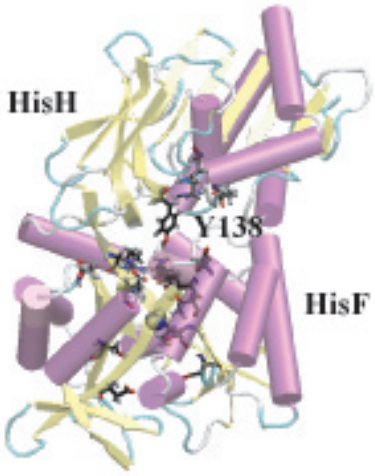
University of Illinois at Urbana-Champaign
Department of Chemistry - Luthey-Schulten Group
Beckman Institute for Advanced Science and Technology
Theoretical and Computational Biophysics Group
Summer School 2003

Determining Force Fields



Rommie Amaro

Felix Autenrieth

Jerome Baudry

Rosemary Braun

Zaida Luthey-Schulten

July 10, 2003.

Contents

1 Biological Background and Chemical Mechanism 5

2 HisH System Setup 7

3 The CHARMM22 Force Field 11

4 Missing Parameter Development 12

5 Semi-empirical parameter generation: SPARTAN 23

6 Minimization with new parameters 30

Introduction

Molecular dynamics (MD) simulations are a powerful scientific tool used to study a wide variety of systems in atomic detail. From a standard protein simulation, to the use of steered molecular dynamics (SMD), to modelling DNA-protein interactions, there are many useful applications. With the advent of massively parallel simulation code such as NAMD2, the limits of computational analysis are being pushed even further.

Inevitably there comes a time in any molecular modelling scientist's career when the need to simulate an entirely new molecule or ligand is necessary. Therefore the technique of determining new force field parameters to describe these novel system components becomes an invaluable skill. Determining the correct system parameters to use in conjunction with the chosen force field is only one important aspect of the process. Equivalently, the ability to use several programs simultaneously to examine and refine the system in question is also a critical element of these kinds of problems. However, it should be noted that the extent of parameterization carried out in this exercise is minimal and designed to emphasize the major points required in a more detailed fitting procedure. Roadmaps for more systematic optimizations that include experimental data can be found in a series of articles for Charmm [1, 2], Amber [3], and other force fields, including OPLS-AA [4, 5] and ECEPP. Additional sources for parameterization can also be found on the web (see [6], [7], [8], [9]). Polarizable force fields that include terms to allow polarization of the charge distribution by environment are under development [10].

This tutorial will walk you through a comprehensive example of how one investigates, sets up, and simulates a small nonstandard ligand bound to a protein system; specifically, we will investigate the glutaminase subunit of the hisH-hisF system and will determine parameters for the nonstandard residue.

Starting from the crystal structure in the protein database, and using the breadth of available biochemical information, we will dock the small ligand (glutamine) to the active site of hisH and develop the missing parameters in accordance with the Charmm22 force field. As a first guess for the parameters, we will try to derive as many of the missing parameters as possible from existing similar molecules already parameterized in the force field. Then we will further refine these new parameters with semi-empirical quantum chemistry calculations. Once the new parameters are finalized, we will minimize the system. The combination of all of these techniques will require the use of at least four different computational biology and chemistry packages.

The entire tutorial takes about 3 hours to complete.

This tutorial assumes that MOE [11], Spartan [12], VMD [13], NAMD [14] and other software has been correctly installed on your computer. All the necessary files must be copied over to a directory called tbss.work in your home directory. Go to this directory by typing:

```
tbss>cd ~/tbss.work
```

Copy the needed directory, but instead of typing `TOP_DIR`, type the location of the Summer School directory tree:

```
tbss>cp -r TOP_DIR/sumschool03/tutorials/06-forcefield/data ./data
```

For instance, if the materials are located at `/mnt/cdrom` or at `~/Desktop`, replace `TOP_DIR` by `/mnt/cdrom` or `~/Desktop`. Use this copy of the tutorial files. Change your current directory to `data` by typing

```
tbss>cd ~/tbss.work/data
```

1 Biological Background and Chemical Mechanism

Living organisms have developed features to ensure their existence in a wide variety of environments on Earth. While details of these features may be unique to particular conditions or species, two minimum requirements are the ability to reproduce and carry out regulated metabolic processes. A common theme in metabolic processes is the synthesis of complex and diverse molecules from a limited number of precursors. Amino acids are not only the building blocks of proteins and peptides, they are also important precursors in the biosynthesis of purines, pyrimidines, and other biomolecules. Additionally, amino acid biosynthesis is an ancient and fundamental process, and these metabolic pathways are represented in a diversity of organisms spanning all three domains of life.

Regulated production of histidine depends on the complex interplay between nine catalytic active sites located on 6-8 polypeptide chains, depending on the organism. High resolution crystal structures of several of the enzymes regulating this vital pathway are now available [15, 16]. Of particular interest is the fifth step of the metabolic pathway, where a protein complex known as hisH-hisF forms a key branch point. At this step, the formation of two products is catalyzed by the heterodimeric enzyme complex, imidazole glycerol phosphate synthase (IGPS), which consists of hisH, a class-I glutamine amidotransferase, and hisF, a synthase subunit that catalyzes a cyclase reaction. One product, imidazole glycerol phosphate (ImGP), is further used in histidine biosynthesis, and the other, 5-aminoimidazole-4-carboxamide ribotide (AICAR), initiates *de novo* synthesis of purines (see [17, 18] and references therein).

Characteristic of the superfamily to which it belongs, hisH has a strictly conserved catalytic triad active site: CYS84, HIS178, GLU180. The cysteine covalently binds glutamine, and the histidine, initially protonated, donates a proton to the amide group of glutamine to produce ammonia and glutamate. The conserved chemical mechanism for another enzyme (Carbamoyl Phosphate Synthetase) belonging to this superfamily is depicted below [19]. Subsequent steps allow the release of glutamate and the reprotonation of the active site histidine (HIS178). The molecule of ammonia then diffuses roughly 10 angstroms across the interface of the two proteins, enters the alpha-beta barrel of hisF through a presently unknown mechanism, and is transported through the barrel of hisF to the active site located at the C-terminal end of the barrel where it is incorporated into the next substrate.

This exercise will lead you through the modelling of the hisH protein. We will investigate the catalytic triad that comprises its active site and determine the correct protonation states of all functionally important residues. After the cysteine has covalently bound glutamine, we are presented with a non-standard amino acid to simulate. It is our task to develop a set of force field parameters for the novel residue for use in MD simulations.

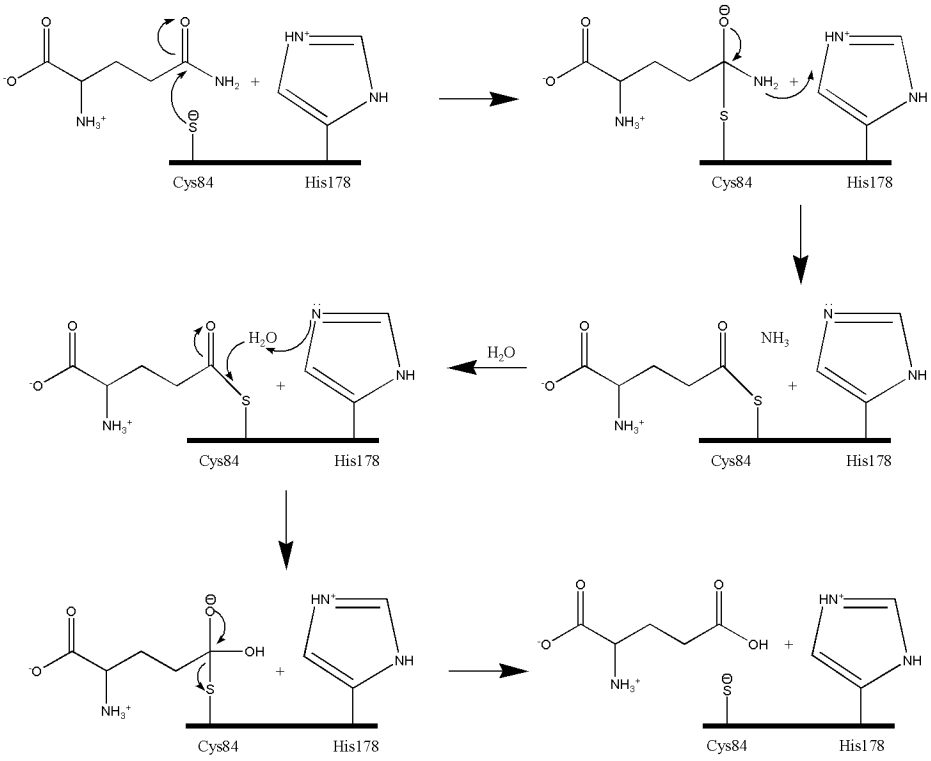


Figure 1: HisH mechanism for glutaminase reaction; we will model the system after step 2 of this mechanism (middle, right).

2 HisH System Setup

Exercise 1: Using VMD to investigate the hisH active site

To get an introduction to the system, load in the structure `HisHmoe_start.pdb` into VMD. Select the ligand glutamine and the catalytic triad (Cys84, His178 and Glu180).

Question: Identify all residues forming the glutamine docking site and describe the chemical properties of them. Using VDW, MSMS, or Surf representation in VMD, investigate the solvent accessibility of the glutamine docking site.

Although the glutamine can be attached in VMD, we will use MOE to make the covalent bond and determine what force field parameters are missing.

Exercise 2: System setup in MOE

For the following three exercises, we will use the molecular modelling program MOE (Molecular Operating Environment). Although the exercise makes exclusive use of the widely-used CHARMM22 empirical force field, any of the calculations may be repeated and applied to other available parameter sets (e.g AMBER, OPLS-AA, etc.).

Exercise 2, step 1: Starting MOE

To run moe, type `moe` at the UNIX command prompt.

Exercise 2, step 2: Loading the protein

Load the starting PDB structure `HisHmoe_start.pdb` into Moe by clicking on:
File: Open and click OK
Check at the load options: *Center molecule* and click OK

In MOE we will work with the *Main* window and the *Sequence Editor*. The *Main* window contains pull-down menus and buttons for the major function groupings. Often there are multiple ways a particular feature of the program can be accessed. For example, you can open the *Sequence Editor* with the buttons on the right or by selecting it under the *Window* pulldown.

You should see something similar to the screen image in figure 2. The *Sequence Editor* is a very powerful tool to select residues, hide them or render

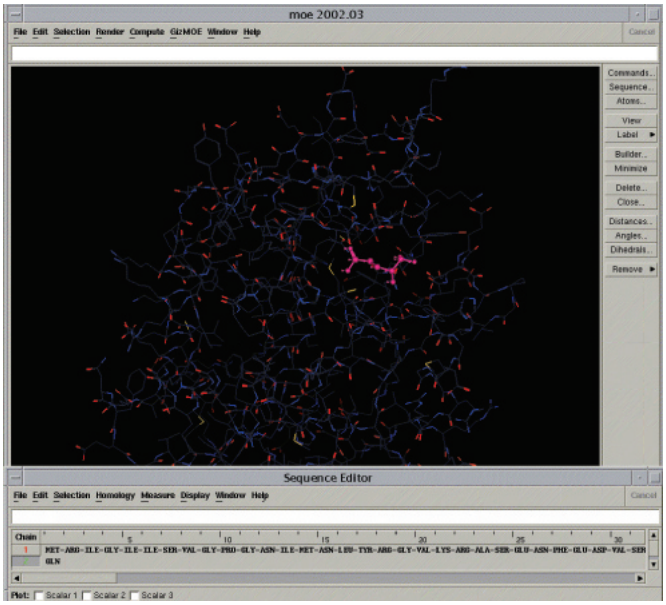


Figure 2: This is what your hisH system should look like in MOE.

them in different representations.

Exercise 2, step 3: Changing the representation of an object

In the *Sequence Editor*, click with the left mouse button on the ligand GLN to select it. After it is selected, right mouse click on it. A popup menu will appear. To make the selection:

Atoms: Select (the atoms will turn pink)

Render : Ball and Stick

Now glutamine should be rendered differently than the rest of the residues. Click in the *Main* window in the open space and the pink arrowheads will disappear (indicating the ligand is deselected).

Questions: Now try changing the protein representation to backbone on your own. (Click with the left mouse button on chain 1 (protein chain) for selecting it, open the pop-up menu (right mouse button) and choose Backbone: Cartoon.)

Exercise 2, step 4: Highlighting CYS84, HIS178 and the ligand

Let's focus in on the catalytic triad to make our work easier. First, we will hide the protein:

Click on Chain 1 to select the entire protein chain
Right click: Atoms: Select
Right click: Atoms: Hide Now only the ligand is shown.

With VMD, you have already determined the residues of the catalytic triad as well as the residues that comprise the docking site for the glutamine ligand. You can access the residue CYS84 rapidly in the *Sequence Editor* by clicking on them with the left mouse button. The residue will highlight in the *Sequence Editor* and clicking with the right mouse button shows you the various options you may apply to the selection.

Right click: Atoms: Show Now you should see just CYS84 and the ligand.
Repeat this to select HIS178

For selecting atoms by element, hybridization or other chemical properties the *Selection* pull-down menu is more suitable and you will see examples for using it. With the *Render* pull-down you can do the same as you have experienced with the pop menus. Play around with different selections and rendering modes to get a feel for the program.

Exercise 2, step 5: Adding missing hydrogens

Question: Does the x-ray structure have hydrogens? Why or why not?

To add the hydrogens, click on:
In the MAIN window, click on *Edit: Add hydrogen*

It is important to remember that MOE adds hydrogens by guessing and does not intrinsically know anything about protonation states of the various amino acids. For example, histidine residues can be in a charged or neutral state depending on the local environment within the protein. This property is very often exploited by enzymes in order to carry out biochemical reactions (e.g. as in our example of the catalytic triad of hisH). Since it is important from a functional perspective, when we are building our model we must ensure the model takes the relevant biochemical information into account.

Question: Study the chemical mechanism for the glutaminase reaction catalyzed presented before. Compare the protonation states of the catalytic residues generated by MOE to those in the mechanism. Make sure you have exactly the same protonation states as after step 2 of the chemical mechanism.

When performing a full system preparation, it is important to check the protonation state of every residue in the protein. However for this particular exercise we will restrict ourselves only to the docking site in hisH.

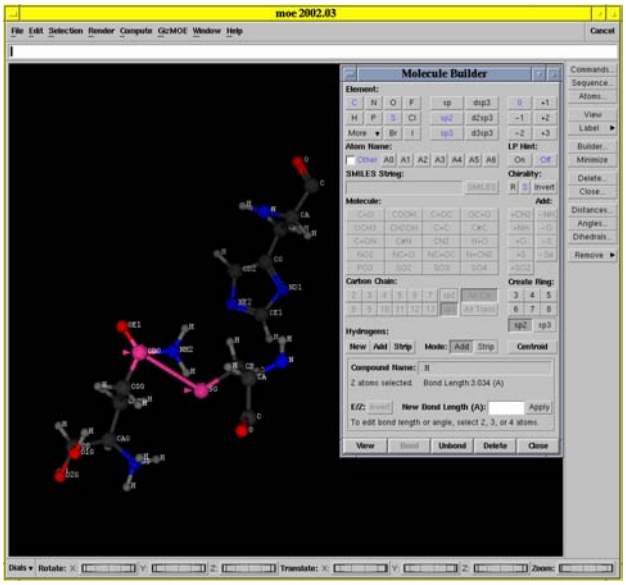


Figure 3: Making a bond in MOE: note the atoms to be bonded are shown in pink (indicating they are selected).

Exercise 3: Connecting the ligand to the protein: MOE

Now we want to make the covalent bond between glutamine and the CYS84 residue of hisH:

In MAIN window, click on *Builder*
In MAIN window, click on *Label: Name*

Now we will build the bond between the sulfur and carbon:

In MAIN window, click on the atom labelled *SG*
to select the second atom, press SHIFT and click on *CDG*
click on *Bond* in the Builder window.

We want to model the system after step 2 of the mechanism (see Figure 1). To do this, we need to delete some atoms. We need to delete the atom NH2 and its hydrogens (doing this will "hydrolize" the ammonia away!). To delete:
In MAIN window, click on atom NH2
Now in *Builder*, click Delete. The attached hydrogens will also be deleted.

Finally select all hydrogens and delete them (we will add the hydrogens back later with another program).
In MAIN window, click Selection : Elements : Hydrogen
In MAIN window, click Edit: Delete

Save your structure:
File: Save as:
Give your file a name, and select *Format: PDB.*

Now we need to close the session:
File: Close

In case you have problems making this file, we have provided you with the structure `HisHmoe_cyg.pdb`.

3 The CHARMM22 Force Field

The form of the potential energy function we will use in this exercise is taken directly from CHARMM22 and given by the following equation [1]:

$$\begin{aligned} V &= \sum_{\text{bonds}} k_b \left(b - b_0\right)^2 + \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} k_\phi \left[1 + \cos\left(n\phi - \delta\right)\right] \\ &+ \sum_{\text{impropers}} k_\omega \left(\omega - \omega_0\right)^2 + \sum_{\text{Urey-Bradley}} k_u \left(u - u_0\right)^2 \\ &+ \sum_{\text{nonbonded}} \epsilon \left[\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon r_{ij}} \end{aligned} \tag{1}$$

The first term in the energy function accounts for the bond stretches where k_b is the bond force constant and $b - b_0$ is the distance from equilibrium that the atom has moved. The second term in the equation accounts for the bond angles where k_θ is the angle force constant and $\theta - \theta_0$ is the angle from equilibrium between 3 bonded atoms. The third term is for the dihedrals (a.k.a. torsion angles) where k_ϕ is the dihedral force constant, n is the multiplicity of the function,

ϕ is the dihedral angle and δ is the phase shift. The fourth term accounts for the impropers (i.e. out of plane bending), where k_ω is the force constant and $\omega - \omega_0$ is the out of plane angle. The Urey-Bradley component (cross-term accounting for angle bending using 1,3 nonbonded interactions) comprises the fifth term, where k_U is the respective force constant and U is the distance between the 1,3 atoms in the harmonic potential. Nonbonded interactions between pairs of atoms (i, j) are represented by the last two terms. By definition, the nonbonded forces are only applied to atom pairs separated by at least three bonds. The VDW energy is calculated with a standard 12-6 Lennard-Jones potential and the electrostatic energy with a Coulombic potential.

Question: Sketch the individual Charmm22 potential energy terms. What is the difference between bonded and nonbonded interactions and what are the energy terms corresponding to them? What terms are likely to dominate the potential? What are the typical barrier heights?

4 Missing Parameter Development

We now have a nonstandard amino acid side group attached to the hisH protein. As is very often the case with nonstandard groups, this residue does not have a full set of parameters describing it.When presented with a nonstandard group, the first task is to determine exactly which parameters are already known, and which will need to be developed from scratch. In this case, only the atoms forming the connection between the ligand glutamine and CYS84 are likely to contain unknown parameters, as this connection is special to the reaction mechanism of hisH.

In order to assign initial values to the missing Charmm22 parameters, we will use the program MOE. These values will be checked and redefined using semi-empirical calculations will be performed with the quantum chemistry package Spartan. Finally we will incorporate the newly created parameters into a preexisting Charmm22 parameter file so that we can perform a minimization of hisH with NAMD2.

When developing the new parameters for this nonstandard residue, we will neglect any solvent effects and consider the molecule to be in the gas phase. The glutamine-cysteine linkage is defined as a new residue in the Charmm22 topology file with the name CYG. We will parameterize a truncated form of CYG in order to simulate step III. The truncated CYG has the following formula: CH3-S-CO-CH3 and its IUPAC name is methyl thioethanoate.

An Introduction to the Charmm22 Topology File

Below is the final topology file entry for the CYG residue. This exercise will show you how to develop the values appearing in it.

```
RESI CYG 0.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |   HB1
ATOM HA   HB      0.09  !     |   |
GROUP                   !  HA-CA--CB--SG
ATOM CB   CT2    -0.11  !     |   |   |
ATOM HB1  HA      0.09  !     |   HB2 |
ATOM HB2  HA      0.09  !   O=C       |
ATOM SG   S      -0.07  !     |        \
!ATOM HG1  HS      0.16 !               \
GROUP                   !                \
ATOM CDG   CC      0.55 !                 \
ATOM OE1   O      -0.55 !                  \
GROUP                    !     HN2G         \
ATOM CGG   CT2    -0.18  !       |            \
ATOM HG1G  HA      0.09  ! HN1G-NG   HB1G HG1G\
ATOM HG2G  HA      0.09  !       |    |     |   \
GROUP                    !  HAG-CAG--CBG--CGG--CDG=OE1
ATOM CBG   CT2    -0.18  !       |    |     |
ATOM HB1G  HA      0.09  !       |    HB2G HG2G
ATOM HB2G  HA      0.09  !  O1G=CG
GROUP                    !       |
ATOM CG    CD      0.75  !      O2G-HO2G
ATOM O1G   OB     -0.55
ATOM O2G   OH1    -0.61
ATOM HO2G  H       0.44
ATOM CAG   CT1    -0.12
ATOM HAG   HB      0.09
ATOM NG    NH3     -0.62
ATOM HN1G  HC       0.31
ATOM HN2G HC      0.31
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB CA   SG CB   N HN  N  CA
BOND C  CA   C +N  CA HA  CB HB1
BOND CB HB2
BOND SG CDG  CDG CGG CGG HG1G CGG HG2G
BOND CGG CBG CBG HB1G CBG HB2G
BOND CBG CAG CAG HAG CAG NG CAG CG
BOND NG HN1G NG HN2G CG O2G O2G HO2G
```

```
DOUBLE O  C CDG OE1  CG O1G
IMPR N -C CA HN  C CA +N O
IMPR CDG OE1 SG CGG  CG O1G O2G CAG
DONOR HN N
DONOR HN2G NG HN1G NG HO2G O2G
ACCEPTOR O C OE1G CDG O1G CG O2G CG NG CAG
```

Note the following features in the Charmm22 topology file:

- In the second column you can find the atom name of each individual atom. The connectivity of the atoms is depicted in the drawing.
- In the third column the atom type is designated for each atom. The parameter files of force fields have not chemical properties for individual atoms of the system. They have parameters only for atom types in order to minimize the number of entries in the file. It is assumed that the chemical properties of i.e. every carbonyl group are the same and are independent from the individual atom properties. From this column the force field extracts the information for assigning the right parameters during the simulation. If you want to distinguish two chemical functional groups from each other you have to give them different atom types in the force field. Assigning only different atom names is not sufficient.

Question: Find the two carbonyl groups in CYG. Will the bond stretching, angle bending and torsional terms be the same for both functional groups?

- The fourth column shows the partial charge of each atom. There are various quantum mechanical methods available for assigning partial charges in force fields. We will calculate partial charges from the electrostatic potential (ESP charges). The atoms are grouped in sets, which the net charge of each set is zero. This allows simulations of large residues when an electrostatic cutoff is applied as every group is a separate entity.
- On the bottom you can find Bond and Improper (IMPR) entries. These entries define the connectivity and the planarity of residues. Angles are autogenerated in Charmm22.

An Introduction to the Charmm22 Parameter File

Below is an excerpt from a Charmm22 parameter file that shows some required parameters necessary for a minimization run. The force field below contains the final missing heavy atom parameters for CYG and can be used as a reference for the assessment of the parameters you develop in the next few exercises.

Due to time constraints, for the following exercises, we will only perform semi-empirical calculations. The final parameter set was developed using a higher level of theory (an ab initio Hartree-Fock calculation with the 6-31G** basis set). Therefore the final parameter values presented below may differ slightly than the ones you determine.

Question: Why does an ab initio calculation for the truncated form of CYG so take so much longer compared to pentane (both systems have the same number of heavy atoms)?

```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
! Modified for CYG residue after 6-31G** geometry optimization Felix
S    CC   240.000      1.7814 ! ALLOW   ALI SUL ION

ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
!atom types     Ktheta    Theta0   Kub     S0
!

! Modified for CYG residue after 6-31G** geometry optimization
 CT2 S    CC    34.000    100.2000 ! ALLOW   ALI SUL ION

 CT2  CC   S     50.000    114.5000 ! ALLOW   ALI SUL ION

 O    CC   S     75.000    122.2000 ! ALLOW   ALI SUL ION
```

Note the following features in the Charmm22 parameter file:

- It contains all numerical values required for Charmm22 energy functions explained in section 2 of this tutorial. The forcefield contains entries for bonds, angles, dihedrals, impropers and nonbonded interactions. The forcefield also contains comments describing where these values have been obtained either by experimental spectroscopic measurements or quantum mechanical calculations.
- The bond and angle entries contain the value for the force constant in the first column in the physical units of kcal/mole/Å. The second column contains equilibrium bond or angle distance in Åor degrees, respectively.
- The next section contains the torsion angles comprised of dihedrals and impropers. The individual values will be explained in the course of the exercise and you will perform a dihedral drive of the torsion CT3-S-CC-CT3.

The values for the nonbonded interactions are the standard VDW parameters in Charmm22 required for the Lennard-Jones potential of your atom types.

Assigning Initial Values for Unknown Parameters with MOE

We will begin parameterizing CYG in this exercise. First you have to build the truncated CYG residue. MOE will use pattern recognition on the Charmm22 force field and the atom types to assign initial values to any unknown parameters. Afterwards, we will set up a minimization run with our newly developed parameters and calculate one dihedral rotational barrier in the truncated CYG.

Exercise 4, step 1: Building CYG

Open the *Builder* in the right column.

MOE has pre-defined templates for many functional groups. Here, we are going to build a chain of C-S-CO-C (hydrogens will be added automatically). In the *Molecule Builder*:

Click on C

Now to add the next atom, click on the very end of one of the hydrogens. When you click with the left mouse button on a hydrogen, a pink arrowhead will appear indicating that it is selected.

Click on S

Repeat this sequence, making sure to select a hydrogen on the previous group is selected before adding a new group

Click on *Close* to exit the builder and deselect all atoms by clicking on empty space in the MOE window. Change the rendering style to ball and stick by:

Render: Ball and Stick

Experiment with different rendering styles. An example of ball and stick is shown in Figure 4.

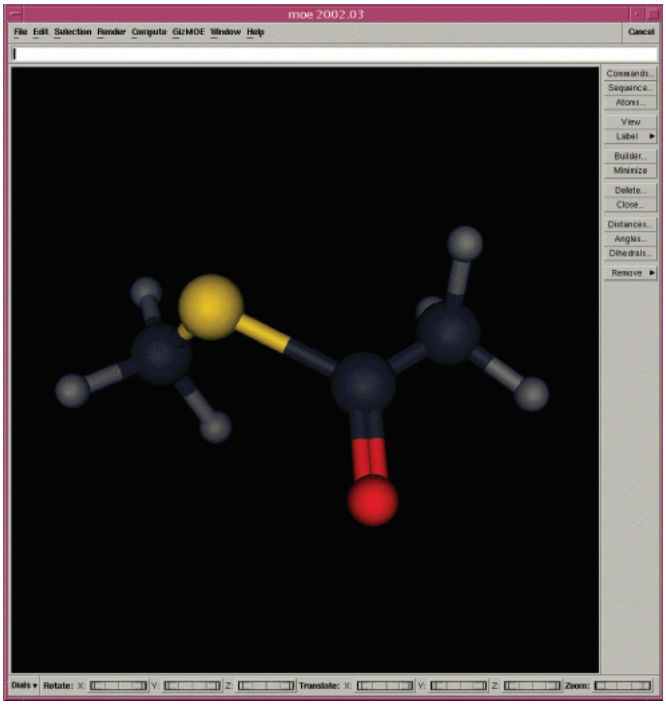


Figure 4: The truncated residue CYG, as it appears in MOE.

Exercise 4, step 2: Specifying the Force Field

Now it is time to specify the force field and options we want to use for our calculations. As mentioned previously, we will perform our calculations in gas phase, i.e., no explicit or implicit solvation models. The dielectric constant used for electrostatic interactions will be set to one with no distance-dependency.

Compute: Mechanics: Potential Control

Make sure that the loaded force field is *charmm22.ff* and click *Close*.

Exercise 4, step 3: Labelling Atoms

Next, display the atom types and partial charges of the atoms.

Label: MM Type

Label: Charges

You will notice a charge of zero on each atom because we have not loaded in the

Charmm22 force field. The atom types are displayed together with the partial charges. Note that the atom types for the terminal methyl groups (CT3) are different from the oxygen bound carbon atom (C). They correspond to carbon atoms that are electronically different. Note also that all hydrogen atoms have the same atom type (HA). All the hydrogens are electronically equivalent, since all are bonded to sp3 carbons.

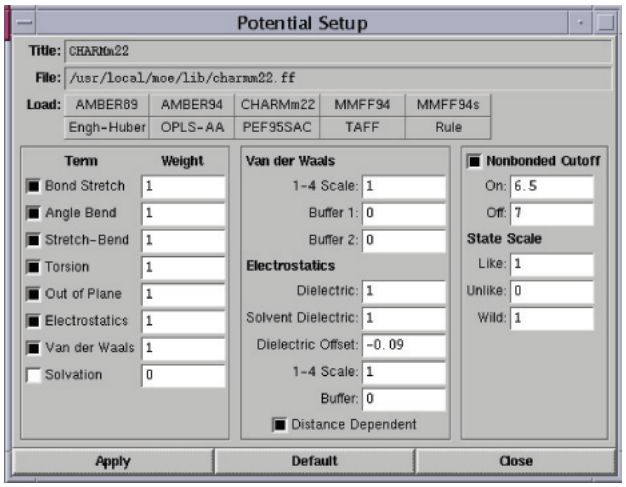


Figure 5: What your potential setup box should look like.

Exercise 4, step 4: Assigning Partial Charges

The partial charges of all the atoms need to be "assigned." In order to do this, click on:

Compute: Partial Charges: Force Field Charges

The charges are now defined according to the Charmm22 force field for the corresponding atoms. Note that all equivalent hydrogens have the same charge, and that the total charge of CYG sums to zero as expected (you can do this by selecting *Label: Charge*.

Exercise 4, step 5: Viewing / Printing Parameter Reports

You can look at the parameters for the truncated CYG molecule by clicking *Commands* in the righthand column and then clicking:

Compute: Mechanics: Parameter Report

Browse through the window to look at the parameters. The parameters are copied below. Additional descriptions of the force field can be found under the

HELP menu in the main window.

```
loading force field '$MOE/lib/charmm22.ff'
            E      str      ang      stb      tor      oop      ele      vdw
  sol
ALL:  216.023  185.641    2.489    0.000    8.629    0.000    0.000   19.263
0.000
loading forcefield '$MOE/lib/charmm22.ff'
FORCE FIELD PARAMETERIZATION SUMMARY
Mon May 12 19:47:45 2003

The following sections summarize the force field parameters used
for bond stretch, angle bending, proper and improper torsions. Van
der Waals and hydrogen bond parameters are not listed.

ATOM SUMMARY

All atoms relevant to the summary are listed below.  Each atom is
listed on a separate line that contains its name, element symbol,
ionization, hybridization, partial charge, mass (amu), radius,
and its potential type.  If the potential type is not known, then
the line is marked a >.

  Chain   1: <noname>

    Residue   1:   *   0  none

          1: CT3  C    +0 sp3   q=-0.220 m=12.0110 r=2.06 MM=CT3
          2: S    S    +0 sp3   q=-0.057 m=32.0660 r=2.00 MM=S
          3: HA   H    +0 sp    q= 0.090 m= 1.0080 r=1.32 MM=HA
          4: HA   H    +0 sp    q= 0.090 m= 1.0080 r=1.32 MM=HA
          5: HA   H    +0 sp    q= 0.090 m= 1.0080 r=1.32 MM=HA
          6: C    C    +0 sp2   q= 0.517 m=12.0110 r=2.00 MM=C
          7: CT3  C    +0 sp3   q=-0.270 m=12.0110 r=2.06 MM=CT3
          8: HA   H    +0 sp    q= 0.090 m= 1.0080 r=1.32 MM=HA
          9: HA   H    +0 sp    q= 0.090 m= 1.0080 r=1.32 MM=HA
         10: HA   H    +0 sp    q= 0.090 m= 1.0080 r=1.32 MM=HA
         11: O    O    +0 sp2   q=-0.510 m=15.9990 r=1.70 MM=O


BOND STRETCH (1-2) PARAMETERIZATION SUMMARY

Each bond stretch interaction is listed below. For each
interaction, the atom numbers and atom names of the atoms involved
are printed along with the equilibrium bond length (A) and the
```

quartic force constants (kcal/mol/A**i).

	A1	A2	N1	N2	len	k2	k3	k4
	1	5	CT3	HA	1.111	322.000	0.000	0.000
	1	2	CT3	S	1.816	240.000	0.000	0.000
>	2	6	S	C	1.732	264.210	-528.420	616.490
	3	1	HA	CT3	1.111	322.000	0.000	0.000
	4	1	HA	CT3	1.111	322.000	0.000	0.000
	7	10	CT3	HA	1.111	322.000	0.000	0.000
	7	9	CT3	HA	1.111	322.000	0.000	0.000
	7	6	CT3	C	1.490	250.000	0.000	0.000
	8	7	HA	CT3	1.111	322.000	0.000	0.000
	11	6	O	C	1.230	620.000	0.000	0.000

ANGLE BEND (1-3) PARAMETERIZATION SUMMARY

Each bond angle interaction is listed below. For each interaction, the atom numbers and atom names of the atoms involved are printed along with the equilibrium bond angle (deg) and the quartic force constants (kcal/mol/deg**i).

	A1	A2	A3	N1	N2	N3	ang	k2	k3	k4
	5	1	2	HA	CT3	S	111.3	46.100	0.000	0.000
	4	1	5	HA	CT3	HA	108.4	35.500	0.000	0.000
	4	1	3	HA	CT3	HA	108.4	35.500	0.000	0.000
	4	1	2	HA	CT3	S	111.3	46.100	0.000	0.000
	3	1	5	HA	CT3	HA	108.4	35.500	0.000	0.000
	3	1	2	HA	CT3	S	111.3	46.100	0.000	0.000
>	1	2	6	CT3	S	C	120.0	63.449	-25.447	0.000
	7	6	11	CT3	C	O	121.0	80.000	0.000	0.000
>	7	6	2	CT3	C	S	120.0	66.643	-26.728	0.000
>	11	6	2	O	C	S	120.0	83.697	-33.568	0.000
	10	7	9	HA	CT3	HA	108.4	35.500	0.000	0.000
	10	7	6	HA	CT3	C	109.5	33.000	0.000	0.000
	9	7	6	HA	CT3	C	109.5	33.000	0.000	0.000
	8	7	10	HA	CT3	HA	108.4	35.500	0.000	0.000
	8	7	9	HA	CT3	HA	108.4	35.500	0.000	0.000
	8	7	6	HA	CT3	C	109.5	33.000	0.000	0.000

TORSION (1-4) PARAMETERIZATION SUMMARY

Each torsion involving 4 atoms bonded in a sequence is listed once.

	A1	A2	A3	A4	N1	N2	N3	N4	v1	v2	v3	v4	v5
>	5	1	2	6	HA	CT3	S	C	0.000	0.000	0.000	0.000	0.000
>	4	1	2	6	HA	CT3	S	C	0.000	0.000	0.000	0.000	0.000
>	3	1	2	6	HA	CT3	S	C	0.000	0.000	0.000	0.000	0.000
>	1	2	6	7	CT3	S	C	CT3	0.000	-2.372	0.000	0.000	0.000
>	1	2	6	11	CT3	S	C	O	0.000	-2.372	0.000	0.000	0.000
	10	7	6	11	HA	CT3	C	O	0.000	0.000	0.000	0.000	0.000
>	10	7	6	2	HA	CT3	C	S	0.000	0.000	0.000	0.000	0.000
	9	7	6	11	HA	CT3	C	O	0.000	0.000	0.000	0.000	0.000
>	9	7	6	2	HA	CT3	C	S	0.000	0.000	0.000	0.000	0.000
	8	7	6	11	HA	CT3	C	O	0.000	0.000	0.000	0.000	0.000
>	8	7	6	2	HA	CT3	C	S	0.000	0.000	0.000	0.000	0.000

Question: Parameters are given by atoms types. What parameters should we be calculating with a quantum chemistry calculation?

A Closer Look at Dihedral Parameters

Note the dihedral (torsional) parameters. The analytical form of a dihedral term in the CHARMM force field:

$$V_\phi = k_\phi \left[1 + (\cos(n\phi - \delta))\right] \tag{2}$$

where ϕ is the value of the dihedral angle, k_ϕ is the force constant, n is the symmetry of the rotor (e.g. 3 for methyl groups) and δ is the phase. MOE follows this analytical form, but lists the parameters in a different format: The first four numbers are the atom numbers in the molecule. The four atom types are the atom types for which a torsional term is defined and will be calculated. Then follow five numerical values. These values correspond to the k_ϕ constant of the analytical form given above. Each of the five terms correspond to n = 1 to 5 in the analytical function given above, i.e., up to five values of n can be used simultaneously, each with a different force constant. The n = 1 term, with a positive C_1 constant, has a single minimum at 180 degrees or the trans conformation. A negative value of C_1 will create a maximum at 180 degrees, and therefore create a minimum at the cis conformation. The n = 2 term, with a positive C_2 value, creates minima at 90 and 270 degrees. A negative C_2 value creates minima at 0 and 180 degrees, which is useful for enforcing planarity. The n = 3 term, with a positive C_3, creates maxima at 0, 120 and 240 degrees, which will emphasize the staggered conformation of sp3 carbons. A negative C_3 will emphasize the eclipsed conformation.

For instance, the line:

```
9    8    5    1  HA   CT2  CT2  CT3    0.000   0.000  0.195  0.000  0.000
```

could also be written, for these four atoms:

$$V_\phi = 0.195 \cdot [1 + \cos(3\phi)] \tag{3}$$

Since we only have a force constant for n=3, the value of the force constant is 0.195, and the phase is zero (d = 0 since we have +0.195 and not 0.195). A sample Matlab plot of this function is presented below. In standard CHARMM notation, the exact same information would be written as follows:

```
HA   CT2  CT2  CT3       3          0.195          0
```

i.e., the four atom types, the symmetry of the rotor (n), the force constant (k_θ) and the phase (d).

Be sure to save your generated CYG molecule for later reference:
File: Save as: File format PDB

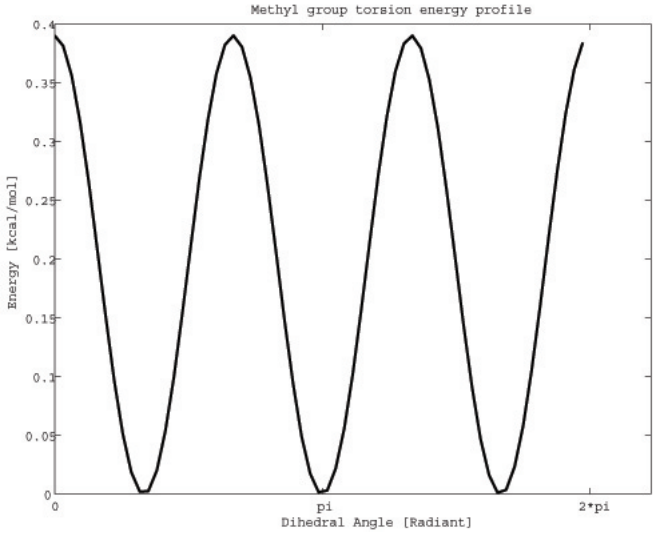


Figure 6: Matlab plot of Eq. (3) presented above.

5 Semi-empirical parameter generation: SPARTAN

In this exercise, we'll be using the quantum chemistry software package Spartan to calculate the force field parameters for CYG. To start, type `spartan -mesa` at the unix command prompt and get ready to build a new molecule by chosing

File: New

You should see the *Builder* interface. You can build molecules by selecting atom (and bond) types at the right, and placing them by clicking on the left portion of the screen. Atoms can be connected by clicking on the open valences. Start by building CYG, $CH_3 - S - CO - CH_3$. All the bonds are single except for the double bond between the carbon and oxygen.

Tip:. In Spartan, you can use the middle mouse button to rotate the molecule, the right mouse button to translate the molecule, and shift+right button to scale the molecule.

You can turn on atom labels by choosing *Model: Labels*. You should have something like this:

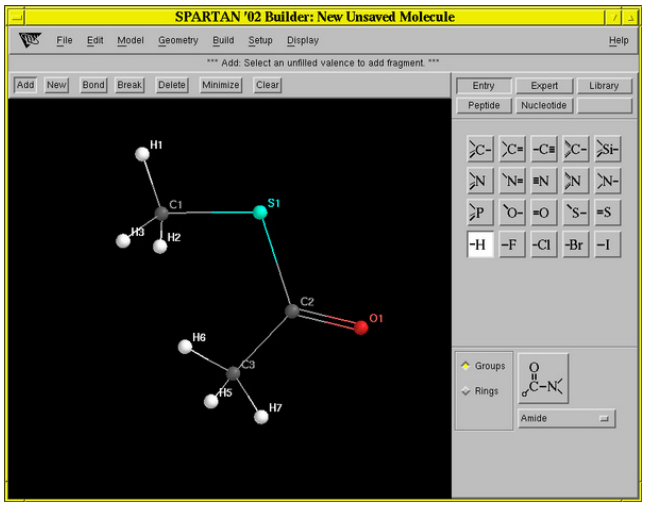


Figure 7: Building CYG in Spartan

Tip:. If you made mistakes in the structure, you can fix them by choosing *Expert* mode at the right and selecting the atoms or bonds to alter.

Minimize the structure by clicking *Minimize* at the top. Once it's done, you should see:

Return to the main menu by choosing

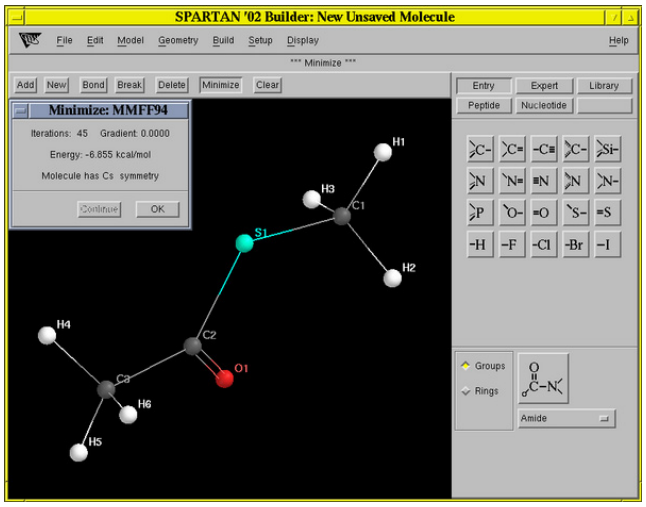


Figure 8: CYG after minimization in Spartan

File: Return to main
Save molecule? *Yes*

Everything is ready now for a semiempirical quantum mechanics calculation. You optimize the geometry of the structure, calculate electrostatic potential (ESP) charges and vibrational frequencies. Finally you perform a coordinate drive for determining the rotation barrier of one of the dihedral angles.

Geometry Optimization and Vibrational Modes

From the main menu, choose *Setup: Calculations* and set the options:
Task: *Equilibrium Geometry*
Model: *Model: Semi-Empirical PM3*
Subject to: *Symmetry*
Compute: *Frequencies, Electrostatic Charges*
Print: *Atomic Charges*

Options: check *Converge*

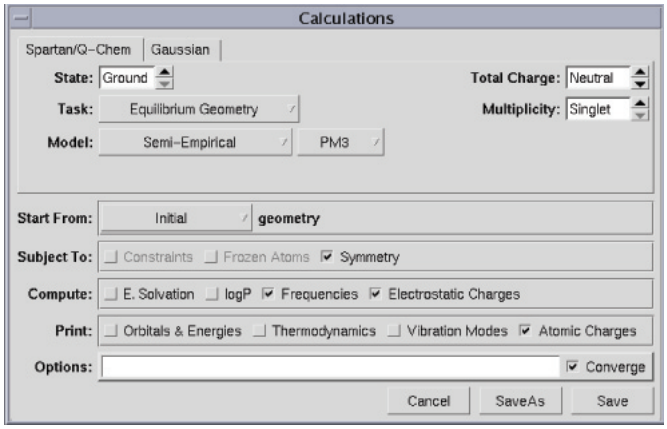


Figure 9: Setting up the energy calculation

Press *Save*. You can submit the job and view its output by choosing:
Setup: Submit
Display: Output
The output file lists details about the method used, and lists the Cartesian coordinates of the atoms. You can measure atomic distances and bond angles by using the *Geometry* pulldown.

Questions: What is the final energy of the CYG system? How many cycles has Spartan performed to converge? Measure the bonds, angles and dihedrals of the missing parameters in the report and compare the values with the proposed values from MOE.

Next, we are going to display the vibrational modes. The modes can be viewed through the animation part of the program. Click: *Display: Vibrations*; when you check the boxes you can watch individual vibrational modes. Be sure to increase the amplitude to see the vibrations clearly.

Question: Look at the measured angles and bond lengths and compare them to experimental values for other thioester linkages found in the PDB database (for general thioester linkage information, see [20]). See thioester linkage for 1ODV.pdb, figure presented below.

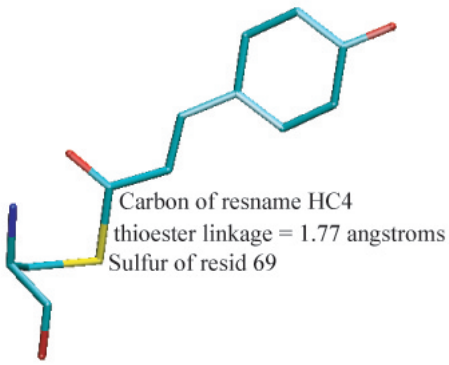


Figure 10: Thioester linkage found in 1ODV.pdb, photoactive yellow protein.

A systematic calculation of the force constants for the bond stretching and angle bending motion can be obtained from the ab initio calculations [9], but is beyond the scope of this exercise.

Atomic Charges

Next, we are going to display a surface around the atoms, on which we will map the ESP charges of the molecule. First check out the numerical values of Mulliken charges and ESP charges and compare the difference between them.

Model: Ball and Wire
Model: Labels
Model: Configure labels

Generally you will see that the partial charge values derived from an ESP calculation are higher than the Mulliken charges. ESP charges are calculated to reproduce the electrostatic potential around an atom, whereas the Mulliken charges are derived from the electron occupancy of orbitals. ESP charges are usually more suitable for force field generation.

You can display a potential surface by clicking on *Setup: Surfaces*. In the menu, set:

Property: Potential
Add

Submit the surface calculation with *Setup: Submit*. When the job completes,

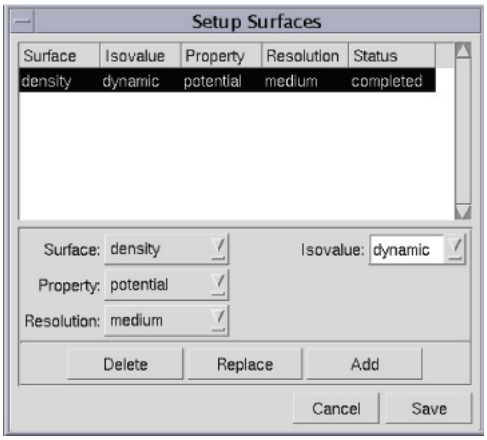


Figure 11: Surface set-up

you can view the surface with *Display: Surfaces*.

Questions: Does the color scheme on the surface fit the ESP partial charges?

Coordinate driving energy profile

The next exercise will be to calculate the energy of rotation around the H1-C1-S1-C2 dihedral. In Spartan, this is called a Coordinate Driving calculation (there are other names for the same task, such as conformational sampling, or dihedral search). First we will have to define the dihedral angle of interest:

Build: Define Profile

Select *Drive Dihedral*. The program expects four continuous atoms to be picked to define the dihedral angle. Click on the same atoms depicted above in the same order. You will rotate the dihedral angle 360 degrees from its initial value in steps of 15 degrees:

From: *initial value*

To: *initial value + 360*

Steps: *24*

Return to the *File: Main* window saving your changes and set up another job with *Setup: Calculations*, this time choosing

Task: *Energy Profile*

Submit it with *Setup: Submit*. It will take several minutes to run, but you

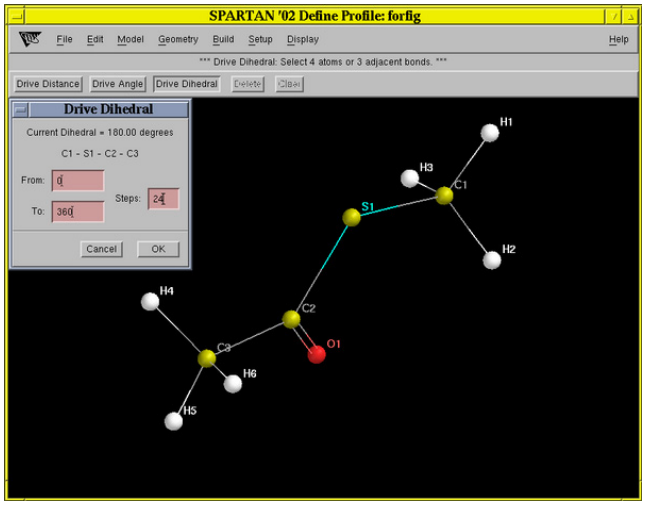


Figure 12: Dihedral selection

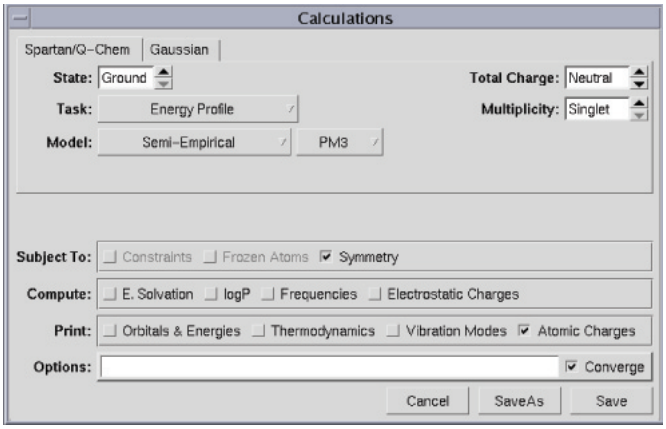


Figure 13: Energy profile calculation

can follow the progress on *Display: Output*. After the job is done, you can see the energy for each sampled conformation by selecting:

Geometry: Measure Dihedral click on same 4 atoms, choose 0 to 360, then click on "ladder-like" button (on right), then close measure dihedral box

Display: Spreadsheet

Column: Add E_gas: in kcal/mol

Plot these values by selecting *Display: Plots: Create*, and choosing *Dih360* for the X-axis and *E_gas (kcal/mol)* for the Y-axis.

Click anywhere on the function (it will look a bit strange; we will smooth is now)

Display: Plots: Edit Choose Fourier Least Squares

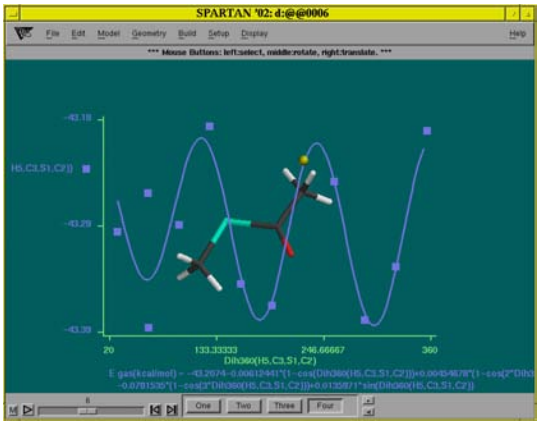


Figure 14: Dihedral energy profile results after smoothing, in Spartan

Question: What is the amplitude and periodicity of the dihedral energy profile? Where are the minimum and maximum points of the energy?

6 Minimization with new parameters

In this unit, you'll use our state of the art molecular dynamics program NAMD2 with the Charmm22 force field. If you did the initial copy command at the beginning of the tutorial, you should have the following files in your directory:

Topology file: `top_all27_prfar_cyg_nh4.inp`
Parameter file: `par_all27_prot_na_lipids_full.inp`
NAMD config file: `HisHmini.namd`
PDB file: `name of your MOE generated PDB file:  filename.pdb`

The parameter file was generated with the same protocol used in the previous section of this tutorial. However, rather than performing a semi-empirical calculation, the parameters in `par_all27_prot_na_lipids_full.inp` were calculated by full ab-initio quantum mechanics using the 6-31G** basis set, a process that takes several hours on a fast machine.

First we have to generate a NAMD2-compatible PDB file. You can generate a clean PDB file using the VMD console. Load the PDB file `HisHmoe_docked.pdb` into VMD and run the following commands in the VMD console:

```
set sel [atomselect top "all"]
$sel set segname HISH
$sel writepdb HisHmoe_cyg.pdb
```

Now open the VMD generated file `HisHmoe_cyg.pdb` in a text editor and change the name of residue 84 from CYS to CYG. Otherwise PSFgen will not recognize the correct topology for the residue.

We also must add hydrogens with the topology file using the program PSFgen in VMD. In the topology file the connectivity and therefore also the correct protonation state for each individual residue is defined. This allows us in contrast to the hydrogen adding procedure in MOE to set up the system with exact biochemical properties resembling glutamine bound to CYS84.

Start PSFgen in the VMD console by running:

```
!Generate protein structure file with PSFgen

package require psfgen
topology top_all27_prfar_cyg_nh4.inp

segment HISH {
   first NTER
   last CTER
   pdb HisHmoe_cyg.pdb
   }

coordpdb HisHmoe_cyg.pdb HISH

guesscoord
```

```
writepsf HisH_gln.psf
writepdb HisH_gln.pdb
```

When you are finished you have PDB file containing all coordinates including hydrogens and PSF file containing all information about connectivity, mass and charge of each individual atom in the structure. Just have a look at both files with an editor and get familiar with the format.

Question: Does the PSF file resemble the topology file?

Finally you will solvate the system for running a minimization in a NPT ensemble with a timestep of 1 femtosecond, a uniform dielectric constant of 1 and periodic boundary conditions. The electrostatic interactions will calculated with the Particle Mesh Ewald Method.

Start Solvate in the VMD console by running:

```
!Solvate system
package require solvate
solvate HisH_gln.psf HisH_gln.pdb
```

Question: Examine HisH_gln.pdb. How many atoms does the file now contain before and after solvation?

Run the minimization on the solvated system using the configuration file `HisHmini.namd`. We will be running all the minimization jobs on the NCSA Platinum Supercomputing Cluster. Be sure you are logged into the platinum.ncsa.uiuc.edu domain before running the minimization job. Once logged in, at the unix command prompt, type:

```
namd2 HisHmini.namd > HisHmini.namd.out
```

Question: Open the NAMD2 config file HisHmini.namd and briefly explain what each command does. The commands are defined in the users guide available online at http://www.ks.uiuc.edu/Research/namd/current/ug/. Look under the Chapter on Basic Simulation Parameters and the NAMD configuration parameters. What minimization method is being used? What is temperature during the minimization run?

The output of your NAMD job can be found in `HisHmini.namd.out`. You can use `namdplot` to look at energy changes during the minimization.

We will also provide you with the final DCD file generated during the minimization run. It can be found in felix/Forcefield (note: if you did the cp -r command in the Introduction, you already have the file). All the minimization files are located in the directory minData (to get there, type "cd minData" at the command line).

Load the DCD file in VMD and check the relevant distances between glutamine and the catalytic triad. You can determine how much the structure has changed during the minimization by using the RMSDscript.

Question: How much do they differ from the equilibrium values in the force field and from the initial unminimized structure? Has the structure converged during your minimization?

References

[1] MacKerell et. al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[2] Jr. A.D. MacKerell and N. Banavali. All-atom empirical force field for nucleic acids: 1) parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, 21:86–104, 2000.

[3] J. Wang, P. Cieplak, and P.A. Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules. *J. Comp. Chem.*, 21:1049–1074, 2000.

[4] W. Jorgensen, S. Maxwell, and J. Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Amer. Chem. Soc.*, 117:11225–11236, 1999.

[5] C. Rizzo and W. Jorgensen. Opls all-atom model for amines: Resolution of the amine hydration problem. *J. Amer. Chem. Soc.*, 121:4827–4836, 1999.

[6] A. Mackerell. Parametrization of molecules for use of charmm. *http://www.psc.edu/general/software/packages/charmm/tutorial/mackerell/parameters.html*, 2003.

[7] A. Mackerell. Workshop on methods and applications of molecular dynamics to biopolymers. *http://www.psc.edu/general/software/packages/charmm/tutorial/index.html*, 2003.

[8] J. Wang, R.M. Wolf, J. Caldwell, P. Kollman, and D. Case. Development of a general amber force field. *http://www.amber.ucsf.edu/amber/amber.html*, 2003.

[9] P. Pulay, G. Fogarasi, F. Pang, and J.E. Boggs. Systematic ab initio gradient calculation of molecular geometries, force constants, and dipole moment derivatives. *J. Amer. Chem. Soc.*, 101:2550–2560, 1979.

[10] H. Stern, G.A. Kaminski, J.L. Banks, R. Zhou, B.J. Berne, and R.A. Friesner. Fluctuating charge, polarizable dipole, and combined models: parametrization. *J. Phys. Chem. B*, 103:4730–4737, 1999.

[11] Chemical Computing Group. Moe. *Montreal, Canada*, 2003.

[12] Inc. Wavefunction. Spartan '02 tutorial and user's guide. *Irvine, CA*, 2001.

[13] W. Humphrey, A. Dalke, and K. Schulten. Vmd - visual molecular dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[14] L. Kale and et. al. Namd2: Greater scalability for parallel molecular dynamics. *J. Comp. Phys.*, 151:283–312, 1999.

[15] A. Douangamath, M. Walker, S. Beismann-Driemeyer, M. Vega-Fernandez, R. Sterner, and M. Wilmanns. Structural evidence for ammonia tunneling across the (beta-alpha)8 barrel of imidazole glycerol phosphate synthase bienzyme complex. *Structure*, 10:185–193, 2002.

[16] B. Chaudhuri, S. Lange, S. Chittur R. Myers, V. Jo Davisson, and J. Smith. Crystal structure of imidazole glycerol phosphate synthase: A tunnel through a (beta-alpha)8 barrel joins two active sites. *Structure*, 9:987–997, 2001.

[17] P. O'Donoghue, R. Amaro, and Z. Luthey-Schulten. On the structure of hish: Protein structure prediction in the context of structural and functional genomics. *J. Struct. Bio.*, 134:257–268, 2001.

[18] R. Amaro, E. Tajkhorshid, and Z. Luthey-Schulten. Developing an energy landscape for the novel function of a (beta-alpha)8 barrel: Ammonia conduction through hisf. *Proc. Natl. Acad. Sci.*, in press, 2003.

[19] J. Thoden, S. Miran, J. Phillips, A. Howard, F. Raushel, and H. Holden. Carbamoyl phosphate synthetase: Caught in the act of glutamine hydrolysis. *Biochemistry*, 37:8825–8831, 1998.

[20] D. Zacharias, P. Murray-Rust, R. Preston, and J. Glusker. The geometry of the thioester group and its implications for the chemistry of acyl coenzyme a. *Archives of Biochemistry and Biophysics*, 222:22–34, 1982.