CHAPTER 1

STRONG AND WEAK HYDROGEN BONDS IN BIOLOGY

1.1 Introduction

Thousands of different molecules make up the intricate internal structures of a cell [1.1]. Each has its characteristic sequence of subunits, its unique three-dimensional structure, and the highly specific selection of binding partners in the cell [1.2–1.5]. The basic constituents of these subunits are made up of amino acids, sugar, nucleic acid and lipids [1.3, 1.4]. These components are held together through many forces out of which the hydrogen bond is the universal glue. Nevertheless the aqueous environment hosts an array of biological events for the smooth functioning of the cell. Water and hydrogen bonds are inseparable in all respects [1.6]. The typical biomolecules carry many groups that form strong and weak hydrogen bonds. The functional groups at the surfaces may be involved in weak hydrogen bonds that operate in water-biomolecule interactions and also in recognition processes and structural stabilization of the molecular peripheries [1.7].

1.2 Definition of a hydrogen bond

To define the hydrogen bonds to its exact term is been a long-standing problem for scientific communities ever since its discovery. Recently the International Union of Pure and Applied Chemistry (IUPAC) have formed a core group of expert to define the hydrogen bond [1.8]. The core group has recommended a modern definition of hydrogen bonds after two meetings, one at Pisa, Italy in 2005 and other at Bangalore, India in 2006. According to the modern definition "The hydrogen bond is an attractive interaction between *X–H* and an atom or a group of atoms *Y*, in the same or different molecule(s), where there is evidence of bond formation'. The most important criteria for a hydrogen bond are: (i) the H in the *X–H* group is more electropositive than *X* and (ii) the physical forces involved in hydrogen bonding should include attractive electrostatic forces, i.e. it should not be primarily dispersive forces". Note that the acceptor will be annotated as A in the present thesis instead of Y, as recommended by the IUPAC core group. Prior to this, definitions proposed by Pauling (1939) and Pimentel, McClellan (1960) are noteworthy [1.9, 1.10]. The

recommended definition is very similar to the definition proposed by Pimentel, McClellan (1960).

1.3 Defining strong and weak hydrogen bond

The hydrogen bonds are manifested in a variety of strengths and geometries. In hydrogen bonds, hydrogen atoms of O−H, N−H, or S−H groups (known as hydrogen bond *donors*) interact with free electrons of *acceptor* atoms (for example, O, N, or S) [1.11–1.14]. The bonding energies of hydrogen bonds (4–40 kcal/mol) are lower than those of covalent bonds. Sometimes C−H is included in the armory of hydrogen bond donors and π electrons of aromatic ring are also included as the hydrogen bonds acceptors [1.15]. The bonding energies of such hydrogen bonds are < 4 kcal/mol. The hydrogen bond is a group-pair interaction, with an energy limit of 4–40 kcal/mol [1.12]. The hydrogen bond involves all three atoms or groups of atoms, X, H and A. In most cases of hydrogen bonding, one of the two bonds formed by the hydrogen atom, namely X–H, is much stronger than the other, H···A. Accordingly hydrogen bonds like O−H···O, N−H···O, O−H···N and N−H···N may be considered to be strong while interactions like C−H···O, C−H···N, O−H···π, N−H···π and C−H···π are taken as weak. The chemical nature of the donor and acceptor species are considered while defining the strong and weak hydrogen bonds rather than on the basis of the distance between them. This is a subjective definition but it is used consistently in this thesis.

1.4 Classification of hydrogen bonds

The wide range of properties of hydrogen bond acceptor and donor species necessitate a classification scheme to emphasize their importance in chemistry and biology. The basis of such classification of hydrogen bonds is geometrical, energetic, thermodynamic and functional in nature. So far three separate attempts have been made independently by Jeffrey and Saenger (1991), Jeffrey (1997), and Desiraju and Steiner (1999) [1.11, 1.12, 1.14]. The properties of hydrogen bonds suggested by Desiraju and Steiner (1999) [1.14] are shown in Table 1.1. They have classified the hydrogen bonds on the basis of the nature of donor and acceptor groups into very strong, strong and weak. This is similar to a proposal made by

Jeffrey (1997) [1.12], who classify hydrogen bonds into strong, moderate and weak types. Such classification should be followed as a guideline rather than on a totalitarian basis.

Table 1.1: Properties of very strong, strong and weak hydrogen bonds*.

	Very strong	Strong	Weak
Bond energy (-kcal/mol)	15–40	4–15	<4
Examples	$[F\cdots H\cdots F]^-$	O−H···O=C	C−H···O
	$[N\cdots H\cdots N]^+$	N−H···O=C	O−H···π
	P–OH···O=P	O−H···O–H	Os–H···O
IR ν_s relative shift	> 25%	5–25%	< 5%
Bond lengths	H−A ~ X−H	H···A > X–H	H···A >> X−H
Lengthening of X–H	0.05–0.2 Å	0.01–0.05 Å	≤ 0.01 Å
X···A (*D*) range	2.2–2.5 Å	2.5–3.2 Å	3.0–4.0 Å
H···A (*d*) range	1.2–1.5 Å	1.5–2.2 Å	2.0–3.0 Å
Bonds shorter van der Waals cutoff	100%	almost 100%	30–80%
Angle (θ) range	175–180°	130–180°	90–180°
*k*T (room temperature)	> 25	7–25	< 7
Effect on crystal packing	dominant	distinctive	variable
Utility in crystal engineering useful	unknown	useful	partly
Covalency	pronounced	weak	vanishing
Electrostatics	significant	dominant	moderate

*Adopted from Desiraju and Steiner, (1999)

1.4.1 *Very strong hydrogen bond*

Very strong hydrogen bonds are formed by unusually activated donors and acceptors, often in an intramolecular situation. Frequently, they are formed between an acid and its conjugate base, $X{-}H\cdots X^-$, or between a base and its conjugate acid, $X^+{-}H\cdots X$. These types of hydrogen bonds are often described in chemistry literature and recently in biology. The recent resurgence of literature on very strong hydrogen bonds in biology is due to the greater advancement in the understanding of enzymatic reactions [1.16–1.34]. Very strong hydrogen bonds are of great importance in the context of enzymatic reactions and are often referred to as low barrier hydrogen bonds (LBHB). Sometimes these hydrogen bonds are also referred as short, strong hydrogen bonds (SSHB). LBHB and SSHB have been postulated to play a crucial role in enzymatic reactions, particularly those that involve a general acid-general base mechanism, by providing substantial stabilization energy (10–20 kcal/mol) for the intermediate or transition state.

1.4.2 *Strong hydrogen bond*

The strong hydrogen bonds are the normal hydrogen bonds described in chemistry and biology [1.2–1.5, 1.14, 1.35]. The energy range of these types of hydrogen bonds is in between 4–15 kcal/mol. The omnipresence of strong hydrogen bonds are topic of interest ever since the structural biology came into existence in 1960s. The strong hydrogen bond is usually counted in many biological phenomena such as stabilizing three-dimensional structure of biomolecules, membrane permeability, and enzyme substrate recognition. The examples of strong hydrogen bonds are O−H···O=C, N−H···O=C.

1.4.3 *Weak hydrogen bonds*

Weak hydrogen bonds in biological structures were observed as early as the 1960s. Sutor (1963) noted the existence of C−H···O interactions in purine and pyrimidine bases, while it was recognized in nucleosides by Shefter and Trueblood (1965) and later by Sundaralingam (1966) [1.36–1.38]. Ramachandran *et al.* (1965, 1966) observed these hydrogen bonds in collagen and polyglycine II [1.39–1.40]. The existence of weak hydrogen bonds in biology has been strengthened by recent papers in various journals [1.41–1.74]. Weak hydrogen bonds are electrostatic but this characteristic is modified by variable dispersive and charge-transfer components that depend substantially on the nature of the donor and acceptor group. The strongest example in this category are hydrogen bonds like C≡C–H···O (–2 to –4 kcal/mol). The methyl groups form the weakest hydrogen bond in this series (about –0.5 kcal/mol).

1.4.4 *Other weak interactions in biology*

There are a growing number of reports related to other weak interactions involving π–acceptors [1.75–1.84], halogen atoms (both as electrophiles and nucleophiles) [1.85–1.88], and sulfur-atoms in biology [1.89–1.92]. Short oxygen···halogen interactions have been known since the 1950s. A recent survey of protein and nucleic acid structures reveals similar halogen bonds as potentially stabilizing inter- and intramolecular interactions that can affect ligand binding [1.86]. The acceptor capability of organic halogen, X (X = F, Cl, Br, I), is also important in macromolecules [1.87]. Another weak interaction is the interaction involving sulfur atoms. Sulfur atoms are larger and have a more diffuse electron cloud than oxygen and nitrogen, but are nevertheless capable of participating in hydrogen bonds and are found in macromolecules. While these interactions are weak they seem to

play a definite role in many biological events. In summary, various aspects of the hydrogen bonding phenomenon observed in chemistry are being implicated in biology as well. Hydrogen bonds exist in all respects in the ever complex but still highly organized living world.

1.5 Methods of studying hydrogen bonds

There exist several methods, both experimental and theoretical, to study hydrogen bonds [1.93]. The experimental techniques include X-ray diffraction study, electron diffraction of protein crystals, small-angle X-ray scattering, small-angle neutron scattering, fibre diffraction, electron microscopy, nuclear magnetic resonance and computation. Discussed here briefly are the implications of X-ray diffraction and statistical methods in studying hydrogen bonds in macromolecules in brief. Various experimental and computational techniques to study hydrogen bonds are mentioned in depth elsewhere by Jeffrey and Saenger (1991), Jeffrey (1997), Scheiner (1997), Desiraju and Steiner (1999), Rossman and Arnold (2001) [1.11–1.14, 1.93].

1.5.1 *Crystallography*

The rapid growth and development of X-ray and neutron crystallography have enriched biological research greatly [1.93]. However, the complexity of biomolecules has always been a challenge in crystallization and structure solution. The sophistication in crystallization techniques in crystallization methods like cryocrystallograhy, high-throughput crystallography, automation in the ‘direct methods’ for crystal structure solution and the subsequent refinement, extremely powerful computing facilities, improvements in quality of diffractometers in past few years have overcome many challenges. All this has simplified the overall methods of macromolecular crystallography. The process which was a Herculean task 25 years ago is now a joy among chemists and biologists practising crystallography.

When talking about the detection of the exact location of hydrogen atoms in a molecule, the neutron diffraction method comes to mind [1.94–1.96]. Neutron diffraction technique which was confined to small molecular crystallography earlier is now being use in macromolecular crystallography. Knowing the exact location of hydrogen atoms removes many ambiguities connected with the existence of hydrogen bonds. This is because from the

first visual inspection, chemists/biologists can decipher the important geometric criteria of a putative hydrogen bond.

1.5.2 *Crystallographic databases*

The results of crystal structure determinations are stored in various file formats. The structural information in MMCIF and PDB file formats are routinely saved [1.97]. This enables the scientists to analyze and visualize the biomolecules in computers using sophisticated software. This helps in the proper understanding of biological intricacies through computer graphics and modeling. The ever increasing number of structures solved through experimental techniques necessitates a proper storage and management system. These requirements are fulfilled by crystallographic databases. Crystallographic databases are very useful resources to study hydrogen bonds in the molecular world. Like the Cambridge Structural Database (CSD) [1.98] for small-molecules, the macromolecular database provides a wealth of information for large biological molecules. The databases related to macromolecules are, (a) The Protein Data Bank at Brookhaven (PDB) [1.99], (b) Nucleic Acid Database (NDB), (c) The Biological Macromolecule Crystallization Database (BMCD) [1.93]. The progress in these databases has later given birth to a second generation of databases e.g. SCOP [1.100], CATH [1.101], PDBSUM [1.102], specialized with respect to the structural and functional aspects of biomolecules.

1.5.3 *Statistical analysis*

Statistics plays a major role in scientific research, especially when the dataset is large [1.103, 1.104]. Provided with a large sample of crystallographic data it is possible to derive meaningful conclusions about the nature and behavior of crystals. With this backdrop, hydrogen bond research has been constantly enriched, with a parallel in the increase in the number of deposits in these databases. Statistical methods play a major role in the study of weak hydrogen bonds. This is because a large sampling of structures is needed to estimate quantitative information on weak interactions. One such problem is the analysis of hydrogen bond geometries in receptor–ligand interaction. Chapters 3, 4, and 5 of the thesis deals with the application of statistical methods, to study the nature of hydrogen bond geometries in X-ray structures of receptor–ligand complexes.

Most of the statistical studies are carried out with the help of specialized computer software. There exist many programs for hydrogen bond analysis like HBPLUS [1.105],

HBEXPLORE [1.106], BNDLST from Richardson lab [1.107], CONTACT from CCP4 [1.108] and web based servers like LPC [1.109] and NCI [1.110] in the public domain. However "pros and cons" lie in each of these softwares while performing studies of hydrogen bonds in macromolecules. This is because of the huge numbers of atoms and the variety of donor/acceptor functional groups. The next chapter describes a software called hydrogen bond analysis tool (HBAT) to study strong and weak hydrogen bonds in a PDB file [1.111]. Thus HBAT is a new generation software for the convenient study of strong and weak hydrogen bonds in macromolecules.

1.6 Some technical glitches

Despite advances in macromolecular crystallography there still lie some problems in the quality of data it produces [1.14, 1.93]. They are: (a) location of hydrogen atom in macromolecules [1.112], (b) macromolecular crystal is not time stable, (c) the crystallographic resolution problem and (d) unavoidable errors during data collection and refinement.

1.7 Geometrical parameters

In this section, the geometrical characterization of hydrogen bonds is discussed. This is because the first section of the thesis deals with the hydrogen bond geometry in receptor–ligand complexes. The energetic implications of hydrogen bond can be referred elsewhere in the book of Jeffrey and Saenger (1991), Jeffrey (1997), Scheiner (1997), and Desiraju and Steiner (1999) [1.11–1.14].

1.7.1 *Distances and angles*

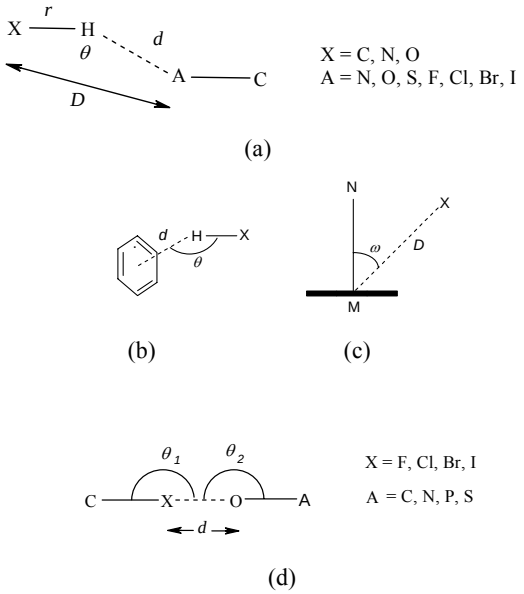
The directional property of the hydrogen bond makes it a unique non-covalent interaction. This directional property is best represented by the hydrogen bond distance and angle [1.11–1.14]. The hydrogen bond is constituted with a donor X−H and an acceptor A, and referred as X−H···A in the present thesis. Scheme 1.1*a* represents a typical hydrogen bond. Various annotations in the scheme are, d (H···A) the distance between hydrogen atom and the pertinent acceptor, D heavy atom distance, θ (angle X−H···A), r is X−H covalent distance. These three parameters are important in identifying hydrogen bonds. However,

another parameter angle ϕ, (angle H···A−C) is also taken into account to provide a strict geometric criterion for the hydrogen bond. In crystal structures of macromolecules, the H-atoms are usually not defined. In those cases, sometimes the distance between the heavy atoms (*D*) is assumed to be a criterion for putative hydrogen bond, which is a crude way to identify hydrogen bonds. However, there is no reasonable alternative in such cases.

1.7.2 *Geometric criteria for other weak interaction*

The geometrical criterion for a multi atom acceptor like phenyl rings is difficult to derive. The usual practice is the measured distances to the centroid of phenyl ring (M) Scheme 1.1*b* and *c*. However $d \leq 3.5$ Å, $\theta \geq 100^0$ and $\omega \leq 40^0$ appear to be satisfactory and this geometric criterion is generally accepted [1.15, 1.75].

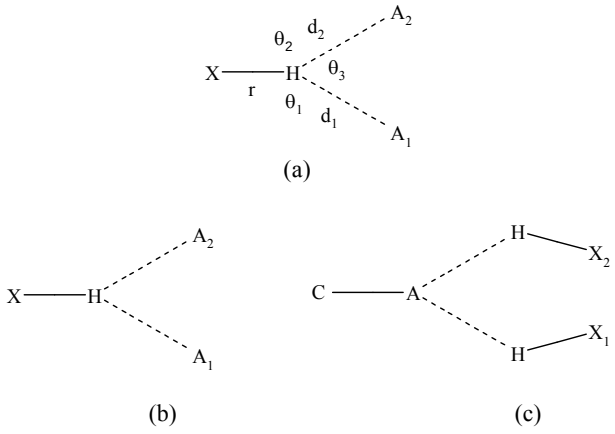
Organic halogens are included in the list of putative donors as well as acceptors (both as electrophiles and nucleophiles). The convention for a halogen bond is retained as C−X···O (Scheme 1.1*d*). A detailed description on the geometry of halogen is described by Auffinger *et al.* [1.86]. The geometry described for halogen as acceptors is in the form of X−H···halogen (here X = O, N, C). The convention for conventional hydrogen bonds can be applied to X−H···halogen interactions, considering the van der Waals criterion. A similarly criteria can be adopted for sulfur acceptors.



Scheme 1.1: (a) Representative hydrogen bond. A–C is a single or double bond, (b), (c) Parameters for X−H···π interactions (d) Parameters for halogen···O interactions. O–A is a double bond.

1.7.3 *Furcation in hydrogen bond*

The hydrogen bond furcation is a situation where a hydrogen atom is shared between more than one acceptor [1.12]. This situation is termed donor furcation. Terms such as bifurcated or '*3-centered*' and trifurcated or '*4-centered*' are also used (Scheme 1.2). Analogously, the converse situation where several donors approach a single acceptor is known as acceptor furcation. The conventions for bifurcated hydrogen bonds are presented by the distances r, d_1, d_2 and the angles θ_1, θ_2, θ_3.



Scheme 1.2: (a) Annotations in a bifurcated hydrogen bond, (b) and (c) represents bifurcated donor and acceptor respectively.

1.8 The weak hydrogen bond in protein–ligand complexes

Hydrogen bonds are instrumental in many molecular recognition events in biology [1.113, 1.114]. Very often, the recognition events are specific with conserved orientation. This reminds one of the exact events followed in a relay race, where the baton is exchanged among the fellow players with great precision and in a time-bound manner. Similar is the process of signal transduction and receptor–ligand interactions. The relaying of message becomes easy due to the weak nature of hydrogen bonds. In this context, the hydrogen bond in the molecular recognition event in the receptor–ligand interface is an attractive research proposition. Further, large numbers of receptor–ligand complexes are being deposited in PDB due to various structural initiatives, and this is an essential prerequisite to statistical studies. Apart from the final recognition events, hydrogen bonds are also responsible for physico-chemical properties of a molecule, like solubility, partitioning, distribution, and

permeability, which are crucial to drug development. Thus hydrogen bond research in drug discovery is been an evergreen area of research.

The topic of hydrogen bonds never limits to a particular type of donor and acceptor, rather refers to a variety of donor acceptor association. Unlike other biological events the protein–ligand interactions involves a plethora of strong and weak hydrogen bond. In fact the active sites manifest almost all the hydrogen bond interaction so far discussed. In particular the reversibility in receptor–ligand interactions is often governed by the weak nature of hydrogen bonds. This is because weaker hydrogen bonds are also easier to break. Presented here is an overview of selected studies describing the importance of weak hydrogen bonds in protein–ligand interactions and drug design.

1.8.1 *Evidence of weak hydrogen bonds in protein–ligand complexes*

The functional importance of C−H···O hydrogen bonds at the ligand binding domain of the retinoic acid receptor RARγ was first discussed by Klaholz and Moras (2002) [1.115]. The ligand binding domain of the retinoic acid receptor RARγ complexed with the retinoid SR11254 was revealed with the help of a 1.4 Å resolution crystal structure. They observed that the hydroxyl group of the ligand acts as a hydrogen bond donor and acceptor, leading to a synergy between the strong O−H···O and weak C−H···O hydrogen bonds. This in turn influences the specificity and affinity of the ligand is within the hydrophobic pocket of the receptor.

With the aim of designing possible kinase inhibitors Pierce *et al.* (2002) [1.116] carried out a statistical and theoretical analysis of C−H donors for aromatic ligands in a data set of 184 kinase crystal structures and 358 high-resolution structures from the PDB. The donor capacity of a variety of C−H donor in the protein–ligand complexes was analyzed. The activated C−H groups adjacent to the positively charged nitrogen of nicotinamide exhibit geometric preferences strongly suggested the hydrogen bond. This was also observed in the heterocyclic C−H groups in kinase ligands. Other aromatic C−H groups did not show such characteristics. *Ab initio* calculations by HF/6-31G** level revealed a considerable range of C−H···O hydrogen bonding potential among the different aromatic ring systems, with nicotinamide and heterocycles preferred in kinase inhibitors. These workers showed in particular many favorable hydrogen bond interactions. Further, in a subsequent work they showed that C−H···O hydrogen bonds play an important role in the binding of several analogues of a pyrazol-3-ylquinazolin-4-ylamine inhibitor to the

glycogen synthase kinase 3 (GSK3) [1.117]. Understanding of these C−H···O and C−H···N hydrogen bonds led them to design a novel quinazolin-4-ylthiazol-2-ylamine inhibitor of GSK3.

Denessiouk and Johnson (2003) [1.118] showed that the molecular recognition events for adenine (and some other nucleotides and nucleotide analogs) occur typically through three key hydrogen bonds. These three hydrogen bonds consists of two conventional (N−H···O and N−H···N) and a weak C−H···O hydrogen bond. Notably the weak C−H···O hydrogen bond was found to be an integral part of the adenine protein interaction. Adenosine as a ligand has a unique donor-acceptor-donor (DAD) pattern and the pertinent motif in the proteins is an acceptor-donor-acceptor (ADA). In fact the work carried out by Pierce *et al.* (2002, 2005) [1.116, 1.117] and Denessiouk and Johnson (2003) [1.118] has prompted me to pursue further study in this direction. This has been dealt in Chapter 4, which deals with a qualitative analysis of the interplay between strong and weak hydrogen bonds in the protein–ligand complexes of kinases. Kinases are a family of proteins having ATP as their natural substrate.

A statistical analysis of N−H···O, O−H···O, and C−H···O hydrogen bonds was carried out by Sarkhel and Desiraju (2004) [1.119] in a group of 28 high-resolution crystal structures of protein–ligand complexes from the PDB. The geometries obtained were compared *vis-à-vis* with the interactions found in small-molecule crystal structures from the Cambridge Structural Database (CSD). Some of the important conclusions derived from this study are, (a) both strong and weak hydrogen bonds are involved in ligand binding, (b) the restrictive geometrical criteria set up for hydrogen bonds in small molecule crystal structures may need to be relaxed in macromolecular structures due to extensive multifurcation, (c) the formation of C−H···O hydrogen bonds is enhanced by the activation of the C_α−H atoms and by the flexibility of the side chain atoms, (d) in contrast to small-molecule structures, anti-cooperative geometries were found to be common in the 28 macromolecular structures, (e) there is a gradual lengthening of hydrogen bond as the extent of furcation increases, (f) the C−H···O bonds formed by Gly, Phe, and Tyr residues are noteworthy, (g) number of hydrogen bond donors and acceptors agree with Lipinski's rule-of-five that predicts drug-like properties, (h) hydrogen bonds formed by water were also seen to be relevant in ligand binding and ligand C−H···O_w interactions are abundant when compared to N−H···O_w and O−H···O_w. This study has initiated my interest to pursue my study on strong and weak hydrogen bonds in receptor-ligand complexes. Chapter 3 of the

thesis is essentially the extension of the work carried out by Sarkhel and Desiraju (2004) [1.119], which describes a holistic view of strong and weak hydrogen bonds in a diverse set of protein–ligand complexes.

Cashin *et al.* (2005) [1.120] carried out a mutation study to elucidate the key interaction between three agonist acetylcholine, nicotine, epibatidine and nicotinic acetylcholine receptor. They suggested that the cation-π interaction is important in acetylcholine and main chain hydrogen bond is instrumental in nicotine binding to the receptor. However in the case of epibatidine both cation-π and main chain hydrogen bonds determine ligand binding. Further they augmented on the basis of theoretical calculations that the binding of agonist epibatidine to the receptor is strengthened by aromatic C−H···O hydrogen bonds.

The importance of strong and weak hydrogen bonds in protein–ligand docking is shown through molecular modeling study [1.121]. This was illustrated through a virtual screening (VS) study, targeted against epidermal growth factor receptor (EGFR) kinase domain. Acceptable results were obtained when the outputs from the commercial software packages were analyzed and modified on the basis of a chemical model that incorporates specific hydrogen bonds. We have shown that for 4-anilinoquinazoline type ligands, inclusion of a hydrogen bonded water molecule was indispensable to obtain meaningful VS results. Consideration of protein–ligand hydrogen bonds of the N−H···N, O_w−H···N and above all the C−H···O type was important to obtain accurate poses and binding affinities in the study.

In summary, the active site of the receptor is a unique environment where macromolecules and small molecules leave their footprints through a variety of strong and weak hydrogen bonds. The better understanding of these interactions in turn will help medicinal chemists and structural biologist to design better and safer drugs in the future.