USER MANUAL
M2 Mining Minima Software for Host-Guest Binding
Version 2.A.2 (beta)

Sarvin Moghaddam
moghaddam@umbi.umd.edu
and
Michael K. Gilson
gilson@umbi.umd.edu

Center for Advanced Research in Biotechnology
University of Maryland Biotechnology Institute
Rockville, MD 20850

Contents

List of Figures

Chapter 1

Introduction

The second-generation version of Mining Minima, M2, provides a systematic method for calculating the binding affinities of host-guest systems by computing the standard chemical potentials of the free host, the free guest, and their complex, and taking the difference to obtain the standard free energy of binding. M2 relies on the predominant states approximation (1; 2) to estimate the configuration integral of a molecule or complex as a sum of contributions from its local energy minima; i.e., its low energy conformations. The potential energies of the host and guest are estimated with an empirical force field. In order to reduce the number of degrees of freedom, and hence of local energy minima, the solvation free energy is estimated with an implicit solvent model. The free energy of each energy well is estimated with the harmonic approximation (HA), adjusted for potential anharmonicities by the Mode Scanning approach (3).

Chapter 2 of this manual summarizes the theory underlying the method. Chapter 3 presents the instructions for installing the components required for running M2. Chapter 4 goes through the steps for computing the binding affinity of a host-guest system with sample files provided along with the software. Chapters 5 and 6 provide greater detail regarding the user interface components of the M2 software. Chapter 8 provides detailed information about compiling the source code, and Chapter 9 provides tips for avoiding pitfalls and getting meaningful results from M2 calculations.

Chapter 2

Overview

The M2 method computes the free energy of host-guest binding as

$$\Delta G^\circ = \mu^o_{\text{complex}} - \mu^o_{\text{host}} - \mu^o_{\text{guest}}, \tag{2.1}$$

where μ^o_{complex}, μ^o_{host}, and μ^o_{guest} are the standard chemical potentials of the complex, host, and guest molecules, respectively. The standard chemical potential of a molecule in solution is approximated as a sum over M local energy minima:

$$\mu^o \approx -RT \ln\left(\frac{8\pi^2}{C^o}\right) - RT \ln \sum_i^M z_i, \tag{2.2}$$

where

$$z_i = \int_i e^{-E(\mathbf{r})/RT} \mathrm{d}\mathbf{r}, \tag{2.3}$$

and R, T, C^o, $E(\mathbf{r})$, and z_i are, respectively, the gas constant, the absolute temperature, the standard concentration, the energy as a function of the internal coordinates $\mathbf{r}$, and the configuration integral over internal coordinates in energy well i. (Factors that will cancel in the final free energy difference have been omitted.) Local energy minima are identified with the Tork search algorithm (4) (carried out by executable `M2_1win.exe` or `M2_1.exe`), and local configuration integrals are computed with the Harmonic Approximation/Mode Scanning (HA/MS) method (3) (`M2_2win.exe` or `M2_2.exe`). Because the Tork search can arrive at the same conformation more than once, duplicate conformations are eliminated with a symmetry-aware algorithm to prevent double-counting (5) (`chkconf1c-win.exe` or `chkconf1c-win.exe`).

For a molecule or complex with N atoms, the HA/MS method works by diagonalizing the Cartesian Hessian matrix of the energy at a local energy minimum

to obtain $3N - 6$ Cartesian modes, each comprising an eigenvalue and an eigenvector. The modes can then be used to approximate the Boltzmann distribution in the energy well as a multidimensional Gaussian with principal axes equal to the eigenvectors and standard deviations along these axes directly related to the corresponding eigenvalues. The free energy of the well is then estimated as sum of $3N - 6$ contributions from the respective modes (3). The software starts by projecting the eigenvectors onto Bond-Angle-Torsion coordinates (3). It then estimates the free energy of the well by integrating the Boltzmann distribution along each mode, limiting the domain of the integral to the lesser of $\pm$ 3 standard deviations of the Gaussian along the mode or that distance which causes any torsion angle to change by more than 60° (6). A number of modes (defined by the user - See Section 5.5) with the lowest force constants are also "scanned" to check and correct for anharmonicity, as follows. For each scanned mode, the scan starts at the local energy minimum, and distorts the molecule along the Bond-Angle-Torsion eigenvector in small steps until the integration limits described above are reached. At each step, the energy and Boltzmann factor are calculated. This scan is done in the positive and negative direction along the mode. The trapezoid rule is then used to estimate the integral of the Boltzmann factor along the mode, and this scanned integral is compared with the harmonic one already obtained (above). If the absolute value of the difference is greater than a cutoff value of 0.5 kcal/mol, then the scanned value is used in place of the harmonic value. The free of the energy well j is thus estimated as a sum of contributions from each mode, where most modes end up being treated harmonically, while a few that prove to be anharmonic, based on their mode scans, are treated by the scanning method. Note that this empirical approach does not account for any anharmonicity that may be associated with distortions along combinations of modes.

The energy $E(\mathbf{r})$ can be decomposed into the sum of the potential energy, $U(\mathbf{r})$, and the solvation energy, $W(\mathbf{r})$, both functions of the conformation $\mathbf{r}$ (7; 8). M2 approximates the potential energy with an empirical force field which, in the present implementation, must follow the same widely used functional form found in current versions of, for example, CHARMM (9; 10; 11; 12) and AMBER (13; 14). During conformational search and HA/MS calculations, a generalized Born model (15; 16) is used for the solvation energy. Solvation energies are subsequently corrected toward the Poisson-Boltzmann/Surface Area model (17; 18), based upon one finite-difference solution of the linearized Poisson-Boltzmann equation and one surface area calculation for each energy minimum i, as previously described (19) (`m2uhbd-win.exe or m2uhbd.exe`, obtainable from Prof. James Briggs, U. of Houston).

The M2 calculations also yield an estimate of the change in the Boltzmann-averaged sum of the potential and the solvation energies on binding, $\Delta \langle U + W \rangle$.

For this purpose, each energy is considered to be purely harmonic in form, so that equipartition applies. Thus, if the energy at the base of energy well i is $U_i + W_i$, then the mean energy in the well is approximated as

$$(2.4) \qquad \langle U + W \rangle_i = U_i + W_i + \frac{3N - 6}{2} RT$$

where the third term is the equipartition contribution, E_{eq}. The mean across all M energy wells then is

$$(2.5) \qquad \langle U + W \rangle = \frac{\sum_i^M z_i (U_i + W_i)}{\sum_i^M z_i} + \frac{3N - 6}{2} RT$$

Applying this expression to the host, the guest, and their complex, yields that

$$(2.6) \qquad \Delta \langle U + W \rangle = \Delta \left[\frac{\sum_i^M z_i (U_i + W_i)}{\sum_i^M z_i} \right] + 3RT$$

where Δ implies the difference between the complex and free species. The $3RT$ correction usually is small in relation to the other quantities and is often neglected.

The estimated change in mean energy provided by Equation 2.6 can be combined with the change in binding free energy to yield the change in configurational entropy (7), which accounts for changes in the mobility of the host and guest on binding, ΔS°_{config}:

$$(2.7) \qquad - T \Delta S^\circ_{config} = \Delta G^\circ - \Delta \langle U + W \rangle.$$

It is important to be aware that the change in configurational entropy includes changes in the rotational, translational, conformational, and vibrational entropy of the host and guest molecules upon binding, but it does not include the change in solvent entropy. The change in mean energy on binding can also be decomposed into Boltzmann-averaged terms plus the equipartition contribution:

$$(2.8) \quad \Delta \langle U + W \rangle = \Delta \langle U_{vdW} \rangle + \Delta \langle U_{Coul} \rangle + \Delta \langle U_{val} \rangle + \Delta \langle W_{elec} \rangle + \Delta \langle W_{np} \rangle + E_{eq}$$

representing, respectively, the changes in mean van der Waals energy, Coulombic energy, valence energy (comprising bond stretches, angle bends, and dihedral rotations), electrostatic solvation free energy from the PB model, nonpolar solvation free energy from the surface area-based approximation, and the equipartion contribution. The change in van der Waals energy, for example, is computed as

$$(2.9) \qquad \Delta \langle U_{vdW} \rangle = \Delta \frac{\sum_i^M z_i (U_{vdW,i})}{\sum_i^M z_i}$$

where $U_{vdW,i}$ is the van der Waals energy at the local energy minimum i.

For a full M2 calculation, we recommend carrying out multiple iterative cycles of Tork search and HA/MS free energy calculations, where each successive cycle starts with the most stable conformations found so far, and continuing this process until the cumulative free energy stabilizes. Sample scripts for carrying out this process under Linux are included with the present distribution; see Chapter 9.

Chapter 3

Installation

This section contains instructions for installing the components required for running Mining Minima on a computer with the Windows XP Professional operating system. The installation is divided into three parts.

3.1 System preparation

You should be logged in with administrator permissions. Before continuing with the current installation, please be sure to uninstall any Python-related software using the ADD/REMOVE PROGRAMS function found under CONTROL PANEL. Mining Minima version 2.A.2 may not perform properly if the required versions of third party software (including Python and Python-related software) are not installed.

3.2 Installation of third party packages

The following third party software packages should be included with this distribution of Mining Minima version 2.A.2. If any or all of these programs are missing, please contact the Gilson Group for replacement copies. The directory THIRD PARTY should contain the following files:

- `blt.4u-for-8.3.exe`
- `bsddb3-3.4.0.win32-py2.2.exe`
- `glut32.dll`
- `Numeric-22.0.win32-py2.2.exe`
- `OpenGL_2.0.1.03_install.rar`

- `plink.exe`
- `Pmw.1.1.tar.gz`
- `PmwBlt.py`
- `psftp.exe`
- `Python-2.2.2.exe`
- `win32all-148.exe`

In order to extract some of the third party software extensions, you will need extraction utilities that can extract `.zip` files as well as `.tar` and `.tar.gz` files. Such software can be found on-line (http://www.winzip.com and http://www.rarlab.com/download.htm, for example).

1. Double click on the `Python-2.2.2.exe` icon and follow all accompanying installation instructions to install Python 2.2.2. MAKE SURE TO NOT USE ANY OTHER VERSION OF PYTHON. While you may not choose to install all of the components of Python 2.2.2, you must install the Python interpreter and libraries, the Tcl/Tk components, and the Python utility scripts. Note: the default repository directory (c:\Python22) is acceptable.

2. Install all win32 extensions for Python 2.2.2 by double clicking the `win32all-148.exe` icon and following all accompanying installation instructions.

3. Install the Numeric extensions for Python 2.2.2 by double clicking the `Numeric-22.0.win32-py2.2.exe` icon and following all accompanying installation instructions.

4. Double click the `bsddb3-3.4.0.win32-py2.2.exe` icon and follow all accompanying installation instructions to install the BSD database extension.

5. Using an appropriate extraction utility, open the `Pmw.1.1.tar.gz` file. Extract the files into the site-packages repository located in the newly built Python 2.2.2 path (for example: c:\Python22\Lib\site-packages).

6. Using an appropriate extraction utility, open the OpenGL_2.0.1.03_install.rar file. Extract the files into the site-packages repository located in the Python 2.2.2 path (see Step 8).
   Now rename the directory called
   c:\Python22\Lib\site-packages\OpenGL_2.0.1.03_install to
   c:\Python22\Lib\site-packages\OpenGL

7. Copy the file `glut32.dll` to the c:\WINDOWS\SYSTEM32 directory.

8. Copy the file `PmwBlt.py` to the c:\Python22\Lib\site-packages\Pmw\Pmw_1_1\lib directory. You must overwrite the copy currently in that directory.

9. Double click on the `blt2.4u-for-8.3.exe` icon to install the BLT packages. The installation directory does not matter; the default directory (c:\Program Files\tcl) is acceptable. After installing the BLT packages, locate the bin subdirectory of the BLT installation directory. Copy the file `BLT24.dll` from the bin directory into the directory c:\Python22\DLLs.

10. Copy the file `psftp.exe` into the c:\Python22 directory.

11. Copy the file `plink.exe` into the c:\Python22 directory.

12. Finally, the newly created Python22 directory must be added to the PATH environment variable. To accomplish this,

    - Click START:SETTINGS:CONTROL PANEL.
    - Double Click SYSTEM.
    - Select the ADVANCED tab.
    - Select the ENVIRONMENT VARIABLES button.
    - Highlight the PATH variable from the list of SYSTEM VARIABLES (found near the bottom).
    - Click the EDIT button.
    - Scroll to the end of the line and add ";c:\Python22" (without the quotation marks).
    - Click OKAY three times to close the SYSTEM window.
    - Close the CONTROL PANEL window.

At this point, the installation of required third party software is complete.

3.3   Installation of M2 codes

This final section provides instructions for installing the Mining Minima version 2.A.2 software. The following files should be included with this distribution of Mining Minima version 2.A.2. If any or all of these programs are missing, please contact the Gilson Group for replacement copies. The provided directory M2_FILES contains the following files and directories:

- figures (dir)

  + `mmlogo.gif`

  + `title_umda.gif`

- parameters (dir)

  + `atomTypes.dat`
  + `bonds.dat`
  + `angles.dat`
  + `dihedrals.dat`
  + `impropers.dat`
  + `LJones.dat`

- Python Files

  + `freeEnergy.pyc`

  + `getAngParam.pyc`

  + `getAtParam.pyc`

  + `getBondParam.pyc`

  + `getDihedParam.pyc`

  + `getImpropParam.pyc`

  + `getPSFAtType.pyc`

  + `M2.py`

  + `numeric_tools.pyc`

  + `numericTools.pyc`

  + `order.pyc`

  + `ProgressBar.pyc`

  + `psf2top.pyc`

  + `ssh.pyc`

  + `telnet.pyc`

  + `togl_opengl_1.pyc`

  + `toolBox.pyc`

- Executable Files

  + `calrmsd.exe` - Executable to calculate RMSD between conformations.

  + `chkconf1c-win.exe` - Executable to filter duplicate conformations. Note that the maximum number of bonds to a single atom can not exceed 8 bonds.

  + `M2_1win.exe` - Main executable to run energy minimization and Tork conformational search.

+ `M2_2win.exe` - Main executable to run HA/MS free energy calculation.

+ `m2uhbd-win.exe` - Executable to run Poisson-Boltzmann calculation. m2uhbd is a version of the program UHBD (20) with our minor modifications for compatibility with M2. The m2uhbd software is available for free noncommercial use from the University of Houston. To obtain it, please email Dr. James Briggs, jbriggs@uh.edu, stating your interest in obtaining the m2uhbd software.

+ `Vcharge.exe` - The executable to calculate Vcharge conformation-independent charges. This can be obtained without charge for noncommercial use from www.verachem.com and the license file (`vcharge.lic` should be obtained accordingly.

Create a repository directory (folder) to contain all of the files associated with Mining Minima version 2.A.2. An example is "c:\Program Files\M2_2.A.2\". Do not use a directory whose pathname includes spaces; for example, do not use a subfolder of the Windows "Documents and Settings" folder. At this point, the installation of Mining Minima version 2.A.2 is complete and you can now try running Mining Minima version 2.A.2 by double-clicking the M2.py icon found in your Mining Minima repository directory. This should activate the M2 GUI.

Chapter 4

Tutorial: A Sample M2 Calculation

This section details the steps to be followed in order to carry out a one-cycle M2 calculation for a sample host-guest system. (A full calculation should normally include multiple cycles, as noted at the end of Chapter 2.) This involves calculating the chemical potential of the free host, the free guest, and their bound complex, and taking the difference to obtain the standard free energy of binding:

$$(4.1) \qquad G^{\circ}_{bind} = \mu^{\circ}_{\text{B5complex}} - \mu^{\circ}_{\text{CB}} - \mu^{\circ}_{\text{B5}}$$

The example provided with this software consists of cucurbit[7]uril (CB) and 1,4-bis(methylamine)bicyclo[2.2.2]octane (B5). A single cycle of M2 calculation yields a binding free energy of about -22 kcal/mol. All the required input files can be found in the "Sample" folder provided with this software distribution. Section 7 describes the various input and output files resulting from this process.

4.1   Procedure

1. Prepare the molecules. Here, you will read in a `.psf` file generated by the program Quanta (Accelrys, San Diego, CA) for each molecule, assign force field parameters to it, and save the resulting topology (`.top`) file for future use.

   - Run M2 by double-clicking `M2.pyc` in the corresponding folder.
   - From the *Tools* menu, select *Make TOP file*.
   - Choose *PSF Converter*, then browse for and open file `CB.psf`.

- By default, the filename `CB.top` should appear in the box labeled *Input the name of the Topology file to be written.* Click the *Convert* button and after the execution, which should be very quick, click the *Return to M2* button, and you will be back in the main M2 window.

- Repeat procedure for the guest, starting with file `B5.psf`.

2. Energy minimize the starting conformations and write the outputs to `.trj` files. This step is done because M2 needs to start from a local energy minimum.

    - Select *Minimization* under the *Steps* menu option, which is the default starting window of the M2.

    - Click the *Load* button and the *System Loading* window will appear on the screen.

    - Click *Load Molecule.* Leave the pull-down menu on *Single Molecule.* You are prompted to browse for the TOP and CRD files of CB molecule sequentially.

    - Click on the *Parameters* button and the *Minimization Parameters* window will open. The default values for the minimization parameters appear on the top section of the window. These values can be used without any change, or may optionally be changed by the user. The user is also allowed to choose which components of the vacuum energy will be taken into account, where, by default, all the energy components are selected. GB is selected by default as the solvation method, but the user may deselect the GB option by un-checking the corresponding box at the bottom section of the window, in which case no solvation model will be used.

    - If GB is selected, the user may click on the *GB Parameters* button and the *Parameters for GB Solvation* window will open. Again, there are some default values assigned for the GB calculation unless the user prefers to revise them.

    - Click *OK* and you will return to *Minimization Parameters* window.

    - Click *OK* in the *Minimization Parameters* window.

    - Click the *Run* button and the *Save Trajectory* window will open.

    - Enter a name in the *File name* box and click the *Save* button. For simplicity, let's call the file `CB.trj`.

    - Click the *Run* button. The GUI will write input file `CB-em.inp` to disk and will invoke `M2_1win.exe` to carry out the minimization specified in this input file. The calculation will generate file `CB.trj`, containing the energy-minimized conformation, and `CB-em.out`, which is the log of the run. (See Section 7 for details regarding these files.)

- Repeat the same procedure to energy-minimize the guest molecule, starting with `B5.psf` to generate files `B5-em.inp`, `B5-em.out`, and `B5.trj` in your folder.

- For the CB-B5 complex (B5complex), after clicking the *Load* button in the *Minimization* window, use the pull-down to select the *Complex* option under the *Select a system type* and you will be prompted to load the TOP and CRD files of CB (using *Load Host* button) and B5 (using *Load Guest* button), sequentially. You will see that the guest lies outside the host, and we need a starting conformation with the guest at least approximately docked into the host. You can address this in one of two ways. The first way is to generate a suitable starting conformation with a separate program and write it as a PDB-format `.trj` file; this can be loaded now via the *Load Trajectory* option provided in the *System Loading* window. The second way is to use *Manual Docking* method available in the *Tools* menu. (Note, however, the docking tool is far from a polished product.) Either way, once a suitable starting conformation is loaded for the complex, the minimization can proceed. In the present example, we used the *Manual Docking* to generate the conformation in `B5complex_ManualDocking.trj`. This file should now be loaded with the *Load Trajectory* option.

3. Conformational search with the Tork algorithm. Normal-mode analysis in Bond-Angle-Torsion coordinates is used to define natural distortions of the molecule or complex. These are projected onto a key subset of torsional coordinates (drivers) selected by the user. Cycles of distortion and energy-minimization are used to discover new local energy minima.

   - Select the *Tork* option under the *Steps* menu and the *Tork* window will open. For simplicity, here we present instructions for the CB host, but corresponding files for the B5 guest and the complex cases can be find in the Sample folder.

   - If you just energy-minimized a molecule or complex (above), it is already loaded and available for this next step. If not, start by loading your molecule and energy-minimized trajectory by using the *Load* button as above. If you are working on the tutorial and you have the `CB.trj` conformation loaded in you system, then as soon as you select the *Tork* option, you will be prompted with *Load selected bonds* window, and the message *A list of selected bonds exists for this structure. Load them?* and if you click the *OK* button, the list of selected bonds exists for CB structure will be loaded and you will see all the selected bonds highlighted. These are the driver torsions mentioned above. However, if this is the

first time running *Tork* on a new structure, you will need to follow the next step (in parentheses) for selecting the key drivers.

- (Choose the bonds that you believe need to be rotated as drivers by Tork in order to generate a good conformational search. This is done by simply clicking on the bonds; your selections will be highlighted. To unselect all and start over, double-click on the background of the graphics window. Completing this process causes the M2 GUI to write two additional files to disk so that your selections will be saved and available for subsequent calculations. These two files have `-bondlist.out` and `-fingerPrint.out` extensions and here are named `CB-bondlist.out` and `CB-fingerPrint.out`.)

- Click on the *Tork Parameters* button and the *Parameters for Tork* window will open. The default values for the Tork conformational search parameters appear, and may be changed if you wish. GB is selected as the solvation method by default and can be removed by un-checking the corresponding box at the bottom section of the window. Then click *OK*. The *Other Parameters* button provides access to the same set of parameters as in the *Minimization* step (above).

- Click the *Run* button.

- You are prompted to enter the name of an output `.trj` file to store the conformations found by Tork. Here, call it `CB_tork.trj` and click *Save*.

- After the job is finished, you will have additional input and output files in your folder. Here, `CB_tork-tork.inp` is the input file read by `M2_1win.exe` and `CB_tork-tork.out` and `CB_tork.trj` are the log and conformational trajectory files. The conformational trajectory file includes the Cartesian coordinates of all the conformations generated by the Tork search, in PDB format. You will also be able to choose and view the various conformations by using the small left and right arrows under the *Conformation viewer* box, or by typing in the desired conformation numbers into the text box. Note that the conformations are sorted by their potential energy, where the first conformation has the lowest potential energy and so forth.

4. Remove duplicate conformations (filtering).

- Select *Filtering* option under the *Steps* menu option and the *Filtering* window will open.

- The trajectory obtained in the prior *Tork* step should be available at this point, unless you closed and reopened the GUI before this step. In this case, load the Tork trajectory output file by choosing the *Load* button and

choosing the `CB.crd`, `CB.top` files and also the Tork trajectory output file `CB_tork.trj`).

- The default values for the conformation overlap checking and filtering are shown and can be modified in the main *Filtering* window (See Section 5.4).

- Click *Run* and you are prompted to enter a name for the filtered trajectory file, for example, `CB_filter.trj`.

- When the filtering process is finished, the number of conformations will have decreased because duplicates will have been removed. As before, the file `CB_filter-opt.inp` includes the input parameters to the filtering algorithm and file `CB_filter-opt.out` is the output log file. The file `CB_filter.trj` includes the Cartesian coordinates of the filtered conformations in PDB format.

5. Calculate the chemical potentials and overall free energy of the filtered conformations.

- Select *Free Energy* under the *Steps* menu option and the *Free Energy* window will open.

- If you just finished the *Filtering* step (above), the filtered trajectory should already be loaded and available for the free energy calculations. Otherwise, load your structure (`.top` and `.crd`) and your filtered conformations (`.trj`) with the *Load* button, as above.

- The *FE Parameters* show the default parameters for use in free energy calculations. Automatically, the *Number of minima* box should show the number of filtered conformations. The *Number of modes by quadrature* may be adjusted by the user, but it defaults to 10 based on our experience. This means that the 10 modes with the lowest eigenvalues will be scanned and potentially corrected for anharmonicity, while all higher modes will be treated harmonically.

- The *Other Parameters* button provides access to the same set of parameters as in the *Minimization* and *Tork* steps.

- Click *Run* and you will be prompted for the name of an output file (e.g, `CB_fe.out`). After the calculation is finished, you will have `CB_fe-im.inp` which includes the input parameters read by `M2_2win.exe` and the `CB_fe.out` output file includes the free energy of the filtered conformations.

6. Adjust the solvation free energies. This step involves reading in the filtered trajectory file a second time, computing the Poisson-Boltzmann (PB)/surface area nonpolar (NP) solvation free energy for each conformation, and using this to adjust the free energies of the minima computed in the prior step.

- Select the *Free Energy Correction* option under the *Steps* menu option and the *Free Energy Correction* window will open.

- If you just calculated the free energy of the conformations obtained following the *Filtering* step, the structures are loaded and available for this step. If not, start by loading your structures by using the *Load* button as above.

- *PB Parameters* and *NP Parameters* windows show the default parameters for use in PB and NP calculations. *GB Parameters* are also shown, because the correction step requires computing and subtracting off the GB solvation energy before adding the PB and NP terms. It is a good idea to use the same GB parameters in this Free Energy Correction step as in the prior steps.

- Click *Run* and you will be prompted to load the file of free energies obtained in previous step, `CB_fe.out`. The calculation then generates several intermediate output files, `CB_fe-gb.out`, `CB_fe-np.out`, and `CB_fe-pb.out`, which include, respectively the bond, angle, dihedral, improper, vdW, Coulombic, Generalized Born (GB), Non Polar Cavity (NP), and Poisson-Boltzmann (PB) solvation energies of the filtered conformations. You will also see a file whose name ends `-new-final.out` (here `CB_filter-new-final.out`) which has all the energy components of the individual conformations, and ends with a summary of overall average energy, entropy and free energy.

7. Repeat the steps described above for the guest (B5) and B5complex (CB·B5) cases and you will have the corresponding output and trajectory files. The final free energy and Boltzmann-averaged energy components for all the 3 cases (CB, B5, and B5complex) can be found on the last block of `CB_filter-new-final.out`, `B5_filter-new-final.out`, and `B5complex_filter-new-final.out` files. To obtain the binding free energy, changes in the mean potential plus solvation energy ($\Delta \langle U + W \rangle$), and changes in Boltzmann-averaged energy components, you need to subtract the corresponding values for the host and guest cases from the complex values. Given the usual standard concentration of 1 mol/liter, an additional contribution of 7.03 kcal/mol must finally be added to the difference in chemical potential, according to Equation 2.2. The configurational entropy change, ΔS°_{config}, is then calculated using Equation 2.7; thus, this entropy change includes the 7.03 kcal/mol standard state adjustment (21). In the present example, the computed binding free energy should be close to -22 kcal/mol.

Chapter 5

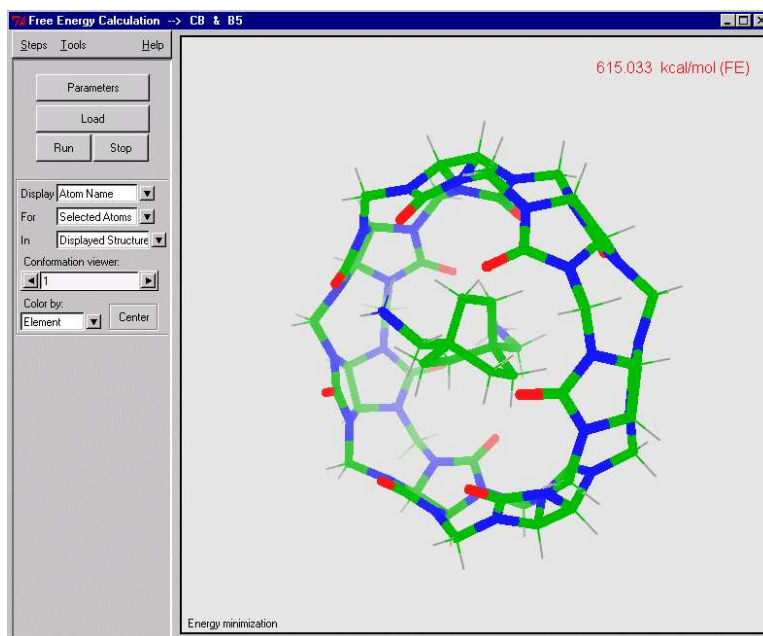# User Interface Components - M2 Steps

Figure 5.1: Main M2 window, with cucurbituril-guest complex.

Figure 5.1 shows the first window that is displayed upon executing the M2 program. The interface has two menu options, *Steps* and *Tools*. The *Steps* option provides the user with the standard series of computational tasks used to complete an M2 calculation:

- Minimization: Starting conformation is subjected to an initial energy mini-

mization.

- Tork: Tork algorithm is used to discover new low-energy conformations.

- Filtering: Duplicate conformations are eliminated.

- Free Energy: Free energies of conformations are calculated.

- Free Energy Correction: Free energies are corrected with a more detailed solvent model.

Typically, a complete analysis consists of successive execution of all processing tasks available under the *Steps* option on a loaded system. The breakdown of the whole process into separate tasks allows the user to skip one or more tasks if desired, and makes it possible for us to provide a set of GUI controls tailored for each task. This chapter details the *Steps* options. The *Tools* menu provides a number of additional capabilities to support the main M2 calculation. These are explained in Chapter 6.

5.1   Shared GUI components

Several controls will always be provided in the left-hand pane of the main M2 window, irrespective of the chosen task. These are the *Load*, *Run*, and *Stop* buttons, along with the molecular graphics options. At any step, a structure to be studied can always be read into the program using the *Load* button, which brings up a new window titled *System Loading* (Figure 5.2). The *Select a system type* menu allows the user to load either a *Single Molecule*, the host or guest; or a *Complex* comprising both host and guest. You will be prompted for topology (TOP) and initial coordinate (CRD) files for each molecule to be loaded. Note that, for a complex, two TOP and two CRD files are needed. These files are loaded using the *Load* button of the *System Loading* dialog. A trajectory file with additional conformations of the molecule or complex can also be loaded with the *Load Trajectory* button of the *System Loading* dialog,

Figure 5.2: Dialog box to load a single molecule or a complex.

as explained below. Note that the trajectory file for a complex includes coordinates of both molecules, even though M2 needs separate `.top` and `.crd` files for them.
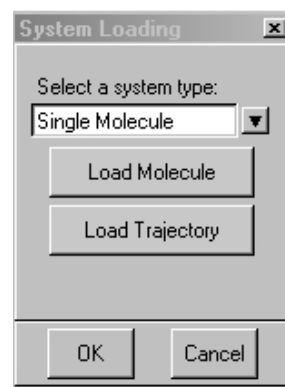
In general, clicking the *Run* button starts the calculation for the chosen Step, and then loads the results into the GUI. Clicking *Run* causes the GUI to gener-

ate an input file (`.inp`) and to invoke the appropriate Fortran executable, such as
`M2_1win.exe`. This in turn reads the `.inp` file and any needed molecular data files
and carries out the calculation. The results are written to files with `.out` and/or
`.trj` extensions, where the `.out` file provides information specific to the type of anal-
ysis selected from the *Steps* menu, and the `.trj` file provides Cartesian coordinates
of any conformation(s) that were generated or selected by the run, in a PDB-style
format. The *Stop* button can be used to terminate the calculation without saving
the results. Note that the `.inp` and molecular data files can also be copied to a
Linux system and processed there by appropriately compiled versions of the chief
M2 Fortran codes, `M2_1.exe` (energy minimization and Tork) and `M2_2.exe` (free
energy calculations).

The lower section of the left pane of the main window provides options for the
molecular graphics.

- Display: Each atom's name, number, type, partial charge, formal charge, and
  vdW radius can be shown on the screen, as can the distance, angle, and
  dihedral between any two, three or four atoms, respectively.

- For: If *Selected Atoms* is displayed, use the left-mouse button to select atoms.
  Alternatively, use the pull-down menu to choose all atoms. Double-click on
  the graphics background to deselect all atoms.

- In: This pull-down allows the display option to apply to the first loaded struc-
  ture, or all the existing structures.

- Conformation viewer: Use the small arrows to step through conformations, or
  enter the index of one or more conformations into the text window to display
  them, or type * in the text window to show all. (See the *Tools* menu for
  methods of computing RMSDs and superposing different conformations.)

- Color By: Color atoms according to the selected rules.

5.2   Energy minimization

The initial window of M2 (Figure 5.1) is by default in *Minimization* mode. Minimization mode can also be selected with the *Steps* menu. Energy minimization is currently carried out in a two-step process. The first step uses the Conjugate Gradient (CG) method and the second uses the truncated Newton (TN) method. The CG method is better at handling the highly distorted conformations generated in the Tork procedure, while TN is better at reaching a true local energy minimum.

The *Parameters* button bring up the *Minimization Parameters* window (Figure 5.3), which allows changes to the default procedural and energy parameters which are defined as:

- TN Minimization Steps: Maximum number of TN iterations to carry out.

- TN Tolerance: Termination criterion based on root mean square (RMS) gradient of the energy.

- TN Frequency of Output: Number of TN steps between updates to the `.out` file.

The parameters set by default here are considered reasonable choices, but they may be changed. One can select the potential energy terms (bond, angle, dihedral, improper, Coulombic, and vdW) to be included. The Generalized-Born (GB) solvation model is selected by default and if it is used, its parameters can be adjusted by clicking on the *GB Parameters* button and using the *GB Parameters* window (Figure 5.4). The GB options are as follows:

- Fixed Dielectric Constant For Solute: The interior dielectric constant of the molecule (default=1).

- Fixed Dielectric Constant For Solvent: The solvent dielectric constant (default 80, for water at 298K).

- Use ideal geometry in radii calculations: Refers to an older implementation of GB in which the bond-lengths used in calculating the GB cavity radii were always kept at their ideal values, based on the force field.

- Output details of GB calculations: Provides greater detail regarding GB calculations in the output screen.

- Atom radii are averaged with probe radius: A method of setting initial atomic cavity radii in which the van der Waals radius of the atom was averaged with the probe radius.
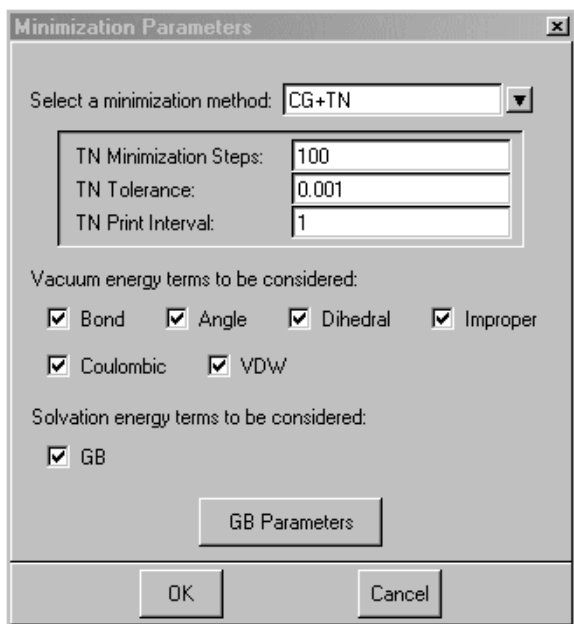
Figure 5.3: Dialog box to set energy minimization parameters.

Note that the energy minimization will process whatever structure is currently loaded. If no structure is loaded, you will need to load one as described above. If only a `.top` and `.crd` file are read in, then the starting conformation will be that in the `.crd` file. However, a different initial conformation can, optionally, be read in from a `.trj` file (above). Note that, although providing an initial trajectory file is optional, if the conformation in the `.crd` file is pathological, minimization may yield a strained and unrealistic conformation.
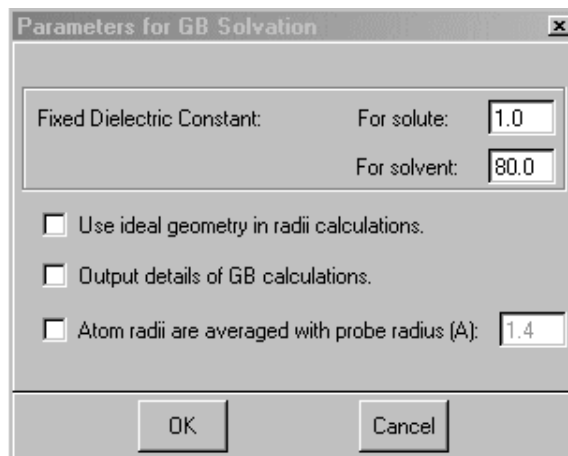

Figure 5.4: Dialog box to set Generalized Born parameters.
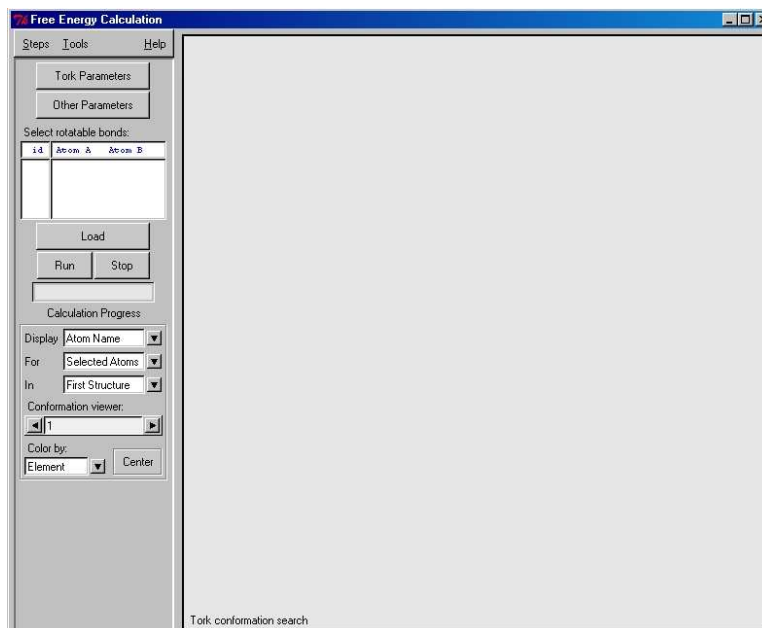
5.3   Tork conformational search

Figure 5.5: User interface for the Tork conformational search step.

Selection of the *Tork* option under the *Steps* menu will change the menu on the left pane of the main window to that in Figure 5.5. By default, the Tork search step uses whatever molecular conformation is already loaded as starting point for its conformational search. However, the user can load other conformations in the form of a trajectory file and the Tork algorithm uses all conformations that are loaded as starting points.

The Tork method involves repeated cycles of conformational distortion and energy minimization, where the distortions focus on normal-mode guided changes in a user-selected set of key torsion angles. These key torsional bonds must be selected by clicking the corresponding bonds in the chemical structure of the system shown in the graphical *Tork conformation search* window, the right-hand pane of the M2 GUI. Each se-
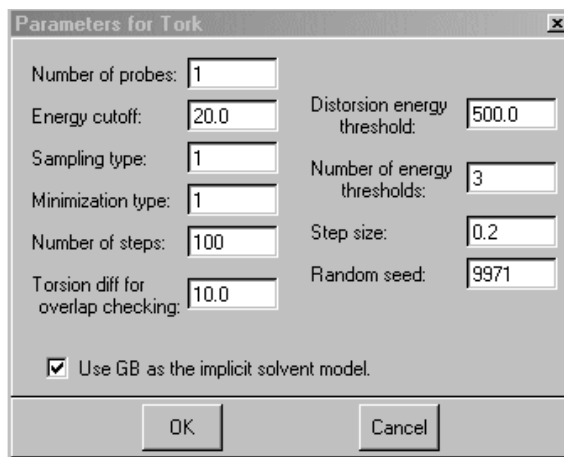

Figure 5.6: Dialog box to set Tork conformational search parameters.

lected bond will be highlighted and
will also be listed in the *Select ro-
tatable bonds* box in the left-hand
pane.

The *Tork Parameters* button brings
up the *Parameters for Tork* window
(Figure 5.6), which provides access to the main parameters of the Tork conforma-
tional search procedure, as follows:

- Energy threshold (kcal/mol): A Tork distortion continues until the energy
  rises above this threshold relative to the starting conformation; then energy
  minimization begins.

- Number of probes: Number of different initial local energy minima from which
  a Tork search will be initiated.

- Number of energy thresholds: If value is greater than 1, then the Tork dis-
  tortion will continue past the first crossing of the energy threshold until the
  desired number of multiples has been passed; a new minimization will be
  started at each crossing.

- Torsion angle for overlap checking: two conformations generated by Tork are
  considered distinct if any of their freely rotating dihedrals differ by more than
  this value. Eliminating repeat conformations in this way during the Tork
  search reduces the amount of redundant searching.

- Number of steps: Maximum number of distortion steps along each distortion
  mode.

- Energy cutoff (kcal/mol): When the rise in potential energy exceeds this value,
  the search will stop.

- Step size: Size of each small distortion step.

- Sampling type: Tork distortions can be carried out along single, pure distortion
  modes along with random combinations of pairs of modes (Type 1); or random
  combinations of pairs of modes only (Type 2). You may want to vary these
  heuristic options in a series of Tork searches to make sure you find the very
  lowest-energy conformations possible.

- Random seed: If the total number of requested distortion directions is large
  then a smaller, tractable number of combined distortions is selected randomly.

This parameter is the random number seed used in selecting the conformations. Changing this seed is a way to generate a new Tork run that may turn up new local energy minima.

- Minimization type: Minimization type, where 1 refers to CG+TN. This option is just a place-holder, as no other types are available in this version of M2.

- Use GB as the implicit solvent model; uncheck to use vacuum.

The *Other Parameters* button provides access to the same energy-minimization and energy parameters offered by the *Parameters* button in the *Minimization* step (above).
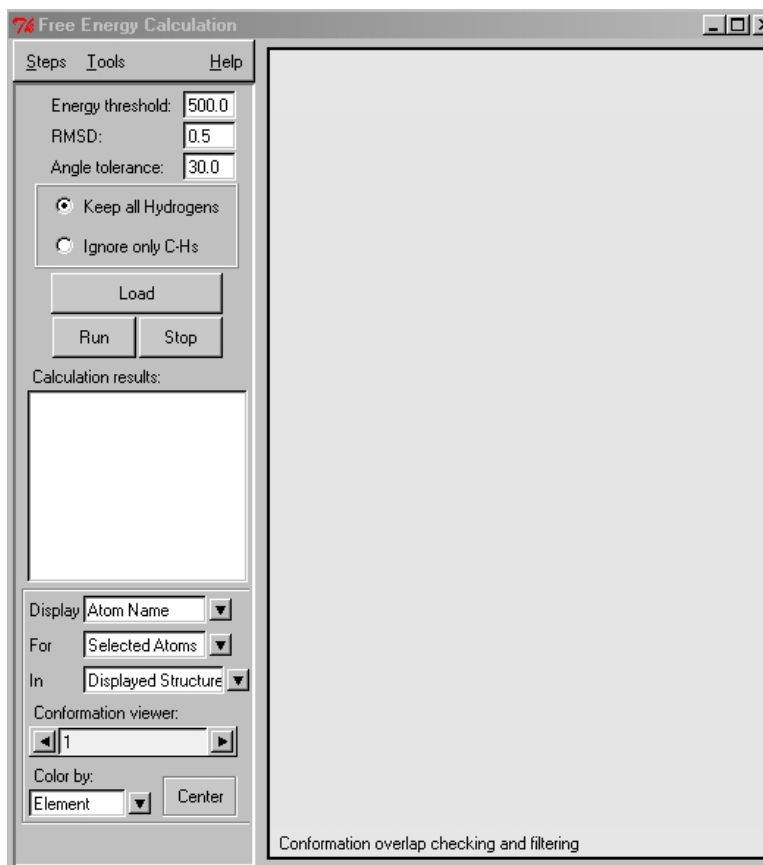
5.4 Conformational filtering

Figure 5.7: User interface for conformational filtering.

The *Filtering* step (Figure 5.7) uses a symmetry-aware algorithm (5) to identify and delete duplicate and high-energy conformations generated by the Tork search, in order to avoid overcounting any conformations in the subsequent free energy calculations. (Note that the current filtering code fails if any atom has more than 8 bonds.) Key parameters for filtering are the *Energy threshold*, *Angle tolerance*, and *RMSD*, whose default values are 100 kcal/mol, 15.0 degrees, and 0.1Å, respectively. The filter deletes all conformations whose energy after Tork search is above the energy threshold relative to the lowest-energy conformation found. The angle tolerance is a criterion needed for the identification of 3-dimensional symmetries (22). For every pair of conformations whose symmetry-corrected RMSD is less than the RMSD limit, the higher-energy conformation will be deleted. By default, only polar hydrogens are included in this symmetry-based conformational filtering, because discarding nonpolar hydrogens saves computer time and does not appear to degrade

the results. Nonetheless, the option is provided to include all hydrogens.

Clicking the *Run* button causes the interface to prompt for the name of an output `.trj` file, which will contain the coordinates of the conformations that survive the filtering step.
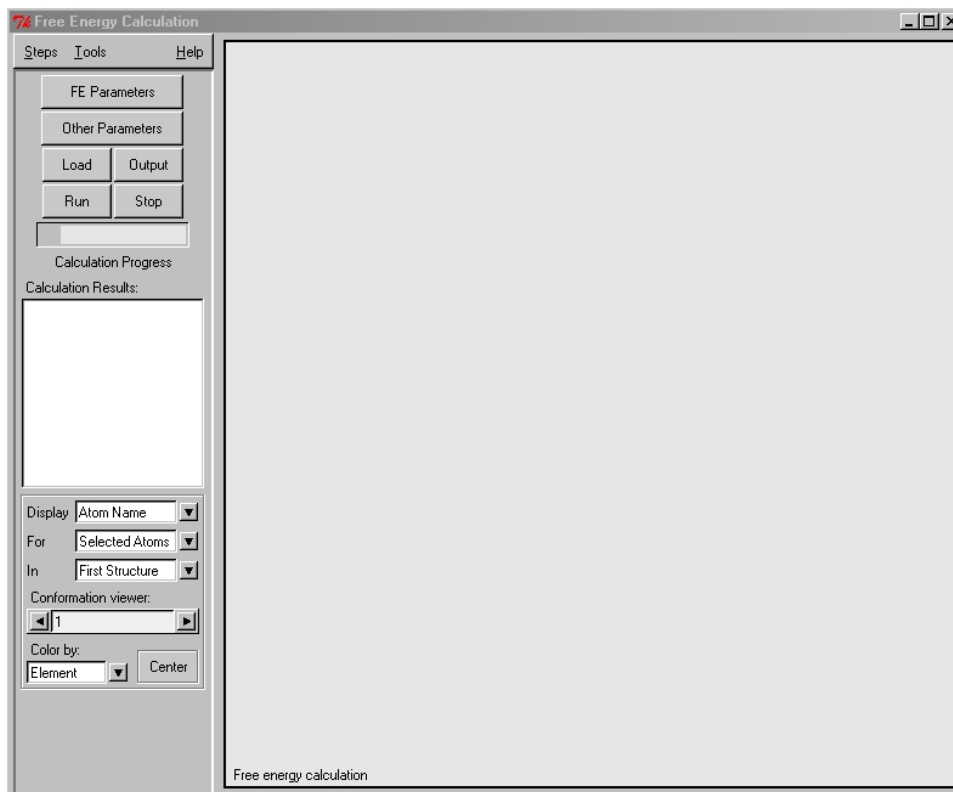
Figure 5.8: User interface for free energy calculation step.

5.5 Free energy calculation

In the *Free Energy* step, each of the filtered set of conformations (above) are processed with the HA/MS method (Chapter 2) to provide an estimate of its free energy (chemical potential). The resulting free energies are also combined to provide the full free energy of the molecule or complex. The calculation will process whatever conformations are currently loaded into the M2 GUI. These might have just been selected in the *Filtering* step (above). Alternatively, the user can load a .trj file with conformations to be processed.

Figure 5.9: Dialog box to set parameters for free energy calculation step.

The interface for the *Free Energy* step (Figure 5.8) provides access to many of the same parameters that appear in other M2 steps, as described above; choosing the *FE Parameters* button provides access to the
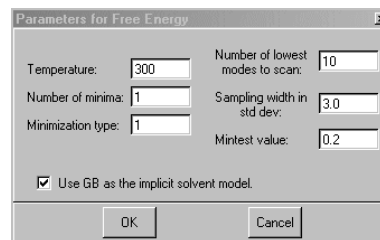
parameters unique to the free energy step through the *Parameters for Free Energy* window (Figure 5.9) as:

- Number of minima: Number of filtered conformations to be processed with HA/MS calculations.

- Number of lowest modes to scan: Number of modes with low force constants to be scanned for anharmonicity and integrated numerically if they prove sufficiently anharmonic (see Chapter 4.1). All remaining modes will be treated harmonically.

- Minimization type: Minimization type where 1 refers to CG+TN. There are no other types available in this version of M2.

- Temperature: Temperature in Kelvin.

- Sampling width in std dev: Each mode will be integrated, either by HA or MS, out to the lesser of this number of standard deviations and a 60° change of any torsion angle.

- Mintest value (kcal/mol): Occasionally mode scanning along a mode yields an energy lower than that of the initial nominal energy minimum. This implies that the starting conformation was not fully energy minimized, and this condition can lead to large errors in the M2 results. If the energy changes by more than this tolerance during mode scanning, the scan is terminated and a warning is output.

- Use GB as the implicit solvent model: GB is used by default.

Clicking the *Run* button causes the GUI to ask for a name for the output file, then generates an input file for `M2_2win.exe` and invokes this program to process the conformations. The results are displayed in the *Calculation Results* box in the left-hand pane of the GUI, and also in the selected output file.
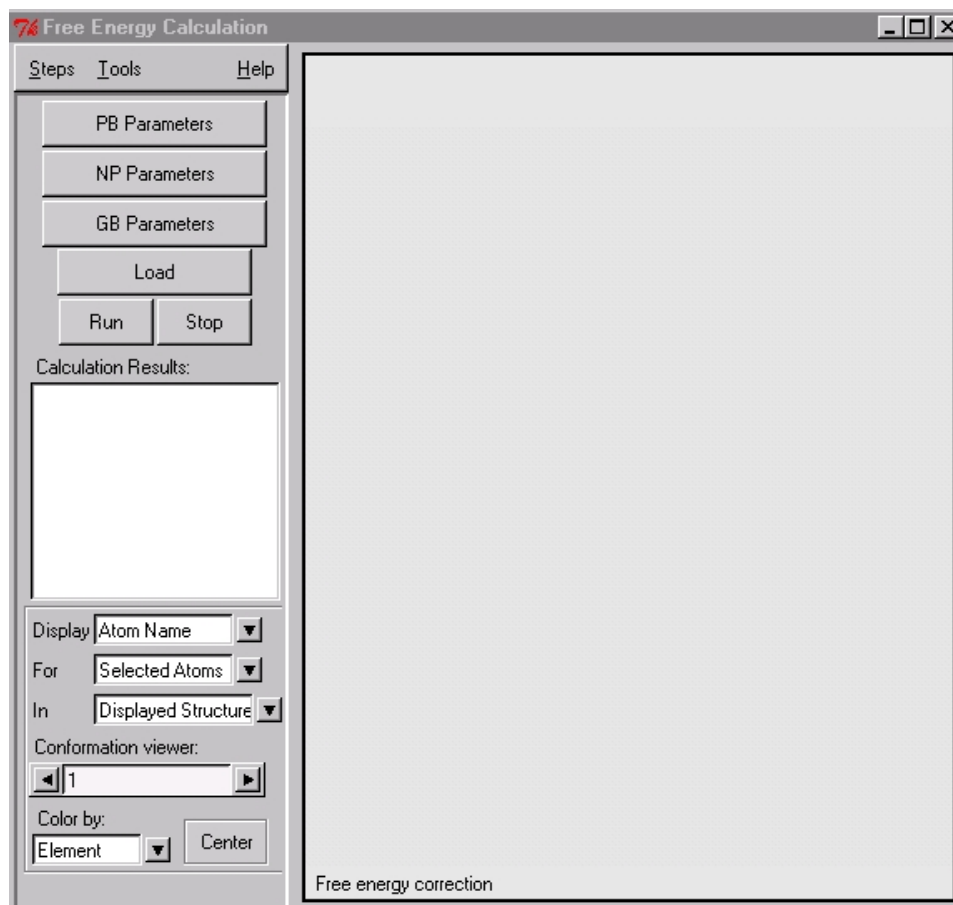
Figure 5.10: User interface to correct free energies with PB/SA solvation model.

5.6 Free energy correction

The final step, *Free Energy Correction*, involves adjusting the free energies from the HA/MS method by subtracting out the GB solvation energy computed at the base of each energy well and adding back the solvation energy computed with the PB/SA solvation model. This estimates the solvation free energy as an electrostatic part computed with the PB continuum model, and a nonpolar part estimated as proportional to molecular surface area with an empirical coefficient.
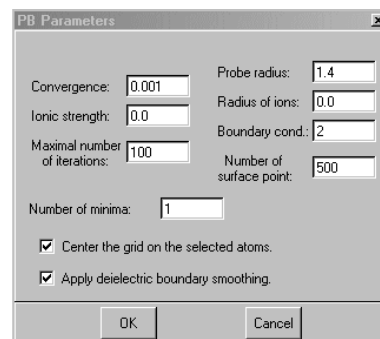

Figure 5.11: Dialog box to set parameters for Poisson-Boltzmann solvation calculations.

The M2 interface for this step (Figure 5.10) pro-
vides access to the parameters of the GB, PB and
NP solvation terms, via the corresponding buttons.
The *PB Parameters* and *NP Parameters* windows are shown in Figures 5.11 and
5.12, respectively and the GB parameters window is described above. Note that the
GB parameters used here must match those used in the free energy calculations, or
else the solvation energy correction will be done incorrectly.

The version of UHBD modified for use with M2, called m2uhbd, automatically
uses the maximal allocated finite-difference grid dimensions. It looks at the length
(x-axis), width (y-axis) and height (z-axis) of the molecule or complex, and scales
a rectangular grid box to fit, such that the molecule occupies 0.7 of each linear dimen-
sion of the grid. See the UHBD manual (http://adrik.bchs.uh.edu/uhbd/index.html)
and publications (20) for details regarding the PB parameters. In brief,

- Convergence: Fractional convergence criterion for the finite difference solver.

- Ionic strength of solvent (mM).

- Maximal number of iterations allowed in finite difference solver.

- Probe radius: Radius (Å) of solvent probe used to define dielectric boundary
  molecular surface.

- Radius of ions: radius (Å) of the dissolved ions making up the ion atmosphere;
  otherwise known as thickness of the Stern Layer.

- Boundary cond: 0- Zero boundary potential; 1- Molecule is single DH sphere
  of radius; 2- Sum of atoms as independent DH sphere (default); 3- Uniform
  applied electric field.

- Number of surface points: number of points on each atom sphere used in
  defining the dielectric boundary molecular surface (Default=500).

The nonpolar surface energy is estimated as
$W_{NP} = C_1 A + C_2$, where A is the solvent accessible
surface area and C_1 and C_2 are empirical parame-
ters based on experimental solvation data. The NP
Parameters window allows the following parameters
to be set:

- Solvent radius: Solvent probe radius (Å) used
  in computing solvent accessible surface area.
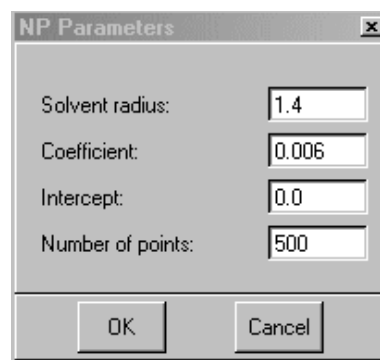
Figure 5.12: Dialog box to set
parameters for nonpolar solva-
tion component.

- Coefficient: Coefficient for surface energy, C_1 (default=0.006 kcal/mol/Å).

- Intercept: Intercept for surface energy, C_2 (default=0 kcal/mol).

- Number of points: Number of points per atomic sphere used in the numerical calculation of the surface area (default=500).

Clicking the *Run* button causes the GUI to ask for the name of a free energy file from the *Free Energy* step (above), which must correspond to the loaded conformations. It will output a file with `-new-final.out` extension with the corrected free energies.

Chapter 6

User Interface Components - Tools

The *Tools* menu provides access to a number of tools that support execution and analysis of M2 results.

- *Manual Docking*: This tool (Figure 6.1) is designed to help set up a reasonable initial conformation of a host-guest complex, which will then be energy minimized and subjected to a thorough Tork conformational search. The basic goal here is to put the guest into the host's binding cleft. The procedure is as follows:

  1. use the *Load system* button to load Topology (`.top`) and initial coordinate (`.crd`) files for the ligand and receptor.

  2. Make sure the "Frame" button is selected so that the docking box is visible, and make sure the "Box" button is selected instead of the "Ligand" button so that the Translate and Rotate controls act on the box rather than the ligand.

  3. Adjust the box size, position and orientation so that it covers the targeted binding site of the receptor.

  4. Click the "Run Docking" button and wait while the software builds a potential grid in the docking box, checks the energy of various positions and orientations of the ligand within the box, and then displays the lowest-energy conformation found.

  5. If you want to further adjust the bound conformation, click the "Ligand" button and use the Translate and Rotate controls to move the ligand around.

  6. If you want to adjust a dihedral angle, click "Show atom numbers", type the numbers of the 4 atoms defining the angle of interest into the text
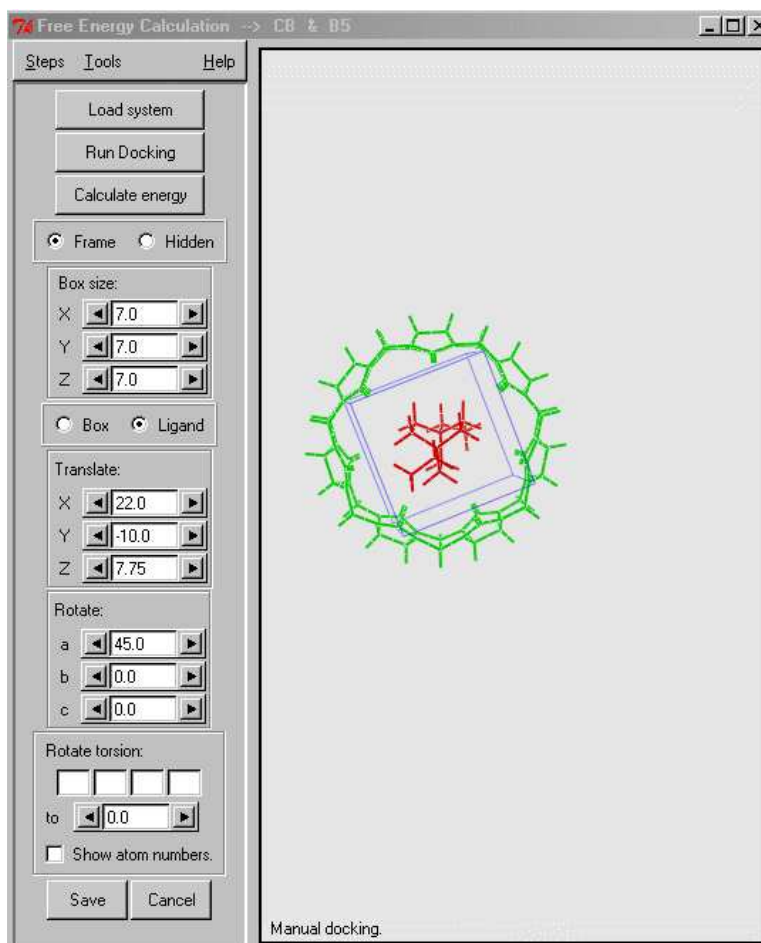
Figure 6.1: User interface for manual docking of guest into host binding site.

boxes below "Rotate torsion", hit "Enter" to visibly highlight the atoms, and use the arrows or text box to change the angle.

7. Repeat above procedures as desired

8. Finally, click the Save button to write a trajectory file with the displayed coordinates of the complex.

- *RMSD*: Once a set of conformations has been loaded into M2, this tool allows calculation of a simple root mean square deviation (RMSD) between any two conformations. Upon selection of the *RMSD* option under the *Tools* menu, the message "Please select TWO structures to calculate the RMSD" will appear. The conformations should then be selected by typing their numbers into the *Conformation viewer* box; their RMSD in Å will appear near the top of the screen.

- *RMSD by symmetry*: Same as *RMSD*, except that the result will be corrected for molecular symmetries, resulting in equal or lower RMSD values.

- *Superimpose*: Any number of loaded conformations can be superimposed with either of two methods, "principal axis" and "anchor atom". The conformations to be superimposed are chosen by typing their numbers on the *Conformation viewer* box. Selecting the *Superimpose* option under the *Tools* menu brings up the *Superposition method selection* box, which allows the desired method to be chosen. Clicking the *OK* button causes the superposition to be carried out. The principal axes method compute the center of mass and principal moments of inertia of each molecule and aligns them. The anchor atom method translates and rotates the molecules to optimally superimpose the selected atoms.

- *De-superimpose*: Undoes a conformational superposition (above).

- *Read MDL*: Reads in the chemical structure of a molecule in MDL Molfile format, in order to enable calculation of partial atomic charges (below).

- *Partial Charges*: If the Vcharge program (23) has been obtained from VeraChem by the user (free for noncommercial use at www.verachem.com), this command allows calculation of Vcharge conformation-independent, "ab initio-like" partial atomic charges for an MDL Molfile. Once the Molfile is loaded (above), choose the *Partial Charge* option under the *Tools* menu to bring up the *Partial Charge* window, and click *Run* button. The charges will be calculated and written into a new file with extension `_vcharge.mol`. The *Calculate* and *Smooth* options here are briefly described in the tooltips visible by hovering the mouse over these buttons. If desired, the user can manually replace the existing charges in a TOP file with the new charges generated via Vcharge.

- *Make TOP file*: Allows a `.top` file with Quanta/CHARMm parameters to be generated from a Quanta `.psf` file. Following the steps below, you can build the necessary files to load your molecule.

  1. Use *QUANTA* (Accelrys, San Diego, CA) to generate `.psf` (Protein Structure File) and `.crd` files.
  2. Choose *Tools* in the main M2 window.
  3. Select *Make TOP file*.
  4. Select *Convert PSF*.
  5. Use the *Browse* button to select the Quanta `.psf` file.

6. The `.top` file name should appear automatically in the same location with the same name as the `.psf` file, except with the `.top` extension. However, you may change the name or location here.

7. Press the *Convert* button. This will generate the files.

8. Select the *Return to M2* button and you will have the corresponding TOP file.

- *TRJ → CRD*: This allows a CHARMM format `.crd` file to be generated from a conformation displayed in the M2 graphics window.

- *CRD → TRJ*: Reverse of the previous option.

- *Link files*: Allows multiple `.trj` files to be linked into a single `.trj` file. This can be useful, for example, when multiple Tork outputs need to be combined into a single `.trj` file for conformational filtering. Note that M2 does not allow multiple `.trj` files to be loaded at the same time, so linking them before loading is the only option. *Save target file* allows the user to specify the desired name of the output `.trj` file. Then the *File Selection Dialog* allows the user to choose the `.trj` files to be linked. Once the output `.trj` file has been written, it can be loaded with the *Load* button for further work.

- *Export TRJ*: Writes out a `.trj` file for the molecular conformations currently displayed in the M2 graphics window.

- *Clear window*: Unload all molecules for a new calculation.

Chapter 7

M2 Files: Contents and Format

1. `.psf` - Protein structure file, a basic data file format used by the program CHARMM.

2. `.top` - M2 molecular topology file with the following contents:

   - ATOM BLOCK: Atom index, CHARMM atom type, mass, charge (au), Lennard-Jones parameters (kcal/mol and Å), and alternate Lennard-Jones parameters for 1-4 interactions.
   - BOND BLOCK: Indices of the bonded atoms, force constant (kcal/mol/Å), and equilibrium bond-length (Å).
   - ANGLE BLOCK: Indices of the atoms forming each angle, force constant (kcal/mol/radian^2), and equilibrium value of the harmonic angle term.
   - PROPER DIHEDRALS: Indices of the atoms forming the dihedral angle, the coefficient of the sinusoidal torsional energy term (kcal/mol), the multiplicity of the torsion, and phase of the torsion (radians).
   - IMPROPER DIHEDRALS: Indices of the atoms forming the dihedral angle, the harmonic force constant (kcal/mol/radian^2), and the equilibrium value of the angle.
   - NBFIX BLOCK: Special Lennard-Jones parameters (kcal/mol Å), which override the standard combination rules for the specified atom-type pairs.
   - NFINAL: numbers of atoms, bonds, angles, proper dihedrals and improper dihedrals. (The final integer can be ignored.)

3. `.crd` - CHARMM coordinate file

4. `.trj` - M2 trajectory file, which holds Cartesian coordinates of the output conformations in a stacked PDB format. (For further information on this format, please see the source code that generates it, `wtrjlm_new.f`.

5. `.inp` - An input file written by the GUI for one of the M2 executables. The lines should be reasonable self-explanatory, based on the parameters visible in the GUI windows and discussed in this manual.

6. `.out` - An output file from one of the M2 steps. The final output file (`*-new-final.out`) contains the chemical potential (free energy) and energy components for all the filtered and sorted conformations, along with the corresponding cumulative results.

Chapter 8

Compilation and Automation

The source code for `M2_1win.exe` and `M2_2win.exe` is provided in the files `M2_1.zip` and `M2_2.zip`, respectively. The code is written in FORTRAN 90 and was developed and compiled on Windows XP with Compaq Visual Fortran. We have also successfully compiled both executables with the Intel(R) Fortran compiler under the Windows XP and Linux operating systems. The IntelWindows and VisualFortranWindows folders contain the source code and workspaces for the corresponding compilers. (You may need to modify the paths in the workspaces to match your usage in order to use these.) The following notes provide additional information for compiling and running the M2 codes.

- If you work on a large system, you may need to increase the sizes of some arrays. This can be done by modifying the file `parameters.inc`:

  + maxatm: maximum number of atoms.

  + maxbnd: maximum number of bonds.

  + maxang: maximum number of angles.

  + maxdih: maximum number of dihedrals.

  + maxnb: maximum size of the nonbonded pairlist

- The default compilation settings only need to be supplemented with the flags /fpp (to enable preprocessing) and /stack:0x1dcd6500 (to increase the stack size).

- To port the code to the Linux environment and compile with the Intel ifort compiler:

+ Place all the source code files provided in the VisualFortranWindows (sic) folder in a single directory on the Linux system, along with the provided Makefile.

+ Type "make all" at the Linux prompt to build the program.

+ `M2_1.exe` and `M2_2.exe` should be run at the command prompt with the names of the desired input and output files on the command line; e.g., "`M2_1.exe B5_tork-tork.inp B5_tork-tork.out`".

+ You may use the input files generated with the Windows GUI as models for your Linux runs.

- The Scripts folder provided with this distribution includes a collection of ancillary scripts that automate the process of iterating the Tork/free energy calculation cycle to convergence. Please refer to the ReadMe file in the Scripts folder for information.

Chapter 9

Tips and Troubleshooting

- If the graphics window does not display a molecular system when you first load it, use the right mouse button to rescale the display and the molecule(s) should show up.

- Starting with a poor conformation can lead to trouble. In particular, make sure that the initial bound conformation of your complex looks basically reasonable. Be aware that a conformation can be energy minimized to a local energy minimum yet be incorrect due, for example, to damaged stereochemistry, a covalent bond thrust through a phenyl ring, a ligand outside the receptor, etc.

- Tork is an aggressive search algorithm and sometimes generates abnormal conformations, especially of macrocyclic hosts. Be sure to view the conformations before computing the free energy. If there are abnormal conformations, you can simply use a text editor to excise them from the `.trj` file to be used in the free energy calculations.

- If the Tork calculations for the bound complex generate conformations with the guest outside the binding site, it is usually a good idea to delete these conformations before proceeding.

- Be aware that the free energy results depend upon the force field you use. Whereas the force field parameters for amino acids have been pretty highly optimized over the years, this is not the case for the varied atom types that appear in host-guest chemistry. If your results are unsatisfactory, it is a good idea to take a critical look at the force field parameters, or try using a different set.

- Especially when working with a new system, it is a good idea to get a sense for the convergence by doing several Tork runs with different random number

seeds.

- The present default parameters are based on our experiences with the M2 software, but it is not a bad idea to experiment with variations, if only to get a sense for what they all do.

- Occasionally, Tork will not fully energy-minimize a conformation. In this case, the Mode Scan that occurs during the HA/MS free energy calculation may scan into a lower-energy conformation. The way this plays out in the math is that the conformation in question will be reported to have an abnormally large (favorable) entropy. Therefore, if you come across a conformation with a surprisingly large entropy, it is a good idea to try re-minimizing the energy and recalculating its free energy, just to be sure it is all right. The code does try to catch these problems in advance: if the energy decreases more than "mintest" during Mode Scanning (Section 5.5), the scan is terminated and a warning message is written to the console.

- The version of UHBD that we adapted for M2 automatically scales the finite difference grid to fit the molecule or complex you are studying. However, if your system is too large, the finite difference grid spacing may become too coarse to yield good PB results. (A grid spacing of 0.3 Angstroms or smaller is best.) It is a good idea to check `*_fe-pb.out` to make sure that UHBD is using a sufficiently fine grid.

- When comparing with experiment, always take a careful look at the experimental methods to make sure they are appropriate and that you match them as closely as possible. For example, if the experiments use cyclodextrins at high concentration, then the binding measurements may reflect not only 1:1 host-guest binding, but also the formation of higher-order complexes, such as a 2:1 cyclodextrin:guest complex. Or, if the solvent is a mixture of water and methanol, it may be hard to settle on appropriate solvation parameters for use in M2.

- In order to avoid file-format problems, it is safest to generate all input files via the M2 GUI. You can use a text editor to revise them, but be sure to maintain an identical format.

- Remember to add the contribution of $-RT \ln \left(\frac{8\pi^2}{C^\circ} \right) = 7.03$ kcal/mol to the computed binding free energy. See Equation 2.2 and Section 4.1.

- Filtering may crash when there is a large number of conformations. If this happens, divide the trj file into several separate trj files with fewer conformations, filter them separately, merge the filtered files, and then filter the merged file to get the final filtered trj file.

- Remember that the present conformational filtering program will fail if an atom has more than 8 bonds. This can happen in the case of some metalloorganic complexes. You have the option of using a third-party filtering program and then loading the filtered conformations back into M2 for the free energy calculations.

- CRD and TOP files should follow the same format as the CRD and TOP files in the Sample folder. Neither of the files should contain any header title at the beginning or any other comment in between or at the end of the files.

- It is important that the pathnames associated with the M2 repository or any of your working directories do not contain any blank space character or be too long.

- When you generate a TOP file using the "Make TOP File" option under the *Tools* menu, check that all the parameters are defined and your TOP file does not include any "n/a" flags. This flag, if present, means that a parameter required by the PSF file is not present in the parameter data files. You will need to provide such parameters yourself by manually editing the TOP file. For example, the " NX CT NX" angle set is not defined in our CB-original.top file (see Sample folder) and a reasonable value has been manually inserted in `CB.top`.

Bibliography

[1] Gilson, M. K. *Prot. Struct. Func. Gen.*, **1993**, *15*, 266–282.

[2] Gilson, M. K.; Given, J. A.; Head, M. S. *Chem. & Biol.*, **1997**, *4*, 87–92.

[3] Chang, C.-E.; Potter, M. J.; Gilson, M. K. *J. Phys. Chem. B*, **2003**, *107*, 1048–1055.

[4] Chang, C.-E.; Gilson, M. K. *J. Comput. Chem.*, **2003**, *24*, 1987–1998.

[5] Chen, W.; Huang, J.; Gilson, M. K. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 1301–1313.

[6] Kolossvary, I. *J. Phys. Chem. A*, **1997**, *101*, 9900–9905.

[7] Chang, C.-E.; Chen, W.; Gilson, M. K. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 1534–1539.

[8] Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem.*, **1997**, *101*, 1609–1618.

[9] Polar hydrogen parameter set for CHARMm, Molecular Simulations Inc. Waltham, MA.

[10] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.*, **1983**, *4*, 187–217.

[11] MacKerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B*, **1998**, *102*, 3586–3616.

[12] MacKerell, A. D. Jr.; Wiókiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.*, **1995**, *117*, 11946–11975.

[13] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.*, **1984**, *106*, 765–784.

[14] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.*, **1993**, *115*, 9620–9631.

[15] Gilson, M. K.; Honig, B. *J. Comput. Aided. Mol. Des.*, **1991**, *5*, 5–20.

[16] Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem.*, **1997**, *101*, 3005–3014.

[17] Gilson, M. K.; Honig, B. *Prot. Struct. Func. Gen.*, **1988**, *4*, 7–18.

[18] Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.*, **1994**, *98*, 1978–1988.

[19] Chang, C.-E.; Gilson, M. K. *J. Am. Chem. Soc.*, **2004**, *126*, 13156–13164.

[20] Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.*, **1991**, *62*, 187–197.

[21] Zhou, H.; Gilson, M. K. *Chem. Rev.*, **2009**, *Articles ASAP July 9, 2009.*

[22] Chen, W.; Huang, J.; Gilson, M. K. *J. Chem. Inf. Sci.*, **2004**, *44*, 1301–1313.

[23] Gilson, M. K.; Gilson, H. S. R.; Potter, M. J. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 1982–1997.