

# **HBexplore 2.0.1**

[http://www.imb-jena.de/www\\_bioc/hbx/hbx.html](http://www.imb-jena.de/www_bioc/hbx/hbx.html)

## **Geometrical Analysis of Hydrogen Bonding Patterns in Biological Macromolecules**

**Klaus Lindauer, Michael Meyer, Friedrich Haubensak, Jürgen Sühnel**

Biocomputing  
Institut für Molekulare Biotechnologie  
Postfach 100813, D-07708 Jena / Germany

Tel.: +49-3641-656200  
Fax: +49-3641-656210  
E-Mail: [jsuehnel@imb-jena.de](mailto:jsuehnel@imb-jena.de)

2002

## Table of contents

1.	Introduction.....	3
2.	Geometrical criteria .....	3
3.	Input files .....	5
4.	Output files .....	10
5.	Analysis tools .....	10
6.	Miscellaneous .....	17
7.	Installation .....	18
8.	Known problems .....	20
9.	New features of HBexplore 2.0 .....	21
10.	Utility programs .....	22
11.	Copyright and Availability .....	22
12.	Citation .....	22
13.	Papers that have made use of HBexplore .....	23
14.	References .....	24

## 1. Introduction

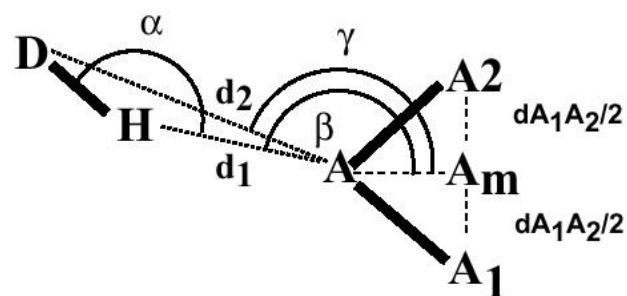
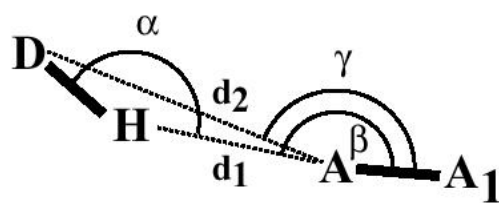
The program HBexplore is a tool for identifying and analyzing hydrogen bonding patterns in biological macromolecules. In addition to classical hydrogen bonds with nitrogen or oxygen donor atoms (Baker and Hubbard, 1984) C-H...O/N interactions (Wahl and Sundaralingam, 1997; Weiss et al., in press) can be investigated. The program selects all potential hydrogen bonds according to geometrical criteria and generates a hydrogen bond table. This table can then be subjected to further automatic or interactive analysis tools. These tools include the calculation of mean values and distributions of geometrical hydrogen bond parameters for parts of a single structure, for complete single structures and for structure sets, the classification of each H-bond according to the participation of backbone, sidechain or base, ligand and water parts of nucleic acids or proteins, identification of Watson-Crick nucleotide pairs and of H-bonded pairs of equal nucleotides, the calculation of the mean number of H-bonds per residue and of the fraction of potential donor and acceptor atoms involved in H-bonds. HBexplore generates further automatically a H-bond residue interaction table. This table lists for all residues of the structure the other residues, ligands or water molecules directly connected via H-bond. By means of a binary tree search algorithm this table is then converted into a H-bond residue cluster table. Clusters are understood here as an uninterrupted network of H-bonded residues. For nucleic acids the secondary structure and tertiary interactions are automatically derived from the hydrogen bonding pattern. Finally, two types of H-bond residue type interaction matrices are calculated. They provide information on the occurrence of hydrogen bond interactions between residue types. In matrix I each interaction between two residues is counted as 1, independent on the actual number of H-bonds. In matrix II each H-bond is considered as one interaction.

## 2. Geometrical criteria

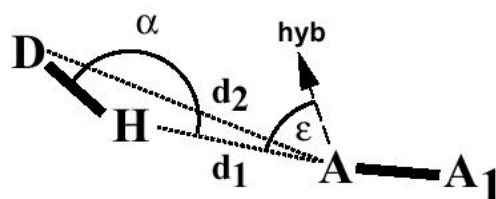
HBexplore requires structural information in the Protein Data Bank (PDB) format (Bernstein et al., 1977). For structures determined by diffraction methods the position of the hydrogen atoms are usually not resolved. We are convinced that the position of H-atoms are required for a reliable identification of hydrogen bonds by geometrical criteria. Therefore, HBexplore generates in a first step hydrogen coordinates according to standard geometrical rules (Cornell et al., 1995). The position of the hydrogen atoms are dependent on the hybridization of the donor atoms and on the atomic environment. For details of the geometrical parameters used see the explanation of the connectivity files. HBexplore can apply two criteria sets for identification of potential H-bonds. Set I uses exclusively atom coordinates and set II takes into account the directions of hybrid orbitals on hydrogen bond acceptor atoms. The information necessary for the calculation of the H-atom coordinates and hybrid vectors of the acceptor atoms is stored in the so-called connectivity files for donor and acceptor atoms. These files can easily be edited to include further residues or ligands with H-bonding capabilities.

The two criteria sets are described in more detail in the following Figure. Criteria set I is identical to the criteria used by HBPLUS (McDonald and Thornton, 1994). If the acceptor atom A is part of a cycle, then there are two bonded neighbours of A, A 1 and A 2 . In this case A m is used as bonded neighbour of A instead of A 1 for calculating the angles. A m is located midway on the line connecting A 1 and A 2 . Criteria set II checks in addition to the distance constraints D...A, H...A and the angle DHA, the angle between the H...A hydrogen bond and the direction of the hybrid orbital of the acceptor atom A. HBexplore calculates the hydrogen bond table with either of the two criteria sets and can also use logical combinations of the two sets. In the Figure the default cutoff values are given. They can easily be changed by the user. When using 2.5 Å as cutoff value for the H...A distance a D...A value of 3.9 Å is too large. The reason for using this value is a practical one. The D...A distance is first used for selecting all bond which fulfill this criterion, at least. To be sure to find all potential H-bonds a rather large value is used. In a second step the other criteria are applied. With a H...A cutoff value of 2.5 Å this leads to an effective D...A cutoff of about 3.5 Å.

I.



II.



Criteria set I (default values):  $d_1 < 2.5 \text{ \AA}$ ,  $d_2 < 3.9 \text{ \AA}$ ,  $\alpha > 90^\circ$ ,  $\beta > 90^\circ$ ,  $\gamma > 90^\circ$ .

Criteria set II (default values):  $d_1 < 2.5 \text{ \AA}$ ,  $d_2 < 3.9 \text{ \AA}$ ,  $\alpha > 90^\circ$ ,  $\beta > 90^\circ$ ,  $\epsilon > 60^\circ$ .

$d_1$	-	distance H...A
$d_2$	-	distance D...A
$\alpha$	-	angle DH...A
$\beta$	-	angle H...AA <sub>1</sub>
$\gamma$	-	angle D...AA <sub>1</sub>
$\epsilon$	-	angle H...Ahyb

HBexplore treats as potential H-bond donor and acceptor atoms: O, N, S.

### 3. Input files

#### *PDB files*

The program HBExplore requires structural information in the Protein Data Bank (PDB) format. This means that all atoms have to be accurately named and ordered according to the PDB format. The results generated by HBExplore may be incorrect, for example, if the coordinates of a heavy atom within a residue is missing. HBExplore needs the exact ordering of the heavy atoms to find the correct neighbours of potential donor or acceptor atoms. In search of the bonded neighbours of a donor atom, for example, the program jumps stepwise starting from the donor atom. If the coordinates of atoms are missing then HBExplore uses the coordinates of wrong atoms.

#### *Connectivity files*

The connectivity files contain all information on the potential hydrogen bond donor and acceptor atoms. The six connectivity files contain eight columns. The first column represents the potential donor or acceptor atom. The second shows the code for the residue type. The next four columns indicate the neighbours of the potential donor or acceptor atoms. The number code is essential for finding the coordinates of the correct neighbour atoms. The program uses this number for jumping stepwise to the coordinates of the appropriate neighbour atoms starting out from the coordinate record of the corresponding donor or acceptor atom. The next column indicates the hybridization of the donor or acceptor atoms and the last column represents a classification according to RNA-DNA (R), proteins (P), water (W) and ligands (L).

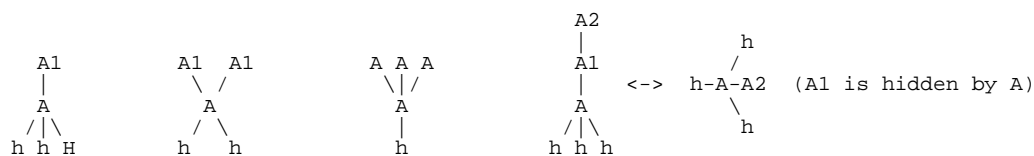
In order to facilitate a flexible analysis of different biopolymer structures the connectivity information has been spread over the six files.:

```
PROTEIN_LIGAND_connect_acceptor
PROTEIN_LIGAND_connect_donor
DNA_connect_acceptor
DNA_connect_donor
RNA_connect_acceptor
RNA_connect_donor
```

The first two files are the default connectivity files for proteins, ligands and water. The remaining files may be loaded additionally to analyse DNA, RNA or complexes of nucleic acids with proteins or ligands. Note, that the program is currently not able to do a standard analysis for DNA-RNA hybrids.



### sp<sup>3</sup>: tetrahedral



code: 321  
angle (A(1)Ah) = 109 deg

32

311

33

The two 33 representations are two views of one and the same conformation which show that h and A2 point in opposite directions.

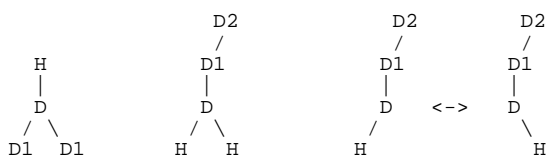
### Connectivity-file (donor)

```
Atom Base Co.b i.b Co.a i.a Hyb
12345123451234512345123451234512345
N6 A C6 -1 C5 -2 22 R
O2* A C2* -1 C1* +1 31 R $ neighbour code:
-----
N1 G C6 -2 C2 +1 21 R # -n indicates that the neighbour atom is n positions
N2 G C2 -1 N3 +1 22 R # before the donor atom (in the PDB file)
O2* G C2* -1 C1* +1 31 R $ # +n indicates that the neighbour atom is n positions
----- # after the donor atom (relating to the PDB file)
N3 U C2 -2 C4 +1 21 R
O2* U C2* -1 C1* +1 31 R # hybridization code for the donor atom:
O2' U C2' -1 C1' +1 31 R $ # the first number is for hybridization type
----- # (e.g 3 for sp3).
N3 T C2 -2 C4 +1 21 R # the second number indicates the number of
O2* T C2* -1 C1* +1 31 R $ # hydrogen atoms (e.g 2 for N6 in A).
----- # the third number indicates the number of
N4 C C4 -1 N3 -2 22 R # hybrid orbitals in the case of the modified
O2* C C2* -1 C1* +1 31 R $ # base, where the atom can be donor and acceptor
----- # (e.g 1 for N6 in 1MA).
N20 YG C15 -5 C21 +1 21 R # $ # for O in water the code is 3.
O2* YG C2* -1 C1* +1 31 R $
-----
N1 M2G C6 -2 C2 +1 21 R R - RNA / DNA
O2* M2G C2* -1 C1* +1 31 R $ P - Protein
----- W - Water
N GLN CA +1 C +2 21 P L - Ligand
NE2 GLN CD -2 OE1 -1 22 P $
```

### Description of the hybridization code

Code	Hybridization	Number of hydrogens to be added	number of bonded neighbours	distance D - H in Å
21	sp <sup>2</sup>	One	Two	1.01
22	sp <sup>2</sup>	Two	One	1.01
211, 210	sp <sup>2</sup>	one	One	1.01
31	sp <sup>3</sup>	one (in the direction of a potential acceptor atom)	One	0.96
311	sp <sup>3</sup>	One	Three	1.09
32	sp <sup>3</sup>	Two	Two	1.09
33	sp <sup>3</sup>	Three	One	1.01
3	sp <sup>3</sup>	one (in the direction of a potential acceptor atom)	Water	0.96

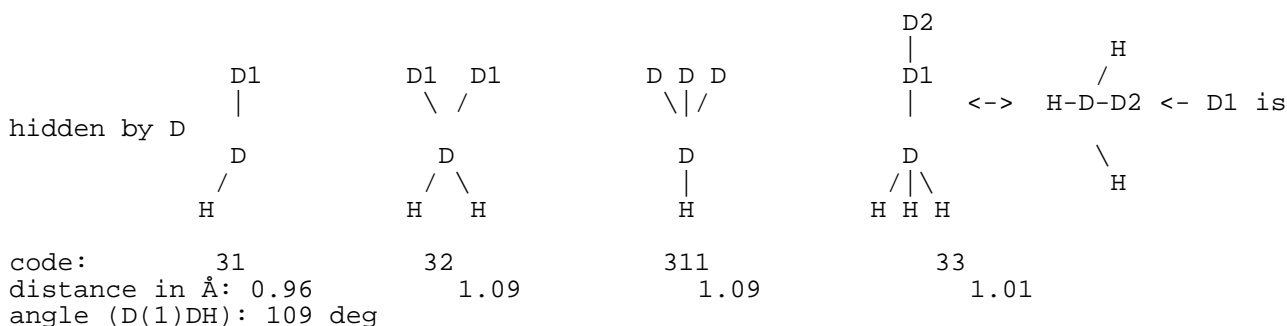
### sp<sup>2</sup>: trigonal planar



plane: D1-D-D1 D2-D1-D D2-D1-D  
code: 21 22 211, 210  
distance in Å: 1.01 1.01 1.01  
angle (D1DH): 120 deg

The two 211 representations represent two alternative conformations. In this case the conformation with the shortest hydrogen - acceptor distance is used.

### sp<sup>3</sup>: tetrahedral



The two 33 representations are two views of one and the same conformation which show that H and D2 point in opposite directions.

## 4. Output files

The output filenames generated by HBexplore are formed according to simple rules. For PDB filenames the PDB code is extracted and HBexplore adds filename extensions. If the input filename is not a PDB filename the program uses the full name and adds to that name extensions.

The following extensions are used:

*_hbx_DAs*	All donor - acceptor interactions which fulfil the distance constraint dDA ( e.g. dDA < 3.9 Å)
*_hbx_all	All potential H-bonds which fulfil all geometrical parameters of both criteria sets
*_hbx_anal	Selection of one of the possible criteria, classification, statistics
*_hbx_freqs.*	Tables of the distribution of geometrical parameters (useful as input files for graphic programs )
*_hbx_freq.*	Tables of the distribution of geometrical parameters of classified H-bonding sets
*_hbx_sequ-idl-arr	All interactions between residues are represented in a matrix
*_hbx_h	Coordinates of hydrogen atoms are given in the PDB format
*_hbx_hyb	Coordinates of the hybrid vectors are given in the PDB format
*_hbx_h2o	List of all potential H-bonds with water
*_hbx_ent	PDB file with coordinates of hydrogen atoms involved in hydrogen bonds

It may happen that the matrices in \*\_hbx\_anal do not find on a page. In this case you can use the command pstext for converting plain text PostScript with an appropriate pointsize. For example,

```
➤ pstext -s7 206d_hbx_anal > 206d_hbx_anal.ps
```

converts the plain text file 206d\_hbx\_anal to the PostScript file 206d\_hbx\_anal.ps with fonts of pointsize 7.

## 5. Analysis tools

The major part of the results of H-bond analysis are given in the files **\*hb<sub>x</sub>\_all** and **\*\_hb<sub>x</sub>\_anal**. The file **\*hb<sub>x</sub>\_all** contains all H-bonds selected according to both criteria sets. On the other hand, the file **\*hb<sub>x</sub>\_anal** provides only H-bonds which fulfill the criteria set adopted (set I, set II or a logical combination of both sets). In addition, it contains results of various types of analysis of the basic set of H-bonds identified. It is our primary working tool. In the following the analysis tools are described in more detail. The results shown are taken from an H-bond analysis of an RNA pseudoknot (PDB code: 1rnk; Shen et al. 1995).

The file **\*\_hb<sub>x</sub>\_all** contains for both criteria (set I and set II) a list of all H-bonds identified. Each list shows all geometrical parameters, see the following Table.

PDB Code / filename "\*\*\*\* 1rnk \*\*\*\*"  
Criteria set 1

-----  
Distance DA (dDA) < 3.9  
Distance HA (dHA) < 2.5  
Angle DHA (aDHA) > 90.0  
Angle HAA (aHAA) > 90.0  
Angle DAA (aDAA) > 90.0  
-----

No.	Donor			Acceptor			DDA	DHA	aDHA	aHAA1	aHAA2	aDAA1	aDAA2
13	N4	C	5	O6	G	15	2.610	1.769(1)	138.2	132.1	153.7	133.6	160.7
14	O2*	A	6	O1P	G	7	2.703	1.734	179.9	98.6	117.0	98.7	117.0
15	O2*	A	6	O4*	G	7	2.616	1.683	162.8	98.6	150.7	96.9	153.2
16	N6	A	6	O1P	C	5	2.730	2.119(1)	117.0	148.4	119.6	130.3	100.7
17	O2*	U	8	O4*	G	9	2.756	1.803	171.6	153.6	96.3	151.7	98.0

The file **\*hb<sub>x</sub>\_anal** contains first a list of all H-bonds (which fulfill the criteria of the particular criterion selected) sorted according to the individual residues. Due to the fact that each H-bonds connects two residues this list contains two times the total number of H-bonds. In this H-bond table only the D...A distance is given.

#### H-bond table

Potential H-bonds sorted according to:

-----  
nucleotide, amino acid, ligand, water

at - atom, rt - residue type, rn - residue number

dDA - distance (D-A) in Å

ty - type:

B - Backbone, b - base, s - sidechain, l - ligand, w -water

No.	Donor			Acceptor			dDA	Ty
	at	rt	Rn	At	rt	rn		
14	O2*	A	6	O1P	G	7	2.7029	B-B
15	O2*	A	6	O4*	G	7	2.6155	B-B
16	N6	A	6	O1P	C	5	2.7299	b-B
14	O2*	A	6	O1P	G	7	2.7029	B-B
15	O2*	A	6	O4*	G	7	2.6155	B-B

The number in the first column refers to the numbering of H-bonds in the file **\*hb<sub>x</sub>\_all**.

#### Classification

The further lists contain the results of subgroups of the identified H-bond set e.g. Backbone - Backbone bonds. For each of these subgroups the distribution of geometrical parameters is calculated and can be stored in separate files. The next Tables provide information on the general H-bond statistics.

#### H-bond statistics

Number of H-bonds:

-----  
total number - Tn, frequency F

	Tn	F
Backbone - Backbone	24	0.39
Base - Base	34	0.55
Backbone - Base	4	0.06
Backbone - Sidechain	0	0.00
Sidechain - Sidechain	0	0.00
Sidechain - Base	0	0.00
Watson-Crick pairs	29	0.47
Pairs with equal bases	1	0.02
Protein - Protein	0	0.00
Nucleic acid - Nucleic acid	62	1.00
Nucleic acid - Protein	0	0.00
With water	0	0.00
Total	62	1.00

Number of residues: 34

Mean number of H-bonds per residue:

With water: 3.647  
Without water: 3.647

Number of potential donor atoms:

	all	involved in H-bonds
Backbone:	34	21
Base:	45	34
Sidechain:	0	0
Water:	0	0
total:	79	55

Number of potential acceptor atoms:

	all	involved in H-bond
Backbone:	202	26
Base:	87	36
Sidechain:	0	0
Water:	0	0
total:	289	62

Note that the mean number of H-bonds per residue is obtained by dividing two times the number of H-bonds by the number of residues.

#### *H-bond residue interaction table*

This table lists for each individual residue to which other residues H-bonds are formed. In this way one can identify at the first glance interesting residues with a large number of H-bonded other residues.

H-bond residue interaction table:

G1	19				
G2	18				
C 3	4	17			
G 4	3	5	26	16	27
C 5	4	27	15	6	
A 6	7	5			
.	.	.			

### *Nucleic acid secondary structures and tertiary interactions*

For nucleic acids the secondary structure and tertiary interactions are automatically derived from the three-dimensional structure via the hydrogen bonding pattern. It is assumed that a nucleic acid structure consists of helix-like and non-helix-like secondary structure elements. Helix-like elements are formed from at least three H-bonded nucleotide pairs of subsequent nucleotides in different sequence regions of one strand or in different strands. Non-helix-like elements are all other parts of a nucleic acid structure. All H-bonds within one and the same secondary structure element are assumed to belong to this element. This includes for example, H-bonds between non-neighbour nucleotides in loops and H-bonds between a nucleotide and more than one other nucleotide in helix-like regions. H-bond tertiary interactions are defined as H-bonded nucleotide pairs for which the nucleotides involved do not belong to one and the same secondary structure element.

### *Protein secondary structures*

The secondary structure information enclosed in PDB files for proteins is used directly by HBExplore. The \*\_**hbx\_**anal file contains a section with potential hydrogen bonds listed according secondary structure element. At the end of each row the assignment of the donor and acceptor atoms to a helix, sheet or turn is given. Upper case letters are used for potential hydrogen bonds within a secondary structure element, lower case letters are used for hydrogen bonds between different secondary structure elements.

### *H-bond residue cluster table*

The residue interaction table is converted to the residue cluster table, where a cluster is understood as an uninterrupted network of H-bonded residues. The residues represents the knots and the H-bonds the interactions (edges). The following table contains two clusters..

H-bond residue cluster table :

-----

* 1 - 19	* 19 - 18	* 18 - 2	* 19 - 20	* 19 - 24	
	* 24 - 23	* 23 - 22	* 24 - 25	* 25 - 26	* 26 - 4
	* 4 - 3	* 3 - 17	* 17 - 16	* 16 - 4	* 16 - 15
	* 15 - 5	* 5 - 4	* 5 - 27	* 27 - 4	* 27 - 26
	* 27 - 17	* 5 - 6	* 6 - 7	* 15 - 14	* 14 - 13
	* 13 - 28	* 28 - 29	* 29 - 12	* 29 - 30	* 30 - 11
	* 30 - 31	* 31 - 10			
* 8 - 9	* 9 - 32	* 32 - 33	* 33 - 8	* 33 - 34	
	* 34 - 8				

### H-bond residue type interaction matrix

HBExplore generates two different residue type interaction matrices. For the first matrix (I) the program utilizes the information of the residue interaction table. This means that each H-bond interaction between two residues is counted as 1 independent of the actual number of H-bonds between them.

#### H-bond residue type interaction matrix I:

-----

(An interaction is counted only one time even if there are more than one H-bonds between a special residue pair. Modified residues are considered as unmodified ones; refers also to T/U)

#### Frequency of the residue types used in the primary structure:

-----

A	G	U	C	SUM
0.24	0.32	0.12	0.32	1.00

#### Total number:

-----

	A	G	U	C
A	3	5	3	6
G	5	1	2	14
U	3	2	1	1
C	6	14	1	2

#### Residue types involved in interactions:

-----

A	G	U	C
20	23	8	25

Total number of interactions: 38  
Number of H-bonds (without water) per interaction: 1.632

For the second matrix all H-bonds are used for calculating the matrix, where now the rows stand for acceptor residues and columns for donor residues.

#### H-bond residue type interaction matrix II:

-----

(All H-bonds are taken into account.  
Modified residues are considered as unmodified ones;  
refers also to T/U)

"x-axis" (horizontal) - acceptor

"y-axis" (vertical) - donor

Total number

	A	G	U	C	SUM
A	5	4	2	2	13
G	2	1	1	21	25
U	2	2	1	1	6
C	5	11	0	2	18
SUM	14	18	4	26	62

So far, there is no statistical evaluation of these data.

### *Interactive search*

H-bonds can be selected interactively according to various atomic, residue and geometrical search criteria. Furthermore, multicentre hydrogen bonds can be found, protein secondary structure element specific searches can be carried out and intermolecular interactions can be analysed. For a single structure the search criteria can be combined in successive cycles using logical operators. Note, that this search uses the H-bond table generated in the automatic run. This means that if one wants to increase cutoff values in the interactive search beyond the values used in the automatic run one has to repeat the calculation with larger cutoff values.

### *Search for specific H-bonds*

#### Geometry:

Distance: Donor - Acceptor (D...A) in Å

\* Default range: 0.00 - 3.90

\* Use default range: 0

\* Select new range: 1

\* Enter selection:

0

--- Intervall: 0.00 -- 3.90 ---

Distance: H - Acceptor (H...A) in Å

\* Default range: 0.00 - 2.50

\* Use default range: 0

\* Select new range: 1

\* Enter selection:

1

Lower limit:

0.0

Upper limit:

2.5

--- Intervall: 0.00 -- 2.50 ---

Angle: Donor - H - Acceptor (D-H...A) in deg

\* Default range: 90.00 - 180.00

\* Use default range: 0

\* Select new range: 1

\* Enter selection:

0

--- Intervall: 90.00 -- 180.00 ---

### *Molecular specification:*

Donor:

\* All atom/residue types: 0

\* Atom type (e.g. O2\*): 1

\* Residue type (e.g. G): 2

\* Enter selection:

1

Input of the atom type

N6

--- atom: N6 ---

Acceptor:

\* All atom/residue types: 0

\* Atom type (e.g. O2\*): 1

\* Residue type (e.g. G): 2

\* Enter selection:  
1  
Input of the atom type  
O4  
--- atom: O4 ---

Number of H-bonds: 1

## 6. Miscellaneous

### *Treatment of water*

If the structure contains water the user is asked if water should be used taken into account.

### *Water as H-bond donor*

In this case the hydrogen is placed directly on the straight line connecting donor and acceptor atoms. This means that the DHA angle is always 180 deg. Note, that this approach may identify mutually exclusive H-bonds.

### *Water as aH-bond acceptor*

The hybrid orbital vector is assumed to point directly to the donor hydrogen. The water oxygen has no bonded neighbour. Therefore, HBexplore uses for criteria set I a restricted set of cutoff values: dDA 3.9 Å, dHA 2.5 Å, aDHA 90deg.

### *Rotation of sp<sup>3</sup> donors*

For sp<sup>3</sup> donor groups the D-H group is rotated around the D-D1 bond. The hydrogen is then placed on the straight lines connecting D with all acceptor atoms fulfilling the criteria adopted. If there is more than one acceptor atom within the donor-acceptor distance limit (dDA 3.9 Å) mutually exclusive H-bonds are identified.

### *Disulfide bonds*

Disulfide bonds are identified if the distance between the sulfur atoms is smaller than 3.0 Å. Potential disulfide bridges are reported in the \*\_**hbx\_all** file. If a disulfide bridge is found according to this criterion the interaction is not classified as a H-bond. However, it may happen that H-bonds with the sulfur atoms of disulfide bonds may occur.

### *Analyzing structure sets*

After having finalized the H-bond analysis of a single structure HBexplore asks for the next structure file. In this way one can prepare scripts for the automatic analysis of structure sets. Distributions of geometrical H-bond parameters and a few other statistical data for the complete are stored in the file **hbx\_total\_distribution**.

### *Further analysis*

The output files can be further processed by own AWK or Perl scripts.

## 7. Installation

HBexplore was developed originally on an SGI Indigo2 under IRIX 5.3/6.1 and tested in addition on SUN SPARCstation under Solaris 5.3/5.5. Programming was done using the GNU gcc compiler version 2.7. Recent testing was done on SGI Octane under IRIX 6.5 and on SUN Ultra under Solaris 8, using gcc version 3 and also the vendor supplied cc compilers.

The program should be portable to any other computer with a standard ANSI/ISO C compiler.

The program is distributed as a compressed file "hbx-2.01.tar.gz" or similar. To uncompress and untar the file, type

```
unix> gunzip -c hbx-2.01.tar.gz | tar xvf -
```

This creates in the current directory a new directory "hbx-2.01" and should extract the following files into this new directory:

```
DNA_connectivity_acceptor
DNA_connectivity_donor
Makefile
PROTEIN_LIGAND_connectivity_acceptor
PROTEIN_LIGAND_connectivity_donor
README
RNA_connectivity_acceptor
RNA_connectivity_donor
hb-analyse.c
hb-analyse-choose.c
hb-analyse-clust.c
hb_count.c
hb-declaration.h
hb-matrix-statistic.c
hb-three-bond-atom.c
hb-two-bond.c
hbx201man.doc
hbx201man.pdf
input-template
make-input.awk
```

To install HBExplore, change directory to "hbx-2.01". If the GNU gcc compiler is not available on your computer, you have to edit "Makefile" to choose, e.g., the vendor supplied cc compiler and flags according to the comments given there.

Then type

```
unix> make install
```

This command creates the new directories "hbexplore" and "hbexplore/connect" in the current directory and, after compilation, copies the executable "HBX" to the directory "hbexplore". The connectivity files will be copied to the directory "hbexplore/connect".

If you want to use other directories than the default directories, you have to edit the Makefile before installing HBExplore. The variable BINDIR defines the directory where the executable file HBX will be installed. The variable PATH\_CONNECT defines the path where HBX will look for the connectivity files (\*\_connectivity\_donor, \*\_connectivity\_acceptor), the variable DIR\_CONNECT defines the path, where these will be installed.

To remove the object files type

```
unix> make clean
```

Run HBExplore from within directory "hbexplore" by typing

```
unix> HBX
```

Note that the program has the option to store the output files in separate directories. Make sure, that these directories were created before starting the program.

HBExplore provides the option to use alternative user specific connectivity files. To use this option, the environment variable PATH\_CONNECT has to be set before starting the program.

E.g., when your shell is like a C-shell, type

```
unix> setenv PATH_CONNECT path-to-connectivity-files
```

If the environment variable PATH\_CONNECT is set, HBX will look for the connectivity files in that directory. If the environment variable PATH\_CONNECT is not set, HBX will look for the connectivity files in the directory defined in the Makefile (default: PATH\_CONNECT = connect), when HBX was compiled.

HBX uses quite some amount of virtual memory (about 300 MB). You may compile a version, that uses much less memory, but may crash on some large pdb files. In the directory "hbx-2.01", edit the file "hb-declaration.h" according to the comments at the beginning, and then type "make install" as described earlier.

## 8. Known problems

Missing atom coordinates in the PDB file

```
pdblser.ent (V. Biou et al 1994)
ATOM 6867 C4* A T 26 83.953 45.262 16.086 0.50 86.52 1SER7196
ATOM 6868 O4* A T 26 83.504 46.075 17.182 0.50 92.25 1SER7197
Missing atom C3*
Missing atom O3*
Missing atom C2*
Missing atom O2*
ATOM 6869 C1* A T 26 83.153 45.258 18.324 0.50 92.31 1SER7198
```

*The neighbour atom C1\* of O4\* will not be found by HBexplore if atom coordinates are missing.*

Different conformations of one residue

```
pdb9rnt.ent (J. Martinez-Oyanedel et al. 1991)
ATOM 174 CB ALYS 25 20.271 19.145 19.965 0.50 10.29 9RNT 284
ATOM 175 CB BLYS 25 20.232 19.112 19.962 0.50 11.36 9RNT 285
ATOM 176 CG ALYS 25 21.501 19.605 20.757 0.50 15.59 9RNT 286
ATOM 177 CG BLYS 25 21.000 19.930 21.015 0.50 19.87 9RNT 287
ATOM 178 CD ALYS 25 21.475 21.016 21.319 0.50 20.15 9RNT 288
ATOM 179 CD BLYS 25 22.066 19.120 21.731 0.50 20.29 9RNT 289
ATOM 180 CE ALYS 25 20.461 21.583 22.262 0.50 33.32 9RNT 290
ATOM 181 CE BLYS 25 22.961 19.926 22.685 0.50 19.71 9RNT 291
ATOM 182 NZ ALYS 25 20.304 23.079 22.185 0.50 25.98 9RNT 292
ATOM 183 NZ BLYS 25 23.071 21.380 22.382 0.50 24.82 9RNT 293
```

*HBexplore cannot treat residues with different conformations given in the PDB file. HBexplore will recognize A-LYS as a residue name „ALYS“ and not as the A conformation of LYS. To overcome the problem one can prepare two different PDB files.*

The residue number has to be an integer

```
pdblmmme.ent (W.G.Scott et al. 1995)
ATOM 576 P G B 31L 10.518 -10.090 45.865 1.00 44.58 1MME 687
ATOM 577 O1P G B 31L 9.553 -9.162 46.510 1.00 45.95 1MME 688
ATOM 578 O2P G B 31L 10.428 -11.553 46.132 1.00 44.22 1MME 689
ATOM 579 O5* G B 31L 12.020 -9.554 46.079 1.00 38.68 1MME 690
ATOM 580 C5* G B 31L 12.861 -10.062 47.130 1.00 37.30 1MME 691
ATOM 581 C4* G B 31L 13.467 -11.416 46.741 1.00 39.98 1MME 692
ATOM 582 O4* G B 31L 14.212 -11.306 45.495 1.00 34.24 1MME 693
ATOM 583 C3* G B 31L 14.446 -12.029 47.743 1.00 42.50 1MME 694
ATOM 584 O3* G B 31L 13.767 -12.892 48.675 1.00 46.60 1MME 695
ATOM 585 C2* G B 31L 15.351 -12.865 46.845 1.00 37.74 1MME 696
ATOM 586 O2* G B 31L 14.709 -14.082 46.508 1.00 39.58 1MME 697
ATOM 587 C1* G B 31L 15.446 -11.985 45.599 1.00 31.75 1MME 698
ATOM 588 N9 G B 31L 16.460 -10.943 45.637 1.00 27.54 1MME 699
ATOM 589 C8 G B 31L 16.185 -9.601 45.691 1.00 34.44 1MME 700
ATOM 590 N7 G B 31L 17.256 -8.853 45.749 1.00 35.17 1MME 701
ATOM 591 C5 G B 31L 18.306 -9.755 45.726 1.00 22.08 1MME 702
ATOM 592 C6 G B 31L 19.704 -9.521 45.789 1.00 19.93 1MME 703
ATOM 593 O6 G B 31L 20.307 -8.449 45.903 1.00 28.20 1MME 704
```

ATOM	594	N1	G B	31L	20.417	-10.693	45.722	1.00	19.92	1MME	705
ATOM	595	C2	G B	31L	19.861	-11.940	45.625	1.00	30.07	1MME	706
ATOM	596	N2	G B	31L	20.756	-12.950	45.581	1.00	35.35	1MME	707
ATOM	597	N3	G B	31L	18.542	-12.186	45.579	1.00	24.22	1MME	708
ATOM	598	C4	G B	31L	17.834	-11.053	45.638	1.00	22.75	1MME	709

HBExplore cannot identify 31L as a residue number. You can again edit the number manually.

## 9. New features of HBExplore 2.0/2.0.1

In addition to classical hydrogen bonds C-H...O interactions can be analysed with HBExplore. The automatic search modus has been supplemented by a PDB file with hydrogen atom coordinates and a protein secondary structure analysis. The distance range has been extended up to 12 Å to enable the determination of statistical potentials. Furthermore, a reduced number of output files for standard calculations is possible. The interactive analysis has new features for the analysis of complexes, protein chains and secondary structures. Multicentre hydrogen bonds can be listed and the availability of logical operators for complex search cycles has been extended.

Connectivity files have been supplemented by additional residues. The files have been split up for a more flexible search for nucleic acids.

The program version 2.0 was not able to cope with very large structures. Version 2.0.1 can now analyze the files (for example ribosome structures (PDB code: 1ffk)).

## 10. Utility programs

*make-input.awk*

HBExplore can be used interactively for the analysis of a few structures, but for the statistical analysis of a large number of PDB files background calculations are more convenient. The script *make-input.awk* generates a common input file for a series of structures using a single set of search parameters. The command

```
make-input.awk input-template pdb.list > HBX.in
```

writes the input to the file *HBX.in* on the basis of a template file with the search parameters and a file with the names and the path of the PDB files to be analysed. An example for the template file is distributed with the program.

For example, the HBExplore background job

```
HBX < HBX.in >& HBX.log &
```

writes the log-file to the current directory and the individual analysis files for each biopolymer structure are written to the directory specified in the template file.

## 11. Copyright and Availability

See the HBExplore website ([http://www.imb-jena.de/www\\_bioc/hbx/hbx.html](http://www.imb-jena.de/www_bioc/hbx/hbx.html)) for further details.

## 12. Citation

If you use HBExplore please cite:

Lindauer, K., Bendic, C., Sühnel, J.

HBExplore - a new tool for identifying hydrogen bonding patterns in biological macromolecules.

*Comput. Appl. Biosci.* 1996, 12, 281-289.

### 13. Papers that have made use of HBExplore

Brandl, M., Lindauer, K., Meyer, M., Sühnel, J.  
C-H...O and C-H...N interactions in RNA structures.  
*Theor. Chem. Acc.* 1999, 101, 103-113

Brandl, M., Weiss, M. S. Jabs, A., Sühnel, J., Hilgenfeld, R.  
C-H... $\pi$  interactions in proteins.  
*J. Mol. Biol.* 2001, 307, 357-377

Lindauer, K., Bendic, C., Sühnel, J.  
HBExplore - a new tool for identifying hydrogen bonding patterns in biological macromolecules.  
*Comput. Appl. Biosci.* 1996, 12, 281-289.

Lindauer, K., Loerting, T., Liedl, K. R., Kroemer, R.  
Prediction of the structure of human Janus kinase 2 (JAK2) comprising the two carboxy-terminal domains reveals a mechanism for autoregulation.  
*Protein Eng.* 2001, 14, 2001.

Meyer, M., Sühnel, J.  
Quantum-chemical ab initio study on the adenine-difluorotoluene complex - a mimic for the adenine-thymine base pair.  
*J. Biomol. Struct. Dyn.* 1997, 15, 619-624

### 14. References

Bernstein F.C., Koetzle T.F., Williams G., Mayer E.F., Bryce M.D., Rodgers J.R., Kennard O., Simanouchi T. and Tasumi M.  
The Protein Data bank: a computer based archival file for macromolecular structures.  
*J. Mol. Biol.* 1977, 112, 535-542.

Biou V., Yaremchuk A., Tukalo M. and Cusack S.  
The 2.9 Å crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA<sup>Ser</sup>  
*Science* 1994, 263, 1404-1410

Brandl, M., Lindauer, K., Meyer, M., Sühnel, J.  
C-H...O and C-H...N interactions in RNA structures  
*Theor. Chem. Acc.* 1999, 101, 103-113

Brandl, M., Weiss, M.S. Jabs, A., Sühnel, J., Hilgenfeld, R.  
C-H... $\pi$  interactions in proteins  
*J. Mol. Biol.* 2001, 307, 357-377

Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W. and Kollman P.A.  
A second generation force field for the simulation of proteins, nucleic acids and organic molecules.  
*J. Am. Chem. Soc.* 1995, 117, 5179-5197.

Lindauer, K., Bendic, C., Sühnel, J.  
HBExplore - a new tool for identifying hydrogen bonding patterns in biological macromolecules.  
*Comput. Appl. Biosci.* 1996, 12, 281-289.

Martinez-Oyanedel J., Choe H.-W., Heinemann U. and Saenger W.  
Ribonuclease T1 with free recognition and catalytic site: Crystal structure analysis at 1.5 Å resolution.  
*J. Mol. Biol.* 1991, 222, 335-352

McDonald I.K. and Thornton J.M.  
Satisfying hydrogen bonding potential in proteins.  
*J. Mol. Biol.* 1995, 238, 777-793.

Meyer, M., Sühnel, J.

Quantum-chemical ab initio study on the adenine-difluorotoluene complex - a mimic for the adenine-thymine base pair.  
*J. Biomol. Struct. Dyn.* 1997, 15, 619-624

Scott W.G., Finch J. and Klug A.

The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage.  
*Cell* 1995, 81, 991-1002

Shen, L.X. and Tinoco I.J. Jr

The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus.  
*J. Mol. Biol.* 1995, 247, 963-978.

Wahl, M. C., Sundaralingam, M.

C-H...O hydrogen bonding in biology.  
*Trends Biochem Sci.* 1997, 22, 97-102.

Weiss, M.S., Brandl, M., Sühnel, J., Pal, D., Hilgenfeld, R.

More hydrogen bonds for the (structural) biologist.  
*Trends Biochem. Sci.* 2001, 26, 521-523.