 1b) and then refines the ∠H-O-H to 104.5° by repositioning the hydrogen in the same plane (Fig. 1c). In the case of more than two proton acceptors present in the vicinity of water oxygen, hydrogen atoms are added along the axis of the least distant proton acceptors and the ∠H-O-H is refined to 104.5°.

Finding hydrogen bonds and structural water

It is well-known that hydrogen bonds display a marked directionality. The criteria suggested by Baker and Hubbard [38] is used to identify the hydrogen bonds which is defined in Fig. 2. In this figure, the A-S_m bond represents the mean direction of the bonds attached to the acceptor atom (A). Unlike many other hydrogen bond identifying tools, instead of considering the distance between the donor (D) and the acceptor atoms (A), we used the distance (d)

Table 1 HIV-1 protease and PTP1B complexes selected for the analysis

PDB Id	Inhibitors	Reference	PDB Id	Inhibitors	Reference
HIV-1 protease			1ODY	LP1-130	[44]
1AAQ	PSI	[45]	1W5W	BE4	[46]
1C6Y	MK1	[2]	2PSU	ACT, MUU, PO4	[47]
1D4H	BEH	[3]	2PSV	ACT, MUV, PO4	[47]
1D4J	MSC	[1]	*PTP1B*		
1D4K	PI8, SO4	[43]	1KAV	FEP	[48]
1D4L	PI9, SO4	[43]	1PH0	418	[49]
1HIH	BME, C20	[50]	1PYN	941	[51]
1HIV	O, NOA	[52]	1Q6P	CL, 213	[53]
1HVI	A77	[42]	1Q6S	CL, MG, MPD, 214	[53]
1HVJ	A78	[42]	2AZR	982	[54]
1HVK	A79	[42]	2B07	598	[54]
1HVL	A76	[42]	2H4G	694	[55]
1HXB	DIQ, QNC	[56]	2H4K	509	[55]
1N49	RIT	[50]	2HB1	512	[55]
1NPW	LGZ	[57]	2NT7	902	[58]

between the hydrogen (H) and the acceptor atoms as a parameter. Since there is no well established parametric definition for the hydrogen bond, a provision is also provided for the user to alter the values of the parameters involved in the criteria given in Fig. 2. HBPredicT also identifies weaker hydrogen bonds [39, 40] like C-H···O and C-H···N with carbon as the donor atom, as well as strong hydrogen bonds such as N-H···N and N-H···O. The program

Table 2 Details of various hydrogen bond (H-bond) interactions identified in HIV-1 Protease complexes

PDB Id	Total no. of H-bonds	No. of H-bonds between				
		Protein-inhibitor	Protein-water	Protein-protein	Inhibitor-water	Water-water
HIV-1 protease						
1AAQ	182	12	3	166	1	0
1C6Y	241	7	51	167	4	12
1D4H	384	8	146	198	4	28
1D4J	348	7	124	201	3	13
1D4K	308	23	81	191	3	10
1D4L	360	10[a]	119	195	4[b]	22
1HIH	401	7	144	213	2	35
1HIV	284	6	86	184	2	6
1HVI	219	6	2	209	2	0
1HVJ	216	6	2	206	2	0
1HVK	219	4	2	211	2	0
1HVL	210	5	2	201	2	0
1HXB	290	8	82	188	2	10
1N49[c]	221	5	16	197	2	1
1N49[d]	220	7	14	194	3	2
1NPW	342	5	104	206	3	24
1ODY	301	15	91	166	2	27
1W5W	396	10	160	202	2	22
2PSU	421	13	163	199	12	34
2PSV	459	9	177	200	11	62
PTP1B complexes						
1KAV	495	3	142	333	1	16
1PH0	592	15	216	303	5	53
1PYN	685	16	293	322	9	45
1Q6P[e]	490	12	132	327	2	17
1Q6P[f]	482	9	134	326	3	10
1Q6S[e]	589	11	189	327	5	57
1Q6S[f]	613	14	194	332	5	68
2AZR	619	5	235	325	3	51
2B07	572	6	195	326	5	40
2H4G	482	5	126	335	4	12
2H4K	496	4	142	333	2	15
2HB1	580	4	179	351	4	42
2NT7	612	5	205	352	5	45

[a] Eight more hydrogen bonds are observed with the modres ABA.

[b] two hydrogen bonds are formed with modres ABA.

[c] 1N49 comprises of 4 chains (A, B, C and D), in which both A,B and C,D pairs contain the inhibitor RIT. All other HIV systems are comprised of A and B chains. The results belong to A, and B chains.

[d] The results belong to C, and D chains.

[e] 1Q6P and 1Q6S exist as dimers, made up of chains A and B while all other systems consist of only the A chain. The results belongs to chain A.

[f] The results belongs to chain B.

Table 3 Structural features of the different types of hydrogen bonds found in HIV-1 Protease 1D4L

S.No.	Type of bond	Count	Avg. distance (Å)		Avg. angle (degree)		
			D-A	H-A	∠D-H-A (β)	∠H-A-D (α)	∠Sm-A-H (γ)
1	N-H···O	186	2.952	2.009	159.3	14.0	138.5
2	O-H···O	93	2.809	1.876	168.3	7.9	133.7
3	C-H···O	70	3.303	2.352	146.3	23.3	133.4
4	O-H···N	4	2.912	1.989	170.3	6.8	150.5
5	C-H···N	7	3.273	2.394	136.6	30.4	121.4
6	N-H···N	0	–	–	–	–	–

can generate histograms showing the distribution of the identified hydrogen bonds based on the distance between proton-acceptor atoms and the angles defined in the above criteria (Fig. 2). Moreover, classification of hydrogen bonds into various types based on (i) donor and acceptor atoms, and interactions such as (ii) protein-protein, (iii) protein-ligand, (iv) protein-water, (v) ligand-water, (vi) water-water, and (vii) protein-water-ligand is generated in the output of the program. The identification of the type (vii) interaction allows the user to locate the possible structural water molecules present in the system.

Finding CH···π interactions

We have implemented a Perl module to locate all the π-ring structures in the complex and the center of mass (X) of π-rings are calculated. The program identifies both five and six-membered π-rings as CH acceptors. In the case of

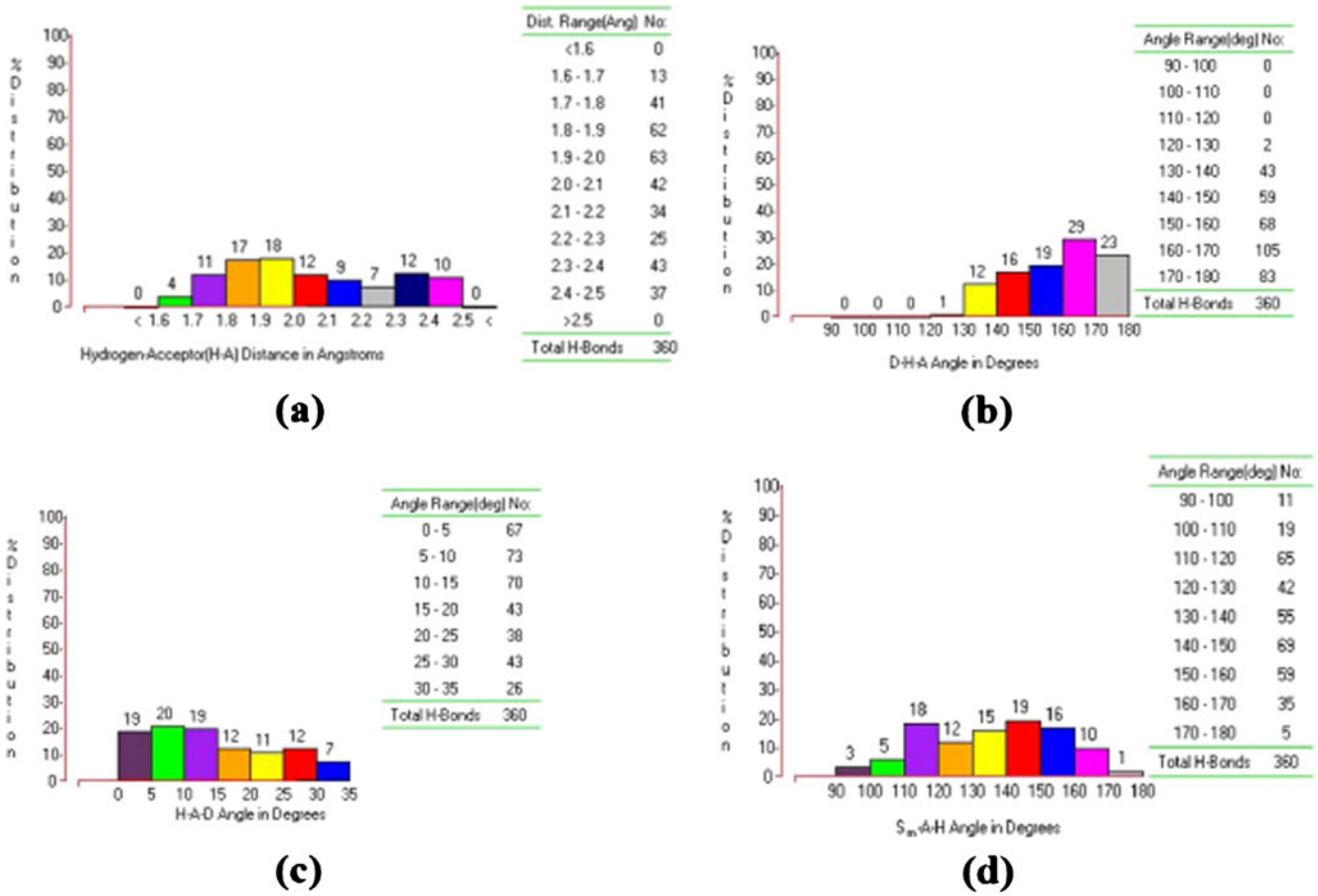


Fig. 4 Histograms obtained in 1D4L using HBPredicT, showing the distribution of the identified hydrogen bonds using the criteria shown in Fig. 2. (**a**) d, (**b**) ∠D-H-A, (**c**) ∠H-A-D, and (**d**) ∠Sm-A-H

Table 4 Distribution of hydrogen bonds among the amino acids in HIV-1 protease 1D4L

Amino acids	Count	Total no. of H-bondsformed by amino acid with			No. of H-bonds per amino acid
		Water	Inhibitor	Rest of protein	
Glycine	26	15	5	25	1.7
Proline	14	6	0	10	1.1
Alanine	6	0	0	15	2.5
Valine	14	2	0	37	2.8
Leucine	20	4	2	42	2.4
Isoleucine	26	3	1	62	2.5
Methionine	4	3	0	4	1.8
Cysteine	0	0	0	0	0.0
Phenylalanine	4	4	0	7	2.8
Tyrosine	2	0	0	7	3.5
Tryptophan	4	4	0	0	1.0
Histidine	2	0	1[a]	6	4.5
Lysine	14	10	0	18	2.0
Arginine	8	14	2	31	5.9
Glutamine	10	7	0	25	3.2
Asparagine	8	7	2	22	3.9
Glutamic acid	8	8	0	22	3.8
Aspartic acid	8	12	1	13	3.3
Serine	0	0	0	0	0.0
Threonine	16	20	2	29	3.2

[a] Histidine also showed two more hydrogen bonds with the modres ABA.

proteins, side chains of phenylalanine, tyrosine, tryptophan and histidine amino acids which have π-ring are also considered in the program. Based on the position of X, three parameters, *viz*. (i) the distance ($d_{C–X}$) between the carbon atom (C) of CH-donor and X, (ii) the angle (∠C-H-X) formed by C and X at the hydrogen H, and (iii) the distance (d_{Hp-X}) of the hydrogen positioned on to the π-plane from X are defined. The criteria to identify the CH···π interactions using these parameters are explained in Fig. 3 [41]. The maximum value of $d_{C–X}$ was set to 4.5 Å. The parameter ∠C-H-X is similar to the one used for conventional hydrogen bonds with values between 120° and 180° to find more linear bonds. In order to confirm whether the hydrogen is above or below the plane, the distance parameter d_{Hp-X} is used with a maximum value of 1.4 Å. Moreover, the program enables the user to change these

Table 5 Distribution of hydrogen bonds based on the classification of amino acids

Classification of amino acids	Count (%)	% of hydrogen bonds formed with		
		Water	Inhibitor	Amino acids itself
Based on charge				
Non polar aliphatic	49	23	44	49
Non polar aromatic	5	7	0	4
Polar uncharged	25	34	22	23
Polar charged positive	12	20	28	15
Polar charged negative	8	17	6	9
Based on hydrophobicity				
Hydrophobic	48	22	17	49
Hydrophilic	38	66	56	44
Amphiphilic	13	13	28	7

Table 6 Structural water identified by HBPredicT in several HIV-1and PTP1B complexes

PDB ID	Distinct protein-water-ligand interactions				
	Amino acid[a]	Ligand molecule	Chain identifier	No. of interactions	Structural water
HIV-1 complexes					
1D4H	ILE(50)	BEH	A	2	'305'
	ILE(150)	BEH	B	2	'305'
	ASP(29)	BEH	A	1	'389'
1D4L[b]	ILE(50)	P19	A	2	'301'
	ILE(50)	P19	B	2	'301'
1HIH	ILE(50)	C20	A	2	'6'
	ILE(50)	C20	B	2	'6'
1HVL	ILE(50)	A76	A	2	'415'
	ILE(50)	A76	B	2	'415'
1HXB	ILE(50)	DIQ	A	1	'1'
	ILE(50)	DIQ	B	1	'1'
1NPW	ILE(50)	LGZ	A	1	'304'
	ILE(250)	LGZ	B	1	'304'
PTP1B complexes					
2AZR	SER(216)	982	A	1	'24'
	ALA(217)	982	A	1	'24'
	ALA(217)	982	A	2	'132'
	ILE(219)	982	A	2	'132'
	GLY(220)	982	A	2	'132'
2H4G	ALA(217)	694	A	2	'4'
	ILE(219)	694	A	2	'4'
	GLY(220)	694	A	2	'4'
	SER(216)	694	A	2	'26'
	ALA(217)	694	A	2	'26'
2H4K	ALA(217)	509	A	2	'91'
	ILE(219)	509	A	2	'91'
	GLY(220)	509	A	2	'91'
2HB1	SER(216)	512	A	2	'47'
	ALA(217)	512	A	2	'47'
	ALA(217)	512	A	2	'58'
	ILE(219)	512	A	2	'58'
	GLY(220)	512	A	2	'58'

[a] The numbers shown in the parentheses corresponds to the numbers in respective PDB files.
[b] Two more interactions with the modres ABA is also observed.

parameters. The user can view all the identified CH···π interactions with the help of RasMol interface.

Results and discussion

For analysis, we have selected 19 HIV-1 protease [42, 43] and 11 PTP1B structures having resolution in the range of 1.75 Å to 2.50 Å from Protein Data Bank (www.rcsb.org/pdb) [35] and they are listed in Table 1. The name of the inhibitors in specific protein complexes and references are also reported in Table 1. Using HBPredicT, the missing hydrogen atoms are added to the protein structures and all the conventional hydrogen bonds are computed. The located structural water, and the CH···π interactions are used for further analysis.

Hydrogen bonds in HIV-1 protease and PTP1B complexes

Distribution of various types of hydrogen bond interactions observed in all HIV-1 and PTP1B complexes selected in this study are reported in Table 2. As expected, most of the hydrogen bonds are located with polar amino acid residues in all the structures. For instance, in the case of 1D4L, a total of 360 hydrogen bonds are identified and among them 195 are observed exclusively for the amino acid residues of the protein, 10 for protein-inhibitor, 119 for protein-water, four for inhibitor-water and 22 for water-water interactions.

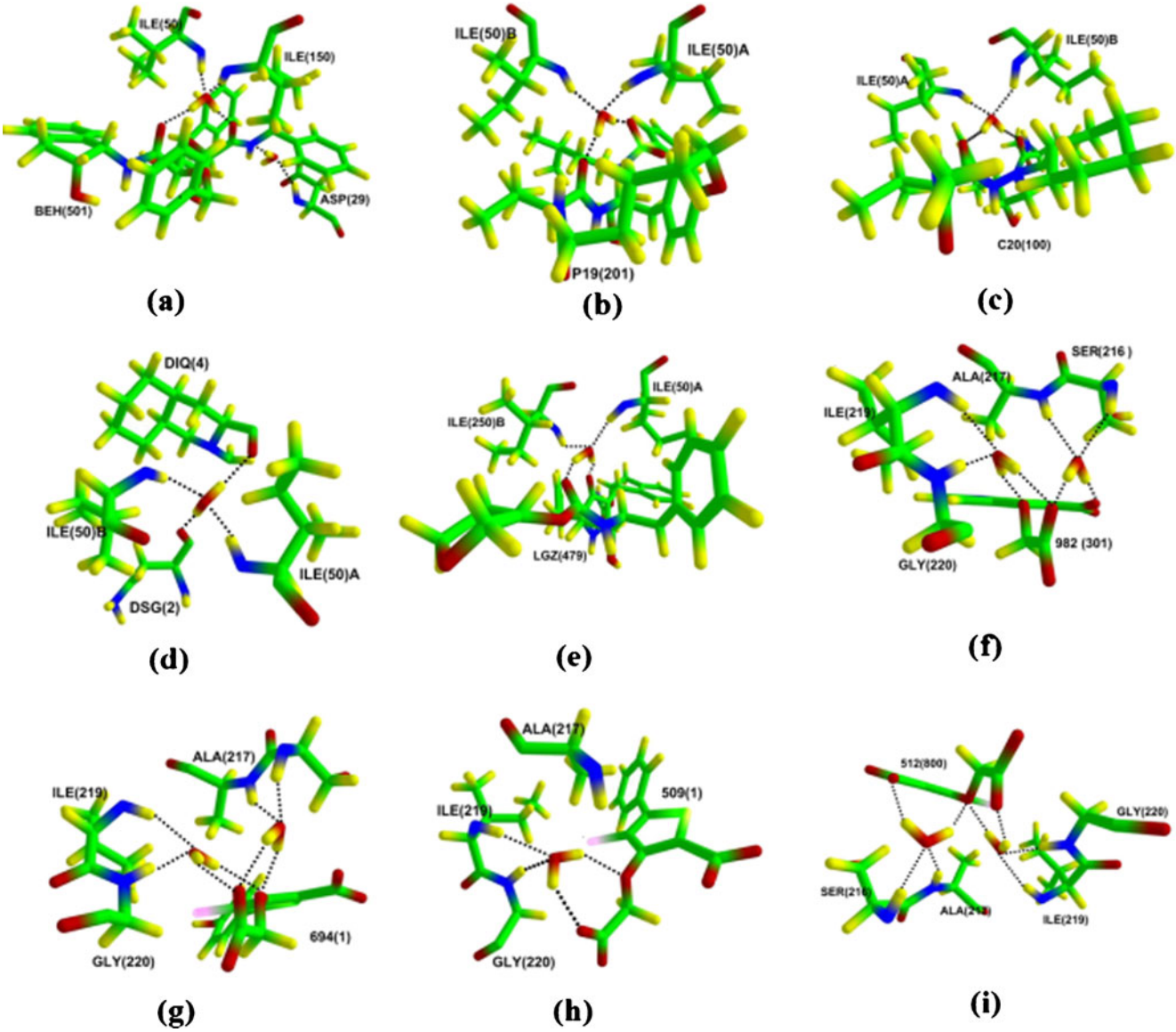


Fig. 5 Structural water identified at the active site of HIV-1 and [(**a**) 1D4H (**b**) 1D4L (**c**) 1HIH (**d**) 1HXB (DSG is a modified residue D-Asparagine) (**e**) 1NPW] and PTP1B [(**f**) 2AZR (**g**) 2H4G (**h**) 2H4K (**i**) 2HB1] complexes with symbols of amino acids and ligand, hydrogen bonds are shown by dotted lines

Among the selected HIV-1 complexes, 1D4K has the highest number of hydrogen bond interactions between protein and inhibitor while 1HVK has the lowest such interactions. Among all the PTP1B complexes, 1Q6P and 1Q6S exist as dimers made up of chains A&B while all other systems consist of only chain 'A'. For direct comparison of these dimers with other PTP1B systems, we split into two monomers with chains A and B, the obtained results are summarized in Table 2. The highest number of 16 protein-inhibitor interactions is found in 1PYN while the lowest numbers of such interactions are located in 1KAV. The protein-protein interactions for one chain are the highest (352 interactions) in 2NT7 whereas it is the lowest (303 interactions) in 1PH0.

Details of various hydrogen bonds present in 1D4L

As a representative example, more detailed analysis of hydrogen bonds on the basis of interacting atoms is given for one of the HIV-1 Protease-Inhibitor complex (PDB code: 1D4L) [43] where the inhibitor ligand PI9 is a macro cyclic peptidomimetic molecule. The salient geometrical features of these hydrogen bonds including the average values of the geometric parameters used are shown in Table 3.

HBPredicT also generates histograms showing the frequency of the hydrogen bonds with all the distance and angle parameters used to identify them (Fig. 4). In the case of 1D4L, a high degree of hydrogen bonds (58%) are identified in the hydrogen-acceptor (H-A) distance range of 1.7 to 2.1 Å

Table 7 CH···π interactions found in HIV-1 protease crystal structures

PDB Id	Total CH···π interactions	CH···π interactions between protein and ligand	CH···π interaction between		
			CH-donor	Chain identifier[a]	π-donor
1AAQ	12	1	VAL(82)	B	PSI(1)
1C6Y	15	0	–	–	–
1D4H	15	0	–	–	–
1D4J	17	0	–	–	–
1D4K	18	0	–	–	–
1D4L	18	2	ALA(28)	A	PI9(201)
			ILE(47)	A	PI9(201)
1HIH	14	0	–	–	–
1HIV	16	2	CAV(3)	I	NOA(1)
			CAV(3)	I	NOA(1)
1HVI	3	0	–	–	–
1HVJ	2	0	–	–	–
1HVK	4	0	–	–	–
1HVL	3	0	–	–	–
1HXB	10	1	HPH(3)	I	QNC(1)
1N49[b]	11	2	ALA(82)	A	RIT(301)
			ALA(82)	B	RIT(301)
1N49[c]	11	1	ALA(82)	C	RIT(401)
1NPW	14	0	–	–	–
1ODY	10	0	–	–	–
1W5W	18	1	VAL(82)	A	BE4(1100)
2PSU	16	1	ALA(28)	A	MUU(200)
2PSV	14	0	–	–	–

[a] 'I' represent CH···π interaction exist within the same ligand or between two different ligands.

[b] 1 N49 is a dimeric unit and the results belongs to A, B chains, the total CH···π interactions are given in Table S3.

[c] Results belongs to chains C, D.

(Fig. 4a). Moreover, about 70% of the hydrogen bonds showed the values of ∠D-H-A and ∠H-A-D angles in the range of 150° to 180° (Fig. 4b) and 58% between 0° to 15° (Fig. 4c), which indicate the domination of linear nature of the identified hydrogen bonds. Furthermore, the distribution of the hydrogen bonds based on the $\angle S_m$-A-H is shown in Fig. 4d.

Table 4 shows the distribution of hydrogen bonds among various amino acids present in 1D4L. Isoleucine has contributed 62 hydrogen bonds with the other amino acid residues. Threonine formed 20 hydrogen bonds with the water molecules whereas serine and cystiene residues have no contribution to hydrogen bonding. Isoleucine and aspartic acid contributed one hydrogen bond each with inhibitor PI9 while leucine, arginine, asparagine, and threonine contributed two such interactions each. Further, the distribution related to the classification of amino acids based on charge and hydrophobicity is shown in Table 5. About half of the hydrogen bonds formed within the amino acids are of the non-polar aliphatic amino acids. Hydrophilic and hydrophobic amino acids respectively formed 66% and 22% of hydrogen bonds with water. The user is able to visualize these hydrogen bonds through the Rasmol interface.

Structural water molecules in HIV-1 protease and PTP1B complexes

In Table 6, six HIV-1 and four PTP1B inhibitor complexes with details of ligand bonded to protein through structural water molecules with respective chain, amino acid residues, and ligands are shown. A complete list of structural water molecules identified for remaining HIV-1 and PTP1B complexes selected in this study are available in supporting information. Figure 5 depicts the structural water identified at the active sites of some typical systems.

In 1D4H the water molecule '305' acts as a bridge between ILE(50) of chain A, ILE(150) of chain B of the protease and the ligand BEH (2,5-dibenzyloxy-3,4-dihydroxy-hexanedioic acid benzylamide (2-hydroxy-indan-1-yl)-amide) through hydrogen bonding. Water molecule '389' is observed as a bridge between ASP(29) and the

Table 8 CH···π interactions found in PTP1B crystal structures

PDB Id	Total CH···π interactions	CH···π interactions between protein and ligand	CH···π interaction between		
			CH-donor	Chain identifier[a]	π-donor
1KAV	39	1	ALA(217)	–	FEP(301)
1PH0	29	1	418(1)	A	TYR(46)
1PYN	36	1	941(1)	A	TYR(46)
1Q6P[b]	32	1	213(801)	A	TYR(546)
1Q6P[c]	29	2	213(1301)	B	TYR(1046)
			GLY(1259)	B	213(1301)
1Q6S[b]	38	4	ALA(717)	A	214(801)
			VAL(549)	A	214(801)
			MET(758)	A	214(801)
			214(801)	A	TYR(546)
1Q6S[c]	29	2	MET(1258)	B	214(1301)
			214(1301)	B	TYR(1406)
2AZR	40	3	ALA(217)	A	982(301)
			PHE(182)	A	982(301)
			982(301)	A	PHE(182)
2B07	38	2	ALA(217)	A	598(301)
			598(301)	I	598(301)
2H4G	40	1	ALA(217)	A	694(1)
2H4K	40	1	ALA(217)	A	509(1)
2HB1	42	1	ALA(217)	A	512(800)
2NT7	39	0	–	–	–

[a] 'I' represent CH...π interaction exist within the same ligand or between two different ligands.

[b] crystal structure exists as dimeric units and the total CH...π interaction between ligand and protein are given in Table S3.

[c] Data belongs to chain B.

BEH inhibitor. These two water molecules contributed in stabilizing the ligand interaction at the active site of the protein through hydrogen bonds as shown in Fig. 5a. In 1D4L, a single structural water molecule is located between ILE(50) of chain A and B with the inhibitor PI9 as shown in Fig. 5b. Similar results for the protein-water-ligand interactions at the active site are clearly seen in other systems (Fig. 5). In the case of PTP1B complexes of 2AZR, 2H4G, and 2HB1, two structural water molecules lying close to each other are located. Thus the program HBPredicT helps the user to view all the structural water molecules identified along with the surrounding amino acid residues and the inhibitor to which it forms the hydrogen bonds. In addition, it will show the H-A distance (d) of those hydrogen bonds and more information

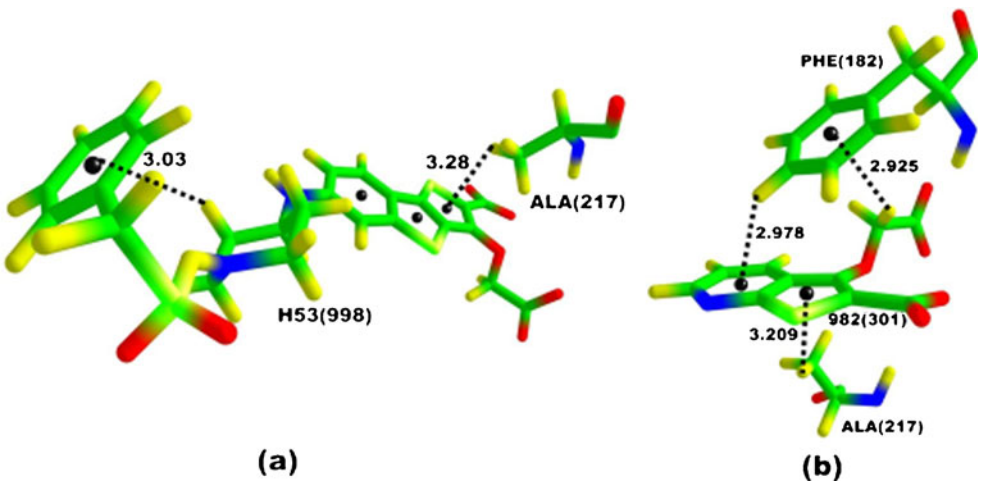


Fig. 6 CH···π interactions present in PTP1B inhibitors. (**a**) 2B07 showing an *intra*-ligand CH···π interaction at 3.03 Å (**b**) In 2AZR, the inhibitor 982 serving as C-H acceptor as well as donor for CH···π interaction. The black dot represents the center of mass of the π-group and the dotted lines represent the CH···π interactions. (Distances in Å)

Table 9 Details of the CH···π interactions found in HIV-1 protease complex of 1D4L7

Sl. No.	CH-donor[a]	π-donor[a]	$d_{C\text{-}X}$ (Å)	∠C-H-X (°)	$d_{Hp\text{-}X}$ (Å)	$d_{H\text{-}X}$ (Å)
1	ARG(57)	TRP(42)	3.769	162.2	1.196	2.716
2	PHE(99)	HIS(69)	4.017	138.9	0.252	3.130
3	PHE(99)	HIS(69)	4.165	139.1	1.009	3.280
4	ALA(28)	PI9(201)	4.348	132.9	1.031	3.532
5	ILE(47)	PI9(201)	4.224	172.1	0.959	3.142
6	THR(91)	TRP(6)	3.634	155.8	0.590	2.613
7	PRO(1)	PHE(99)	4.166	148.9	1.291	3.195
8	ILE(93)	PHE(99)	4.197	172.0	1.049	3.115
9	TRP(42)	TYR(59)	3.602	157.8	0.676	2.569
10	LEU(38)	TYR(59)	3.887	133.8	1.021	3.053
11	THR(91)	TRP(6)	4.313	149.9	1.216	3.336
12	ILE(93)	PHE(99)	4.004	135.4	0.557	3.154
13	PRO(1)	PHE(99)	3.926	153.0	1.241	2.923
14	LEU(38)	TYR(59)	4.399	138.5	1.315	3.524
15	TRP(42)	TYR(59)	3.743	158.4	0.814	2.708
16	ARG(57)	TRP(42)	3.578	145.6	0.981	2.626
17	ARG(57)	TRP(42)	3.574	166.8	0.500	2.505
18	GLY(48)	PHE(53)	3.458	128.6	0.373	2.672

[a] Protein residue numbers in respective PDB files are shown in parenthesis.

regarding the atom numbers, residue numbers and chain identifiers in "outfile.hbp." These details may be vital information in structure based drug discovery to develop potential inhibitors.

CH···π interactions in HIV-1 Protease and PTP1B complexes

We have identified an average of 12 CH···π interactions in HIV-1 complexes and 36 in PTP1B complexes. Interestingly, several structures have shown CH···π interactions between the protein residues and the ligand as well as *intra*-ligand CH···π interactions. The CH···π interactions found in selected HIV-1 protease and PTP1B complexes are listed in Tables 7 and 8, respectively. 1HIV and 1HXB, of HIV-1 protease showed inter ligand CH···π interactions. In the case of 1HIV and 1HXB, there are no CH…π interactions present between the protein and the ligand. Among the HIV-1 protease, the 1N49 system with two identical inhibitors namely RIT(301), RIT(401) showed a total of three protein-ligand CH···π interactions with ALA(82) of chains A, B, and C (Table 7). Among the PTP1B complexes except 2NT7, all the complexes showed CH···π interactions between protein and ligand. 1Q6P is also characterized by three protein-ligand CH···π interactions. Interestingly, in chain A of 1Q6S four CH···π interactions are observed from the inhibitor. In Fig. 6, the CH···π interactions found in 2AZR and 2B07 of PTP1B inhibitors are depicted. 2B07 shows *intra*-ligand CH···π interactions as well as CH···π interactions between ALA (217) and π-donor of ligand H53 (Fig. 6a). In 2AZR, the aromatic phenyl ring of PHE(182) serves as a π-donor to the inhibitor 982 and at the same time, its para hydrogen atom serves as the acceptor of π-density from the six membered heterocyclic ring of the inhibitor 982 at a distance of 2.978 Å (Fig. 6b).

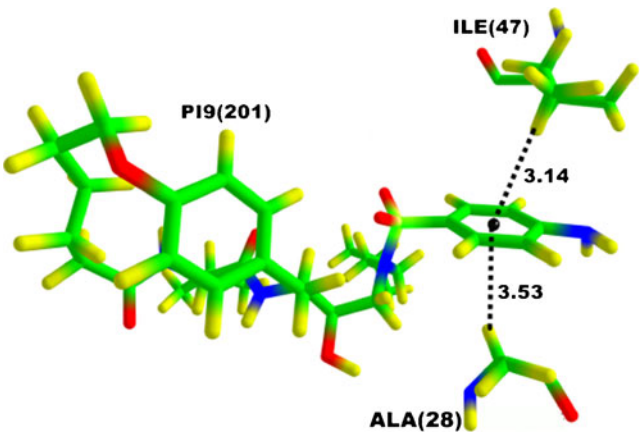


Fig. 7 CH···π interactions found between the inhibitor PI9 and the amino acid residues, alanine, isoleucine, of the HIV-protease system (1D4L). The black dot represents the center of mass of the π group and the dotted lines represent the CH···π interactions. (Distances in Å)

The CH···π interactions in 1D4L, including the details of the groups and the values of the parameters used for identifying the interactions are listed in Table 9. A total of 18 CH···π interactions are located and out of which two are between the protein and ligand (Fig. 7). Interestingly, the one phenyl ring acts as the π-donor for both the ligand-protein CH···π interactions (the one above the ring is from ILE(47) and the other from below the ring is from ALA

(28)). The CH···π interaction distance of ILE(47) is 3.14 Å and the ∠C-H-X of 172° shows its linear nature. Further, the $d_{Hp\text{-}X}$ value of 0.959 Å confirms that the CH bond is directed toward the π-face of the aromatic ring.

Conclusions

HBPredicT is a computational tool developed for identifying and analyzing hydrogen bonds in protein-ligand complexes. On the basis of specific geometric criteria it detects all conventional hydrogen bonds, structural water molecules and non-bonded CH···π interactions. The program gives a classification of identified hydrogen bonds based on distances, angles, hydrophobicity and type of bonds (CH···O, NH···O, NH···N, OH···O *etc*.) between the interacting protein residues. It also gives the details of the contribution of individual amino acids toward the protein-protein, protein-water, and protein-ligand hydrogen bonds. In order to test the application potential of HBPredicT, several HIV-1 and PTP1B complexes have been analyzed for hydrogen bonds, structural water molecules and CH...π interactions. All the systems showed presence of structural water molecules and in many cases of PTP1B complexes, close lying dimeric units of structural water molecules were located. HIV-1 complexes showed an average of 12 CH···π interactions while the PTP1B complexes exhibited an average of 36 CH···π interactions. The computational tool HBPredicT is expected to be useful for structure based drug designing, especially to understand the role of weak non-covalent interactions in ligand binding in proteins.

Acknowledgments The authors gratefully acknowledge the support from Council of Scientific and Industrial Research (CSIR), Government of India. F. B. S is thankful to CSIR for the senior research fellowship.

References

1. Jackson RC (1997) Contributions of protein structure-based drug design to cancer chemotherapy. Semin Oncol 24:164–172
2. Parlow JJ, Case BL, Dice TA, Fenton RL, Hayes MJ, Jones DE, Neumann WL, Wood RS, Lachance RM, Girard TJ, Nicholson NS, Clare M, Stegeman RA, Stevens AM, Stallings WC, Kurmbail RG, South MS (2003) Design, parallel synthesis, and crystal structures of pyrazinone antithrombotics as selective inhibitors of the tissue factor VIIa complex. J Med Chem 46:4050–4062
3. Rowland RS (2002) Using X-ray crystallography in drug discovery. Curr Opin Drug Discov Develop 5:613–619
4. Terasaka T, Kinoshita T, Kuno M, Seki N, Tanaka K, Nakanishi I (2004) Structure-based design, synthesis, and structure-activity relationship studies of novel non-nucleoside adenosine deaminase inhibitors. J Med Chem 47:3730–3743
5. Blundell TL (1996) Structure-based drug design. Nature 384S:23–26
6. Kellogg GE, Abraham DJ (2000) Hydrophobicity: is LogP(o/w) more than the sum of its parts? Eur J Med Chem 35:651–661
7. Nishio M, Umezawa Y, Hirota M, Takeuchi Y (1995) The CH/π interaction: significance in molecular recognition. Tetrahedron 51:8665–8701
8. Jeffrey GA, Saenger W (1991) Hydrogen bonding in biological structures. Springer, Berlin
9. Poornima CS, Dean PM (1995) Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. J Comput Aided Mol Des 9:500–512
10. Fornabaio M, Spyrakis F, Mozzarelli A, Cozzini P, Abraham DJ, Kellogg GE (2004) Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 3. The free energy contribution of structural water molecules in HIV-1 protease complexes. J Med Chem 47:4507–4516
11. Ladbury JE (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. Chem Biol 3:973–980
12. Connelly PR (1994) Acquisition and use of calorimetric data for prediction of the thermodynamics of ligand-binding and folding reactions of proteins. Curr Opin Biotechnol 5:381–388
13. Suresh CH, Vargheese AM, Vijayalakshmi P, Neetha M, Koga N (2008) Role of structural water molecule in HIV protease-inhibitor complexes: a QM/MM study. J Comput Chem 29:1840–1849
14. Fischer S, Verma CS (1999) Binding of buried structural water increases the flexibility of proteins. Proc Natl Acad Sci U S A 96:9613–9615
15. Nishio M, Umezawa Y, Honda K, Tsuboyama S, Suezawa H (2009) CH/π hydrogen bonds in organic and organometallic chemistry. Cryst Eng Comm 11:1757–1788
16. Ozawa T, Tsuji E, Ozawa M, Handa C, Mukaiyama H, Nishimura T, Kobayashi S, Okazaki K (2008) The importance of CH/π hydrogen bonds in rational drug design: an ab initio fragment molecular orbital study to leukocyte-specific protein tyrosine (LCK) kinase. Bioorg Med Chem 24:10311–10318
17. Harigai M, Kataoka M, Imamoto Y (2006) A single CH/π weak hydrogen bond governs stability and the photocycle of the photoactive yellow protein. J Am Chem Soc 128:10646–10647
18. Muraki M (2002) The importance of CH/π interactions to the function of carbohydrate binding proteins. Protein Pept Lett 9:195–209
19. Huber RE, Hakda S, Cheng C, Cupples CG, Edwards RA (2003) Trp-999 of β-galactosidase (Escherichia coli) is a key residue for binding, catalysis, and synthesis of allolactose, the natural Lac operon inducer. Biochemistry 42:1796–1803
20. Poole DM, Hazlewood GP, Huskisson NS, Virden R, Gilbert HJ (1993) The role of conserved tryptophan residues in the interaction of a bacterial cellulose binding domain with its ligand. FEMS Microbiol Lett 106:77–84
21. Davis R, Saleesh Kumar NS, Abraham S, Suresh CH, Rath NP, Tamaoki N, Das S (2008) Molecular packing and solid-state fluorescence of alkoxy-cyano substituted diphenylbutadienes: structure of the luminescent aggregates. J Phys Chem C 112:2137–2146
22. Radhakrishnan KV, Anas S, Suresh E, Koga N, Suresh CH (2007) Molecular recognition in an organic host-guest complex: CH···O and CH···π interactions completely control the crystal packing and the host-guest complexation. Bull Chem Soc Jpn 80:484–490
23. Spiwok V, Lipovova P, Skalova T, Buchtelova E, Hasek J, Kralova B (2004) Role of CH/π interactions in substrate binding by Escherichia coli β-galactosidase. Carbohydr Res 339:2275–2280
24. Chakrabarti P, Samanta U (1995) CH/π interaction in the packing of the adenine ring in protein structures. J Mol Biol 251:9–14
25. Umezawa Y, Nishio M (2005) CH/π hydrogen bonds as evidenced in the substrate specificity of acetylcholine esterase. Biopolymers 79:248–258

26. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graphics 14:33–38
27. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18:2714–2723
28. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 285:1735–1747
29. Tiwari A, Panigrahi SK (2007) HBAT: a complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. In Silico Biol 7:651–661
30. Lindauer K, Bendic C, Sühnel J (1996) HBexplore-a new tool for identifying and analysing hydrogen bonding patterns in biological macromolecules. Comput Appl Biosci 12:281–289
31. McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. J Mol Biol 238:777–793
32. Kaur H, Raghava GPS (2006) Prediction of Cα-H···O and Cα-H···π interactions in proteins using recurrent neural network. In Silico Biol 6:111–125
33. Evans JL, Jallal B (1999) Protein tyrosine phosphatases: Their role in insulin action and potential as drug targets. Exp Opin Invest Drugs 8:139–160
34. Hashimoto N, Zhang WR, Goldstein BJ (1992) Insulin receptor and epidermal growth factor receptor dephosphorylation by three major rat liver protein-tyrosine phosphatases expressed in a recombinant bacterial system. Biochem J 284:569–576
35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
36. Sayle RA, Milner-White E (1995) RASMOL: Biomolecular graphics for all. J Trends Biochem Sci 20:374–376
37. Budayova-Spano M, Fisher SZ, Dauvergne M, Agbandje-McKenna M, Silverman DN, Myles DAA, McKenna R (2006) Production and X-ray crystallographic analysis of fully deuterated human carbonic anhydrase II. Acta Crystallogr F 62:6–9
38. Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 44:97–179
39. Desiraju GR (2002) Hydrogen bridges in crystal engineering: Interactions without borders. Acc Chem Res 35:565–573
40. Desiraju GR (1991) The C-H-O hydrogen bond in crystals: What is it? Acc Chem Res 24:290–296
41. Brandl M, Weiss MS, Jabs A, Sühnel J, Hilgenfeld R (2001) C-H···π-interactions in proteins. J Mol Biol 307:357–377
42. Hosur MV, Bhat TN, Kempf D, Baldwin ET, Liu B, Gulnik S, Wideburg NE, Norbeck DW, Appelt K, Erickson JW (1994) Influence of stereochemistry on activity and binding modes for C2 symmetry-based diol inhibitors of HIV-1 protease. J Am Chem Soc 116:847–855
43. Tyndall JDA, Reid RC, Tyssen DP, Jardine DK, Todd B, Passmore M, March DR, Pattenden LK, Bergman DA, Alewood D, Hu S, Alewood PF, Birch CJ, Martin JL, Fairlie DP (2000) Synthesis, stability, antiviral activity, and protease-bound structures of substrate-mimicking constrained macrocyclic inhibitors of HIV-1 protease. J Med Chem 43:3495–3504
44. Kervinen J, Lubkowski J, Zdanov A, Bhatt D, Dunn BM, Hui KY, Powell DJ, Kay J, Wlodawer A, Gustchina A (1998) Toward a universal inhibitor of retroviral proteases: comparative analysis of the interactions of LP-130 complexed with proteases from HIV-1, FIV, and EIAV. Protein Sci 7:2314–2323
45. Dreyer GB, Lambert DM, Meek TD, Carr TJ, Tomaszek JTA, Fernandez AV, Bartus H, Cacciavillani E, Hassell AM, Minnich M, Petteway JSR, Metcalf BW (1992) Hydroxyethylene isostere inhibitors of human immunodeficiency virus-1 protease: structure-activity analysis using enzyme kinetics, X-ray crystallography, and infected T-cell assays. Biochemistry 31:6646–6659
46. Lindberg J, Pyring D, Lowgren S, Rosenquist A, Zuccarello G, Kvarnstrom I, Zhang H, Vrang L, Classon B, Hallberg A, Samuelsson B, Unge T (2004) Symmetric fluoro-substituted diol-based HIV protease inhibitors: ortho-fluorinated and meta-fluorinated P1/P1′-benzyloxy side groups significantly improve the antiviral activity and preserve binding efficacy. Eur J Biochem 271:4594–4602
47. Chellappan S, Kiran Kumar Reddy GS, Ali A, Nalam MN, Anjum SG, Cao H, Kairys V, Fernandes MX, Altman MD, Tidor B, Rana TM, Schiffer CA, Gilson MK (2007) Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis. Chem Biol Drug Des 69:298–313
48. Jia Z, Ye Q, Dinaut AN, Wang Q, Waddleton D, Payette P, Ramachandran C, Kennedy B, Hum G, Taylor SD (2001) Structure of protein tyrosine phosphatase 1B in complex with inhibitors bearing two phosphotyrosine mimetics. J Med Chem 44:4584–4594
49. Liu G, Xin Z, Liang H, Abad-Zapatero C, Hajduk PJ, Janowick DA, Szczepankiewicz BG, Pei Z, Hutchins CW, Ballaron SJ, Stashko MA, Lubben TH, Berg CE, Rondinone CM, Trevillyan JM, Jirousek MR (2003) Selective protein tyrosine phosphatase 1B inhibitors: Targeting the second phosphotyrosine binding site with non-carboxylic acid-containing ligand. J Med Chem 46:3437–3440
50. Prabu-Jeyabalan M, Nalivaika EA, King NM, Schiffer CA (2003) Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: Structural insights for better antiviral therapy. J Virol 77:1306–1315
51. Pei Z, Li X, Liu G, Abad-Zapatero C, Lubben T, Zhang T, Ballaron SJ, Hutchins CW, Trevillyan JM, Jirousek M (2003) Discovery and SAR of novel, potent and selective protein tyrosine phosphatase 1B inhibitors. Bioorg Med Chem Lett 13:3129–3132
52. Thanki N, Rao JK, Foundling SI, Howe WJ, Moon JB, Hui JO, Tomasselli AG, Heinrikson RL, Thaisrivongs S, Wlodawer A (1992) Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: Comparisons with molecular modeling. Protein Sci 1:1061–1072
53. Scapin G, Patel SB, Becker JW, Wang Q, Desponts C, Waddleton D, Skorey K, Cromlish W, Bayly C, Therien M, Gauthier JY, Li CS, Lau CK, Ramachandran C, Kennedy BP, Asante-Appiah E (2003) The structural basis for the selectivity of benzotriazole inhibitors of PTP1B. Biochemistry 42:11451–11459
54. Moretto AF, Kirincich SJ, Xu WX, Smith MJ, Wan ZK, Wilson DP, Follows BC, Binnun E, Joseph-McCarthy D, Foreman K, Erbe DV, Zhang YL, Tam SK, Tam SY, Lee J (2006) Bicyclic and tricyclic thiophenes as protein tyrosine phosphatase 1B inhibitors. Bioorg Med Chem 14:2162–2177
55. Wan ZK, Lee J, Xu W, Erbe DV, Joseph-McCarthy D, Follows BC, Zhang YL (2006) Monocyclic thiophenes as protein tyrosine phosphatase 1B inhibitors: capturing interactions with Asp48. Bioorg Med Chem Lett 16:4941–4945
56. Krohn A, Redshaw S, Ritchie JC, Graves BJ, Hatada MH (1991) Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. J Med Chem 34:3340–3342
57. Smith AB III, Cantin LD, Pasternak A, Guise-Zawacki L, Yao W, Charnley AK, Barbosa J, Sprengeler PA, Hirschmann R, Munshi S, Olsen DB, Schleif WA, Kuo LC (2003) Design, synthesis, and biological evaluation of monopyrrolinone-based HIV-1 protease inhibitors. J Med Chem 46:1831–1844
58. Wan ZK, Follows B, Kirincich S, Wilson D, Binnun E, Xu W, Joseph-McCarthy D, Wu J, Smith M, Zhang YL, Tam M, Erbe D, Tam S, Saiah E, Lee J (2007) Probing acid replacements of thiophene PTP1B inhibitors. Bioorg Med Chem Lett 17:2913–2920