Parameterizing a Novel Residue

Rommie Amaro

Brijeet Dhaliwal

Zaida Luthey-Schulten

November 2005

Contents

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 programs 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 arises. The technique of determining new force field parameters to describe these novel system components therefore 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. 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. 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. Road-maps 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. A recent paper by Autenrieth et al. reports both AMBER and CHARMM parameterizations for the reduced and oxidized forms of cytochrome C and is a useful reference for parameterization methodologies [6]. Additional sources for parameterization can also be found on the web (see [7], [8], [9], [10]). Polarizable force fields that include terms to allow polarization of the charge distribution by environment are under development [11].

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 its covalently bound substrate. Starting from the crystal structure in the protein database and using the breadth of available biochemical information, we will dock the substrate (glutamine) to the active site of hisH and develop the missing parameters in accordance with the CHARMM 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 three different computational biology and chemistry packages.

The entire tutorial takes about 2-3 hours to complete.

This tutorial assumes that VMD [12] (version 1.8.3), Spartan [13] (version '04), NAMD2 [14] (version 2.5) and a text editor has been correctly installed on your computer.

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. As amino acid biosynthesis is an ancient and fundamental process, 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, 17]. 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 (IGP synthase), 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), is utilized in the *de novo* synthesis of purines (see [18, 19] 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 [20, 21]. The conserved chemical mechanism for another enzyme (carbamoyl phosphate synthetase) belonging to this superfamily is depicted in Fig. 1 [?]. Subsequent steps allow the release of glutamate and the reprotonation of the active site histidine (HIS178). An investigation of the available crystal structures, coupled with experimental kinetic and mutagenic studies and molecular dynamics simulations, strongly suggests that the following reaction scenario: the nascent ammonia 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 [19, ?].

This exercise will lead you through the steps required to create the covalently bound substrate within the active site and develop a set of parameters to use in molecular dynamics simulations.
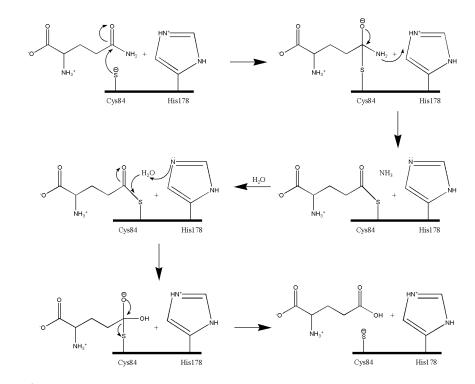
Figure 1: Putative hisH mechanism for glutaminase reaction; we will model the system after step 2 of this mechanism (middle, right).

Figure 2: IGP synthase, also called hisH-hisF, with bound substrates and ammonia is depicted. A molecule of glutamine binds in the active site of hisH (top, Rossman fold) and is hydrolized to form a molecule of free, reactive ammonia and glutamate. The ammonia molecule diffuses approximately 10 angstroms through the interface of the two proteins. HisF (bottom, $(\alpha/\beta)_8$ barrel) exploits its barrel structure to transport ammonia another 12 angstroms from the interface to the next active site, where it then reacts with the ribonucleotide substrate of hisF, PRFAR. PRFAR is shown docked to the C-terminal active site end of hisF.

2   HisH System Setup

Exercise 1: Creating the covalently bound substrate

Open a terminal window, and move into the directory for this tutorial by typing:

```
> cd forcefield-tutorial/forcefield-files
```

To get an introduction to the system, load in the structure `HisH_glutamine.pdb` into VMD. To do this, at the unix prompt type:
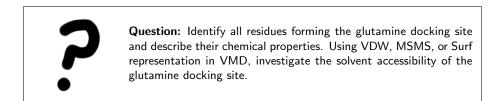
```
> vmd HisH_glutamine.pdb
```

The above command will launch VMD and simultaneously load in the molecule we specify. Now take some time to investigate HisH. First we will display the glutamine substrate. Make one atom selection for the ligand glutamine by selecting *Graphics: Representations* from the VMD Main menu and typing:

```
> segname LIG
```

in the Selected Atoms box. You can change the *Drawing Method* of the selection if you like (CPK is a nice view). Now let's look at the catalytic triad. To do this, click on *Create Rep* and type:

```
> segname HISH and resid 84 178 180
```

This creates a new representation for the catalytic triad of hisH. For this selection, change the *Drawing Method* to *licorice*.

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

Zoom in on the catalytic triad. Looking at the mechanism depicted in Fig. 1, identify which carbon of the substrate glutamine will covalently bind to the cysteine sulfur. Measure the distance between the two atoms. Identify which nitrogen group will be hydrolized off by this enzyme to become free, reactive ammonia.

Now we are going to perform "open-heart surgery" on the active site. The first step to making a covalently bound substrate is to append the glutamine substrate to the active site cysteine in hisH. This will take several steps in VMD

and then the use of a text editor.

To begin, open a TK Console window (under *Extensions* in the VMD Main window)and perform the following selections and commands:

```
> set sel [atomselect top "resid 84"]
> $sel set resname CYG
> set sel2 [atomselect top "segname LIG"]
> $sel2 set segname HISH
> set all [atomselect top "all"]
> $all writepdb hisH_cyg.pdb
```

After you have created the new pdb file, please quit VMD.

Let's take a look at what you've done. To do this, at the unix prompt in the working directory, type:

```
> more HisH_glutamine.pdb
```

Look at this file; see where the ligand glutamine is specified and look at the resname of the first residue 84. Now do the same for your `hisH_cyg.pdb` file. What did the above commands do to the file?

Now, we want to open the `hisH_cyg.pdb` file in a text editor. To do this, open a new terminal window and move into the working directory. Type:

```
> cd forcefield-tutorial/forcefield-files
```

Then type:

```
> nedit hisH_cyg.pdb
```

(You may use a text editor of your choice instead of `nedit`)

Our CYG residue still exists as two pieces; we want to combine them and will do so right now. Scroll down to residue 84. Notice that this residue (atoms 670 to 675) has all the atoms of a cysteine residue, but now has the resname field `CYG`. Now scroll to the end of the file; the last residue appearing is the glutamine molecule (atoms 1628 to 1637). Notice that it now has the segname HISH and it has a resname field called CYG. Copy and cut out this part of CYG by cutting all the lines specifying atoms 1628 to 1637. Be sure you cut out the entire line, not just a part of it! Leave the END statement intact.

We need to insert this cut text (*i.e.* the substrate) into the active site cysteine. To do this, scroll back up to the real residue 84 and paste the rest of the CYG
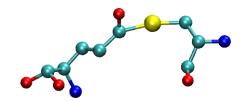
Figure 3: The novel CYG residue in the hisH active site. This residue was created by merging a glutamine and a regular cysteine residue.

residue in after atom 673 (the sulfur of CYG).

Now it is time to hydrolyze the ammonia molecule! To do this, delete out the entire line specifiying atom name NH2 (atom 1630).

Finally, save this modified file as `hisH_cyg_mod.pdb` in the folder: `forcefield-tutorial/forcefield`.

It's time to admire your handiwork. Open the `hisH_cyg_mod.pdb` file in VMD, and highlight residue 84 (hint: it helps if you set the representation to CPK). You just created the covalently bound substrate! (See Figure 3 for comparison – what you made should look identical to the picture.) When you are finished admiring your fine work, close VMD.

We have renamed the entire residue CYG, so it is easy to select and identify this novel residue throughout your simulations. Although we have parameters for both cysteine and glutamine, this residue will require some parameterization because they are connected in a new way.

> **?**
>
> **Question:** Will all the necessary parameters already exist in the CHARMM27 force field? What part(s) of this novel residue will need parameters?

3 Testing out your new residue

Now that you have the covalently bound substrate attached to the protein, let's try to simulate it. The first thing we need to do is create the .psf (protein structure file) and .pdb (protein data bank file) files we will need for the MD simulation. These files will be explained in the next few sections. For now, follow these instructions to use PSFGEN and make the necessary files.

Open a new terminal window and go to the directory:
> `forcefield-tutorial/forcefield-files`

This is where you have created your `hisH_cyg_mod.pdb` file. If not, move it into
this directory now. Look at the contents of the directory by typing:
> `ls`

Note that we have provided you a topology file called `top_all27_prot_na.inp`;
this is the normal CHARMM topology file distributed by the MacKerrell group.
Here we will use it to build our .psf and .pdb files.

Also within this directory you will see a PSFGEN input file called `build_covsub.pgn`.
At this point, we assume you are already familiar with PSFGEN (if not, please
see the NAMD tutorial for more detailed explanations). Take some time to look
at the file with a text editor, and read through all the lines in the file. The first
line specifies which topology file we will be using, the next 3 lines specify `alias`
commands, the next group of commands denotes the segname, any special start
or end commands, and which pdb file to use to build the segment. More often
than not, in simulations, you have many different segments in one huge pdb file,
so it is common to see many grouping listing the different segments. Finally,
we again specify which files will be used for the coordinates and the command
`guesscoord` tells the program to build in any missing coordinates (hydrogens
and any other atoms that might be missing). Lastly, it writes out the .pdb and
.psf files.

At a prompt, type:
`psfgen < build_covsub.pgn > myoutput`

This command should take a few seconds to run. When it is finished, it will
return the prompt.

> **Question:** What happened when you tried to run PSFGEN? Did
> you actually make the .psf and .pdb files? Why or why not?

When you tried running PSFGEN, it should have returned an error. To deter-
mine what the problem is, you need to read through the `myoutput` file. You can
do this by typing the following at a unix prompt within the working directory:

> `more myoutput`

Now read through the `myoutput` file. You will see that there is a line that
says:

```
unknown residue type CYG
```

You have made a new residue, but there is no information about a residue CYG in the standard topology file! Now you see why we need to build a topology file entry, as well as develop parameters for this novel residue. The rest of this tutorial will take you through some of the basics of doing this.

4   The CHARMM Force Field

There are several versions of the CHARMM force field available from the MacKerrell website. Don't let this confuse you! Most commonly you will see CHARMM19, CHARMM22, or CHARMM27. CHARMM22 (released in 1991) and CHARMM27 (released in 1999) are the most recent versions of the force field. For purely protein systems, the two are equivalent. However, if you are simulating any nucleic acids, be aware that CHARMM27 has been optimized for simulating DNA and should be used.

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, that is 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 van Der Waals (VDW) energy is calculated with a standard 12-6 Lennard-Jones potential and the electrostatic energy with a Coulombic potential. In the Lennard-Jones potential above, the $R_{min_{ij}}$ term is not the minimum of the potential, but rather where the Lennard-Jones potential crosses the x-axis (*i.e.* where the Lennard-Jones potential is zero). In order to run the MD simulations, we need to be sure that our parameter file has all the terms specified in the energy function we are using (*i.e.* the equation specified above).

> **Questions** 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?

5 Developing Topology and Parameter Files

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. The unparameterized part of the CYG residue is the connection between the sulfur and the carbon. The connection between the sulfur of CYS and the carbonyl carbon of the glutamine is called a thioester bond. By attaching the substrate in such a way, you have created a *glutamyl thioester intermediate*. In order to use this CYG residue in MD simulations, we need to make a topology file entry for it as well as develop a complete set of parameters describing it.

To do this, we will integrate two pre-existing topology/parameter file entries – a cysteine entry and glutamine entry. From a chemical and biological standpoint, this should make intuitive sense: though connected in a new way, we can still expect the two bonded residues to behave (for the most part) as they would when they are separate entities. The chemical and physical behavior of CYG should be similar to CYS and GLN.

The topology file specifies the partial charge of each atom and the connectivity of the atoms. We will then use the topology file when we create the .psf and .pdb file for our simulations. The parameter file specifies the specific equilibrium bond lengths and angles, bond and angle force constants, dihedrals, and impropers. We need bond, angle, and dihedral parameters for the atoms around this new residue, and we will generate them in the next exercise. The creation of both the topology file entry and the missing parameters are discussed in detail now.

5.1 An Introduction to a CHARMM Topology File

Below is the final topology file entry for the CYG residue. The CYG parameters appearing in it were created by combining a regular CYS entry (with the HG1 atom of CYS deleted) with a GLN entry (with the NH2 atom of GLN deleted). When merging the two residues into one, at the place where you connect the groups you will have to delete atoms (in this case a hydrogen on cysteine's S and the NH2 atom of the glutamine) and alter the charges. Following the standard CHARMM protocol, the HG1 atom on CYS was deleted and its charge was moved to SG (-0.23 + 0.16 = -0.07). In the CYG entry, you can see the line for HG1 has been commented out (comments are denoted by ! marks), and

the charge added to the S. Note that the group still has an overall neutral value.

Take a few minutes and look through one of the topology files we have supplied you. To do this, at a unix prompt, type:

`more top_all27_prot_lipid_prfar_cyg_cyn.inp`

Look at a regular CYS and GLN residue entries, and be sure you understand how we have pasted them together to make CYG below.

```
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
```
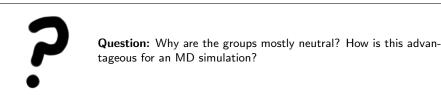
Note the following features in the CHARMM27 topology file:

- In the second column you can find the *atom name* of each individual atom. The connectivity of the atoms is depicted in the drawing and uses the *atom name*. This name must match the atom name in the pdb file.

- In the third column the *atom type* is designated for each atom. The parameter files of force fields do not have chemical properties for individual atoms of the system. Instead, they have parameters for *atom types* in order to minimize the number of entries in the file. It is assumed that the chemical properties of similar atoms are the same, i.e. every carbonyl group of the backbone has similar chemical and physical properties. From this column the force field extracts the information for assigning the right parameters during the simulation. If you want to distinguish two chemically functional groups from each other, then you have to give them different atom types in the force field. Assigning only different atom names is not sufficient.
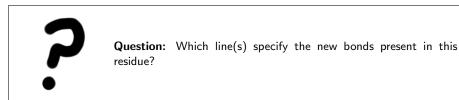
> **?**
>
> **Question:** Find the two carbonyl groups in CYG. Will the bond stretching, angle bending and torsional terms be the same for both functional groups?

- 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 discuss the determination of partial charges in more detail later. The atoms are divided into groups; with the exception of charged residues (such as arginine or glutamate), the net charge of each group is neutral.

Question: Why are the groups mostly neutral? How is this advantageous for an MD simulation?

- At the end of the entry, you will find Bond (`BOND`)and Improper (`IMPR`) entries. These entries define the connectivity and the planarity of residues.

Question: Which line(s) specify the new bonds present in this residue?

You may have noticed that although the bonds and atoms are defined in the topology file, the entry is missing angle and dihedral assignments. The angle and dihedrals are never assigned in the topology file. PSFGEN automatically determines the angles and dihedrals based on the connectivity of the atoms. This should make sense – every 3 connected atoms generates an angle and every 4 generates a dihedral term. By automatically generating the terms, PSFGEN saves you a lot of work!

5.2 An Introduction to a CHARMM Parameter File

Below is an excerpt from a CHARMM27 parameter file that shows some required (and previously missing!) parameters necessary for MD simulations. The force field excerpts below contain 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 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
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


DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types            Kchi    n    delta

CC   S    CT2  CT1     0.2400  1    180.00
CC   S    CT2  CT1     0.3700  3      0.00
HA   CT2  S    CC      0.2800  3      0.00
CT2   S   CC   CT2     2.05    2    180.00
CT2   S   CC   O       2.05    2    180.00
```

Note the following features in the CHARMM27 parameter file:

- It contains all numerical values required for CHARMM27 energy functions explained in section 2 of this tutorial. The force field contains entries for bonds, angles, dihedrals, impropers and nonbonded interactions. The force field 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 Å or degrees, respectively.

- The next section contains the torsion (*a.k.a.* dihedral) angles. The individual values will be explained in the course of the exercise and you will perform a dihedral drive of the dihedral HA CT2 S CC.

The values for the nonbonded interactions are the standard VDW parameters in CHARMM27 required for the Lennard-Jones potential of your atom types.

5.3   Assigning Initial Values for Unknown Parameters

Before we get into the details of parameterizing, it is useful and often very helpful to be aware of programs that are able to assist you in the process. CHARMm, AMBER, and MOE are all programs with modules that will identify which atoms are missing parameters and even give you an initial first guess of values. These programs almost always use a pattern recognition algorithm to guess for you, and any parameters they supply you will have to be checked with your chemical intuition and at least a basic level of quantum mechanics. If you choose to use one of these packaged programs, *caveat emptor*! While it seems like an easy solution to a sometimes difficult and always tedious process, it rarely, if ever, provides you with a suitable and/or reliable set of parameters.

5.4   A Closer Look at Dihedral Parameters

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

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

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. In the CHARMM energy function, you can specify multiple dihedrals for the same bond. Often one needs to use multiple dihedrals in order to generate energetic barriers of proper height and location.

You can use the following table to help you generate the correct dihedrals in
CHARMM.

Dihedral Parameters in CHARMM format			
Multiplicity, n	Phase δ	Location of Minima	Useful Notes
1	0	180	yields trans conformation
1	180	0	yields cis conformation
2	0	90, 270	
2	180	0, 180	useful for enforcing planarity
3	0	60, 180, 300	emphasizes staggered conformation of sp3 carbons
3	180	0, 120, 240	emphasizes eclipsed conformations of sp3 carbons

For instance, the line:

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

could also be written, for these four atoms:

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

i.e., the four atom types, the symmetry of the rotor (n), the force constant (k_ϕ)
and the phase (d). Since we only have a force constant for n=3, the value of the
force constant is 0.195, and the phase is zero. A sample Matlab plot is shown
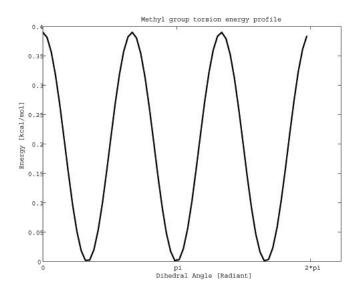in Fig. 4.

Figure 4: Matlab plot of Eq. (3) presented above.

6 Minimization with new parameters

In this unit, you'll use our state of the art molecular dynamics program NAMD2 with the CHARMM27 force field and the parameters we developed. You should be in the following directory (if not, go there now):

```
forcefield-tutorial/forcefield-files
```

From now on, we will call this directory the *working directory*.

When you look at the contents of the directory, you should have the following files:

Topology file: `top_all27_prot_lipid_prfar_cyg_cyn.inp`
Parameter file: `par_all27_prot_na_lipids_full.inp`
NAMD config file: `hisHmini.namd`
PDB file: `hisH_cyg_mod.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 NAMD2-compatible PSF and PDB files. We will do this with PSFGEN. PSFGEN will add hydrogens to the atoms; in the original pdb file we supplied you, the correct protonation state has already been specified for each of the residues.

As we did previously, we will be running PSFGEN directly from the terminal. This time we will run PSFGEN exactly as we had before, only we will use our new topology file. To do this, open the PSFGEN input file called `build_covsub.pgn` in a text editor. You can see that the topology file we specified is the old one, not the new one. We need to update the PSFGEN input file so that it uses our new topology file.

Change the first line of the PSFGEN input file from:
```
topology top_all27_prot_na.inp
```

to:
```
topology top_all27_prot_lipid_prfar_cyg_cyn.inp
```

Save the new file with the same name (`build_covsub.pgn`) in the working directory.

Try running PSFGEN again by typing the following at the prompt:

```
psfgen < build_covsub.pgn > myoutput2
```

Again, PSFGEN will take a few seconds to run. When it is finished, it will return the prompt. Look at the output file `myoutput2`. This time it worked! You can see it has generated a PDB file containing all coordinates including hydrogens (called `hisH_covsub.pdb`), a PSF file containing all the information about connectivity, mass and charge of each individual atom in the structure (called `hisH_covsub.psf`), and a logfile of the PSFGEN program's activities in the file called `myoutput2`.

? **Question:** Does the PSF file resemble the topology file?

Now we are going to solvate the system so we can minimize it. To solvate the system, launch VMD from a terminal window. Be sure you are already in the working directory before you start VMD. If you open VMD from the start bar or another directory, open a TK Console window (select *Extensions: TK Console* from the VMD Main menu) and `cd` into the directory where your new pdb and psf files are.

Open the TK Console in VMD (select *Extensions: TK Console* from the main VMD menu). At the prompt, type:

```
package require solvate
solvate hisH_covsub.psf hisH_covsub.pdb -t 9 -o hisH_solv
```

This will run SOLVATE, a program which creates a box or water molecules around the protein/object of interest. Notice that it created two new PDB and PSF files: `hisH_solv.pdb` and `hisH_solv.psf`.

? **Question:** Examine your new solvated system file, `hisH_solv.pdb`. How many atoms does the file now contain before and after solvation?

You just created a rather large and well solvated globular protein system. Let's take a look at the newly solvated system you just created. Load the solvated system by selecting *File: Load New Molecule*. Browse for your new PDB and PSF file and load them both into VMD.

Now you should be looking at your solvated system. Before we minimize the system, we need to be sure that the system has an overall charge of zero. To do this, in the TK Console window, type:

```
> set sel [atomselect top "all"]
> eval vecadd [$sel get charge]
```

The last command should return a number– this number is the overall charge. It should be something very close to zero. This is good, because it means that our system is already neutral and we will not need to add any ions to counter the charge. We will run the minimization on the solvated system using the configuration file hisHmini.namd, which we supply you.

We will do the minimization in the NPT ensemble with a timestep of 1 femtosecond, a uniform dielectric constant of 1 and periodic boundary conditions. The electrostatic interactions will be calculated with the Particle Mesh Ewald Method.

Go back to a normal terminal window and be sure you are once again in the working directory. At the unix command prompt in a terminal shell, type:

```
> namd2 hisHmini.namd > hisHmini.namd.out
```

This command begins the minimization, which should take about 10 minutes to run. You can use the tail command to track its progress. To do this, type this at the unix prompt:

```
> tail hisHmini.namd.out
```

While it is running, open the NAMD2 configuration file 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? How many steps are we minimizing for?

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. The minimization job you just ran was for only 100 steps. Minimizing for 10,000 steps would likely require over 2 hours of compute time for a single processor! We have provided you with the full minimization run results for a much longer run. All the minimization files are located in the directory Long_min. To get there from the working directory, at the unix prompt, type:

```
> cd Long_min
```

Use `namdplot` to look at the energy changes over the longer minimization run
we provide you: `hisH_solv_min_long.namd.out`. To so this, at a terminal win-
dow in the `Long_min` directory, type:

   `> namdplot TOTAL hisH_solv_min_long.namd.out`

We will need to load the DCD file into VMD in order to monitor the move-
ment of the CYG residue over the course of the minimization run. To do this,
in the VMD Main window, click on the hisH_solv molecule, and then select
File: Load Data Into Molecule. Browse into the `Long_min` directory and load
the `hisH_solv_min_long.dcd` file into the `hisH_solv.pdb` or `hisH_solv.psf` file
you already have loaded in VMD. Once you've done that, you will see that there
are 100 frames.

Highlight the CYG residue (VMD selected atoms phrase "resname CYG") and
play the trajectory. Watch how the residue moves. You might also want to
highlight the other residues in the catalytic triad (VMD selected atoms phrase
"protein and resid 84 178 180") and watch how the 3 residues move with respect
to eachother.

You can determine how much the structure of the covalently bound substrate
intermediate has changed during the minimization by using the RMSDscript
you learned about in the NAMD tutorial. Compare the first frame and the last
frame to see how the novel residue has changed.

?

Question: How much do the bond lengths and angles differ from
the equilibrium values in the force field and from the initial unmini-
mized structure?

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] Felix Autenrieth, Emad Tajkhorshid, Jerome Baudry, and Zaida Luthey-Schulten. Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Comp. Chem.*, 25:1613–1622, 2004.

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

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

[9] 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.

[10] 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.

[11] 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.

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

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

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

[15] 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.

[16] 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.

[17] Barnali N. Chaudhuri, Stephanie C. Lange, Rebecca S. Myers, V. Jo Davisson, and Janet L. Smith. Toward understanding the mechanism of the complex cyclization reaction catalyzed by imidazole glycerolphosphate synthase: Crystal structures of a ternary complex and the free enzyme. *Biochemistry*, 42:7003 – 7012, 2003.

[18] 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.

[19] 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.*, 100:7599–7605, 2003.

[20] T Klem and V Davisson. Imidazole glycerol phosphate synthase: The glutamine amidotransferase in histidine biosynthesis. *Biochemistry*, 32:5177–5186, 1993.

[21] T. J. Klem, Y. Chen, and V. J. Davisson. Subunit interactions and glutamine utilization by *Escherichiacoli* imidazole glycerol phosphate synthase. *J Bacteriol*, 183:989–996, 2001.