Amino Acids (2008) 34:617–633
DOI 10.1007/s00726-007-0015-4

ORIGINAL ARTICLE

Strong and weak hydrogen bonds in protein-ligand complexes of kinases: a comparative study

Sunil K. Panigrahi

Received: 9 October 2007 / Accepted: 8 December 2007 / Published online: 8 January 2008

Abstract Strong and weak hydrogen bonds between protein and ligand are analyzed in a group of 233 X-ray crystal structures of the kinase family. These kinases are from both eukaryotic and prokaryotic organisms. The dataset comprises of 44 sub-families, out of which 35 are of human origin and the rest belong to other organisms. Interaction analysis was carried out in the active sites, defined here as a sphere of 10 Å radius around the ligand. A majority of the interactions are observed between the main chain of the protein and the ligand atoms. As a donor, the ligand frequently interacts with amino acid residues like Leu, Glu and His. As an acceptor, the ligand interacts often with Gly, and Leu. Strong hydrogen bonds N–H···O, O–H···O, N–H···N and weak bonds C–H···O, C–H···N are common between the protein and ligand. The hydrogen bond donor capacity of Gly in N–H···O and C–H···O interactions is noteworthy. Similarly, the acceptor capacity of main chain Glu is ubiquitous in several kinase sub-families. Hydrogen bonds between protein and ligand form characteristic hydrogen bond patterns (supramolecular synthons). These synthon patterns are unique to each sub-family. The synthon locations are conserved across sub-families due to a higher percentage of conserved sequences in the active sites. The nature of active site water molecules was studied through a novel classification scheme, based on the extent of exposure of water molecules. Water which is least exposed usually participates in hydrogen bond formation with the ligand. These findings will help structural biologists, crystallographers and medicinal chemists to design better kinase inhibitors.

S. K. Panigrahi (✉)
School of Chemistry, University of Hyderabad,
Hyderabad 500 046, India
e-mail: panigrahisk@yahoo.com

Keywords Hydrogen bond · Kinase · Protein Data Bank · Supramolecular synthon · Water

Abbreviations

PKs	Protein kinases
PKR	Protein kinase resource
PDB	Protein Data Bank
HBAT	Hydrogen bond analysis tool
SVL	Scientific vector language

Introduction

Hydrogen bond is one of the most important interactions between biologically important molecules (Jeffrey and Saenger 1991; Baker and Hubbard 1984; Desiraju and Steiner 1999). The three-dimensional architecture of proteins and nucleic acids is stabilized by hydrogen bonds, biological recognition operates through hydrogen bonding, and the molecular mobility required for biological processes is directly connected with the rapid formation and breaking of hydrogen bonds. The importance of hydrogen bonds in substrate/ligand recognition in macromolecules is an active area of research (Sarkhel and Desiraju 2004; Panigrahi and Desiraju 2007; Aparna et al. 2005; Glusker 1995; Bartlett et al. 2002). Hydrogen bonds are observed with a variety of strengths and geometries in the active sites of protein–ligand complexes (Sarkhel and Desiraju 2004; Panigrahi and Desiraju 2007; Steiner 2002; Bartlett et al. 2002).

The existence of strong and weak hydrogen bonds in the protein–ligand complexes has been demonstrated by us earlier in a dataset of 251 protein–ligand complexes (Panigrahi and Desiraju 2007). The following are some of

the important conclusions of our earlier study: (a) the ubiquitous presence of strong and weak hydrogen bonds in the protein–ligand interface; (b) the linearity of N–H···O and O–H···O bonds; (c) the occurrence of multifurcated hydrogen bonds; (d) the resolution limits are crucial in studying hydrogen bond geometries; (e) the hydrogen bond geometry of water and amino acid residues like Gly and Tyr are significant in the active sites. These important conclusions derived from this dataset were also validated against 233 protein–ligand complexes of kinase family. In this context, there is a growing amount of literature which emphasizes the importance of strong and weak hydrogen bonds in the protein–ligand interface in protein kinases (PKs) (Pierce et al. 2002; Pierce et al. 2005). The present work aims at analyzing the strong and weak hydrogen bonds in PKs in greater detail. A comprehensive study of hydrogen bond patterns, between the main/side chain and ligand across the kinase sub-families, was carried out. The hydrogen bonds between protein and ligand form characteristic hydrogen bond patterns, which have been described through supramolecular synthon approach. The term supramolecular synthon is well known in small molecular crystallography and crystal engineering literature (Desiraju 1995). In the present study, the characteristic hydrogen bond patterns forming synthons between various scaffolds present in the ligand molecules and the main/side chain of proteins in the active site are considered. The importance of conserved residues and interactions forming hydrogen bonds in kinase sub-families are studied through the synthon approach. The water environment in the active sites was also studied.

The therapeutic usages of PKs have recently opened up many research avenues (Cohen 2002; Levitzki 2003; Bridges 2001; Vieth et al. 2004). Till date there are around 450 entries of kinases available in the PDB (Berman et al. 2000; http://www.kinasenet.org/pkr). The abundance of structural information for PKs provides an ideal background for structure-based drug discovery (Vieth et al. 2004; Williams and Mitchell 2002). Similar to various structural initiatives, online resources specific to PKs on various aspects of the kinase family are emerging constantly (http://198.202.68.14/human/kinome/phylogeny.html, http://kinasedb.ontology.ims.u-tokyo.ac.jp, Naumann and Matter 2002). A typical catalytic domain of kinase has 250–300 amino acids and is bilobal in nature (Hanks and Hunter 1995; Dar et al. 2005; Panigrahi and Desiraju 2004). This two-lobed structure can be further subdivided into 12 sub-domains (I–XI). The N-terminal lobe constituting sub-domains I–IV primarily has antiparallel β-sheets with the important exception of α-helix C. Sub-domain V is a single polypeptide chain known as the linker region connecting the N-terminal lobe to the larger C-terminal lobe comprising sub-domains VIA–XI. The C-terminal is predominantly helical in nature. A full explanation of the structure and function of each region is available elsewhere at the protein kinase resource (PKR) website (http://www.kinasenet.org/pkr). The ATP binds in the cleft formed between the N- and C-terminal lobes of the PKs, forming several key interactions conserved across the protein kinase family (Hanks and Hunter 1995; Denessiouk and Johnson 2003). The adenine moiety lays in a hydrophobic region between the β-sheet structure of sub-domains I and II and residues from sub-domains V and VIb. A large number of protein kinase inhibitors have been observed to mimic the donor–acceptor pair of hydrogen bonds made between the protein backbone and adenine. The present study attempts to understand the nature of donor–acceptor hydrogen bonds in the expanded context of hydrogen bonding—one wherein both strong N–H···O and O–H···O and weaker C–H···O interactions are considered.

With this backdrop I assume that the protein–ligand interaction in the active site of PKs provides valuable information to understand the basis of molecular recognition through strong and weak hydrogen bonds. I have discussed the qualitative aspects of hydrogen bonds while describing synthon phenomena within and among the kinase sub-families. This aspect should not be weighed strictly on the basis of the quality of crystallographic data and the over-representation of data for any particular sub-family of PKs. Instead, it should be treated as a qualitative analysis of strong and weak hydrogen bonds present in the protein–ligand complexes of kinases, and its importance in drug discovery. It is hoped that these findings will help structural biologists, crystallographers and medicinal chemists to design better kinase inhibitors.

Materials and methods

Dataset

A set of 233 X-ray structures of kinase protein–ligand complexes from the PDB was used (Berman et al. 2000). This is the same dataset used as the test set in our earlier study (Panigrahi and Desiraju 2007). These structures have a resolution limit of 1.2 to 3.5 Å. The structures are classified into 44 sub-families of kinase. The classification of structure was carried out based on publicly available resources (Vieth et al. 2004; Berman et al. 2000, http://www.kinasenet.org/pkr, Wang et al. 2005). These kinases are from both eukaryotic and prokaryotic organisms. A detailed verification of the sub-family information was verified across various online databases like PKRs, PDBbind database, sugen, structure and phylogeny of the protein kinases home page. The active site was defined by

selecting all amino acid residues within a 10-Å radius of the ligand molecule. The active site also includes water molecules. The ligand selection method was taken from an earlier study (Sarkhel and Desiraju 2004).

Geometry optimization

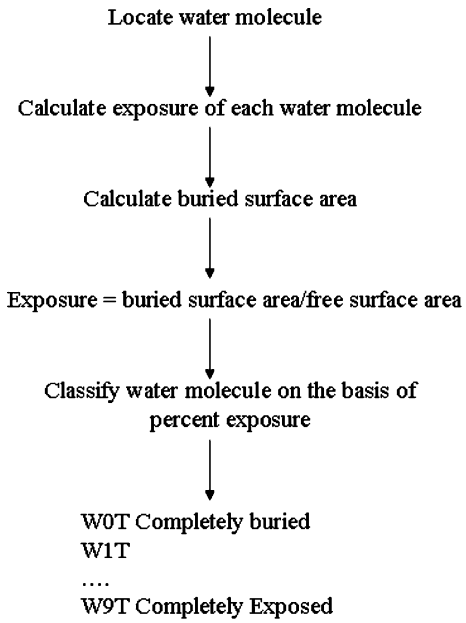
Macromolecular crystal structures rarely contain H-atom positional data with the precision required to properly evaluate hydrogen bond geometry. Therefore a method must be found to add or modify all the H-atom positions. H-atoms were added to the protein, water and ligand using the MOE Program (2006). The H-atom positions were then refined (energy minimization) keeping the position of the non-H atoms fixed using the MMFF94x force field (Halgren 1996). All the optimized structures were exported to the hydrogen bond analysis tool (HBAT) for hydrogen bond analysis (Tiwari and Panigrahi 2007). To evaluate the accuracy of method adopted here, I have used HBAT to reproduce the published geometries in several recent papers (Sarkhel and Desiraju 2004; Brandl et al. 2001; Koellner et al. 2002; Klaholz and Moras 2002; Auffinger et al. 2004). I hope that a similar exercise applied to any system of choice will reproduce reasonably acceptable results.

Hydrogen bond analysis

Strong and weak hydrogen bonds were analyzed with an in-house developed program, HBAT, which analyzes and tabulates all hydrogen bonds present in a PDB file. The output file provides distance-angle distributions across various geometry ranges while tabulation of frequencies for each residue, ligand, water, and nucleic acids is done easily for any kind of interaction. HBAT is a user-friendly desktop tool, which operates both with default and user-selected parameters. The standard H-bonding criteria were set as d (H···A) $\leq$ 2.8 Å and θ (X–H···A) $\geq$ 90°. The hydrogen bond synthon analysis was carried out manually. A single structure from each sub-family was randomly selected for sequence alignment. The active site sequence alignment was carried out with the help of ClustalW (Higgins et al. 1994).

Water in the active sites of kinase

Water molecules in the active sites were classified into ten categories (W0T–W9T) based on solvent accessibility (exposure). The categorization of water is on the basis of percent exposure; for example, 0% exposure (completely buried) is included in W0T, 0–10% exposure are included in W1T and so on up to 90–100% in W10T. This approach is very similar to the method adopted by Williams et al. (Williams et al. 1994). The authors have described a hierarchical classification of the extent of burial of water molecules in the protein cavity. In the present context the active sites are considered as the cavity for ligand binding. This classification scheme was carried out with the help of a scientific vector language (SVL) code written in the MOE software. An overview of the method is given in the flow sheet (Scheme 1).



Scheme 1 Flow sheet for water classification in the active site

Results and discussion

Hydrogen bond is the key interaction for ATP and inhibitor binding in the kinase family. This is especially true of the hydrogen bonds that involve the main chain. The initial aim of the present paper is to elucidate strong and weak hydrogen bond geometries between the main chain, side chain and the ligand. However, interactions between the main chain and the ligand are more important. This is because (a) in kinases, the hydrogen bond geometries (d and θ) involving main chains are more consistent than those formed by the side chains (Panigrahi and Desiraju 2007); (b) hydrogen bond geometries for the side chain are more flexible compared to the main chain; (c) the majority of the hydrogen bonds in the present dataset is observed between the main chain and the ligand atoms (195 out of 233 complexes); (d) main chain hydrogen bonds are crucial for proper a positioning of ligands (Aparna et al. 2005).

Residue frequency

The total number of residues present in these complexes is 66,405, that is, on average there are 285 residues present in each structure. The percentage distribution of nonpolar, charged and polar residues in the dataset is 50, 27 and 23%, respectively. The active sites contain 18,997 residues, so that each active site has around 80 residues on average. The percentage distribution for nonpolar, charged and polar residues in the active sites is 53, 27 and 20%, respectively. Among the nonpolar residues, the most frequently observed are Leu, Val, Ala and Gly. Charged residues like Lys, Glu, Asp are abundant in the active site. This finding suggests that for kinases the amino acid composition of both the entire protein and the active site is hydrophobic in nature, which is expected. However, the higher occurrence of charged residues when compared to the polar residues is uncommon (Panigrahi and Desiraju 2007). The probable reason for this observation might be the importance of electrostatic interaction in ligand binding and the nature of bound ligand. In fact, Bartlett et al. have shown in their study that, the majority of catalytic residues are charged residues in 178 enzyme active sites (Bartlett et al. 2002). Generally the charged residues are present at the surface and active sites of proteins (Gitlin et al. 2006). The percentage occurrence of various amino acid residues in the total dataset and in the active sites is shown in Fig. 1.

The total number of hydrogen bonds observed between the ligand and the main chain is 2,073. Out of the 2,073 interactions, the ligand donates hydrogen bonds in 933 cases and accepts them in 1,140 cases (Table 1). Among these interactions, 53% interactions belong to the nonpolar, 15% to the polar and 29% to the charged amino acid residues. As a donor, the frequently interacting residues are

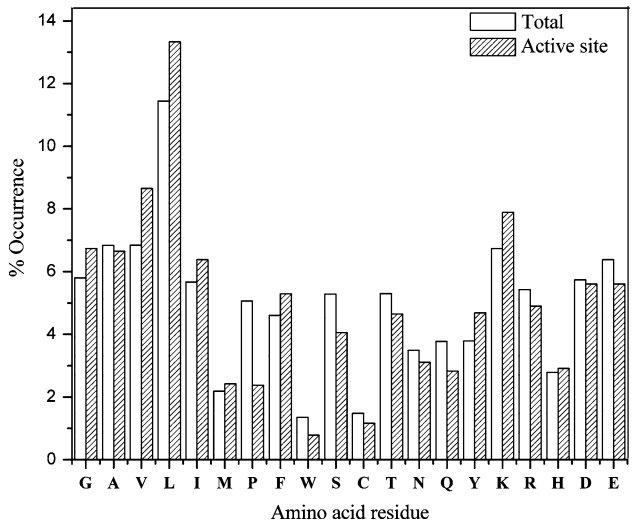


Fig. 1 Distribution of amino acid residues in total and active sites of kinases

Leu (17%), Glu (20%) and His (13%). As an acceptor, the ligand interacts with the residues, Gly (24%), and Leu (10%). These figures represent the two aspects of the donor/acceptor functional groups present in the active sites of kinases. First, the number of donor and acceptor in case of ligand suggests that there are more number of acceptor functional groups present in ligands than in donors. This is an attribute of the donor acceptor aspect in Lipinski's rule of five (Lipinski et al. 1997). Second, the percentage of amino acid types interacting with the ligand indicate that the hydrogen bond observed in the active sites of kinases are mostly contributed by the nonpolar followed by the charged and polar residues. This is because of the hydrophobic nature of the active sites and contribution of the charged residues in providing the anchoring points for ligand binding in kinases. This second aspect is represented by the nonpolar residues like Leu, Gly and charged residues like Glu and His.

The ligand interacts with the main chain through strong and weak hydrogen bonds (Table 2). For reasons of clarity, the hydrogen bond abbreviation consists of three parts: hydrogen bond type, donor, and acceptor. B stands for

Table 1 Percent contribution of main chain donor/acceptor in ligand interaction

	Ligand donor main chain acceptor (%)	Main chain donor ligand acceptor (%)
Nonpolar		
Gly	4.5	24.4
Val	5.0	5.3
Leu	17.0	10.4
Ilu	5.7	1.4
Met	3.7	4.4
Pro	1.5	0.7
Phe	2.3	7.0
Trp	0.4	0.1
Ala	4.2	5.3
Polar		
Ser	3.2	6.0
Cys	1.9	1.8
Thr	1.4	5.7
Asn	0.4	1.4
Gln	4.9	2.0
Tyr	2.4	5.0
Charged		
Lys	2.6	4.5
Asp	3.3	5.0
Arg	2.9	4.1
Glu	19.9	4.2
His	13.0	1.0
Total no.	933	1,140

Table 2 Percent distribution of strong and weak hydrogen bonds among amino acid residues

	{NHO BD LA} (%)	{CHO BD LA} (%)	{NHN BD LA} (%)	{CHN BD LA} (%)	{NHO LD BA} (%)	{CHO LD BA} (%)	{OHO LD BA} (%)
Nonpolar							
Gly	**20.0**	**41.4**	0.7	17.8	2.4	4.7	10.9
Ala	8.9	3.4	2.0	–	1.9	5.1	3.6
Val	6.1	2.4	12.2	2.5	4.7	5.0	3.6
Leu	4.3	3.4	**40.9**	20.3	27.4	14.8	5.4
Ilu	0.2	3.9	0.7	1.7	0.5	7.6	1.8
Met	1.9	0.5	26.5	–	3.3	4.0	–
Pro	–	2.1	–	–	0.9	1.5	1.8
Phe	3.4	5.3	0.7	**34.7**	0.5	2.6	5.4
Trp	0.2	–	–	–	0.5	0.4	–
Polar							
Ser	8.3	7.2	–	–	4.2	2.8	5.4
Cys	2.1	0.5	4.8	–	2.3	1.6	–
Thr	12.1	1.0	–	–	0.5	1.6	–
Asn	1.6	1.3	–	2.5	–	0.4	1.8
Gln	0.8	4.2	4.8	0.8	1.4	5.3	12.7
Tyr	1.9	8.2	–	14.4	0.9	2.9	3.6
Charged							
Lys	8.3	2.1	1.4	–	5.2	1.3	7.3
Arg	4.1	5.3	2.0	2.5	2.8	2.6	5.4
His	0.2	1.9	0.7	0.8	8.0	15.6	1.8
Asp	6.7	5.3	1.4	0.8	3.3	3.2	5.4
Glu	8.3	1.0	1.3	0.8	**28.9**	**16.8**	**23.6**
Total	514	397	167	118	211	611	55

The bold indicates percentage majority

backbone, L is ligand, W is water, D is donor, and A is acceptor. For example {NHO BD LA} signifies an N–H···O hydrogen bond involving a backbone N–H donor and a ligand O-atom acceptor. As an acceptor, the ligand interacts with the main chain through strong and weak hydrogen bonds, of which {NHO BD LA} 514 interactions, {CHO BD LA} 397, {NHN BD LA} 167, {CHN BD LA}118 are noteworthy. Similarly the ligand donates hydrogen bonds to the main chain, {NHO LD BA} 211, {CHO LD BA} 611, {OHO LD BA} 55 hydrogen bonds.

The percentage of {NHO BD LA} and {CHO BD LA} hydrogen bond interactions is highest for Gly. This reiterates the importance of Gly in the active sites of protein–ligand complexes. The possible reason for the occurrence of Gly in active sites has been discussed by us earlier and has to do with the small size and large conformational flexibility of this residue (Sarkhel and Desiraju 2004; Panigrahi and Desiraju 2007). The percentage of hydrogen bonds involving {NHN BD LA} is highest for Leu. This is because Leu is one of the frequently interacting residues in the active sites of CDK2 and also because the sub-family CDK2 represents the highest number of entries (49) in the dataset. For other types of hydrogen bonds, the percentage contribution of Leu is among the highest. The same is the case for Glu, wherein {NHO LD BA}, {CHO LD BA} and {OHO LD BA} hydrogen bonds are the most frequent. Glu is also one of the residues present at the linker region of CDK2 kinase. Apart from the CDK2 sub-family, Glu interacts commonly with ligand in other sub-families like cAMP-dependent protein kinase, PKA, FGFR1 and INSR. The importance of these two residues can be observed in Table 3. Phe interacts most frequently through {CHN BD LA} hydrogen bonds. In summary, various types of hydrogen bonds involving Gly, Leu and Glu are important in ligand binding in kinases. The Hydrogen bond geometries for {NHO BD LA} in Gly and {CHO LD BA} in Glu are shown in Fig. 2. For strong N–H···O hydrogen bonds in Gly, the median H···O distance, *d*, is less than 2.2 Å. The cone-corrected angular distributions for N–H···O has maxima in the range 160–165°. For the weak C–H···O hydrogen bond in Glu, the cone-corrected angular distributions have maxima in the range 165–170°. The inverse length-angle correlations are accordingly well behaved in both N–H···O and C–H···O hydrogen bonds. The weak

Table 3 Kinase protein–ligand complexes in this study. Main chain and side-chain interactions are represented by M and S, respectively. Hydrogen bond synthons are mentioned. Synthons and residues follow same order. Donor and acceptor are denoted as 'd' and 'a', respectively. Fields left blank represent complexes where other weak interactions like halogen bonds and water mediated protein–ligand interactions are observed

S. no.	Kinase sub-family	Source	PDB ID	M/S	Synthon type	Residue
1	ABL1	Human	PDB ID	M	NHN	MET318d
2	ABL1	Human	*1FPU*	M	NHN	MET318d
3	ABL1	Human	1IEP	M	CHO:NHN:NHO	GLU136a:MET318d:MET318a
4	Adenylate kinase	*Escherichia coli*	1M52	M	NHN:CHO	VAL59d:VAL59a
5	Adenylate kinase	*Saccharomyces cerevisiae*	1AKE	M	NHN:CHO	VAL63d:VAL63a
6	Adenylate kinase	*Saccharomyces cerevisiae*	1AKY	M	NHO	GLN204a
7	Adenylate kinase	*Zea mays*	1DVR	M	NHO	ALA195a
8	BTK	Human	1ZAK	M	NHO	GLN15d
9	BTK	Human	*1B55*	M	NHO	GLN15d
10	cAMP-dependent protein kinase	*Bos taurus*	1BWN	M	NHN:CHO	VAL123d:GLU121a
11	cAMP-dependent protein kinase	*Bos taurus*	1SVE	M	NHN: CHO:CHO	VAL123d: VAL123a:GLU121a
12	cAMP-dependent protein kinase	*Bos taurus*	1VEB	M	NHN:CHO	VAL123d:GLU121a
13	cAMP-dependent protein kinase	*Bos taurus*	1SVH	M	NHN:CHO	VAL123d:GLU121a
14	CK2	Human	1SVG	M	CHO	VAL116a
15	CSK	Human	*1OM1*	M	NHO	MET269d
16	c-Src	Human	*1BYG*	M	NHO	HIS60a
17	c-Src	Human	*1O4A*	M	NHO	HIS60a
18	c-Src	Human	1O4B	M	NHO:NHO	ARG34a:ARG34d
19	c-Src	Human	1O4D	M	NHO:NHO	ILE63d:TYR61a
20	c-Src	Human	1O4E	M	NHO	GLU37d
21	c-Src	Human	1O4F	M	NHO	GLU37d
22	c-Src	Human	1O4G	M	NHO	GLU37d
23	c-Src	Human	1O4H	M	NHO	GLU37d
24	c-Src	Human	1O4I	M	NHO	GLU37d
25	c-Src	Human	1O4J	M	NHO	GLU37d
26	c-Src	Human	1O4K	M	NHO	GLU37d
27	c-Src	Human	1O4L	M	NHO	GLU37d
28	c-Src	Human	1O4M	M	NHO:NHO	GLU37d:THR38d
29	c-Src	Human	1O4N	M	NHO	GLU37d
30	c-Src	Human	1O4O	M	NHO	GLU37d
31	c-Src	Human	1O4P	M	NHO	GLU37d
32	c-Src	Human	1O4Q	M	NHO:NHO	GLU37d:THR38d
33	c-Src	Human	1O4R	M	NHO	GLU37d
34	c-Src	Human	1O41	M	NHO:NHO	GLU37d:HIS60a
35	c-Src	Human	1O42	M	NHO:NHO	GLU37d:HIS60a
36	c-Src	Human	1O43	M	NHO:NHO	GLU37d:HIS60a
37	c-Src	Human	1O44	M	NHO:NHO	GLU37d:HIS60a
38	c-Src	Human	1O45	M	NHO:NHO	GLU37d:HIS60a
39	c-Src	Human	1O46	M	NHO:NHO	GLU37d:HIS60a
40	c-Src	Human	1O47	M	NHO:NHO	GLU37d:HIS60a
41	c-Src	Human	1O48	M	NHO:NHO	GLU37d:HIS60a
42	c-Src	Human	1O49	M	CHO:NHN:NHO	GLU339a:MET341d:MET341a
43	SRC_RSVSA	Human	1Y57	M	NHO:NHO	GLU284d:HIS307a
44	SRC_RSVSA	Human	*1NZL*	M	NHO:NHO	GLU284d:HIS307a

Table 3 continued

S. no.	Kinase sub-family	Source	PDB ID	M/S	Synthon type	Residue
45	CDK2	Human	1NZV	M	NHO	LEU83d
46	CDK2	Human	1FVT	–	–	–
47	CDK2	Human	1P5E	M	NHO:NHN:NHO	GLU81a:LEU83d:LEU83a
48	CDK2	Human	1OGU	M	CHO:NHO:NHN	GLU81a:LEU83a:LEU83a
49	CDK2	Human	1OI9	M	NHO:NHN	GLU81a:LEU83d
50	CDK2	Human	1OIY	M	NHN:NHO	LEU83d:LEU83a
51	CDK2	Human	*1GZ8*	M	NHN:NHO:CHO	LEU83d:LEU81a: LEU81a
52	CDK2	Human	1JVP	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
53	CDK2	Human	1H00	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
54	CDK2	Human	1OIT	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
55	CDK2	Human	1H01	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
56	CDK2	Human	1H08	M	NHO:NHN:NHO	GLU81a:LEU83d:LEU83a
57	CDK2	Human	1E1X	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
58	CDK2	Human	1H07	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
59	CDK2	Human	2BHE	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
60	CDK2	Human	1OIR	M	NHN:NHO	LEU83d:LEU83a
61	CDK2	Human	1E1V	M	NHO:NHO	GLU81a:LEU83d
62	CDK2	Human	1AQ1	M	NHN:NHO	LEU83d:LEU83a
63	CDK2	Human	1B38	M	NHO:NHO	GLU81a:LEU83d
64	CDK2	Human	1R78	M	NHO:NHO	VAL83d:VAL83a
65	CDK2	Human	1GII	M	NHN:NHO	LEU83d:LEU83a
66	CDK2	Human	1H1S	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
67	CDK2	Human	1Y8Y	M	NHN:NHO	LEU83d:LEU83a
68	CDK2	Human	1H1R	M	CHO:NHO:NHN	GLU81a:LEU83a:LEU83d
69	CDK2	Human	1PYE	M	NHO:NHO:CHO	GLU81a:LEU83d:LEU83a
70	CDK2	Human	1KE6	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
71	CDK2	Human	1KE7	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
72	CDK2	Human	1KE8	M	NHO:NHO	GLU81a:LEU83d
73	CDK2	Human	1KE9	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
74	CDK2	Human	1CKP	M	NHO:NHN	GLU81a:LEU83d
75	CDK2	Human	1B39	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
76	CDK2	Human	1DM2	M	NHO:NHN	LEU83a:LEU83d
77	CDK2	Human	1H1P	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
78	CDK2	Human	1Y91	M	CHO:NHN:CHO	GLU81a:LEU83d:LEU83a
79	CDK2	Human	1DI8	M	NHO:NHO	VAL83d:VAL83a
80	CDK2	Human	1GIJ	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
81	CDK2	Human	1W0X	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
82	CDK2	Human	1KE5	M	NHO:NHN:NHO	GLU81a:LEU83d:LEU83a
83	CDK2	Human	1VYZ	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
84	CDK2	Human	1V1K	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
85	CDK2	Human	1OIQ	M	NHO:NHO:NHO:CHO	GLU81a:LEU83d:LEU83a:HIS84a
86	CDK2	Human	1P2A	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
87	CDK2	Human	1E9H	M	NHN:NHO	LEU83d:LEU83a
88	CDK2	Human	1H1Q	M	NHO:NHO	GLU81a:LEU83d
89	CDK2	Human	1PF8	M	NHO:NHO:NHO	GLU81a:LEU83d:LEU83a
90	CDK2	Human	2BHH	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a
91	CDK2	Human	1G5S	M	CHO:NHO:NHO	GLU81a:LEU83d:LEU83a
92	CDK2	Human	1GIH	M	CHO:NHN:NHO	GLU81a:LEU83d:LEU83a

Table 3 continued

S. no.	Kinase sub-family	Source	PDB ID	M/S	Synthon type	Residue
93	CDK2	Human	1URW	M	NHO	LEU83a
94	CDK5	Human	1H0W	M	CHO:NHN:NHO	GLU81a:CYS83d:CYS83a
95	CDK6	*Herpvirus saimri*	*1UNL*	M	NHO	VAL101d
96	CHK1	Human	1XO2	M	NHO:NHO	GLU85a:CYS87d
97	CHK1	Human	*1NVQ*	M	NHO:NHO	GLU85a:CYS87d
98	CHK1	Human	1NVR	M	NHO:NHO	GLU85a:CYS87d
99	C-KIT tyrosine kinase	Human	1NVS	M	CHO:NHN:CHO	GLU671a:CYS673d:CYS673a
100	Death-associated protein kinase	Human	*1T46*	M	CHO:CHO	GLU94a: VAL96a
101	EGFR	Human	*1P4F*	M	NHN:CHO	MET769d:GLN767a
102	EGFR	Human	*1M17*	M	NHN:CHO	MET793d:GLN791a
103	FGFR1	Human	1XKK	M	NHO:NHO	ALA564d:GLU562a
104	FGFR1	Human	1AGW	M	CHO:NHO	GLU562a:ALA564d
105	FGFR1	Human	1FGI	M	NHO:NHN	ALA564a:ALA564d
106	Glycerol kinase	*Escherichia coli*	*2FGI*	M	NHO	ARG83d
107	Glycerol kinase	*Escherichia coli*	1BO5	M	NHO	GLY411d
108	GSK3b	Human	1BWF		–	–
109	GSK3b	Human	1GNG	M	NHO	VAL135d
110	GSK3b	Human	1Q3D	M	NHO:NHO	VAL135a:VAL135d
111	GSK3b	Human	1Q3W	M	NHO:NHO:NHO	VAL135a:VAL135d:ASP133a
112	GSK3b	Human	*1Q41*	M	NHO:NHO	VAL135d:ASP133a
113	GSK3b	Human	1Q4L	M	NHO:NHN:NHO	VAL135a : VAL135d: PRO136a
114	GSK3b	Human	1Q5K	M	NHO:NHO	VAL135d:ASP133a
115	GSK3b	Human	1R0E	M	NHO:NHO:NHO	VAL135a:VAL135d: ASP133a
116	HCK	Human	1UV5	M	NHO	MET341d
117	Hexokinase type I	*Rattus norvegicus*	*2HCK*	S	OHO:OHO	ASP209a:ASP209a
118	JNK3	Human	1BG3	M	NHN:CHO	MET149d:GLU147a
119	JNK3	Human	*1PMN*	M	NHN:CHO	MET149d: MET149a
120	JNK3	Human	1PMU	M	NHN:CHO	MET149d:GLU147a
121	JNK3	Human	1PMQ	M	CHO:NHN:CHO	MET149a:MET149d:GLU147a
122	INSR	Human	1PMV	M	NHN:NHO	MET1079d:GLU1077a
123	INSR	Human	*1RQQ*	M	NHN:NHO	MET1079d:GLU1077a
124	ERK2	Human	1GAG	M	NHN:NHO	MET108d:ASP106a
125	ERK2	Human	1TVO	M	NHN:CHO	MET108d:HIS106a
126	ERK2	*Rattus norvegicus*	*1PME*	M	NHN:NHO	MET106d:ASP104a
127	ERK2	*Rattus norvegicus*	3ERK	M	CHO:NHN:NHO	ASP104a:MET106d:MET106a
128	MAPK14	Human	4ERK	M	NHN:CHO	MET109d:HIS107a
129	MAPK14	Human	1BL6	M	NHN:CHO	MET109d:HIS107a
130	MAPK14	Human	1BL7	M	NHN:CHO	MET109d:HIS107a
131	MAPK14	Human	1BMK	M	NHO	ASP168d
132	MAPK14	Human	1KV1	M	NHO	ASP168d
133	MAPK14	Human	1KV2	M	NHN:CHO	MET109d:HIS107a
134	MAPK14	Human	1OUK	M	NHN	MET109d
135	MAPK14	Human	1W84	M	NHN	MET109d:HIS107a
136	MAPK14	Human	1OZ1	M	NHO	ASP168d
137	MAPK14	Human	1W82	M	NHO	ASP168d
138	MAPK14	Human	1W83	M	NHO	ASP168d
139	MAPK14	Human	1WBN	S	OHO	ASP168a
140	MAPK14	Human	1WBO	M	NHN:CHO	MET109d:HIS107a

Table 3 continued

S. no.	Kinase sub-family	Source	PDB ID	M/S	Synthon type	Residue
141	MAPK14	Human	1WBS	M	NHN:CHO	MET109d:HIS107a
142	MAPK14	Human	1WBT	M	NHO	ASP168d
143	MAPK14	Human	1WBV	M	NHN:NHO	MET109d:HIS107a
144	MAPK14	Human	1WBW	M	NHN:CHO	MET109d:HIS107a
145	MAPK14	Human	1YQJ	S	OHO	ASP150a
146	MAPK14	Human	1ZZ2	M	CHO:NHN	HIS107a:MET109d
147	MAPK14	Human	*1A9U*	M	CHO:NHN:CHO	HIS107a:MET109d:MET109a
148	MAPK14	Human	1DI9	M	NHO:NHO	MET109d:HIS107a
149	MAPK14	Human	1M7Q	M	NHO:NHO:NHO	HIS107a:MET109d:GLY110d
150	MAPK14	Human	1OUY	M	NHO:NHO	MET109d:GLY110d
151	MAPK14	Human	1OVE	M	NHN:NHO	MET109d:HIS107a
152	MAPK14	*Mus Muusculus*	1W7H	M	NHN:NHO	MET109d:HIS107a
153	MET	Human	1YWR	M	NHO:NHO	PRO1158a:MET1160d
154	Nucleoside-diphosphate kinase	*Dictyostelium discoideum*	*1R0P*	S	NHO:NHO	LYS16d:HIS122d
155	P56-LCK	Human	1BUX	M	NHO:NHO:NHO	LYS60a:LYS60d:HIS58a
156	P56-LCK	Human	1IJR	M	CHO:NHN:NHO	MET319a:MET319d:GLU317a
157	PDK1	Human	*1QPE*	M	NHO:NHO	SER160a:ALA162d
158	PDK1	Human	1UU3	M	NHO:NHO	SER160a:ALA162d
159	PDK1	*Spodoptera frugiperda*	1UVR	M	NHO:NHO	SER160a:ALA162d
160	PDK1	Human	1UU9	M	NHO:NHO	SER160a:ALA162d
161	PDK1	Human	1UU7	M	NHO:NHO	SER160a:ALA162d
162	PDK1	Human	1UU8	M	NHO:NHO	SER160a:ALA162d
163	Phosphoinositide 3 Kinase Gamma	*Sus scrofa*	*1OKZ*	M	OHO: NHO	GLU880a:VAL882d
164	Phosphoinositide 3 Kinase Gamma	Human	1E8W	M	NHO	VAL882d
165	Phosphoinositide 3 Kinase Gamma	*Sus scrofa*	*1E8Z*	M	NHO	VAL882d
166	Phosphoinositide 3-Kinase Gamma	*Sus scrofa*	1E7V	M	NHO	VAL882d
167	Phosphoenolpyruvate Carboxykinase	Human	1E90	M	NHO	PHE530d
168	Phosphoenolpyruvate Carboxykinase	Human	*1M51*	M	NHO	PHE530d
169	Phosphoglycerate kinase	*Sus scrofa*	1NHX	M	NHO	GLY312a
170	Phosphoglycerate kinase	*Trypanosoma brucei*	1KF0	M	NHO	ALA314a
171	PKA	Human	16PK	M	NHO	VAL123d
172	PKA	Human	1BX6	M	NHO	VAL123d
173	PKA	Human	1RE8	M	NHO	VAL123d
174	PKA	Human	1REK	M	NHN:NHO	VAL123d:GLU121a
175	PKA	Human	*1BKX*	M	NHO	VAL123d
176	PKA	Human	1REJ	M	NHN:NHO	VAL123d:GLU121a
177	PKA	Human	1FMO	M	NHN:NHO	VAL123d:GLU121a
178	PKA	Human	1JBP	M	NHN:NHO	VAL123d:GLU121a
179	PKA(alpha)	Human	1RDQ	M	NHN:NHO	VAL123d:GLU121a
180	PKA(alpha)	Human	*1CDK*	M	NHN:NHO	ALA123d:GLU121a
181	PKA(alpha)	Human	1Q24	M	NHO:NHO	ALA123d:GLU121a
182	PKA(alpha)	Human	1SZM	M	NHN:CHO	VAL123d:GLU121a
183	PKA(alpha)	Human	1Q8T	M	NHN:CHO	VAL123d:GLU121a
184	PKA(alpha)	Human	1Q8U	M	NHN:CHO:CHO	VAL123d:GLU121a:VAL123a
185	PKA(alpha)	Human	1Q8W	M	NHN:NHO	VAL123d:GLU121a

Table 3 continued

S. no.	Kinase sub-family	Source	PDB ID	M/S	Synthon type	Residue
186	PKA(alpha)	Human	1STC	M	NHN:CHO	VAL123d:GLU121a
187	PKA(alpha)	Human	1YDR	M	NHN:CHO	VAL123d:GLU121a
188	PKA(alpha)	Human	1YDS	M	NHN:CHO	VAL123d:GLU121a
189	PKA(beta)	*Sus scrofa*	1YDT	–	–	–
190	Protein kinase C-iota	Human	1CTP	M	NHO:NHO:CHO	GLU324a:VAL326d:VAL326a
191	Protein kinase C-iota	Human	*1ZRZ*	M	CHO	HIS42a
192	PLK1	Human	1KPF	–	–	–
193	PIM-1	Human	*1UMW*	M	NHO	GLU121a
194	TGFbetaR1	Human	*1XWS*	–	–	
195	TGFbetaR1	Human	1RW8	M	NHN:CHO	HIS283d:ASP281a
196	Thymidine kinase	Herpes simplex virus	*1PY5*	S	NHO	GLN125d
197	Thymidine kinase	Herpes simplex virus	1E2K	S	NHO	GLN125d
198	Thymidine kinase	Herpes simplex virus	1E2N	S	NHO	GLN125d
199	Thymidine kinase	Herpes simplex virus	1E2P	S	NHO:NHN	GLN125d: ARG176d
200	Thymidine kinase	*Ureaplasma urealyticum*	1QHI	M	NHO:NHO	PHE128d:LYS1880a
201	Thymidine kinase	Human herpesvirus 1	1XMR	S	NHO	GLN125d
202	Thymidine kinase	Human herpesvirus 2	2KI5	S	NHO	GLN125d
203	Thymidine kinase	Herpes simplex virus	1P7C	S	NHO	GLN125d
204	Thymidylate kinase	Human	1E2L	S	NHO	ARG76a
205	Thymidylate kinase	Human	*1E9A*	S	NHO	ARG76a
206	Thymidylate kinase	Human	1E9B	S	NHO	ARG76d
207	Thymidylate kinase	Human	1E9C	S	NHO	ARG76d
208	Thymidylate kinase	Human	1E9D	S	NHO	ARG76d
209	Thymidylate kinase	Human	1E9E	S	NHO	ARG76d
210	Thymidylate kinase	Human	1E98	S	NHO	ARG76d
211	Thymidylate kinase	*Saccharomyces cerevisiae*	1E9F	S	NHO	ARG73d
212	Thymidylate kinase	*Escherichia coli*	3TMK	S	NHO:NHO	ARG78d:GLN109d
213	Thymidylate Kinase	Human	5TMP	S	NHO	ARG76d
214	Thymidylate Kinase	Human	1E2D	S	NHO	ARG76d
215	Thymidylate Kinase	Human	1E2E	S	NHO	ARG76d
216	Thymidylate Kinase	Human	1E2F	S	NHO	ARG76d
217	Thymidylate Kinase	Human	1E2G	S	NHO	ARG76d
218	Thymidylate Kinase	*Mycobacterium tuberculosis*	1E2Q	S	NHO	ARG74d
219	Thymidylate Kinase	*Mycobacterium tuberculosis*	1MRN	S	NHO	ARG74d
220	Thymidylate Kinase	*Mycobacterium tuberculosis*	1MRS	S	NHO	ARG74d
221	Thymidylate Kinase	*Mycobacterium tuberculosis*	1G3U	S	NHO	ARG74d
222	Thymidylate kinase	*Escherichia coli*	1W2G	S	NHO	GLN109d
223	Tyrosine kinase LCK	Human	4TMK	M	NHO	HIS58a

Table 3 continued

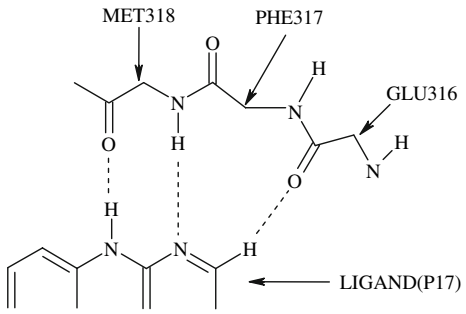
S. no.	Kinase sub-family	Source	PDB ID	M/S	Synthon type	Residue
224	Uridine-cytidine kinase 2	Human	*1FBZ*	S	NHO	ARG176d
225	Uridine-cytidine kinase 2	Human	*1UEI*	S	NHO	ARG176d
226	Uridine-cytidine kinase 2	Human	1UDW	S	NHO	ARG176d
227	Uridylmonophosphate/ cytidylmonophosphate kinase	*Dictyostelium discoideum*	1XRJ	M	NHO	VAL65d
228	Uridylmonophosphate/ cytidylmonophosphate kinase	*Dictyostelium discoideum*	1QF9	M	NHO	VAL65d
229	Uridylmonophosphate/ cytidylmonophosphate kinase	*Dictyostelium discoideum*	3UKD	M	NHO	VAL65d
230	VEGFR2	Human	4UKD	M	CHO:NHN:NHO	GLU915a:CYS917d:CYS917a
231	VEGFR2	Human	1Y6A	M	CHO:NHN:NHO	GLU915a:CYS917d:CYS917a
232	VEGFR2	Human	*1Y6B*	M	NHN:NHO	CYS917d:GLU915a
233	Wee1A kinase	Human	1YWN	M	NHN:NHO	CYS379d:GLU377a

C–H···O hydrogen bond geometry demonstrated here is thus similar to the strong N–H···O hydrogen bond in its directional properties.

Hydrogen bond motif and synthons

The concept of supramolecular synthons is well known in small molecule crystallography and crystal engineering (Desiraju 1995). Synthons are structural units within supermolecules, which can be formed and/or assembled by known or conceivable synthetic operations involving intermolecular interactions. If synthon is formed between the same functional group it is called as homosynthon, and if it is formed between two different functional groups it is referred to as a heterosynthon (Vishweshwar et al. 2003; Walsh et al. 2003). The existence of synthons in protein–ligand complexes has been reported earlier (Sarkhel and Desiraju 2004). In the present context, synthons between various scaffolds present in the ligand molecules and the main/side chain of proteins in the active site (Table 3) are considered. The abbreviation for synthons is according to the short form of the hydrogen bond types. For example "CHO:NHN:NHO" represents a synthon formed by the combination of C–H···O:N–H···N:N–H···O types of hydrogen bonds (Scheme 2).

Both homosynthons and heterosynthons exist in the protein–ligand complexes of kinases. The synthons are formed with a combination of N–H···O, C–H···O, and N–H···N hydrogen bonds (Table 3). Rarely do O–H···O hydrogen bonds participate in the synthon formation. The percentages of synthons formed with one hydrogen bond (single-point) is 36%, two hydrogen bonds (two-point), 42%, three hydrogen bonds (three-point), 21% and four-point, 0.4%. Certain residues in each sub-family frequently



Scheme 2 Typical CHO:NHN:NHO synthon between main chain and ligand (PDB ID 1M52)

interact with the ligand. Residues frequently interacting with ligands in different sub-families are: cAMP-dependent protein kinase sub-family, VAL123 and GLU121; CDK2 sub-family, GLU81 and LEU83; CHK1 sub-family CYS87 and GLU85; PKA sub-family, VAL123 and GLU121. These are the conserved residues present in the particular sub-family. These residues have N–H···O, C–H···O or N–H···N hydrogen bonds while interacting with various ligands within a sub-family. For example, GLU81 in the CDK2 sub-family interacts with ligands through N–H···O, C–H···O and N–H···N hydrogen bonds. Therefore in a particular kinase sub-family, the conserved amino acid residues always participate in hydrogen bond interactions with a variety of ligands, so that a particular synthon is retained intact. This reflects the multifaceted hydrogen bond capability of key residues in any particular kinase sub-family. Sometimes, these types of hydrogen bonds are retained in a synthon of a particular subfamily, the best example being sub-families GSK3b, EGFR, INSR, PDK1, and a few members of PKA (Table 3).

Synthons in GSK3b sub-family are composed of a similar hydrogen bond patterns (Fig. 3). VAL135 and

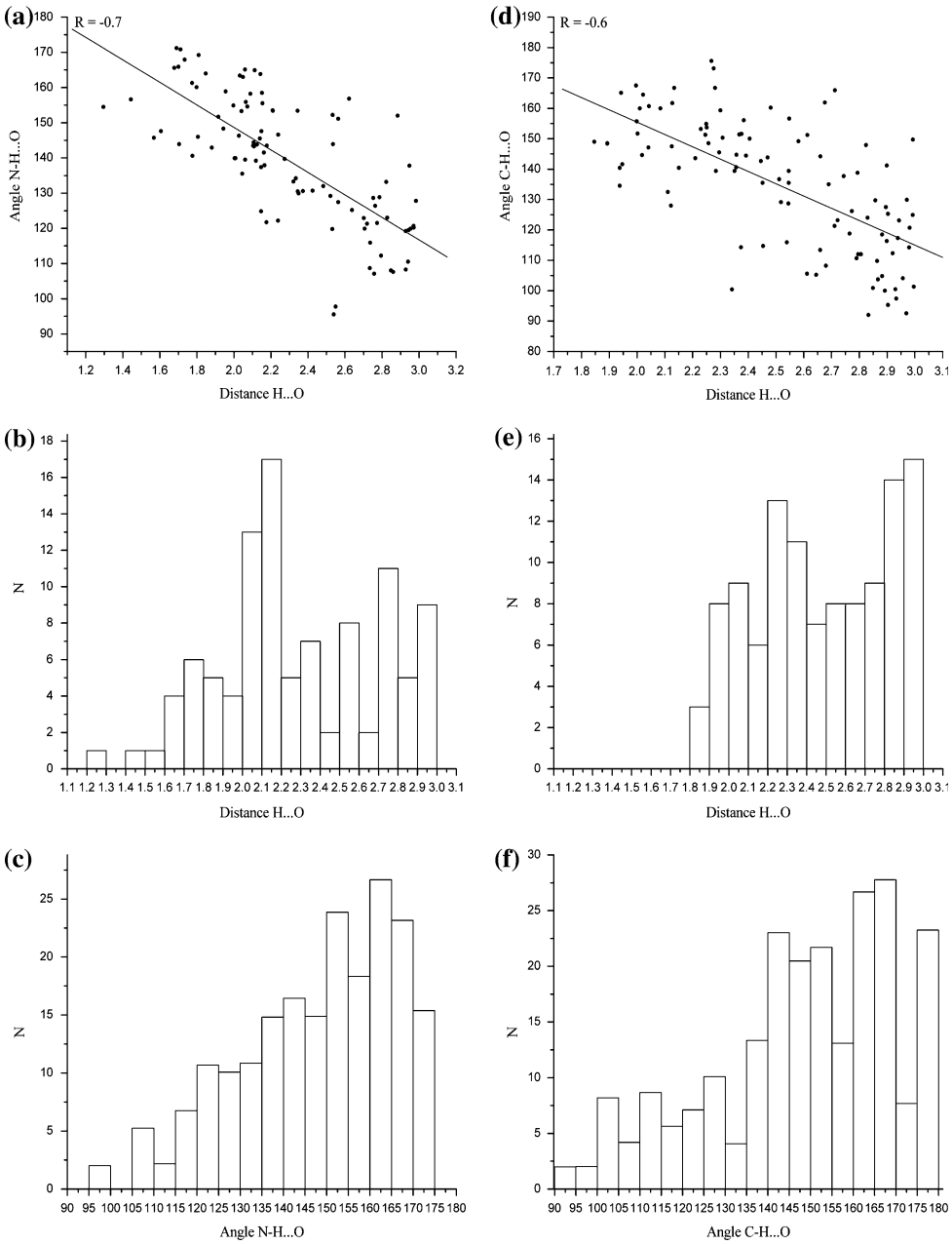


Fig. 2 Hydrogen bond geometry for {NHO BD LA} in Gly (**a**–**c**) and {CHO LD BA} in Glu (**d**–**f**). In each case the inverse length-angle scatterplot is followed by *histograms* of distances and cone-corrected angular distributions

ASP133 are observed to have NHO:NHO:NHO synthons, which are retained across a variety of ligands. However, swapping of donor-acceptor pairs is observed between ligand and protein. Significantly, however, this swapping never hinders the formation of desired synthons. In summary, in each particular sub-family of kinase, various ligands bind to the key residues of the main chain with a typical synthon pattern. Of course, these inhibitors certainly have similar functional groups and scaffolds, which support such interaction consistency.

The thymidine/thymidylate kinase is unusual because ligands for this sub-family generally form the NHO synthon with the side chain residue GLN125 (Fig. 4).

Role of conserved residue

The kinase domains across many sub-families share a great percentage of similarity in the active site residues. Hence there is a chance that residues at the same position might be involved in ligand binding. To verify this assumption an active site sequence alignment for 35 sub-families of human source was carried out, each structure representing a sub-family (Table 3, PDB ID with italics). It was observed that residues interacting with ligands are aligned at a particular position across all sub-families with very few exceptions. The exceptions (Fig. 5) are interactions involving side chains and adjacent residues. This finding

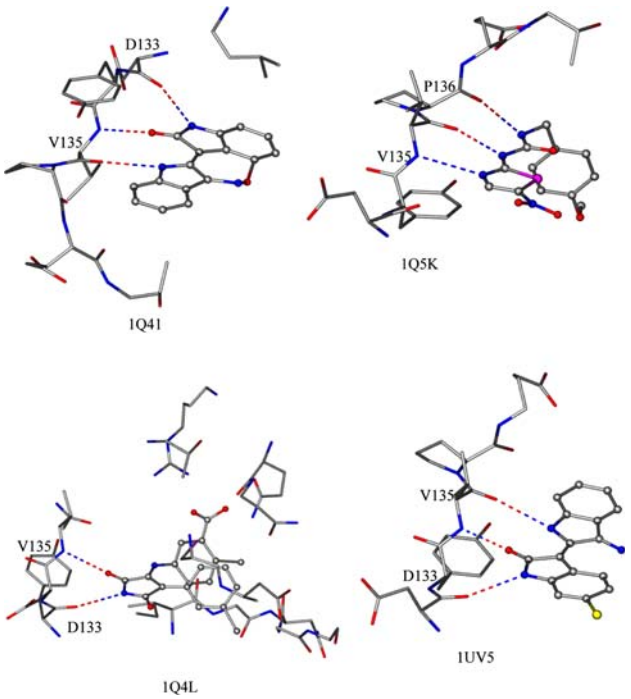


Fig. 3 Synthons formed by N–H···O hydrogen bonds in key residues, are shown for the GSK3b sub-family with the respective PDB ID

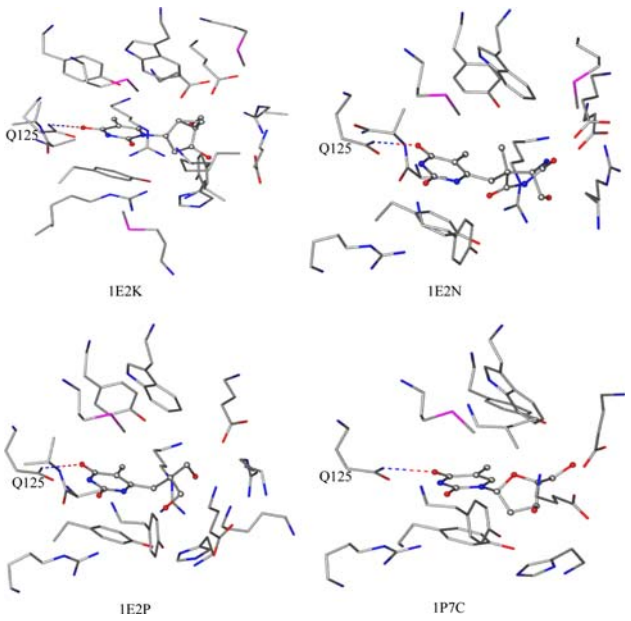


Fig. 4 Side chain participation in synthon formation in Thymidine kinase sub-family

reveals that residues determining the ligand interaction with the main chain are conserved across all the kinase sub-families. Despite the variation in amino acid substitutions across kinase sub-families, the hydrogen bond signature pattern is always retained in the conserved residues. The consistency of the strong and weak hydrogen bonds between protein–ligand complexes across various sub-families is definitely a matter of great interest. The formation of supramolecular synthons with the amalgamation of strong and weak hydrogen bonds in protein–ligand complexes is vital for ligand specificity and the interplay between the strong and weak hydrogen bonds in supramolecular synthons may be a determining factor for the diversity of ligand interaction to the hinge region of kinase. To verify this hypothesis, the sub-families of receptor tyrosine kinases are discussed here.

The receptor tyrosine kinases considered in the 233 datasets are EGFR, FGFR1, VEGFR2 and INSR. For these receptor tyrosine kinases, the conserved residues interact with ligands in a similar manner (Fig. 6). The amino acid residues involved in ligand binding in these structures are aligned at a particular position (Fig. 5). In EGFR, the interacting residues are MET769 and GLN767. For the FGFR1 sub-family the pertinent residue is ALA564. In VEGFR2, the residues are GLU915, and CYS917. In sub-family INSR, the residues are MET1079 and GLU1077. These residues interact with the ligands through N–H···O, N–H···N hydrogen bonds. Therefore it can be assumed that the sample receptor tyrosine kinases are similar on the basis of supramolecular synthon formed by N–H···O, N–H···N hydrogen bonds. These findings can be extrapolated to the other kinase sub-families also. This finding has a potential role in multikinase targeting, where a single molecule can be targeted against several kinases (Wilhelm et al. 2006).

Active site solvation

Solvation and desolvation are the crucial phenomena, which occur in the process of ligand receptor interaction (Ladbury 1996). During binding, the active site undergoes a rapid desolvation process to accommodate the ligand. This process finally results in the release of ordered water favoring the formation of the ligand–receptor complex. While being separate, both enzyme and substrate force the neighboring water molecules into an ordered shell. Binding of substrate to enzyme releases some of the ordered water, and the resulting increase in entropy provides a thermodynamic push toward the formation of the ligand–receptor complex. Some of the ordered water molecules in the receptor are unaffected by this rapid desolvation process and are retained in the active site due to their structural and functional importance. These structural/functional waters usually form hydrogen bonds with the neighboring residues or/and ligand and even with water molecules, thus increasing the enthalpy of the final complex and all this stabilizes the ligand–protein complex. A classification of water molecules in the active sites of kinases was carried out, based on the accessible surface area (exposure). Buried

```
1okz.pdb  KVPYVTRERDVMSRLDHPFFVKLYFTFQDDEKLYFGLSYAKNGELLKYIRKIGSFDETCT M
1bkx.pdb  ----------------ELFLVKLEF---------MVMEYVAGGEMFSHY---------- M
1cdk.pdb  ----------KVLQHTELFLVKLEY---------MVMEYVPGGEMFSHYR---------- M
1zrz.pdb  ----------VEKVFASFLVGLHS------CLFFVIEYVNGGDLMFHMQ---------- M
1om1.pdb  ----------------KEIILCNIVKL-----LDSLIFEYVNNT-DFKVHD---------- M
1p4f.pdb  ---------------VELIVIT---------LHELILELVAGG-ELD-HD---------- M
1nvq.pdb  ---------------IEIINLNVVKF-----YGYLFLEYCSGGELFDRHDI--------- M
1py5.pdb  -----WREAEIY------QILGFIAAD-----LWLVSDYHEHGSLFDYHDL--------- M
1x8b.pdb  -----SNAREVA------HVVRYFSAW-----MLIQNEYCNGGSLADAHD---------- M
1fpu.pdb  ------LKEAAVMKEIKHNLVQLLGVC-----FYIITEFMTYGNLLDYA---------- M
1qpe.pdb  ------LAEANLMK-----LVRLYAVV-----IYIITEYMENGSLVDF---------- M
2hck.pdb  ------KLHA----------------------YIITEFMAKGSLLDFK---------- M
1m17.pdb  ------DEAYVM-----AVVCRLLGIC-----VQLITQLMPFGCLLDYR---------- M
1xws.pdb  ---------EVL-----LSVIRLLDWF-----VLILERPEPVQDLFDF---------- M
1rqq.pdb  ---------EEF-----EMVVRLLG---------VVMELMAHGDLKSYR---------- M
1r0p.pdb  ----------------FMVLSLLGI-------LVVLPYMKHGDLRNFN---------- M
2fgi.pdb  ---DLISEMEMMKIG-KHNIINLLGACT----LYVIVEYASKGNLR----EYLQA----- M
1y6b.pdb  ---ALMSELKILIV------VNLLGACT----LMVIVEFCKFGNLS----TYLR------ M
1t46.pdb  ----LMSELKVLSYLGNHNIVNLLGACT----TLVITEYCCYGDLLNFRVGMFLA----- M
1byg.pdb  ----LVQLLG----------------------IVTEYMAKGSLV----DYLRS----- M
1unl.pdb  ----------------EILNIVR---LHDVLTLVFEFCDQDLKKY-------------- M
1gz8.pdb  ----------------------IVK---L---LDLVFEFLHQDLKK-------------- M
1q41.pdb  ------------------ELMLIVR---LR--YNLVLDYVPETVY--------------- M
1pme.pdb  ----------------EIILLEI---IG--INDIIVYLVTHLMGAD------------ M
1pmn.pdb  ----------------HRAELL----MIISLLNVFVYLVMELMDANLC---------- M
1a9u.pdb  ----------------HRTELLLKH-MVIGLLDVFVYLVTHLMGAD------------ M
1o4a.pdb  ----------------CLSVSD--------KNVKHYKIR--YITSR---------- M
1nzl.pdb  ----------------CLSVSDD-------KNVKHYKIRF-YITSR---------- M
1fbz.pdb  ----------------CLSVRDF-------DVVKHYKIRNFYISPR---------- M
1b55.pdb  ---------------------------------LFLSYYEYS----------------- M
1udw.pdb  ----------QFNFDHPAFDVYDFVSHSRKFEGILAFVDTTRLSRRVR---------- S
1e8z.pdb  ----------------QDMLILLRLPYGCISKIGMIEIVKDATTIAKIQ---------- M
1e9a.pdb  ---------------VDHSVHLLFSAN----RWEQVDRYAFSGVAFTGA---------- M
1umw.pdb  -------------------------------NYMSEHLLKAGAWFR------------- M
1m51.pdb  ----------------------PLSEAVNWFRKDGKFLWPGFGENSRVLMLGHIL----- M
```

Fig. 5 Active site sequence alignments for 35 entries of human kinases, each representing a subfamily. Key residues are colored *red*. '*M*' stands for main chain and '*S*' for side chain

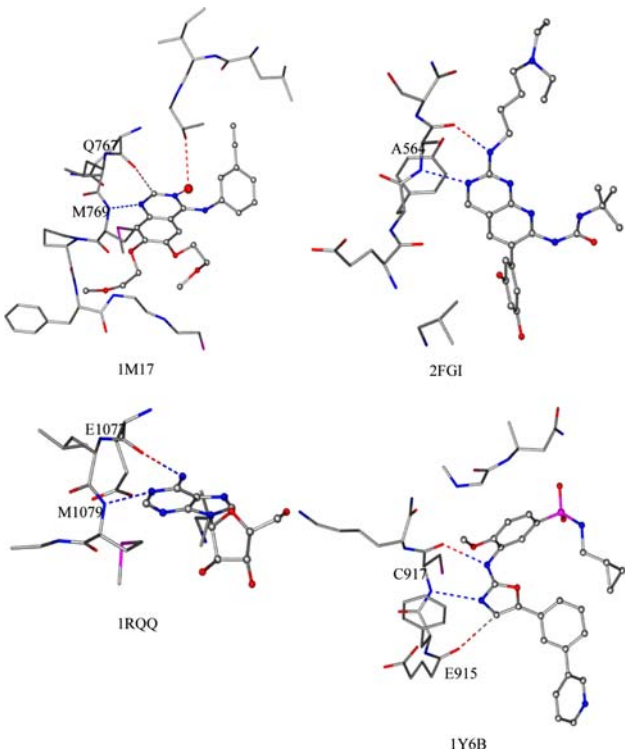


Fig. 6 Conserved synthon across receptor tyrosine kinases (EGFR;1M17, FGFR1; 2FGI, INSR;1RQQ, VEGFR2;1Y6B)

water is of structural and/or functional significance while partially accessible or completely accessible water (category W1T to W9T) is less important. Structurally, the buried water molecules (W0T) are important because, very often they form hydrogen bonds, and sometimes they mediate hydrogen bonds between protein and ligand (Bottoms et al. 2002). Functionally, these water molecules may also participate in enzyme catalysis.

The dataset contains a total of 40,944 water molecules, out of which 6,066 (15%) are present in the active sites. This means that, on average, there are 176 and 26 water molecules in the entire protein and in the active site, respectively, per structure. The percentage distribution of these water molecules in various categories based on percentage of exposure, and also the hydrogen bond between water and ligand is shown in Fig. 7.

The majority of the active site waters are inaccessible (category W0T). This suggests that the active sites are present deep inside the proteins. Very often, the water molecules are enclosed even at the protein–ligand interface. The hydrogen bonds considered here are confined to those between ligand and water. The various types of hydrogen bonds are {OHN WD LA}, {OHO WD LA}, {CHO LD WA}, {NHO LD WA}, {OHO LD WA}. Approximately 60% of the hydrogen bonds are manifested by the W0T category of the active site water molecules. The hydrogen bonds between ligand and water are

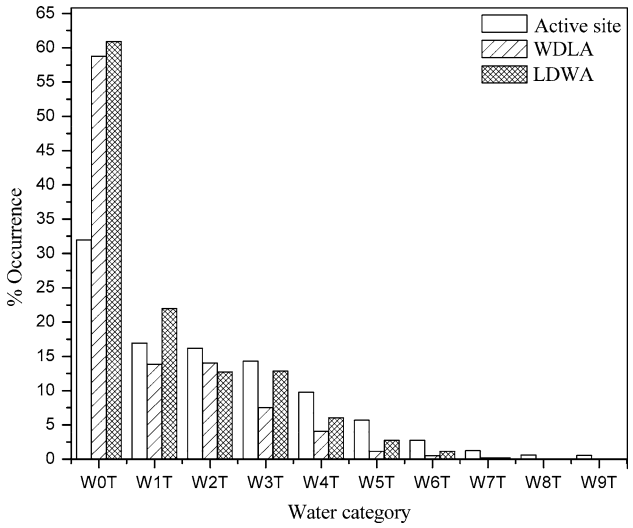


Fig. 7 Hierarchical classification of active site water molecules. Also shown is the percentage of interacting water as an acceptor and as donor

observed up to the category W6T. Beyond this an exposed water molecule rarely interacts with the ligand. The hydrogen bond geometry also depends on the water category. Hydrogen bonds geometries are better represented in the W0T category (Fig. 8). The quality of hydrogen bonding gradually decreases from W1T to W6T. For example, the d-θ scatterplot for {OHO WD LA} is better represented in the W0T category, with the correlation becoming poorer as one moves from W1T to W3T (Fig. 8). This suggests that the buried water molecules in the active sites very often participate in hydrogen bonds with the ligand. Hence they are structurally important.

Conclusions

Strong (N–H···O, O–H···O, N–H···N) and weak (C–H···O C–H···N) hydrogen bonds in the active sites of the kinase family have been studied in a dataset of 233 protein–ligand complexes. The kinase family is dominated by nonpolar and charged residues. Residues like Leu, Glu and His, frequently accept hydrogen bonds from the ligand, while Gly and Leu are favoured donors. Both strong and weak hydrogen bonds are of comparable importance in ligand binding. In this context, the acceptor capacity of main chain Glu is noteworthy and the geometry of C–H···O hydrogen bonds to Glu is on par with the strong hydrogen bonds in the active sites. The hydrogen bonded supramolecular synthons formed between main/side chain and ligand atoms is a typical characteristic of the kinase family. These synthons are formed by an amalgamation of strong and weak hydrogen bonds. The synthon patterns are unique to kinase

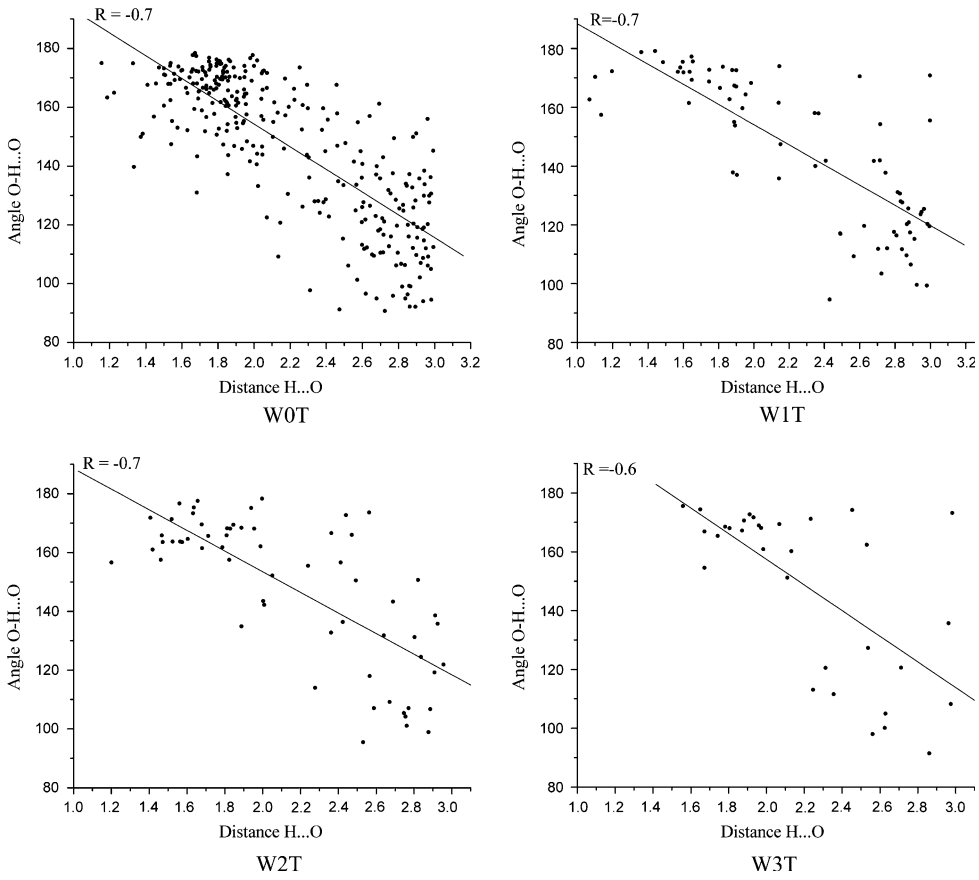


Fig. 8 Inverse length-angle (d-θ scatterplots for {OHO WD LA} for various categories of water (W0T–W3T). Notice the poor correlation in W3T

sub-families. The relationships between the sub-families are established on the basis of similar synthon patterns. The similarities among synthon patterns across sub-families arise due to the conserved residues in the active sites of kinase. The active site water molecules exist in a variety of environments. Water molecules, which are least exposed, usually take part in hydrogen bonding with the ligands. It can be concluded that along with strong hydrogen bonds, weak hydrogen bonds are also important in the kinase family. This information is a valuable asset for kinase inhibitor design, especially in the realm of multikinase inhibitor design.

Acknowledgments S. K. P. thanks Prof. Gautam R. Desiraju, School of Chemistry, University of Hyderabad, Hyderabad, for his helpful guidance in preparing the manuscript. S. K. P. also thanks the chemical computing group (CCG) for technical support and CSIR for fellowship.

References

Aparna V, Rambabu G, Panigrahi SK, Sarma JARP, Desiraju GR (2005) Virtual screening of 4-anilinoquinazoline analogs as EGFR kinase inhibitors: importance of hydrogen bonds in the evaluation of poses and scoring functions. J Chem Inf Model 45:725–738

Auffinger P, Hays FA, Westhof E, Ho PS (2004) Halogen bonds in biological molecules. Proc Natl Acad Sci USA 101:16789–16794

Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. Progr Biophys Mol Biol 44:97–179

Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. J Mol Biol 324:105–121

Bikadi Z, Demko L, Hazai E (2007) Functional and structural characterization of a protein based on analysis of its hydrogen bonding network by hydrogen bonding plot. Arch Biochem Biophys 461:225–234

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank Nucleic Acid Res 28:235–242. http://www.rcsb.org/pdb

Brandl M, Weiss MS, Jabs A, Sühnel J, Hilgenfeld R (2001) C–H···π interactions in proteins. J Mol Biol 307:357–377

Bridges AJ (2001) Chemical inhibitors of protein kinases. Chem Rev 101:2541–2571

Bottoms CA, Smith PE, Tanner JJ (2002) Protein Sci 11:2125–2137

Cohen P (2002) Protein kinases—the major drug targets of the twenty-first century. Nature Rev Drug Discov 1:309–315

Dar AC, Wybenga-Groot LE, Sicheri F (2005) The eukaryotic protein kinase domain. In: Cesareni G, Gimona M, Sudol M, Yaffe M (eds) Modular protein domains, chap 9. Wiley-VCH, Verlag GmbH & Co, KGaA

Desiraju GR, Steiner T (1999) The weak hydrogen bond in structural chemistry and biology, chaps 1 and 5. Oxford University Press, Oxford

Desiraju GR (1995) Supramolecular synthons in crystal engineerings A—new organic synthesis. Angew Chem Int Ed 34:2311–2327

Denessiouk KA, Johnson MS (2003) "Acceptor–donor–acceptor" motifs recognize the watson–crick, hoogsteen and sugar "donor–acceptor–donor" edges of adenine and adenosinecontaining ligands. J Mol Biol 333:1025–1043

Gitlin I, Carbeck JD, Whitesides GM (2006) Why are proteins charged? Networks of charge–charge interactions in proteins measured by charge ladders and capillary electrophoresis. Angew Chem Int Ed 45:3022–3060

Glusker JP (1995) Intermolecular interactions around functional groups in crystals: data for modeling the binding of drugs to biological macromolecules. Acta Cryst D51:418–427

Halgren T (1996) Merck molecular force field. I. Basis, form, scope, parameterisation and performance of MMFF94. J Comput Chem 17:490–519

Hanks SK, Hunter T (1995) Protein kinases the eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J 9:576–596

Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673–4680. ClustalW WWW Service at the European Bioinformatics Institute http://www.ebi.ac.uk/clustalw

Jeffrey GA, Saenger W (1991) Hydrogen bonding in biological structures. Springer, Berlin

Klaholz B, Moras D (2002) C–H···O hydrogen bonds in the nuclear receptor RARγ-a potential tool for drug selectivity. Structure 10:1197–1204

Koellner G, Steiner T, Millard CB, Silman I, Sussman JL (2002) A neutral molecule in a cation-binding site: specific binding of a PEG-SH to acetylcholinesterase from Torpedo californica. J Mol Biol 320:721–725

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

Lipinski CA, Lombardo F, Dominy BW, Feeney PL (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25

Levitzki A (2003) Protein kinase inhibitors as a therapeutic modality. Acc Chem Res 36:462–469

Mao L, Wang Y, Lie Y, Hu X (2004) Molecular determinants for ATP-binding in proteins: a data mining and quantum chemical analysis. J Mol Biol 336:787–807

MOE Program, Version 2006–08 Chemical Computing Inc, Sherbrooke St. W, Montreal, Quebec, Canada, H3A 2R7

Naumann T, Matter H (2002) Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. J Med Chem 45:2366–2378

Panigrahi SK, Desiraju GR (2007) Strong and weak hydrogen bonds in the protein–ligand interface. Proteins 67:128–141

Panigrahi SK, Desiraju GR (2004) Homology modelling in protein structure prediction: epidermal growth factor receptor kinase domain. Nat Acad Sci Lett (India) 27:1–11

Pierce AC, Sandretto KL, Bemis GW (2002) Kinase inhibitors and the case for CH···O hydrogen bonds in protein–ligand binding. Proteins 49:567–576

Pierce AC, ter Haar E, Binch HM, Kay DP, Patel SR, Li P (2005) CH···O and CH···N hydrogen bonds in ligand design: a novel quinazolin-4-ylthiazol-2-ylamine protein kinase inhibitor. J Med Chem 48:1278–1281

Sarkhel S, Desiraju GR (2004) N–H···O, O–H···O, and C–H···O hydrogen bonds in protein–ligand complexes: strong and weak interactions in molecular recognition. Proteins 54:247–259

Steiner T (2002) The hydrogen bond in the solid state. Angew Chem Int Ed 41:48–76

Sugen (2002) Structure and Phylogeny of the Protein Kinases, Salk Institute, La Jolla, CA. http://198.202.68.14/human/kinome/phylogeny.html and http://kinasedb.ontology.ims.u-tokyo.ac.jp

The Protein Kinase Resource, SDSC, UC, San Diego, MC, 2004, http://www.kinasenet.org/pkr

Tiwari A, Panigrahi SK (2007) HBAT: a complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. In Silico Biol 7:0057

Vishweshwar P, Nangia A, Lynch VM (2003) Molecular complexes of homologous alkanedicarboxylic acids with isonicotinamide: X-ray crystal structures, hydrogen bond synthons, and melting point alternation. Cryst Growth Des 3:783–790

Vieth M, Higgs RE, Robertson DH, Shapiro M, Gragg EA, Hemmerle H (2004) Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. Biochim Biophys Acta 1697:243–257

Wang R, Fang X, Lu Y, Yang CY, Wang S (2005) The PDBbind database: methodologies and updates J Med Chem 48:4111–4119. http://www.pdbbind.org

Walsh RDB, Bradner MW, Fleischman S, Morales LA, Moulton B, Rodríguez-Hornedo N, Zaworotko M (2003) Crystal engineering of the composition of pharmaceutical phases. J Chem Commun 2:186–187. doi:10.1039/b208574g

Williams DH, Mitchell T (2002) Latest developments in crystallography and structure-based design of protein kinase inhibitors as drug candidates, Curr Opin Pharmacol 2:567–573

Williams MA, Goodfellow JM, Thornton JM (1994) Buried waters and internal cavities in monomeric proteins. Protein Sci 3:1224–1235

Wilhelm S, Carter C, Lynch M., Lowinger T, Dumas J, Smith RA, Schwartz B., Simantov R, Kelley S (2006) Discovery and development of sorafenib: a multikinase inhibitor for treating cancer. Nat Rev Drug Discov 5:835–844