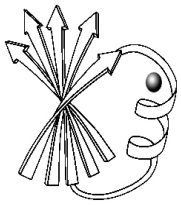
FIRST Installation and User's Guide (Floppy Inclusion and Rigid Substructure Topography, version 4.0)



Protein Structural Analysis and Design Laboratory
Department of Biochemistry
Michigan State University
(517) 353-8745 KuhnL@msu.edu

Table of Contents

1. Introduction

FIRST (Floppy Inclusions and Rigid Substructure Topography) is software developed to predict and analyze protein flexibility.

The following sections explain how to set up and compile FIRST code and run FIRST. In all the commands below, <FIRST> refers to the directory you create in which FIRST software would be installed.

2. FIRST: Algorithm in Brief

FIRST identifies flexible and rigid regions in three-dimensional bond molecular frameworks. The algorithm and the underlying mathematical rigidity theory have been detailed elsewhere [2,5,6,7]. Here, we describe the approach briefly to introduce the terminology used below. FIRST applies the pebble game algorithm [7] to identify and count the bond-rotational degrees of freedom in a directed graph, whose vertices represent protein atoms and whose edges represent covalent and non-covalent (hydrogen-bond and hydrophobic) constraints within the protein [3,6,2]. Flexibility in this network results from dihedral rotations of bonds that are not locked in by other bonds ("hinge joints"). Each bond is assigned by FIRST to be part of either a rigid cluster or a flexible (underconstrained) region. A rigid cluster forms a collection of interlocked bonds in which no relative motion can be achieved without a cost in energy. If a rigid cluster does not contain redundant bond constraints, it is minimally rigid or "isostatic". Conversely, if a rigid cluster contains redundant bond constraints, stress is introduced within this region. Such an overconstrained region is more stable than an isostatic (just rigid) region, in that it remains rigid even if one of the bonds is broken. Underconstrained regions typically are flexible links between rigid clusters. A number of degrees of bond-rotational freedom (so called "floppy modes") are associated with each underconstrained region. Note that this number of internal degrees of freedom within that region is usually much smaller than the actual number of rotatable bonds, because not all the rotatable dihedral angles associated with hinge joints are independent (as they are usually part of a ring of constraints formed by covalent and non-covalent interactions). Finally, FIRST identifies distinct collective motions, each of which consists of coupled rotatable bonds. These motions occur within a particular underconstrained region without affecting internal coordinates outside of this region.

A continuous flexibility index f_i has been defined [2] that characterizes the degree of flexibility of the ith bond in the network (eq. 4). Going beyond the qualitative distinction of regions as overconstrained (stressed), isostatic (just rigid), or underconstrained (flexible), this index allows quantification of *how much more* flexible an underconstrained region is compared to an isostatically rigid region or *how much more* stable an overconstrained region is.

$$f_i = \begin{cases} \dfrac{F_j}{H_j} & \text{in an underconstrained region} \\ 0 & \text{in an isostatically rigid region} \\ -\dfrac{R_k}{C_k} & \text{in an overconstrained region} \end{cases} \qquad (4)$$

Here, F_j and H_j are the number of independent degrees of bond-rotational freedom and the number of potentially rotatable bonds (independent or not) within the jth underconstrained region (i.e. the region containing bond i). R_k and C_k are the number of redundant bonds and the total number of bonds, respectively, in the kth overconstrained region (again, the region containing bond i). Since $F_j \le H_j$, it follows that $0 < f_i \le 1$ for underconstrained regions. Similarly, for overconstrained regions, $R_k \le C_k$, and f_i is bounded by 0 and –1.

In previous studies with FIRST [3,6,2,8], protein structures determined by X-ray crystallography have been used. Here, we perform FIRST analyses for a series of snapshots extracted from MD trajectories. Instead of providing a unique assignment for each atom as part of a rigid or flexible region based on a single input structure, we can now define the probability $P_j(i)$ that atom i belongs to the jth largest cluster of mutually rigid atoms.

$$P_j(i) = \frac{n_j(i)}{N} \qquad (5)$$

$n_j(i)$ is the number of occurrences of atom i as part of the jth rigid cluster, determined over all N snapshots. $P_j(i)$ is expected to provide a more accurate picture of local rigidity compared to the "all-or-nothing" answer given by only one input structure. This is expected to be particularly valuable for substructures that are close to isostatic and, thus, may change from flexible to rigid or *vice versa* upon formation or breaking of one or a few non-covalent bonds as a result of the inherent mobility of proteins.

Along these lines, a flexibility index $\Phi(i)$ for C_α atom i is calculated by averaging over all flexibility indices $f_{i,j}(k)$ (eq. 4) of the two backbone bonds j originating from this atom as well as averaging over all snapshots k.

$$\Phi(i) = \frac{1}{2N} \sum_{j=1}^{2} \sum_{k=1}^{N} f_{i,j}(k) \qquad (6)$$

The sum over j is necessary because the flexibility index is a property of bonds, not of atoms. In most cases, the N-C_α and C_α-C' will belong to the same under- or overconstrained region. If this is not the case, the flexibility index $\Phi(i)$ assigned to the ith C_α is the average of the two backbone bonds that connect to it.

The above is excerpted from H. Gohlke, L. A. Kuhn, and D. A. Case (2004) "Change in Protein Flexibility Upon Complex Formation: Analysis of Ras-Raf Using Molecular Dynamics and a Molecular Framework Approach", *Proteins: Struct. Funct. Bioinformatics,* in press. See also the Methods sections of references [2-4] and the primer on constraint counting at:

http://www.pa.msu.edu/people/rader/data/flexprimer.html

3. Setting up FIRST directories

Before proceeding, please note that the manual adheres to following conventions in print:

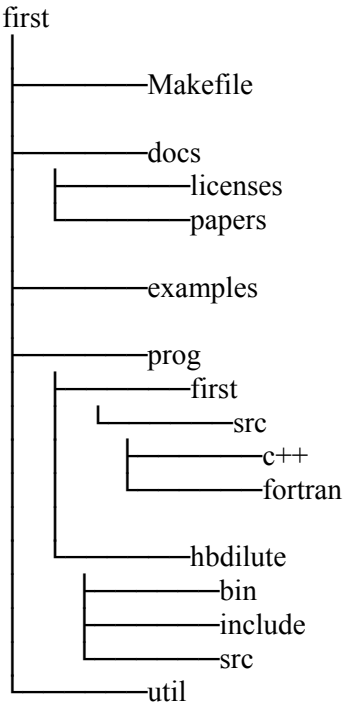
- `Courier font for commands or interactive messages.` `Executable commands are colored blue.`
- [argument] – option argument

- < argument> – mandatory argument
- * - asterisk character represents 0 or more characters in its place in a filename (like regular expression usage on Unix and Windows)

Following steps explain how to set up and execute FIRST:

1. Create a directory with a suitable name (e.g. FIRST) and move the tarred-gzipped file (with extension '.tar.gz') into this directory
2. Enter this directory, now it is the current working directory (CWD).
3. Extract the source code from this file.
   - Un-zip the file using gzip or gunzip, by typing one of the commands below at the command prompt:
     ```
     $ gunzip first.tar.gz
     $ gzip -d first.tar.gz
     ```
     The filename must have '.gz' extension. This will replace the file with '.gz' extension with a new file with same name except the 'gz' extension.
   - Un-tar the resulting file to extract the FIRST file-tree into your current working directory, as shown below:
     ```
     $ tar xvf <filename>
     ```
     where filename is the file created after the unzip operation above.

This should create the following directory structure in your current working directory:



4. Set the environment variable FIRSTPTB: (in csh or tcsh)

   For csh or tcsh:

```
$ setenv FIRSTPTB <CWD/first>/prog
```
(in csh or tcsh)

For other shells, you may have to type:

```
$ FIRSTPTB=<CWD/first>/prog
$ export FIRSTPTB
```

where CWD is the current working directory in which unzip and untar operations were performed.

This environment variable is required for FIRST execution. This can be done by performing the setenv (or FIRSTPTB and export) command in your current window before running FIRST, or by including the setenv (or FIRSTPTB and export) command in your Unix shell initialization file (e.g., in your .cshrc file if you are using c shell), so the command is executed automatically.

Henceforth in the manual, **<FIRST>** will refer to the absolute path of directory '**first'** which is the CWD directory described in the gzip/tar section above

Compilation

FIRST software has been tested to work well on following platforms using indicated compilers

	gcc version	g++ version	FORTRAN
Sparc	2.8.1	2.8.1	f77 WorkShop Compilers 4.2
Solaris/intel	2.95.2 19991024 (release)	2.95.2 19991024 (release)	g77 2.95.2 19991024 (release)
Linux/intel (Redhat 3.2.2-5)	3.2.2 20030222	3.2.2 20030222	g77 3.2.2 20030222
SGI	2.95 19990728	2.8.1	MIPSpro Fortran

To compile the entire code tree, we have provided make change directory to <FIRST> and build the code by typing make

```
$ make
```

'make' is a standard UNIX command that expects a file named 'Makefile' or 'makefile' in the directory from which it is invoked, and compiles the source code according to instructions specified in 'Makefile' (or 'makefile').
A file named 'Makefile' should be present in the <FIRST> directory. The 'make' command reads this file and invokes another 'make' in relevant source code sub-directories to recursively build respective code trees. Individual code trees may also be compiled by typing '**make'** in the directories containing following make files:

- $FIRSTPTB/first/src/c++/Makefile
- $FIRSTPTB/first/src/fortran/Makefile
- $FIRSTPTB/hbdilute/src/Makefile

Following executables, with indicated path, would be created:

- $FIRSTPTB/first/bin/runfirst: application driver program providing the user interface and driving the application,
- $FIRSTPTB/first/bin/first: program implementing the pebble game algorithm [1,2], and

- $FIRSTPTB/hbdilute/bin/hbdilute: program to generate hydrogen-bond dilution plots using the flexibility and rigidity analysis files generated by first. This is internally invoked by first program.

To be able to execute the above programs, from any directory, you must include the absolute path of directories containing these executables in your PATH environment variable. E.g.:

In cshell or tcsh

```
$ setenv PATH $FIRSTPTB/first/bin:$FIRSTPTB/hbdilute/bin:$PATH
```

For other shells, you may have to type

```
$ PATH=$FIRSTPTB/first/bin:$FIRSTPTB/hbdilute/bin:$PATH
$ export FIRSTPTB
```

4. PDB file Pre-processing

FIRST supports the PDB file format for proteins and their complexes (see PDB format information at http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html).

FIRST models the protein as a network of atoms linked by covalent and noncovalent bonds, with each bond assigned a strength based on geometric and chemical properties. The bonds and their strengths affect the rigidity (or flexibility) of a region and hence should be carefully considered. For instance, highly networked water molecules (forming many hydrogen bonds to the protein) will rigidify the protein structure according to FIRST, whereas this effect may not occur in nature because of the rapid exchange of water molecules that are not buried within the protein. Similarly, not including all polar hydrogen atoms or mispositioning them will result in too few hydrogen bonds being identified by FIRST, leading regions being interpreted as more flexible than they actually are. Finally, some crystal and NMR structures have not been subjected to rigorous stereochemical and energetic refinement of atomic positions and have inaccurate bond lengths and angles, in particular. The latter can be problematic, especially when main-chain bond angles are involved, because poor main-chain geometry results in significant loss of structurally important hydrogen bonds in the protein network, and greatly increased apparent flexibility in the network. This is an artifact of an inaccurate initial protein model, rather than a problem in identifying good hydrogen bonds. For this reason, stereochemical validation of structures, and using structures with similarly good stereochemistry and only including buried waters is important, particularly when FIRST results will be compared between protein structures.

Hence, PDB files usually require some preprocessing before they are input to FIRST. A sequence of pre-processing steps is suggested below, followed by more detailed descriptions in section 4.1:

- Remove exposed water molecules (strongly recommended)
- Add polar hydrogen atoms (mandatory)
- Resolve multiple occupancies (optional)
- Remove Anisotropic records (mandatory)
- Verify quality of protein structure (strongly recommended)

4.1. Pre-processing Steps

1. Remove exposed water molecules

Protein crystal structures may have many bound water molecules, which can be buried or exposed. Each water molecule can participate in up to four strong hydrogen bonds, and hence typically affect the flexibility analysis by creating a web of noncovalent interactions that apparently make the neighborhood more rigid. However, protein surface waters often can rapidly exchange with surrounding water molecules, and this currently cannot be modeled in FIRST. Hence, it is recommended to exclude exposed water molecules and include only the buried waters, which usually cannot readily exchange. The PRO_ACT software (http://www.biochem.ucl.ac.uk/~williams/pro_act/summary.html) can classify protein-associated water molecules as surface, cleft, or buried, thus providing an object way to define buried water molecules for inclusion in the PDB structure file for use with FIRST. The file named *_W0T.xray generated by PRO_ACT lists the water molecules that are completely buried. Notes: The file to be processed by PRO_ACT must have the extension .pdb. It was found that PRO_ACT sometimes failed to handle atoms having multiple occupancies.

2. Add polar hydrogen atoms to the PDB file, if needed

Crystal or NMR structure files often do not have polar hydrogen atom positions defined. Absence of hydrogens will significantly affect the flexibility results produced by FIRST, since many of the hydrogen bonds and salt bridges will not be identified by FIRST. Hydrogen atoms placed in energetically favorable positions for hydrogen bonding can be added using tools like WHAT IF. (http://www.cmbi.kun.nl/WHAT IF) Molecular mechanics software packages such as AMBER (http://amber.scripps.edu) can also place the hydrogen atoms in energetically favorable positions. Note: When using WHAT IF, the output file will need corrections – renaming terminal oxygen atoms as OXT, renaming hydrogen atoms to have standard PDB atom names, re-inserting chain identifiers, etc. (Hydrogen atoms added by WHAT IF are given numbers instead of names in the atom name column.) FIRST expects a file with valid PDB records and appropriate hydrogen atom names for input (see format specification at http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html). Note: It is recommended to remove surface waters from the PDB file before running WHAT IF, because the run time of WHAT IF increases greatly when many water molecules are included, due to needing to sample the many possible hydrogen positions for optimizing hydrogen-bond networks involving water molecules.

3. Resolve multiple occupancies

Whenever FIRST encounters atoms with multiple, alternative conformations (identified by multiple ATOM records for a given atom, each with an alternate location indicator given in the record, and a partial occupancy value), it will prompt the user to specify which of atom position should be considered.

This may lead to lot of user interaction if there are many residues with multiple conformations (as may occur in very high resolution structures). To avoid this, user can scan process the PDB file to keep only one of such atom position in the PDB file. (This atom position can be simply the one with the highest occupancy value.)

4. Remove Anisotropic records

Sometimes "ANISOU" records are present, detailing the anisotropic (directional) temperature factors. Include only the ATOM record for each atom, and delete the corresponding ANISOU record. Note that FIRST will only parse PDB records beginning with ATOM or HETATM.

Note: You may want to create new (physical) file in each of the steps above, for keeping track of the pre-processing done, for any future reference or verification.

5. Verify quality of protein structure

Because FIRST is sensitive to bond geometry when it identifies hydrogen bonds, salt bridges, and hydrophobic interactions, we strongly recommend using a tool such as Procheck (http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html) for validating the quality of the protein structure, in particular its main-chain bond lengths and angles (using a phi,psi plot), "goodness" value (specific to Procheck), and noting any close contacts (which could result in hydrogen bonds being disallowed, or too many hydrophobic tethers being created), before running FIRST.

5. Execution

As mentioned before, FIRST software package comes with 3 executable programs, namely: *runfirst, first, hbdilute*.

runfirst is the front-end program which presents menu options to the user, collects bond selection criteria and creates a file called FIRSTdataset file, since its name ends with "FIRSTdataset". FIRSTdataset file holds the bond-network information, along with PDB ATOM and HETATM records. This is implemented using C/C++. *runfirst* also creates some temporary files which serve as input to

first implements the pebble game algorithm (bond network constraint counting) in FORTRAN and uses the FIRSTdataset file (and some other temporary files written by *runfirst* as input to *first*) to analyze the rigid substructures in the bond network. If hydrogen bond dilution is chosen as an option, *first* also creates a file named "decomp_list", which can be separately processed by program *hbdilute* to create hydrogen-bond dilution plots.

Input to *first* are FIRSTdataset and temporary files created by *runfirst*. *first* deletes these temporary files under normal termination. It is meaningless to execute *first* without these files, and hence should not be executed independently; *first* will automatically be invoked by *runfirst*.

hbdilute uses the decomp_list file created by first to generate the hydrogen-bond dilution plots representing how the flexible and rigid regions of the protein change as the hydrogen-bond network is successively diluted, by removing the weakest hydrogen bond (or salt bridge) in the network, then recalculating which regions of the network are rigid or flexible. This mimics thermal denaturation of the protein; hydrophobic regions are believed to increase in strength with moderate increases in temperature, so they remain intact throughout the simulation.

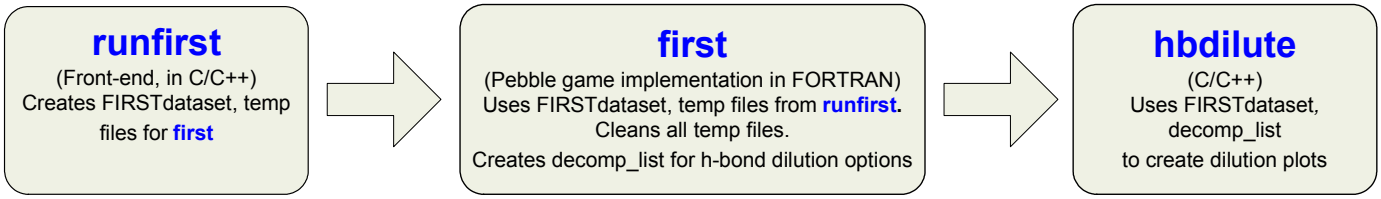


Figure 1. Executable programs and call sequence. *runfirst* invokes *first after creating the FIRSTdataset. first* reads the network information from FIRSTdataset, executes the pebble game. *first* may invoke hbdilute if dilution option was chosen, or perform pruning or output flexibility and rigidity analysis file.

Under abnormal terminations, some temporary files may not get deleted. Their names contain look visibly strange meaningless strings "qXyZaB", "ijkl", "wxyz" or files named fort.n, where n is a number. Please delete these files before proceeding.

5.1. Executing 'runfirst'

Command Line arguments

```
Usage: runfirst [-r] [-e] <-h|{-p|-pw}> <filename>
```

- *filename* This is the file input to FIRST. It can either:
  - a PDB file with explicitly listed hydrogen atoms
  - a FIRSTdataset file , which would have be created in a prior run of FIRST

Option definitions:

- *-non* Use for non-interactive hydrogen bond dilution (explained below)
- *-h* Use when the PDB file has hydrogen atoms explicitly listed. Hydrogens can be added explicitly by programs like WHAT IF. This should be considered the default mode of running first, which will not add hydrogen atoms itself; explicit hydrogen atom positions allow first to check for good hydrogen-bond and salt-bridge geometry. Program would complain if sufficient numbers of hydrogens are found missing.
- *-r* Specify the maximum distance (Å) between the van der Waals surfaces of two carbon/sulfur atoms for a hydrophobic interaction to be modeled between them; changing this value can strongly influence the prediction of rigid/flexible regions, and we recommend using the default value of 0.5 (Angstroms)
- *-p* Use when the original PDB file has been previously processed. FIRSTdataset file from previous run must exist in the same directory.
- *-pw* Same as [-p] except warning messages are summarized
- *-e* Maximum (highest, least favorable) energy a hydrogen bond can have to be included in the analysis. (E.g. -0.1 for including only H-bonds with energy < -0.1 kcal/mol. Default = -0.1 kcal/mol) Thresholds in the neighborhood of –0.6 kcal/mol can be justified on theoretical grounds; at room temperature, this bond energy corresponds to a reasonable probability of the bond being either broken or formed. Note that weaker hydrogen bonds (between –0.1 and 0 kcal/mol) should only be included with caution, because including many very weak hydrogen bonds will overly rigidify the network. This is not so much a concern in hydrogen-bond dilution (which will probe the effect of breaking the hydrogen bonds

according to strength), but it can result in an overly rigid prediction for the protein's native state when a single first run is done on the structure.

Note that incorrect or unsupported command line options will be ignored instead of warned. No more than 13 command line arguments are currently processed, of which the last argument must be the input file.

5.1.1. Interactive Processing

Once started in the interactive mode (i.e. without the –non option), FIRST presents to user various options from which to choose. The flow chart below shows these options and their sub-options, and is followed by explanation of each.

Analysis Options:

1. Flexibility and Rigidity Analysis: Using constraints for covalent bonds, torsions (double bond angular locking), hydrophobic interactions, and salt bridges and hydrogen bonds with energy values more favorable than (less than) the user-specified energy level, this option builds the bond-bending network from the input. The pebble game algorithm calculates the independent bond rotational degrees of freedom from this network. With this information, mutually rigid clusters of atoms, bonds that remain rotatable, and an index of relative flexibility of each bond or residue (based on comparing the number of degrees of bond rotational freedom with the number of bonds) can be determined. Output files allowing visualization of the rigid clusters and flexible regions in 3D, using Rasmol or InsightII.

2. Hydrogen Bond Dilution: The covalent and noncovalent bond network is built as in the Flexibility and Rigidity option. Hydrogen bond dilution then incrementally breaks the hydrogen bonds and salt bridges in the network, from weakest to strongest, and executes the pebble game on the bond network each time a bond is removed. Hydrophobic regions are believed to increase in strength with moderate increases in temperature, so they remain intact throughout the simulation. No hydrophobic tether is removed in dilution. Only changes in the rigid decomposition are output.

Hydrogen bond dilution has the following sub-options

- *Standard Hydrogen Bond Dilution*: Hydrogen bonds are removed in order of energy (from weakest to strongest), thus simulating thermal denaturation. Rigid decompositions after every bond-removal is archived in a hydrogen-bond dilution plot (explained in output files section).

- *Random Dilution over all H-bonds*: Order of hydrogen bond removal is random, where a 9-digit random seed is to be provided by the user. *This is more a research option, allowing the user to probe to what extent the density of the hydrogen-bond network alone accounts for the distribution of rigid/flexible regions, rather than the strength and density of H-bonds (which is probed by the Standard Hydrogen Bond Dilution option).*

- *Flexibility and Rigidity Analysis at a specific point in standard dilution*: This option asks user to enter a hydrogen-bond number from the REMARK:HB section of the FIRSTdataset file. The hydrogen-bond number can be identified in the FIRSTdataset file The hydrogen bonds and salt bridges are numbered/ranked according to their relative strengths. All hydrogen bonds weaker

(having lower absolute energy) than this bond are removed. Flexibility and Rigidity Analysis is then performed on the remaining network.

3. *Hydrogen Bond Stripping*: This option generates quantitative, atom-specific flexibility/rigidity data for the constraint network formed by covalent bonds and bond coordination angles, torsional constraints, hydrogen bonds and salt bridges. Starting with the native structure, this option performs network pruning after each removal of hydrogen bond (from weakest to strongest). Specifically this option prunes the protein down to a skeleton of atoms that are all 2 or higher-fold coordinated on each removal and passes the pruned network to the pebble game for analysis. Pebble game outputs:

    1. the mean (average) coordination over all atoms in the system
    2. Fraction of floppy modes (independent bond-rotational degrees of freedom) assigned to each atom in the network; this value can be fractional because the degrees of freedom are distributed over all bonds/atoms in that particular rigid or flexible region
    3. Fraction of all atoms which are part of the largest rigid cluster
    4. Energy of the hydrogen bond broken

    Note that the so-called pruning of dangling ends in the network (chemical groups that do not form a ring structure by bonding to other groups) is done before distributing and counting the degrees of freedom over the network, to stay consistent with the mathematical theory. Useful relationships can be observed by plotting data from the last three items above with respect to the first. Please refer to [3] for an example of exploring the relationship between the fraction of floppy modes and mean coordination

    Before performing rigidity calculations for the above data, the network pruning to remove atoms that do not contribute to network rigidity can be done in three ways:

    - *No pruning* – all atoms or groups are considered in rigidity calculations, including any singly-coordinated atoms or unconnected side groups
    - *Standard Pruning* - remove all singly-coordinated atoms and unconnected side-groups (which do not connect to the rest of the network via hydrogen bonds or hydrophobic tethers)
    - *Full Pruning* – apart from removing singly-coordinated atoms and unconnected side groups, also remove all dead-end side chains (a dead-end side chain may be connected but not bonded to any other group – in which case it does not contribute to rigidity)

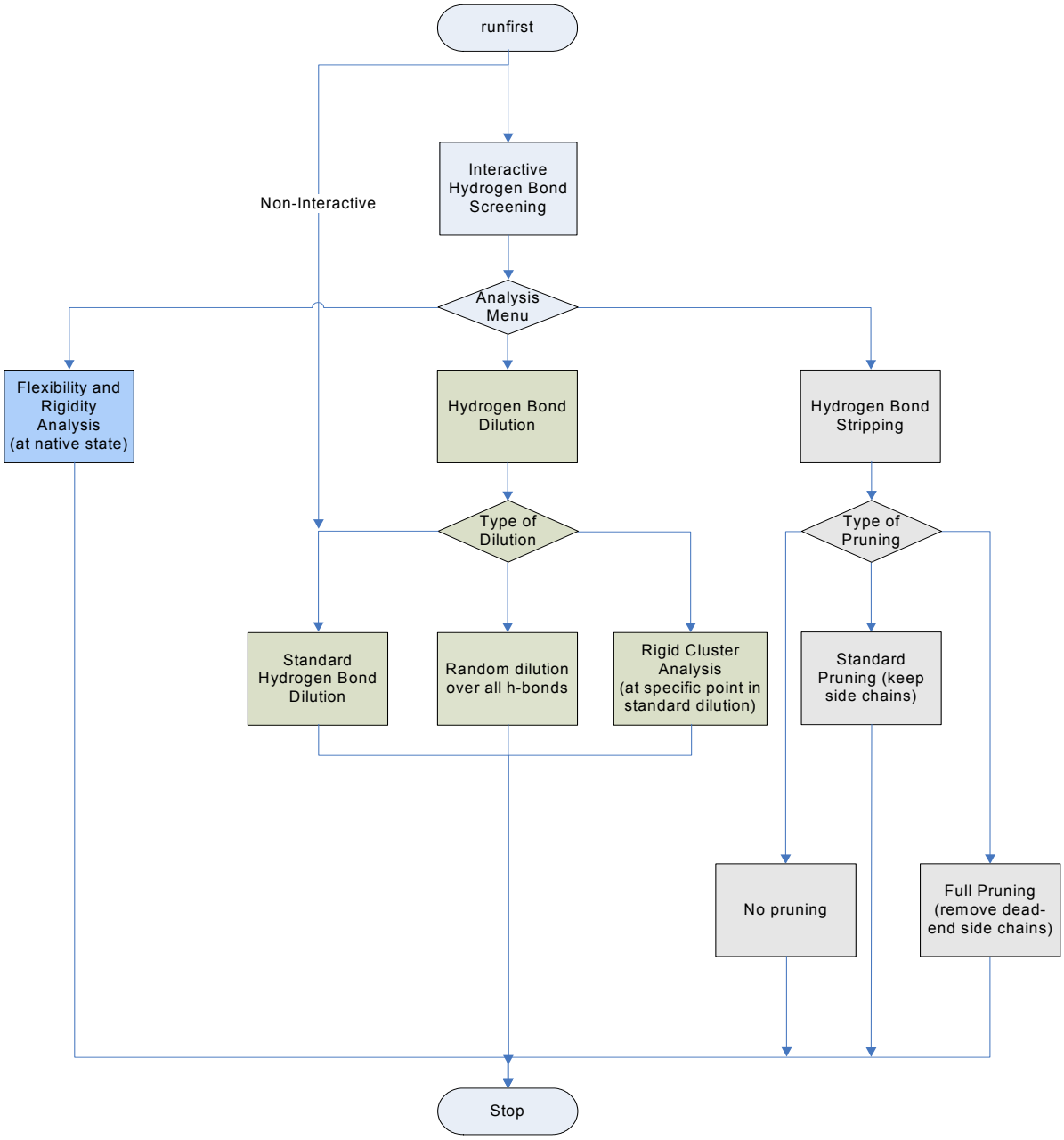


Figure 2. Interactive Control Flow

5.1.2. Interactive Hydrogen Bond Screening

FIRST parses the atom and heteroatom (non-protein atom) records from the PDB file and validates the atomic valencies and geometry of bonds. During this process, it may display caution messages about distances between atoms, or inaccurate valencies it interpreted. In most cases, FIRST will present corresponding queries to the user, responses to which help FIRST correctly determine connectivity between atoms. Most of these queries are explained in the12. Common Warning Messages section.

As mentioned before, the pebble game algorithm within FIRST determines the flexibility of the protein by analyzing the degrees of freedom in the bond network built by modeling bonds and interactions as distance constraints [1], while allowing dihedral rotations. The torsional forces associated with peptide and partial double bonds prevent dihedral rotations and are also represented as angular constraints. After covalent bonds and locked bond angles, salt bridges and hydrogen bonds form the next strongest interactions within the protein network. Salt bridges and hydrogen bonds are assigned energies based on their geometry, using an energy function based on that of Mayo and colleagues (see [4] for more details). Step-by-step analysis of protein flexibility can then be carried out by breaking these bonds in increasing order of energy, thus simulating thermal denaturation.

To define the hydrogen bonds and salt bridges for inclusion in the analysis of rigid and flexible regions, the geometry and energy of these interactions is assessed. The energy assigned is calculated using the geometry of the bonds[1]. Initially, a superset of possible hydrogen bonds is assigned based on meeting the following geometric criteria:

- Donor-acceptor distance $\leq$ 3.6 Å
- Hydrogen-acceptor distance $\leq$ 2.6 Å
- Donor-hydrogen-acceptor angle between 90° and 180°

Salt bridge (ion-pair) interactions are considered as a special case of hydrogen bonds with no angular dependence due to their Coulombic nature.

FIRST provides various options to refine this list of hydrogen bonds and salt bridges to define those to be included in the flexibility analysis. The following screen options appear in a menu in the interactive run. Multiple options can be chosen and the bonds satisfying all chosen options will be included in the analysis.

These screen options appear in a menu in the interactive run. Multiple options can be chosen and the bonds satisfying all chosen options would be included in the analysis.

1.Keep all hydrogen bonds ----> Default assumption
2.Keep mainchain-mainchain hydrogen bonds
3.Keep mainchain-sidechain hydrogen bonds
4.Keep mainchain-HET_NWatm hydrogen bonds (HET_NW means heteroatom, excluding water molecules)
5.Keep mainchain-WATERatom hydrogen bonds
6.Keep sidechain-sidechain hydrogen bonds
7.Keep sidechain-HET_NWatm hydrogen bonds (HET means any heteroatom, including water molecules)
8.Keep sidechain-WATERatom hydrogen bonds
9.Keep HET_atom-HET_NWatom hydrogen bonds
10.Keep WATERatom-WATERatom hydrogen bonds
11.Keep manually selected hydrogen bonds
12.Filter on Donor-Hydrogen-Acceptor angle
13.Filter on Donor-Acceptor distance
14.Filter on Hydrogen-Acceptor distance
15.Filter on Hydrogen bond energy
18.Throw away manually selected hydrogen bonds (discard hydrogen bonds selected through other options in this interactive menu)

19. `Throw away all hydrogen bonds`

These options fall into two categories:

Keep Options: Options 1 – 11 determine which bonds will be included in the FIRST analysis
Filter Options: Options 12 – 15 allow further screening (discarding) of the bonds selected using the "keep" options, based on the criteria of distance, angle, and energy (which will be provided by the user in the next set of queries by the program)

Bonds selected through options 1-11 are subjected to filter criteria, irrespective of the order in which the options are selected.

If any manually selected bonds are to be kept irrespective of the filter options, then you can choose:

17. `Override filter on manually selected hydrogen bonds`

This which becomes available only after choosing one of the filter criteria.

After all the desired selection and filtering options have been supplied, FIRST will include only those hydrogen bonds and salt-bridges which satisfy all the selected criteria.

Notes:
- The distance filters specify the maximum distance; an error will result if the user specifies a value for options 13 or 14 that are greater than the initial, generous criteria of donor-acceptor distance ≤ 3.6 Å and hydrogen-acceptor distance ≤ 2.6 Å
- The angle filters specify the minimum donor-hydrogen-acceptor angle, and an error will result if a value less than the initial, very generous value of 90 degrees is specified by the user
- For typical flexibility analysis, option 15 is the most suitable in our experience
- Most of the threshold values are constants in the configuration files (See section 9. Configurable Parameters)

5.1.3. Interactive handling of Ligand and Metal Ions

Non-water Hetero-atoms
FIRST classifies atoms based on their activity in a bond. For common atoms found in amino acids, FIRST recognizes the atom's chemical nature. However, when processing a crystal structure with for ligands or and metal-ions for the first time, it needs to know from the user, the chemical nature of each hetero-atom. It prompts the user to choose one of the following categories options to classify such hetero-atoms:

- H-bond Donor – if the hetero-atom can donate one or more proton
- H-bond Acceptor – if the hetero-atom can accept one or more proton
- Both H-bond Donor and H-bond Acceptor – if the hetero-atom can both donate or accept one or more proton
- Charged Donor – if the hetero-atom is charged and can donate one or more proton
- Charged Acceptor – if the hetero-atom is charged and can accept one or more proton
- None – if none of the above categories apply

While answering these queries, it would be helpful to refer to the structure of the ligand.

Metal ions

Though FIRST asks the above queries for metal ions, it does not bond the metal ions in hydrogen bonds since metal ions do not have protons to donate. This functionality would be provided in future versions. For now, one work around is to manually introduce bonds between the metal and user-determined neighbors. This can be done by inserting a CF (central force) record in REMARK:CF section of the FIRSTdataset file using the corresponding atom numbers from the FIRSTdataset. E.g.

REMARK:CF 1234 5678

where 5678 can be metal's atom number, and 1234 be the atom number of the bonded atom.

5.2. Non-interactive execution option

So far, a non-interactive option is available only for hydrogen-bond dilution plot generation as this has been found to be the most-used option in FIRST.

Non-interactive execution may be invoked using the **–non** flag in the command line, in combination of other command line arguments.

This option skips prompting user to clarify about valency checks, metal ions' or ligand atoms' acceptor or donor role in the protein complex. Due to this, this option generates reliably accurate hydrogen-bond dilution plots when non-water hetero-atoms are not present or a prior interactive run has already clarified these issues and created the FIRSTdataset file. Hence it is *strongly recommended* to execute interactively at least once if ligands and/or metal ions are present and correctly identify roles of hetero-atoms. FIRSTdataset so produced may be subsequently be used in non-interactive hydrogen bond dilution.

5.3. Executing 'hbdilute'

hbdilute program creates the hydrogen bond dilution plot using the files decomp_list and FIRSTdataset. Order of dilution is derived from decomp_list, which itself is created in a prior dilution analysis run and hence has the order from that prior run.

Command line options

```
Usage: hbdilute decomp_list <output type: b|e|i|s|t >
       <*_FIRSTdataset>

Output types are essentially flags to inform hbdilute about format of
hydrogen bond dilution plot.
```

b –Try as best to fit the entire dilution plot on single page.

e –Print a stripe, representing rigid cluster decomposition for every hydrogen bond removed. Plot may span multiple pages

i – Print only one stripe in every interval of specified energy. 4th (last) argument must be a +ive energy interval (e.g. 0.5). Plot may span multiple pages.
s – Print multiple-page plot. This is the default option.
t – Instead of a stripe, print output the sequence of flexible-region residue-ranges in text. Each such range-record corresponds to a stripe in multiple-page plot (using option 's')

Note that two or more options cannot be combined.

6. Output Files

Following is an explanation of relevant files generated from various run-options:

FIRSTdataset

The first time any FIRST analysis is executed on a protein structure, FIRST renumbers the atoms, and stores the PDB records in the FIRSTdataset along with information about the bond-network. Bond information is stored in following format:

- central-force bonds - Appearing in records start with string "REMARK:CF". These represent distance constraint between nearest-neighbor atom pairs. C*entral force constraints are never diluted during any analysis.*
- torsional-force bonds - records start with string "REMARK:TF". Specifying torsional constraints, this record represent locked dihedral angles between the specified pair of atoms.
- hydrogen bonds - records start with "REMARK:HB" and end with description strings like "HB Dsp2 Asp2" (both h-bond donor and acceptor are sp2 hybridized),
- salt bridge - records start with "REMARK:HB" and end with description string as "SB no energy" – (energy is actually listed in 3rd column from left)
- hydrophobic tether - records start with "REMARK:HB" and end with description string as "PH hydr phob". Tendency of pair of hydrophobic atoms to remain relatively near to each another is modeled as hydrophobic tether. These hydrophobic tethers restrict local motion but allow one degree of freedom. Hydrophobic tethers are modeled by introducing 3 pseudo atoms for each tether between hydrophobic atoms [3]. These pseudo atoms appear as HETATM records with chain Id 'T' and atom name 'X' and residue name as 'XXX'.
- warning records – while creating the bond network, FIRST may display warning messages when some geometric criteria are either not met or deserve special attention. Information about such cases is recorded in the FIRSTdataset file and starts with "REMARK:W". (See 12. Common Warning Messages for more details)
- Pseudo atoms used to model hydrophobic interactions appear as HETATM records

As would be clear from the first line in the "REMARK:HB" section of *FIRSTdatast:

- Column 3 represents the energy of the bond
- Columns 4-6 specify the atom numbers assigned by FIRST to Donor, Hydrogen, Acceptor of the hydrogen bond or the salt bridge.
- Last column holds any remark on the nature of bond

This dataset file is needed by, and is common to, all analysis options.

Once created, it can be used for *any* analysis option (with –p, -pw options at command line). It also stores various warnings that FIRST displays the FIRST time, as they help to remind the structural limitations.
Contents of FIRSTdataset do not change for different energy cutoff chosen for including hydrogen bonds during FIRST analysis and hence can be reused with –p or –pw options at command line.
But the contents do change if the threshold distance between van der Wal's surfaces of hydrophobically interacting atoms changes (-r option at the command line). This makes criterion to determine hydrophobic interactions between atoms tunable from the command line. For more information on hydrophobic interactions modeling, please see the published material[2-4].

Hence multiple datasets may be stored in the same directory under different names. However, you will need to suffix the file's name with "FIRSTdataset" to be used in the current run.

Below is the list of different files created by various run options and their purpose

Flexibility and Rigidity Analysis
_graphic.*, *_Rscript. - Together these files help visualize the protein's flexibility in native state using Rasmol in 3D. **_Rscript.** is the Rasmol script which loads the **_graphic.** and colors the protein regions to indicate rigid clusters (same color) or flexible regions (alternating red and yellow). Whenever two colors join, there is a rotatable dihedral angle. The most flexible parts of the protein are shown as the red and yellow regions.

**_analysis.all* – summary of bond selection criterion and other structural information

_bond_wt.* , *_h-bonds. - these files carry information about hydrogen bonds and may be used as input to ROCK[1]. *_bond_wt.* file may also be used for generating a flexibility index to color the 3-D backbone based on flexibility determined by FIRST and view in Insight[2]. Please contact us if you would like to know how this can be done and have access to Insight.

**_fig_*.pdb, *_fig_*.htm, *_txt_*.htm* - Together, these files, enable viewing the 3-D protein structure in browser using Chime[3] plug-in.

_rdecomp.*, *_sdecomp.* ,*_fdecomp. - These files carry the rigid cluster decomposition information analyzed by FIRST and, if present may be used by FIRST for internal use.

Hydrogen Bond Dilution
*.ps – This postscript file holds the Hydrogen bond dilution plot, which presents a 1-dimensional visualization of protein's flexibility as each hydrogen bond is removed from the protein's hydrogen bond network(For explanation, see 7. Results Visualization and Interpretation).

decomp_list – this is an important intermediate file holding the raw information about rigid-clusters. Each record corresponds to rigid-cluster decomposition after each breakage of a hydrogen-bond in increasing

[1] Another software from PSA Lab, MSU, for generating conformations
[2] 3D molecular modeling from Accelrys. http://www.accelrys.com/insight
[3] Chime is freely downloadable plug-in from MDL. Visit www.mdlchime.com/chime or www.umass.edu/microbio/chime to download correct version for your setup

order of strength. In other words, each record holds the information about which regions are flexible or rigid and of the rigid regions which are mutually rigid.

bond_wts – this is a temporary file and can be ignored. It is kept sometimes for internal usage in some subsequent options.

Hydrogen Bond Stripping

_meancoord.c
This analysis option essentially generates global rigidity data for the constraint network formed by hydrogen bonds and salt bridges. Each of the sub-options creates the respective output file

- No Pruning creates * _meancoord.c1
- Standard Pruning creates * _meancoord.c2
- Full Pruning create files creates * _meancoord.c3

with data in the following format :

<r>	F_Floppy_Modes	F_Atms_Lgst_Clst	HB_Enrgy
2.456235	0.004030560	0.883009851	-0.10311
2.455874	0.004030560	0.883009851	-0.10311
2.455513	0.004030560	0.883009851	-0.10590
2.455153	0.004030560	0.883009851	-0.10699

<r> Mean coordination number, is the average measure for every atom in the network. It approximates the average number of bonds per atom in the network and hence relates to global rigidity of the network

F_Floppy_Modes Fractional number of floppy modes in the (equals F/3N, where F is the floppy modes or the number of independent bond-rotational degrees of freedom remaining in the protein, N is the number of sites in the protein). Typically this fraction increases with decreasing mean coordination

F_Atms_Lgst_Clst is the fraction of the number of atoms which are part of the largest rigid cluster. Typically, as mean-coordination decreases, the large rigid clusters break and this fraction decreases (though, being a ratio, it may increase at times too)

HB_Enrgy is the energy of the hydrogen bond stripped before calculating the above network rigidity variables. Again, starting from the input structure, hydrogen bonds are broken from weakest to strongest.

For more details, refer to Methods section of [3].

7. Results Visualization and Interpretation

This section describes various output files help visualize the flexibility in the protein structure.

A. Hydrogen bond dilution plot (requires postscript viewer)

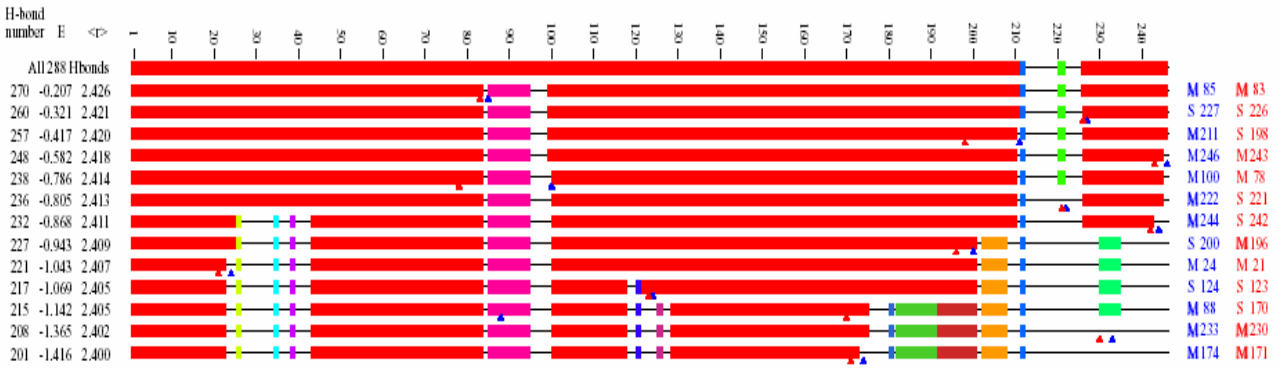


Figure 3. Portion of Hydrogen-bond dilution plot of PDB entry 1ahc.
(See complete plot in file <FIRST>/examples/1ahc/runs/HBDilute/1ahc.ps)

As mentioned before, hydrogen bond dilution plot is generated in
One way to visualize the results of the flexibility analysis is by generating a hydrogen-bond dilution plot (figure above). The numbering along the top, from left to right, represents the amino acid sequence of the protein (residues 1-246). Stripe below is colored to indicate the rigid and flexible parts present in the structure after FIRST included all the hydrogen bonds of energy lower than specified by the user(in this case E = –0.1 kcal/mol, which is also a default value). Hydrogen bond energies are calculated using modified Mayo Potential[4].

Thin black lines represent flexible regions, while thick bars represent rigid regions, with identical colors for mutually rigid regions that belong to the same rigid cluster (which may or may not be contiguous in sequence, and can be across chains, as in the example 2 below for PDB entry 1swa).

The consecutive lines illustrate the changes in flexibility of the protein as hydrogen bonds are removed in order of their increasing energy, a process analogous to thermal denaturation. These energy values should be considered a reasonable ranking of relative energies, rather than an absolute scale.

Moving down the plot, as the energy increased and hydrogen bonds incrementally broke, certain regions became flexible, as represented by black lines intervening between colored blocks. Hydrophobic interactions are kept intact during the thermal dilution, because hydrophobic interactions actually become stronger over moderate increases in temperature. Eventually, systems typically fragment into two or more independent rigid regions (which are internally stable but could move as rigid bodies relative to one another) represented by segments of different colors connected by flexible regions (residues 180 to 200 undergo this change in plot above)

The three numeric columns on the left, from left to right, are:

- remaining number of unbroken hydrogen bonds in structure
- energy of the hydrogen bond just broken (E)
- mean coordination number, <r>, represents the average number of covalent and non-covalent bonds for the atoms in the protein, and provides an overall description of the protein bond network that is useful when comparing rigid to flexible transitions in different proteins[45]

The last two columns specify the residue numbers of the hydrogen donor (blue) and acceptor (red) of the respective hydrogen bond (denoted S for side chain, M for main chain, and W for buried water molecules). The donor and acceptor positions are also shown by carets beneath the bars in the plot.

B. Rasmol script (requires Rasmol or derivatives)

During the normal run, FIRST also creates Rasmol script (file named *Rscript*) file which uses the loads and colors the *_graphic.* file. Graphic file holds the PDB atom records of real atoms as well as pseudo atoms FIRST introduced to model hydrophobic interactions [3] is coloring scheme shows the rigid-cluster decomposition of the protein structure while excluding all the hydrogen-bonds with strengths lower specified energy-cutoff value.

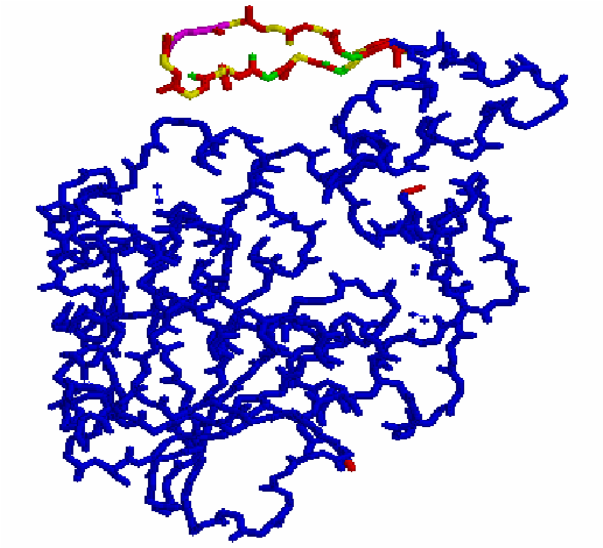


Figure 4. Image snapshot from Rasmol rendering of a Rasmol script generated during Flexibility and Rigidity Analysis

Region with alternating yellow and red colors is flexible region while solid colored region represents a rigid cluster. Also, due to limited 'named' colors available in Rasmol, mutually independent rigid clusters may get colored same. Information about mutually-rigid clusters is readily inferred from hydrogen-bond dilution plots.

C. Flexibility Index (requires InsightII)

Based on degrees of freedom available at a site, pebble game algorithm determines if a site is underconstrained (flexible), overconstrained (rigid) or isostatic (neither rigid, nor flexible). Hydrogen bond dilution plot conveys rigid or flexible regions, but does not convey the degree of flexibility, which can be determined by degrees of freedom information.
<FIRST>/util/flex_index is a program that using the bond information (from *bond_wt* file produced during rigid cluster analysis, along with FIRSTdataset file), creates a PDB file with a measure of degree of flexibility value, called flexibility index, stored in the B-factor column in the PDB file.
Using InsightII and a flexibility index map spectrum (available as file <FIRST>/util/flexindex.spect), this PDB file can be colored to view the degrees of flexibility and rigidity.

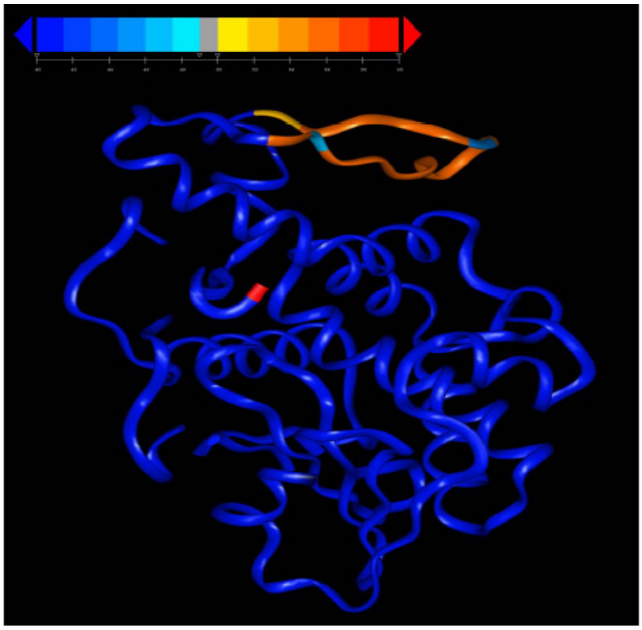


Figure 5. Native state of alpha-momorcharin (1ahc) colored by flexibility index

Figure 6. Flexibility index spectrum

Note: Flexibility-index values in the B-factor columns of PDB files represent degree of flexibility or rigidity transformed onto the scale of flexibility spectrum. Hence the colors represent *relative degree of flexibility or rigidity*, not absolute. So it may be inappropriate to compare degree of flexibility across proteins using flexibility spectrum only.

D. HTML (requires Chime-enabled web browser)

File named *_txt_*.htm produced in a Flexibility and Rigidity Analysis option can be viewed in a Chime-enabled browser. This html file loads the 3D structure and allows coloring by rigid cluster decomposition as well as flexibility index, though coloring by flexibility index is best viewed in InsightII, as explained above.

8. Examples

There are 5 examples provided in the <FIRST>/examples directory.

Following examples would walk you through the process of carrying out FIRST analysis on different proteins.

The table below on the next page shows the structure used to demonstrate specific run options of FIRST.

Example	Protein Structure	PDB Code	Ligand	Metal Ions	Waters	Remarks	FIRST Run Option illustrated
1	ALPHA-MOMORCHARIN	1ahc	No	No	Yes	Monomer	Flexibility and Rigidity Analysis, Non-interactive Hydrogen Bond dilution
2	APO-CORE-STREPTAVIDIN	1swa	No	No	Yes	Tetramer (biological unit)	Standard hydrogen bond dilution
3	ALPHA-MOMORCHARIN	1ahb	Yes	No	Yes	Monomer	Random dilution, Flexibility and Rigidity analysis at a specific point in dilution
4	D-XYLOSE ISOMERASE	1xib	No	Yes	Yes	Monomer	Different dilution plots from single decomposition data, No Pruning
5	D-XYLOSE ISOMERASE	1xid	Yes	Yes	Yes	Monomer	Standard Pruning, Full Pruning

Table 1. Illustrated execution options on specific crystal structures

We are providing five different examples, but primary steps for pre-processing do not vary. For each example, the relevant files after each pre-processing step are stored in the appropriate directory. Files created by FIRST are stored in the directory named after the run-option chosen.
Run these examples following the step–by–step instructions, and compare your results and output files with the ones provided. In case of discrepancies, contact first@sol.bch.msu.edu.

Warning messages appearing during the run are explained in section 12. Common Warning Messages

8.1. Example 1

ALPHA-MOMORCHARIN (1ahc)

This example illustrates following analysis options:

A. **Flexibility and Rigidity Analysis**, which helps visualizing flexibility of the APO protein structure
   - Execution location : <FIRST>/ examples/1ahc/runs/Rigid_Cluster

B. **Non-interactive hydrogen bond dilution,** which helps visualize the flexibility changes as hydrogen bonds are broken in increasing order of strength
   - Execution location : <FIRST>/ examples/1ahc/runs/Non_Interactive_Dilution

Preprocessing

a. *Remove exposed waters*: To determine the buried waters in the protein structure, we have used PRO_ACT (See section 4. PDB file Pre-processing). PRO_ACT outputs a file *_W0T.xray which has the list of buried waters. All the remaining waters should be removed from the structure for realistic flexibility analysis by FIRST. (File with exposed waters removed is 1ahc/waters/1ahc_with_buried_waters.pdb)

b. *Add hydrogen atoms*: We have used WHAT IF [4] for adding hydrogens to protein structures and found it to work well for FIRST analysis's purpose. Output from WHAT IF (file 1ahc/hydrogens/whatif_output.pdb) may need following corrections:
   i. WHAT IF may write atom numbers instead of 'H' for hydrogen atoms it adds. Please rename these numbers as 'H'[5].
   ii. WHAT IF may rename terminal oxygen atoms (OXT) as 'O2' and change their residue number too. Please rename these atoms back to OXT and correct the residue number with the help of original PDB file
   iii. WHAT IF substitutes 'HETATM' keyword by 'ATOM'. Please rename it back to 'HETATM' without disturbing the column alignment
   iv. Place an 'END' as the last line in the file.

   Corrected file is 1ahc/hydrogens/whatif_corrected.pdb. This file is ready for FIRST analysis.

c. *Choose alternate atom locations*: For atoms with alternate locations, choose one. In this case there are no atoms with alternate locations.

d. *Verify quality of protein structure*: As mentioned before, we have used Procheck[6] software to determine the quality of structure (file pdb1ahc.ent) , by analyzing various plots, including Ramachandran plots, generated by Procheck. (Generated files are available in directory first/examples/1ahc/quality)

Careful execution of pre-processing steps is crucial for correct results before carrying executing FIRST on the protein structure.

FIRST Analysis

A. Flexibility and Rigidity Analysis

(This execution was carried out in <FIRST>/examples/1ahc/runs/Rigid_Cluster)

FIRST provides number of analysis options (See section **Error! Reference source not found.**)
This example illustrates the Flexibility and Rigidity Analysis on the **native state** of the protein, executed in the folder <FIRST>/ examples/1ahc/runs/Normal_Run.

The input file is the corrected output from WHAT IF. The command line is :

```
runfirst -h 1ahc.pdb
```

Following WARNING MESSAGES are given:

```
Some heavy atoms were not connected to anything:
   See skip_nonhydrogen in *FIRSTdataset for details.
```

[4] Please visit http://www.cmbi.kun.nl/gv/whatif for more details about the software.
[5] Script util/rename_whatif_hyd may be used for this step.
[6] http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html

See 12. Common Warning Messages to know more about this and other warning messages.

Menu options

- From menu : `H-BOND SCREENING OPTIONS MENU`
  - choose options:
    `15` – to filter on Hydrogen bond energy
    `f` – to finish selection of bonds

    Confirm the selection criteria by typing **`y`**

  - Input the maximum energy of hydrogen-bond (in kcal/mol units). Type:
    `-0.1`
    This would exclude all hydrogen bonds having energy greater than -0.1 kcal/mol. Note that energy cut-off value other than -0.1 may also be specified. The range of the value may be determined either from an already generated hydrogen-bond dilution plot or by looking at the energy values in hydrogen bond records in the FIRSTdataset file.

- From menu : `ANALYSIS MENU`
  choose `1` for `Flexibility and Rigidity Analysis`

  The files available in directory <FIRST>/examples/1ahc/Rigid_Cluster correspond to the default option of saving all the decomposition files

Relevant output files

- *1ahc_Rscript.0001* is the Rasmol script that uses *1ahc_graphic.0001* to display the protein's flexibility in its native state in Rasmol.
- *1ahc_analysis.all* summarizes the geometric criteria used to determine hydrogen bonds and other structural features like degrees of freedom, hinge-joints FIRST found in the structure
- *1ahc_bond_wt.0001* lists the bonds between atoms and the weight allocated to the bond during the pebble-game runs within FIRST. E.g.
  2225 2224 1.00000

The atom numbers correspond to the atom number assigned by FIRST in the 1ahc_FIRSTdataset. This can be used as used to create flexibility index as below

```
<FIRST>/util/flex_index
```

The input can be typed as with the default option

```
1ahc_bond_wt.0001
```

This would result in the file

```
1ahc_flex0001.pdb
```

which can be used to view the protein's flexibility in InsightII using the spectrum file

```
<FIRST>/util/flexindex.spect
```

Please refer to InsightII manual to determine the steps to install a new color spectrum.

B. Non-interactive hydrogen bond dilution

(This execution was carried out in <FIRST>/examples/1ahc/runs/Non_Interactive_Dilution)

To generate hydrogen bond dilution plot non-interactively, type

```
$ runfirst -h 1ahc.pdb -non -e-0.5
```

Simulating hydrogen bond dilution, this command would consider all hydrogen bonds with energy less than or equal to -0.5 kcal/mol. It would create a hydrogen-bond dilution plot, 1ahc.ps with each stripe indicating a significant change in flexibility while h-bonds are broken in increasing order of strength (increasing order of absolute energy)

Output files

- 1ahc.ps – is the hydrogen bond dilution plot.
- decomp_list – holds the rigid cluster decomposition information after each hydrogen bond removal, in the order increasing bond-energy. This file may be further used to generate different hydrogen bond dilution plots using the executable **hbdilute**, in the same order.

8.2. Example 2

APO-CORE-STREPTAVIDIN (1swa)

This example illustrates following analysis options:

A. **Standard Hydrogen Bond Dilution**
   - Execution location : <FIRST>/ examples/1swa/runs/Standard_Dilution

Through this option, FIRST creates hydrogen-bond dilution plot, which helps visualize, in 1-D, flexible, rigid and/or mutually rigid regions of the protein as each hydrogen bond is broken in increasing order of strength.

Preprocessing

Pre-processing steps are the same as followed in the Example 1.
Files resulting from various stages of pre-processing are
(in directory <FIRST>/ examples/1ahc/) :

- Remove exposed waters : waters/1ahc_with_buried_waters.pdb
- Add hydrogen atoms : hydrogens/whatif_corrected.pdb
- Structure quality : Ramachandran plots in directory Procheck

Note that WHAT IF may remove the chain-ids while adding hydrogens, which you would need to put in back before analyzing the protein structure with FIRST.[7]

FRIST Analysis

FIRST was executed from within the directory <FIRST>/examples/1swa/HBDilute.

```
$ runfirst -h 1swa.pdb
```

Note that `1swa.pdb` is the protein-structure file *after* the pre-processing.

In the process, FIRST seeks following clarifications:

```
                    Checking valencies
-------------------------------------------------------------------

Valency of atom 1376 is 4, greater than it's maximum valency (3)
         Atom#  Atom  Res    Res#    Chain
         1376    N    ASN      49       B

Possible incorrect bond(s) with atom(s):
         Atom#  Atom  Res    Res#    Chain
         1380    CA   ASN      49       B
Do you want to keep the bond? (y/n): y
...
...
```

Note that the specified atom number (1376) relates to the number assigned by FIRST as it would appear (or already appears in an existing) in the FIRSTdataset. Despite this renumbering, atom should be identifiable using the residue number (which remains unchanged in the FIRSTdataset).

The cause seems to be the presence of a gap in the crystal structure between residues 45 and 49 in each of the chains B,C and D. Hence the program used to add hydrogen (WHAT IF in this case) has protonated the backbone nitrogen, besides placing two hydrogen atoms. This causes FIRST to inform the user about the incorrect detected valency and allows the user to choose the bonds which user wants to keep.

In such cases, as with the case of terminal nitrogen, keeping all bonds despite warnings has been found not to affect results in a significant way. Still user should choose bonds after having considered the 3-D structure carefully to get accurate results.

Standard Hydrogen Bond Dilution

Menu options

- From menu : `H-BOND SCREENING OPTIONS MENU`

[7] Note that WHAT IF may remove the chain-ids while adding hydrogens, which you would need to put in back before analyzing the protein structure with FIRST.

- choose options:
  15 – to filter on Hydrogen bond energy
  f – to finish selection of bonds
  Confirm the selection criteria by typing **y**
- Input the maximum energy of hydrogen-bond (in kcal/mol units). Type:
  -0.1

♦ From menu : ANALYSIS MENU
Choose **2** for Hydrogen Bond Dilution.
Choose **1** for Standard Hydrogen Bond Dilution.

1. **Relevant output file**
   decomp_list, 1swa.ps – is the hydrogen-bond dilution plot. See section 7. Results Visualization for explanation

8.3. Example 3

ALPHA-MOMORCHARIN (1ahb)

This example illustrates following analysis options:

B. Random Dilution
- Execution location : <FIRST>/ examples/1ahb/runs/Random_Dilution

B. Flexibility and Rigidity Analysis at a particular point in dilution process
- Execution location : <FIRST>/ examples/1ahb/runs/RC_Analysis_at_a_point

Preprocessing

Pre-processing steps are the same as followed in the Example 1.
Files resulting from various stages of pre-processing are
(in directory <FIRST>/ examples/1ahb/) :
- Remove exposed waters : waters/1ahb_with_buried_waters.pdb
- Add hydrogen atoms : hydrogens/whatif_corrected.pdb
- Structure quality : Ramachandran plots in directory Procheck

FIRST Analysis

A. Random Dilution

Execute
runfirst -h 1ahb.pdb

This complex has the ligand Formycin Monophosphate (FMP). Hence FIRST prompts the user to identify the role of each atom of the ligand by asking following queries (responses in **bold red**)

```
Identification of H-bond Donors, Acceptors etc.
```

```
      (d) H-bond Donor
      (a) H-bond Acceptor
      (b) Both H-bond Donor and H-bond Acceptor
      (c) Charged Donor
      (e) Charged Acceptor
      (n) None
-----------------------------------------------------
      Enter (a,b,c,d,e or n) for the following atoms.

       Atom#  Atom  Res    Res#    Chain
       2402    N1   FMP     339          Enter (a,b,c,d,e or n): a
       2404    N3   FMP     339          Enter (a,b,c,d,e or n): a
       2408    N6   FMP     339          Enter (a,b,c,d,e or n): d
       2409    N7   FMP     339          Enter (a,b,c,d,e or n): d
       2410    N8   FMP     339          Enter (a,b,c,d,e or n): a
       2414    O2*  FMP     339          Enter (a,b,c,d,e or n): b
       2416    O3*  FMP     339          Enter (a,b,c,d,e or n): b
       2418    O4*  FMP     339          Enter (a,b,c,d,e or n): a
       2420    O5*  FMP     339          Enter (a,b,c,d,e or n): a
       2421    P    FMP     339          Enter (a,b,c,d,e or n): e
       2422    O1P  FMP     339          Enter (a,b,c,d,e or n): a
       2423    O2P  FMP     339          Enter (a,b,c,d,e or n): a
       2424    O3P  FMP     339          Enter (a,b,c,d,e or n): a
```

These queries are also stored in file <FIRST>/examples/1ahb/lig/ligQueries.txt
Answering these queries would create the FIRSTdataset, which can be re-used for the Flexibility and Rigidity Analysis option.

Menu options

- H-BOND SCREENING OPTIONS MENU
  - choose options:
    15 – to filter on Hydrogen bond energy
    f – to finish selection of bonds
    Confirm the selection criteria by typing **y**
  - Input the maximum energy of hydrogen-bond (in kcal/mol units). Type:
    -0.1
- ANALYSIS MENU
  Choose **2** for Hydrogen Bond Dilution.
  Choose **2** for Random dilution over all bonds
  Enter **928371746** as a random seed

Relevant output files

- 1ahc.ps – is the hydrogen bond dilution plot.
- decomp_list – holds the rigid cluster decomposition information after each hydrogen bond removal, in the *random* order. This file may be further used to generate different hydrogen bond dilution plots using the executable **hbdilute**, in the 'same random' order. See section 7. Results Visualization for explanation

B. Flexibility and Rigidity Analysis at a particular point in dilution process

We will re-use the FIRSTdataset file created in the Random Dilution option above. Copy over the 1ahb_FIRSTdataset and execute

```
runfirst -pw 1ahb_FIRSTdataset
```

Menu options

- `H-BOND SCREENING OPTIONS MENU`
  - choose options:
    - **`15`** – to filter on Hydrogen bond energy
    - **`f`** – to finish selection of bonds
    - Confirm the selection criteria by typing **`y`**
  - Input the maximum energy of hydrogen-bond (in kcal/mol units). Type: **`-0.1`**
- `ANALYSIS MENU`
  Choose **`2`** for Hydrogen Bond Dilution.
  Choose **`3`** for Produce Flexibility and Rigidity Analysis output files at a given point in standard dilution.

- `Enter the H-bond number at which you would like to stop (All h-bonds weaker than this bond would be removed and the RC analysis run on the network with remaining bonds):` **`280`**

This results in all the Flexibility and Rigidity Analysis files for the bond-network consisting of all hydrogen bonds stronger than the user-specified bond, which is.
REMARK:HB 280 -0.53145 616 617 511 HB Dsp3 Asp2 (file 1ahb_FIRSTdataset)

See section 6. Output Files for explanation about the files created.

8.4. Example 4

D-XYLOSE ISOMERASE (1xib)

This example illustrates FIRST processing with *metal ions* present in the structure. Options demonstrated are:

A. Different dilution plots from single decomposition data – usage of program **hbdilute**
- Execution location : <FIRST>/ examples/1xib/runs/hbdilute_plots. Different command line options tell hbdilute to create different dilution plots from the same information. Note that these command line options *cannot* be clubbed together.

B. Hydrogen Bond Stripping with No Pruning – an analysis option
- Execution location : <FIRST>/ examples/1xib/runs/No_Pruning

Preprocessing

Pre-processing steps are the same as followed in the Example 1.
Files resulting from various stages of pre-processing are
(in directory <FIRST>/examples/1xib/) :

- Remove exposed waters : waters/1xib_with_buried_waters.pdb
- Add hydrogen atoms : hydrogens/whatif_corrected.pdb
- Structure quality : Ramachandran plots in directory Procheck

FIRST Analysis

Execute from the directory <FIRST>/1xib/runs/Standard_Dilution

```
runfirst -h 1xib.pdb
```

FIRST seeks following clarifications (responses in **bold red**) :

```
-------------------------------------------------------
        Identification of H-bond Donors, Acceptors etc.

        (d) H-bond Donor
        (a) H-bond Acceptor
        (b) Both H-bond Donor and H-bond Acceptor
        (c) Charged Donor
        (e) Charged Acceptor
        (n) None
-------------------------------------------------------
        Enter (a,b,c,d,e or n) for the following atoms.

         Atom#  Atom  Res    Res#    Chain
         3722   MN    MN      390          Enter (a,b,c,d,e or n): c
         3723   MN    MN      391          Enter (a,b,c,d,e or n): c
```

Some heavy atoms were not connected to anything:

See skip_nonhydrogen in FIRSTdataset for details.

Please see section Common Warning Messages for more details about this error message.

A. Different dilution plots from single decomposition data – usage of program **hbdilute**

To create hydrogen bond dilution plots, **hbdilute,** needs the topology and the rigid cluster decomposition information after each hydrogen bond is removed. This data are available in FIRSTdataset and decomp_list respectively and can be generated by hydrogen bond dilution option. We choose Standard Dilution, but other dilutions can also be chosen.

Menu options

- H-BOND SCREENING OPTIONS MENU
  - choose options:
    15 – to filter on Hydrogen bond energy
    f – to finish selection of bonds
    Confirm the selection criteria by typing **y**
  - Input the maximum energy of hydrogen-bond (in kcal/mol units). Type:
    -0.1
- ANALYSIS MENU
  Choose **2** for Hydrogen Bond Dilution.
  Choose **1** for Standard Dilution

Copy the resulting files (1xib_FIRSTdataset, decomp_list) into each subdirectory of <FIRST>/ examples/1xib/runs/hbdilute_plots.
Execute following command in the following directories for the desired output:

All_Stripes

`hbdilute decomp_list e 1xib_FIRSTdataset`

This option outputs a stripe representing rigid cluster decomposition for each hydrogen bond removal. (Default hydrogen bond dilution process does not print output a flexibility stripe in the dilution plot for every bond dilution – but only when flexibility change occurs.)

Energy_Intrvl

`hbdilute decomp_list i 1xib_FIRSTdataset 0.5`

This option prints the rigid cluster decomposition strip whenever flexibility changes over an interval of 0.5 kcal/mol

Multi_Page

`hbdilute decomp_list s 1xib_FIRSTdataset`

This option prints a stripe whenever flexibility changes, spanning the plot over multiple pages

Single_Page

`hbdilute decomp_list b 1xib_FIRSTdataset`

This option prints a stripe whenever flexibility changes, but makes best effort to squeeze the plot into single page

Text

`hbdilute decomp_list t 1xib_FIRSTdataset`

Apart from creating the multi-page oplot, this option also creates a text file, where corresponding to each stripe, is a record of ranges of flexible residues.

B. Hydrogen Bond Stripping with No Pruning

Menu options

- `H-BOND SCREENING OPTIONS MENU`
  - choose options:
    `15` – to filter on Hydrogen bond energy
    `f` – to finish selection of bonds
    Confirm the selection criteria by typing **`y`**
  - Input the maximum energy of hydrogen-bond (in kcal/mol units). Type: **`-0.1`**
- `ANALYSIS MENU`
  Choose **`3`** for Hydrogen Bond Stripping.
  Choose **`1`** for No Pruning of dangling ends, singly-coordinated atoms of isolated groups before rigidity data calculations

8.5. Example 5

D-XYLOSE ISOMERASE (1xid)

This example illustrates following analysis options:

A. Standard Pruning

- Execution location : <FIRST>/ examples/1xid/runs/Standard_Pruning

B. Full Pruning

- Execution location : <FIRST>/ examples/1xid/runs/Full_Pruning

Preprocessing

Pre-processing steps are the same as followed in the Example 1.
Files resulting from various stages of pre-processing are
(in directory <FIRST>/examples/1xib/) :

- Remove exposed waters : waters/1xidwith_buried_waters.pdb
- Add hydrogen atoms : hydrogens/whatif_corrected.pdb
- Structure quality : Ramachandran plots in directory Procheck

Pre-processing steps are the same as followed in the Example 1.

FIRST Analysis

Menu options

- `H-BOND SCREENING OPTIONS MENU`
  - choose options:
    `15` – to filter on Hydrogen bond energy
    `f` – to finish selection of bonds
    Confirm the selection criteria by typing `y`
  - Input the maximum energy of hydrogen-bond (in kcal/mol units). Type: `-0.1`
- `ANALYSIS MENU`

Choose `3` for Hydrogen Bond Stripping.
Choose `2` for Standard Pruning, which prunes off, singly-coordinated atoms of isolated groups before rigidity data calculations, but not dead-end side chains (directory Standard_Pruning)
Choose `3` for Full Pruning, which prunes off, singly-coordinated atoms of isolated groups including dead-end side chains.(directory Full_Pruning)

Output files

For explanation of output files 1xid_meancoor.*, see section 6. Output Files

9. Configurable Parameters

Various geometric and chemical criteria are read from specified in these files as constants.
These can be modified as desired in following files:

- $FIRSTPTB/first/include/class.h

- $FIRSTPTB/first/src/fortran/set_parameters.
- $FIRSTPTB/first/lib/dist_lookup.lib

These files also provide default values for other important constants used.

$FIRSTPTB/first/include/class.h

Important geometric parameters in this file are:

- *max_energy* : maximum energy a hydrogen bond can have to be included in the hydrogen-bond dilution process. Bonds with energy greater than this value would be kept intact in the dilution process. Default value is set to -0.1 kcal/mol. But it can also be specified at command line of runfirst using '–e' option
- *s_admaxd* : maximum distance allowed in H-bond between acceptor-donor pair, with at least one of the atoms being Sulfur.
- *s_hamaxd* : maximum distance allowed in H-bond between H-acceptor pair with at least one of the acceptor or donor atoms being Sulfur.
- *sb_admaxd* : max distance allowed in salt bridge between acceptor-donor pair
- *sb_hamaxd* : max distance allowed in salt bridge between H-acceptor pair
- *admaxd* : maximum distance allowed in H-bond between acceptor-donor pair.
- *hamaxd* : maximum distance allowed in H-bond between H-acceptor pair.
- *dha_ang* : minimum allowed donor-hydrogen-acceptor angle not involving sulfur
- *s_dha_ang*: minimum allowed donor-hydrogen-acceptor angle involving sulfur
- *sb_dha_ang*: minimum allowed donor-hydrogen-acceptor angle in salt bridge
- *maxdist* : maximum length of a bond between any pair of atoms (default set to 2.5Å). This does not imply that any pair of atoms with lesser inter-atomic distance would be linked through a bond.

$FIRSTPTB/first/src/fortran/set_parameters

Most of the parameters in this file are relevant to pebble game algorithm. Important ones are:

- *izcf* : maximum central force bond coordination number. Default value is 12. This ensures maximum central force constraints on an atom is not more than 12.
- *nb*1 : number of bonds to be placed between two sites acting as rigid bodies

$FIRSTPTB/first/lib/dist_lookup.lib

Apart from hydrogen-bond specific distances, FIRST creates a bond between two adjacent atoms based on minimum and maximum bond lengths. As explained above, maximum bond length is governed by parameter maxdist.

But minimum bond lengths vary depending upon atom-pairs and are stored in a matrix in the distance lookup file $FIRSTPTB/first/lib/dist_lookup.lib. These can be modified depending upon the resolution of the structure.

The code-trees dependent on these files must be recompiled for changes to take affect.

10. Utilities

- <FIRST>/util/flex_index
  Used to color the protein structure according to relative flexibility and visualize in InsightII using spectrum <FIRST>/util/flexindex.spect.

Program needs FIRSTdataset files as well as hydrogen-bond weights file, named, ***bond_wt*** as input. Bond weights file is created in the Flexibility and Rigidity Analysis. **flex_index** must be executed from within the directory containing these files. It produces a modified PDB file, named *flex*.pdb, which has flexibility index values in the B-factor coloumn. These values are transformed on the scale specified in the spectrum file, so comparing flexibility degrees across proteins may not be meaningful.

11. How To's

This section gives quick pointers or explanations to typical capabilities sought by users. We recommend steps in Example 1 should be understood, since few of the How To's refer to example processing.

- How to create a hydrogen-bond dilution plot, simulating thermal denaturation, for a protein, *interactivley*?
  - Pre-process the protein file
  - Execute runfirst, select bond-screening criteria and select option 2 for hydrogen bond dilution. (See example 2)
- How to create a hydrogen-bond dilution plot, simulating thermal denaturation, for a protein, *non-interactivley*?
  - Use **–non** option in the command line and specify cut-off energy below which will be the minimum energy a hydrogen-bond must have to be considered in the network Default is –0.1 kcal/mol (See example 1)

- How to interpret the hydrogen-bond dilution plots?
  - See Results visualization and interpretation section

- How to view the flexibility mapped onto the 3D structure?
  - Execute the Flexibility and Rigidity Analysis either on the native state (Example 1) or at a particular point in hydrogen bond dilution (Example 3)
  - File *_txt_*.htm can be viewed in a Chime-enabled web-browser to view regions of proteins colored according to flexibility index.
  - To view flexibility mapped 3D structure in InsightII, execute *<FIRST>/util/flex_index* program, from within the directory containing FIRSTdataset, ***bond_wt**** files. Color the protein structure using the spectrum file <FIRST>util/flexindex.spect.

- How to approximately determine transition state of the protein from the hydrogen-bond plot
  - [3] determines transition state from the inflection point in the change in the number of independent bond-rotational degrees of freedom (floppy modes) of the protein as its mean atomic coordination decreases. Floppy modes and mean-coordination data is generated in all the sub-options of Hydrogen bond pruning. The mean-coordination at the peak in plot of double derivative of fraction of floppy modes (column two in *meancoord* file) vs. mean coordination (columns 1) indicates transition state mean-coordination.

12. Common Warning Messages

Following warning messages appear when some bonds have found to

1. Low or high distance warning
Message seems like:

```
Caution: The distance of [d = 1.37 Ang.] seems low!
         Atom#  Atom  Res    Res#    Chain
         3110    CG   PHE     106
         3111    CD1  PHE     106
```

Or

```
Caution: The distance of [d = 3.37 Ang.] seems high!
         Atom#  Atom  Res    Res#    Chain
         3110    CG   PHE     106
         3111    CD1  PHE     106
```

Reason: bond-length between pair of atoms was found to be out of range – either smaller than minimum threshold or greater than maximum distance (maxdist).

Resolution: Verify distance between the atoms in structure. Despite these warnings, FIRST will have bonded the atom pair. Verify that a central-force (CF) record exists in the FIRSTdataset, listing the atom number pairs in one single line, e.g.:

```
REMARK:CF       3110 3111
```

In most of the cases it is reasonable to proceed ahead, since distances slightly lesser than threshold minimums do not affect the flexibility analysis much. However, user should be wary of too small distances as that may indicate overlaps or poor structures. Hence, results should be interpreted while keeping such structural characteristics in mind.

2. Some covalent bonds were identified by distance
Message looks like:

```
Some covalent bonds were identified by distance
      criteria: See check_bonding in *FIRSTdataset for details
```

Reason: This warning message usually appears when a covalent bond is formed between heteroatoms whenever the bond-length meets the distance criteria.

Resolution: Please look at the "check_bonding: records in the FIRSTdataset file and verify the bonds as well as their mentioned bond lengths . These bonds would already be created between the corresponding atom-pairs as 'REMARK:CF <atom number1> <atom number 2>' records in the FIRSTdataset file. If you wish to remove certain bond(s), manually delete the REMARK:CF record line(s) from the dataset file as well as the warning record line(s)

3. Incorrect valency warning
The message looks like:

```
Valency of atom 19 is 4, greater than it's maximum valency (3)
         Atom#  Atom  Res    Res#    Chain
           19    N      VAL       3
```

```
Possible incorrect bond(s) with atom(s):
         Atom#  Atom  Res    Res#    Chain
           23    CA   VAL       3
```
…

Reason: Based on chemical and geometric criteria met by the atom in question, FIRST assigned bonds but found resulting valency unusual for the atom.

Resolution: Verify the atom's neighborhood in 3D and choose the bonds to be kept or discarded. FIRST can proceed even if bonds are more than the maximum valency, but user should be aware when to permit this. If user does permit higher valency, FIRST would again show warning message – but would not ask any user-confirmations.

Mostly this warning comes for terminal nitrogen atom, which is protonated by the program adding hydrogen atoms to the crystal structure. This may be fixed in future releases.

4. Dihedral Angle Constraints:

The message looks like:

```
Should the following bond be considered as Non-Rotating?
         Atom#  Atom  Res    Res#    Chain
         3229    C    ARG     110
         3263    OXT  ARG     112
        Type y to LOCK the dihedral angle or any key to keep unlocked:
```

Resolution: FIRST may detect some unknown non rotatable bonds. Please clarify after considering the 3D structure. In the above case, it happens to be terminal oxygen, hence choosing either way may not affect flexibility.

5. Identification of Acceptor or Donor status of an atom

The message looks like:

```
-------------------------------------------------------
        Identification of Donors, Acceptors etc.

        (d) H-bond Donor
        (a) H-bond Acceptor
        (b) Both H-bond donor and H-bond acceptor
        (c) Charged donor
        (e) Charged acceptor
        (n) None
-------------------------------------------------------
        Enter (a,b,c,d,e or n) for the following atoms.

         Atom#  Atom  Res    Res#    Chain
         3260   ZN    ZN      112          Enter (a,b,c,d,e or n):
```

Reason: For non-peptide ligands atoms and metal ions, FIRST cannot determine possible role of the atom as a proton donor or acceptor

Resolution: Please choose the correct category. H-bond donor means proton donor, acceptor means proton acceptor. Metal ions usually classify in one of the charged categories. You should consult the 3D structure to be sure.

6. Some heavy atoms were not connected to anything

Message looks like:

```
Some heavy atoms were not connected to anything
           See skip_nonhydrogen in *FIRSTdataset for details.
```

Reason: This warning may appear for two known reasons:
(1) FIRST detects a hydrogen bond only when a proton can be shared. Since metal ions are not protonated, they go un-connected to any atom.
(2) Second reason could be when there are two or more residues which have same residue numbers but no chain id's to differentiate. In this case, FIRST connects the heavy atoms from the first encountered residue.
E.g. Two same-numbered residues from different chain-ids, where chain-ids got removed by one of the pre-processing softwares. Or, one heteroatom residue's residue (such as water) number clashes with some other residue number of the protein.

Resolution (1) Reason and a work-around is explained in section 5.1.3. Interactive handling of Ligand and Metal Ions
(2) Either add chain-ids before running FIRST on the PDB structure or assign unique residue numbers

7. Some [Expected] covalent bonds were too poor

Message looks like:

```
Some [Expected] covalent bonds were too poor
      to connect: See poor_bond in *FIRSTdataset for details
```

Reason: This happens when bond-length found to be is greater than maximum bond-length. This may happen if FIRST creates covalent bond across a gap in the peptide chain (say across missing residues in the original protein structure) or across chains in proteins structures having multiple chains.

Resolution: FIRST does not connect such atom pairs. User should verify that high bond-length specified in poor_bond records in FIRSTdataset file is really due to this reason and expect the atoms to remain disconnected.

8. Some hydrogen atoms were not connected to anything

Message looks:

```
Some hydrogen atoms were not connected to anything:
See skip_hydrogen in *FIRSTdataset for details.
```

Reason: This may happen if some hydrogen atoms were placed such that they could be bonded to two or more atoms. FIRST skips connecting such hydrogen atoms.

Resolution: Consult the 3D structure. If you want a hydrogen atom to participate in a dangling bond or a hydrogen bond, manually insert appropriate records using atom numbers (from FIRSTdataset) in the FIRSTdataset file and re-process.

9. Some [Expected] heavy atoms were missing from

Message looks:

```
Some [Expected] heavy atoms were missing from
known groups: See missing_atoms in *FIRSTdataset for details.
```

Reason: Some atoms have been found missing from one or more residues

Resolution: Find the missing_atom records from the FIRST dataset to know which atoms were found missing. Check the PDB structure provided to FIRST as well as the original PDB file.

10. Multiple Occupancies

These are information messages and may look like:

```
Multiple conformations were present in the original
PDB file: Default option selected conformations
with maximum occupancy or minimum mobility.
```

Or

```
Multiple conformations were present in the original
PDB file: Each conformation was selected by user
See conformation_user in *FIRSTdataset for details.
```

Reason: This happens when FIRST detects multiple occupancies for one or more atoms. These are likely to appear when user had been previously prompted to either allow FIRST to pick default conformations or choose from presented options.

Resolution: Best resolution is handle multiple occupancies as part of pre-processing itself.

13. Notes & Limitations

Notes

- Modeling of hydrophobic interactions has changed from what has been published in the papers. Hydrophobic interactions are modeled as non-covalent interactions, and only between carbon or sulphur atoms which are bonded only to other carbon, sulphur or hydrogen atoms.

Limitations

- C++ code does not conform to ANSI standard.
- Non-interactive processing is currently supported only for hydrogen bond dilution.
- The default Θ angle cutoff to determine hydrogen bonds and salt bridges is currently 80˚ (in file class.h) and may be too low, especially for hydrogen bonds.
-

Issues

- The hydrogen-bond dilution plots may have erroneous numbering at the top in cases where the PDB file had unusual residue insertions, deletions or errors. Also, while creating multi-page dilution plot for proteins with large number of chains, last chain-id can get misplaced in the plot.
- A structure may two or more residues having same residue numbers as well as same chain-id. This may happen when one residue consists of HETATMs while other doesn't. Or rarely when two or more distinct chains have same chain-id. FIRST distinguishes same numbered residues by chain-ids. If chain-ids are same or missing altogether, FIRST will not connect the residue which was read later from the PDB file. Hence distinct chains should be assigned to residues with same number.
- Hydrogen atom names starting with a number, like 3HZ, may not be correctly interpreted. Also atom names like HG2 may be read in as hydrogen
- When the program finds an h-bond between two sp3 hybridized atoms, it includes an angular term phi in the energy function. This phi angle is the hydrogen-acceptor-base atom angle. If the acceptor is bonded to two neighbors, then there are 2 possible phi angles to measure, one for each base atom. In most of the code, if multiple angles are possible for computing the energy, the angle that results in the lowest (best) energy is used. However, for sp3-sp3 bonds, the angle resulting in the higher (worse) energy is returned. Also, for sp3-sp3 h-bonds in which the acceptor has more than 2 bonded neighbors, the code uses the smallest angle it finds.
- When FIRST finds an sp2 acceptor, it needs to calculate the sp2 "plane" formed by the three sp2 hybridized orbitals. To calculate the plane you need three points. In the code, if the sp2 acceptor has only one neighbor (for example, a carbonyl oxygen), then it uses the acceptor atom, the one neighbor of the acceptor, called the base, and then one atom connected to the base. If there are more than one atom connected to the base, it computes all the planes and returns the best energy. The default assumption here is that if the acceptor has only one neighbor (the base), then the base atom is also sp2 and it's neighbors are in plane with the acceptor. A counter example is the phosphate ion, $PO4^{-3}$. The phosphorus has four sp3 hybridized orbitals, but the oxygen is treated as sp2 in FIRST. There is no plane formed by the acceptor oxygen-phosphorus-second oxygen that corresponds to an equivalent sp2 orbital plane on the acceptor oxygen, because the phosphorus is tetrahedral. For such cases the energy function for these types needs to be changed to reflect the fact that the base atom (P or S) is tetrahedral.

14. References

Following is the list of published related to FIRST. Soft-copies of these are provided in the directory <FIRST>/docs/papers:

1. D. J. Jacobs, L. A. Kuhn and M. F. Thorpe, Flexible and Rigid Regions in Proteins, in "Rigidity Theory and Applications", Ed. By M. F. Thorpe and P.M. Duxbury (Kluwer Academic/Plenum Press, New York, 1999) pages 357384.
2. D. J. Jacobs, A. J. Rader, L. A. Kuhn, and M. F. Thorpe (2001) "Protein Flexibility Predictions using Graph Theory", Proteins 44, 150-165.
3. A. J. Rader, B. M. Hespenheide, L. A. Kuhn, and M. F. Thorpe (2002) "Protein Unfolding: Rigidity Lost", Proceedings of the National Academy of Sciences USA 99, 3540-3545.
4. B. M. Hespenheide, A. J. Rader, M. F. Thorpe, and L. A. Kuhn (2002) "Observing the Evolution of Flexible Regions during Unfolding", J. Molec. Graphics and Modelling 21, 195-207.
5. Tay T-S, Whiteley W. Recent advances in generic rigidity of structures. Struct Topol 1985;9:31-38.

6. Jacobs DJ, Kuhn LA, Thorpe MF. Flexible and rigid regions in proteins. In: Thorpe MF, Duxbury PM, eds. Rigidity Theory and Applications. New York: Kluwer Academic/Plenum Publishers; 1999.
7. Jacobs DJ, Thorpe MF. Generic rigidity percolation: The pebble game. Phys Rev Lett 1995;75:4051-4054.
8. Hespenheide BM, Rader AJ, Thorpe MF, Kuhn LA. Identifying protein folding cores from the evolution of flexible regions during unfolding. J. Molec. Graph. Model. 2002;21:195-207.

15. Support Contact

Please email first@sol.bch.msu.edu for more information or queries.