CHAPTER 2

HBAT: A COMPLETE PACKAGE FOR THE ANALYSIS OF STRONG AND WEAK HYDROGEN BONDS IN MACROMOLECULAR CRYSTAL STRUCTURES

2.1 Introduction

Recent progress in structural biology with advanced techniques like high-throughput crystallography, cryogenic cooling for crystallization and data collection with synchrotron radiation has enabled researchers to determine large numbers of crystal structures of proteins at atomic resolution. The number of structures deposited in the Protein Data Bank (PDB) at or near atomic resolution is around 241 [2.1]. The availability of such high quality structural data has meant that it is possible to accurately characterize the geometrical properties of hydrogen bonds. The hydrogen bond is an attractive electrostatic interaction of the type X−H···A, where the H atom is attached to a donor atom, X (assumed to be more electronegative than H), and is directed towards an acceptor, A [2.2]. The acceptor A is normally an electronegative atom, usually O or N. Although most hydrogen bonds in proteins and nucleic acids are of the N−H···O or O−H···O (less often, N−H···N) type, weak hydrogen bonds such as C−H···O, N−H···S, O−H···S are also known sometimes to play significant roles in biological processes. Likewise, the π-electron clouds of aromatic rings can also act as acceptors for appropriately oriented X−H groups [2.3]. In general, there is a one to one correspondence between a hydrogen bond donor and acceptor. There however, also exist several examples where a hydrogen atom is shared between more than one acceptor. This situation is termed donor furcation. Terms such as bifurcated or '*3-centered*' and trifurcated or '*4-centered*' are also used. Analogously, the converse situation where several donors approach a single acceptor is known as acceptor furcation [2.2]. Cooperativity is another important stabilization feature in hydrogen bond arrangements and it occurs whenever the formation of a hydrogen bond facilitates the formation of another, so as to lead to sequences like O−H···O−H···O−H and so on. Cooperativity is well exemplified in the three-dimensional networks formed by water [2.4]. Halogen atoms are found to act as both electrophiles and nucleophiles interacting with electronegative O, N and π systems [2.4]. In summary, there is a plethora of strong and weak interactions observed in

macromolecules. Analyzing these interactions manually in any macromolecular structure is not practicable and the analysis of interactions in groups of crystal structures rapidly becomes impossible.

Hydrogen Bond Analysis Tool (HBAT) is an attempt to automate the analysis of all non bonded interaction present in a PDB file. There exist many programs like HBPLUS [2.6], HBEXPLORE [2.7], CONTACT [2.8], BNDLST and many web based servers like LPC [2.9], NCI [2.10], which are available in the public domain for analyzing these interactions. The advantage of HBAT over the other available software is its compactness to deliver useful information pertaining to all types of interactions in a more user-friendly way. The program does not require additional information to deal with ligand bound PDB structures. The result can be easily exported to a MSOFFICE Excel file for further analysis. The program finds a wide range of applications from crystallography to structure based drug design and molecular dynamics simulations.

2.2 Program description

HBAT is written using PERL language and the Graphical User Interface (GUI) was built using the Tk library [2.11]. The phenomenon of cooperativity is addressed using a graph theory approach and the visualization of the cooperativity graph is obtained using the open source Graphviz software which is included in HBAT [2.12].

2.2.1 *Hardware and software requirement*

HBAT can be easily installed and run on any personal computer. Operating systems: Windows 98/2002/Me/NT/XP with minimal memory requirement of 512MB RAM.

2.2.2 *Program availability*

The program is available free of charge for academic users in non-profit organization after the license agreement has been signed and faxed. More about the program can be found at HBAT home page (http://202.41.85.161/~grd/HBAT.html).

2.2.3 *Input details*

HBAT requires structural information in the form of a PDB format. The crystal structure should be free from major errors, i.e. it should not contain any missing residues or atoms. The input structure should contain H-atoms for a full analysis of the interaction. A quick force field based minimization of the hydrogen atoms before using HBAT is recommended in order to get more reliable results. The program is not responsible for accurate generation of H-atom positions.

2.3 Methodology

Initially the nearest neighbor list for each atom is generated with a default search radius (5.0 Å) or any user defined cut-off value. The nearest neighbor list is further screened for hydrogen bonds and van der Waals interactions based upon the sum of the van der Waals radii of the interacting atoms. The putative hydrogen bonds are identified using default or user defined cut-off values for distance and angles. A separate cut-off parameter is defined to search for X−H···π bonds, halogen bonds, and cooperative interactions in proteins.

HBAT gives a user-friendly, click-and-use facility to calculate matrices for hydrogen bonds from a PDB file. It allows users to easily define geometric parameters for all types of interactions based upon individual user requirements. It also gives a quick summary of various types of interactions in the form of pie charts and histograms so that the user can easily visualize and compare the results.

2.3.1 *Identification and classification of interactions*

Hydrogen bonds are tabulated more informatively according to the type of hydrogen bond, donor atom, acceptor atom, donor residue name/number, acceptor residue name/number, distances r, d and D, and hydrogen bond angle θ.

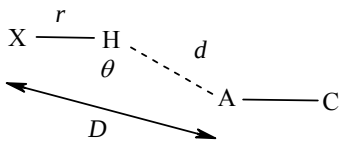
All the hydrogen bonds are further subjected to statistical analysis by classifying them in terms of backbone donor/acceptor, side chain donor/acceptor, ligand donor/acceptor, water donor/acceptors, and nucleic acid donor/acceptors. The respective frequencies of their occurrence are tabulated for all types of hydrogen bonds. Further, a detailed frequency distribution for each type of bond is listed in a 10º degree interval starting at $90º < \theta < 180º$. Similarly the distance d, distribution is specified from 1.2 Å to 3.2 Å with an interval of 0.4 Å. In a macromolecular structure the interacting partners are

protein, ligand, water and sometimes nucleic acid. Along with distance versus angle distributions, are also listed the frequencies with which particular amino acid residues, water, ligands, or nucleic acids residues are involved in various type of hydrogen bond (e. g. N−H···O, C−H···O etc.). Accordingly the user can easily identify the number of contributions from each donor and/or acceptor. For interactions involving π−acceptors and halogen bonds only a list of interactions is given because the number of interactions is rather small.

The following brief description of the interaction types is meant to define the various geometrical parameters used in the program.

2.3.2 *Hydrogen bonds with single atom acceptors*

The general hydrogen bond of this type may be described in terms of the *d*, *D*, *θ*, and *r* values as shown in Scheme 2.1.

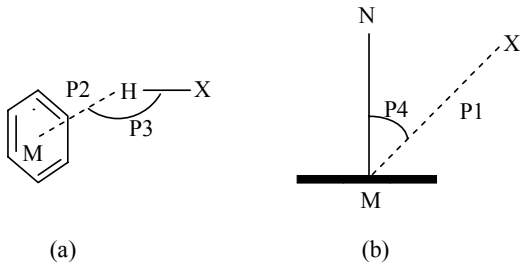


X = C, N, O
A = N, O, S, F, Cl, Br, I

Scheme 2.1: Representative hydrogen bond.

2.3.3 *Hydrogen bonds with multi-atom acceptors*

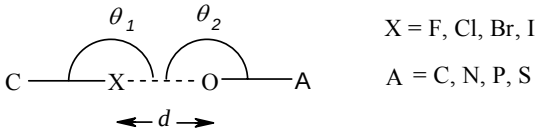
In a π-electron acceptor, the distances are usually measured to the centroid (M) of multiple bonds or phenyl rings. The unit vector is defined by MN. An aromatic ring is represented as a hexagon. In the case of Trp, the five-membered and six-membered ring systems are considered separately. The donor group is represented as X−H, where X is similar to the situation described above for single acceptor hydrogen bonds. For a halogen "π" interaction, the H atom may be replaced by a halogen (F, Cl, Br, I). P1 and P2 are distances from X and H respectively to M. P3 is the angle between vectors X−H and H−M while P4 is the angle between the XM and MN. The geometry is adapted from earlier work of Babu, 2003 [2.10]. Scheme 2.2.



Scheme 2.2: Parameters for X−H···π interactions.

2.3.4 *Halogen bonds*

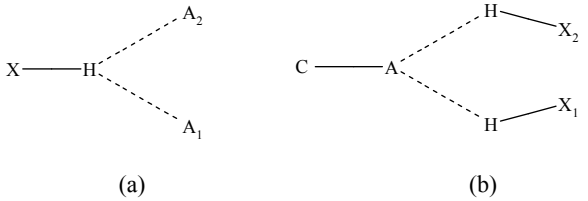
Organic halogens are included in the list of putative donors as well as acceptors (both as electrophiles and nucleophiles). The convention for a halogen bond is retained as C−X···O, Scheme 2.3 [2.5]. The geometry adopted for halogen as acceptors is as in the case of hydrogen bonds.



Scheme 2.3: Parameters for halogen···O interactions

2.3.5 *Donor and acceptor furcation*

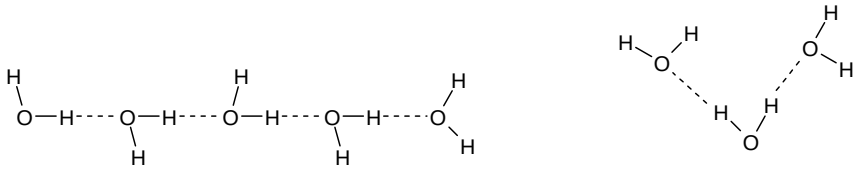
Donor and acceptor *'n-centered'* furcations are listed separately. The list contains '*2-centered'* (not furcated) interaction to '*7-centered'* (six furcated) interactions. A schematic representation of *3-centered* (bifurcated) donor/acceptor furcation is represented in Scheme 2.4.



Scheme 2.4: (a) Bifurcated donor (3-*centered)*, (b) Bifurcated acceptor (3-*centered)*

2.3.6 *Cooperativity*

Cooperative and anti-cooperative geometries for water molecules are shown in Scheme 2.5*a* and *b* respectively. However similar arrangements can be observed in a system involving protein, ligand and water.



(a) (b)

Scheme 2.5: Cooperative and anticooperative geometries.

2.4 Example

To evaluate the accuracy of the program, HBAT was used to reproduce the published geometries in several recent papers [2.12–2.16]. Some of the utilities of the program are listed in the following section taking the example of PDB ID 6RSA.

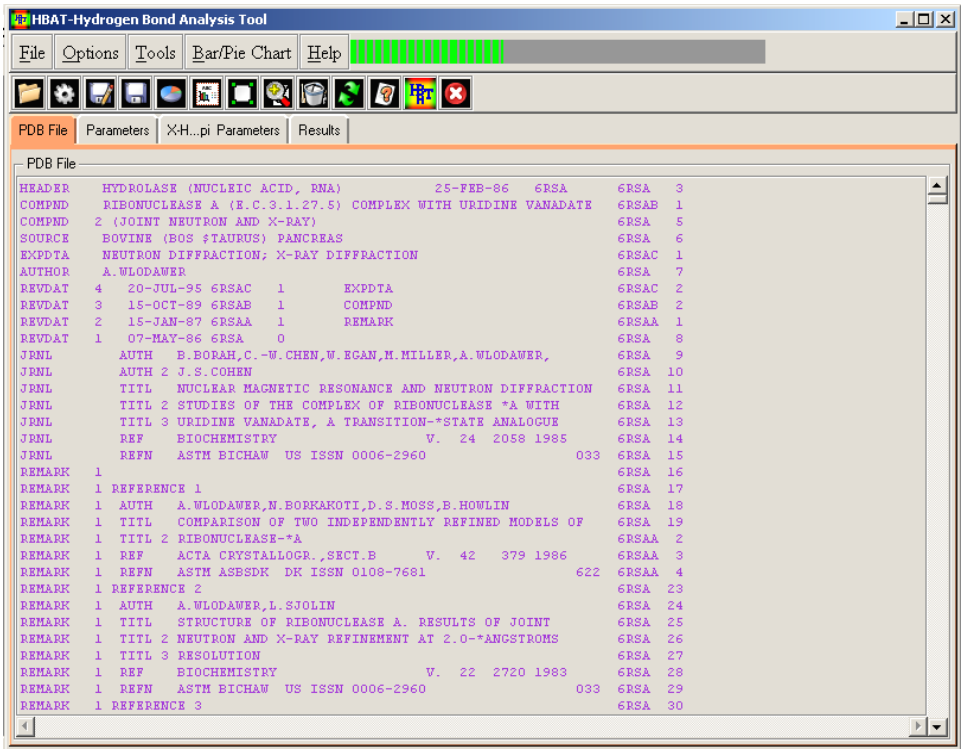


Figure 2.1: Input PDB file and the content of the HBAT input window. Also shown is the main GUI in the process of calculating various interactions.

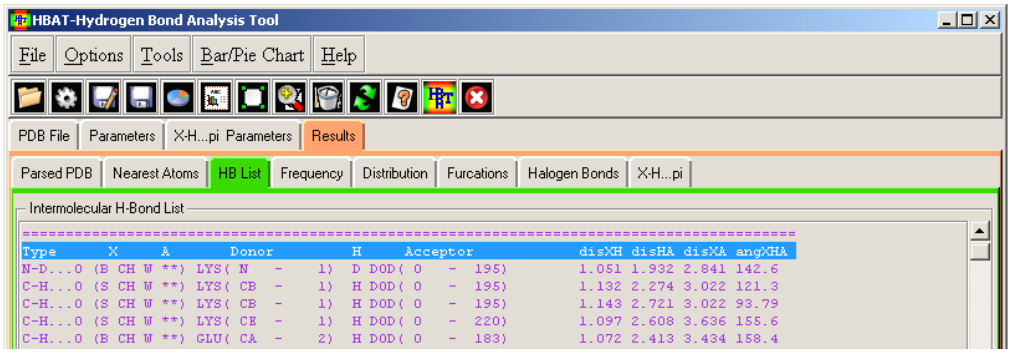


Figure 2.2: Section of result window. The active sub page is highlighted in green listing the segment of observed interactions. Other inactive sub pages contain results like frequency, geometry distribution, furcation, cooperativity, halogen bond, interactions involving π-acceptors.

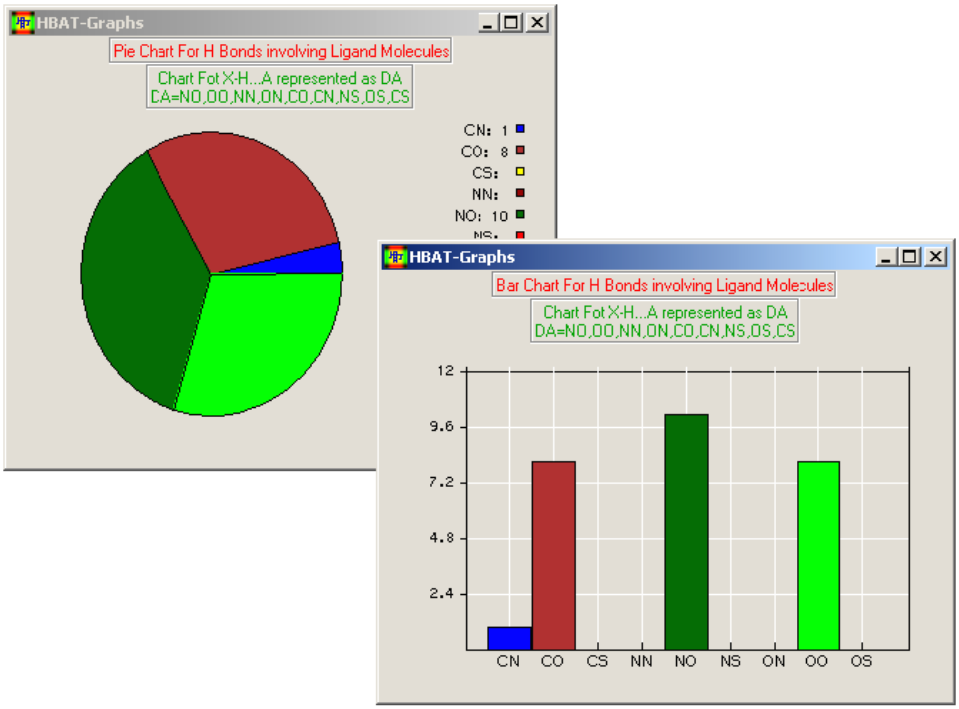


Figure 2.3: Various types of hydrogen bond interaction between protein and ligand are presented through pie chart and histogram depictions.

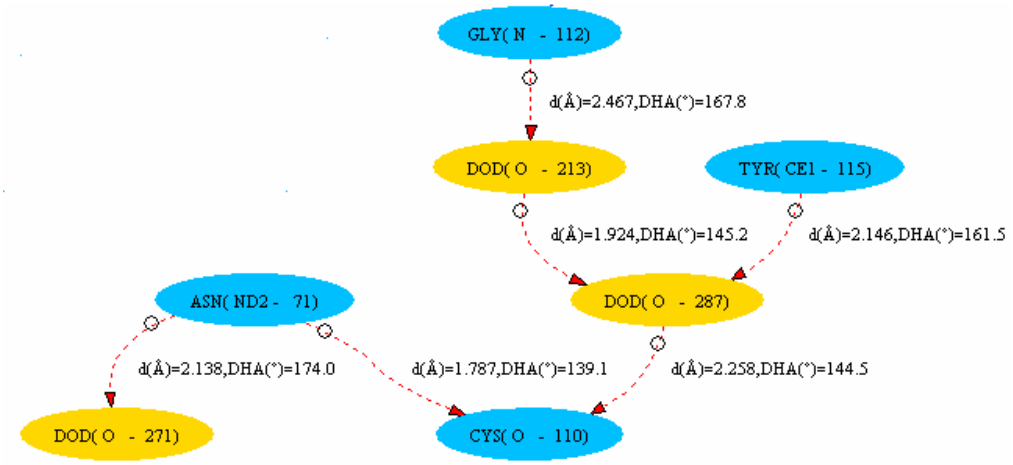


Figure 2.4: Cooperative interactions in 6RSA. Amino acid and water are represented in blue and yellow colored boxes respectively. Two cooperative chains are shown: (1) consists of the sequence Gly 112, Water 213, Water 287, Cys 110, (2) Thr 115, Water 287, Cys 110.

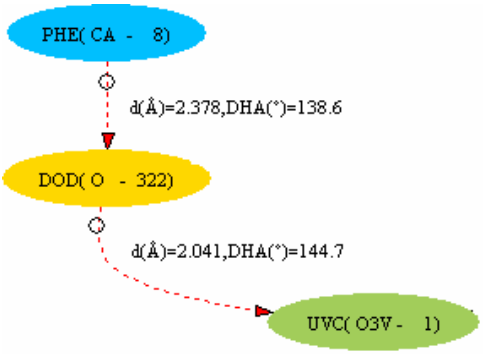


Figure 2.5: Water mediated protein–ligand cooperative interactions. The ligand is colored green.

2.5 Conclusions

Described here is a new software that allows easy analysis of a variety of hydrogen bonds and other interactions in macromolecular crystal structures.