BUFF: A Biological Universal Forcefield Derived from Quantum Mechanics

Thesis by

Matt J. Carlson

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2000

(Submitted May 1 , 2000)

Acknowledgements

I don't like to move much. I spent the first 18 years of my life living in central Minnesota. I got this letter in the mail talking about a place called Caltech in Pasadena, California. What impressed me the most, were the average test scores of its students. They seemed to score very well. Hmm, I could go to college in California. I could learn to surf, maybe learn to ride a skateboard. It could be fun. I came out and investigated and found a community of lots of smart people who liked science. And so, I came. In almost 12 years here, I haven't learned to skateboard or surf (surprise, surprise), but I sure have learned a lot of other stuff. I've found that studying science day in and out hasn't made me lose my love and enthusiasm for science. I've learned that I like to interact with people and discuss things. I've learned I like to be part of a team.

For being my home for the past 12 years, I want to thank the entire Caltech community. Some of my friends were burnt out on Caltech after surviving 4 years as an undergraduate, but I was not. Sure, there were times when things were bad, but there is nowhere else on earth that has the unique properties of Caltech. I attribute it mostly to the people. People who understand that science is fun. People who know that bureaucracy is there to get things done, not a way to shuffle papers. People who like a good prank. I could go on, but I would not have stayed in Los Angeles as long as I have if Caltech hadn't made it a great place to live and work.

The best part of Caltech was clearly meeting my wife, Erica. Being the great physicist she is, I was sometimes jealous. I've long since given that up and am enjoying

the ride following her career as she turns into a great scientist. She has given me encouragement when I needed it, and, most importantly, put up with my roller-coaster life. Its been a blast and I'm looking forward to the new venue.

I might not be thanking Erica at all if it weren't for my advisor, Bill. I need to thank you for letting me back into Caltech after graduating as an undergraduate. I know you officially didn't do anything, but if you weren't willing to take me on as a student if I were accepted to grad school, I'm not sure I would have got in again. And being able to attend Caltech's grad school allowed me to stay in the area and get married! Of course, you've done much more than give me a place to study and work for the past few years. Your enthusiasm is catching and would often pick up my spirits as my experiments would get frustrating. I will always remember your teaching style and think of you whenever I try to make sure my audience is paying attention by asking them direct questions.

Now we come to the Goddard group. I can't imagine my grad student career anywhere else. For Jim Gerdy and Jan Peters in the early years, thanks for keeping the learning fun. Thank you, Darryl, for keeping the computers running and answering all my obscure programming and software questions. I need to thank Ken Brameld, Derek Debe, and lately Joe Danzer and Changmoon Park for being great officemates and for the research and scientific discussions we've had. I need to specifically thank Vaidahi for her help with all things biological and Cecco for almost everything I actually understand about quantum mechanics. I've interacted with the rest of the Goddard group to a greater and lesser extent. If I have any complaints about the group, its that I didn't get to know all of the group members better.

No man can live on bread alone and there are many other people here in Pasadena who kept me sane as I thought about forcefields every day. Thanks to the Graduate Bible study group who met in our apartment almost as long as I was a grad student. We had great fun and great discussions. I know I benefited from Gary's wisdom many times. I know Habakkuk and PERL so much better thanks to his frequent advice. Thanks to my friends at Lake Avenue Church for keeping me on track and all the encouragement. Thanks in particular to Jerod, John & Betsy, Marko, Michael, Dave, Tim, Lanny, all my guys, and pretty much everyone else. You were my creative outlet when research sometimes dragged me down.

My parents and family always supported me from afar and its great to know they love me and are supporting me in whatever I do. I know it was hard for my folks to let their son head all the way out to California to go to school, but I think things have turned out alright. I won't mention my in-laws specifically because I think of them as part of my family, so they should feel included in the first sentence of this paragraph!

Finally, we come to the purpose of life. I wouldn't know what to do with myself if I didn't think there was something greater going on with life than just running around and trying to do interesting stuff. I have to thank God for clueing me in on His existence and His great patience as I slowly figure out who He is and what He wants me to do. I'm not sure where I'm headed in life, but I know who I'm following and why. Thanks for giving me a Why, God.

Thesis Abstract

Molecular mechanical simulations of biomolecules require an accurate potential energy function (forcefield) in order to produce meaningful results. Most current forcefields are highly parameterized in order to correctly reproduce high level theory and experiment. Increasingly, new biomolecules are designed and studied that have atypical configurations such as metal centers and nonstandard amino acids. To avoid a lengthy process to develop new parameters for each new system encountered, a generic forcefield is desired. A hierarchical approach is undertaken herein to achieve this flexibility and accuracy.

Building upon the rule based generic forcefields UFF and Dreiding, a new biological universal forcefield, BUFF, is presented for the simulation of proteins and other biological molecules. In addition to its UFF and Dreiding based terms, the BUFF has additional hydrogen bond terms, specialized protein backbone torsions, and a process for deriving charges for amino acids that is independent of other parameterization. These additional parameters have been fit to *ab initio* quantum mechanical calculations carried out on model systems.

Validation studies of peptide trimers demonstrate that the BUFF accurately reproduces the quantum mechanical torsional energies. Several other common, highly parameterized forcefields are also applied to the same tripeptide systems, as well as short α-helical chains and other model systems in order to make a comparison to the BUFF. These studies show that while the BUFF is universal and can be quickly deployed on new

systems, such as unnatural amino acids or metal containing systems, it is also at least as accurate as other commonly employed, but highly parameterized, forcefields. The biological universal forcefield described herein is presented as complementary to the MSC forcefield derived for simulations of DNA and other nucleic acids.

Contents

List of Figures

List of Tables

Chapter 1: Approximating Chemistry

Introduction

The basic laws of nature have the unpleasant feature that they are expressed in terms of equations we cannot solve exactly, except in a few very special cases. For example, if we wish to study the motion of more than two interacting bodies, even the simple laws of Newtonian mechanics become essentially unsolvable using analytic methods. We must resort to numerical methods to find the answer. Using a computer, we can get the answer to any desired accuracy.

Most interesting molecular systems of interest contain many atoms or molecules, so there is no hope of finding the exact answer using only pencil and paper. Prior to the arrival of computer simulation, properties could only be predicted by using a theory that provided a crude description of the material of interest. From this period, we have the van der Walls equation for dense gases and the Boltzmann equation to describe the transport properties of dilute gases. Given enough information, these theories can provide us with an estimate of the properties of interest. However, we do not know enough about most intermolecular interactions to test the validity of a particular theory by direct comparison to experiments. If theory and experiment disagree, our theories may be wrong, or our estimate of the intermolecular interactions is wrong, or both.

Computer simulations save the day by providing a means to acquire exact results for a given model system. If the calculated properties of a model system do not agree with the experimentally observed properties, we know the model is inaccurate and we must improve the approximation of the intermolecular interactions. However, if we find

disagreement between a simulation and predictions from an approximate analytical theory, we know that the theory itself is flawed. Thus, the computer simulation can also be used as the experiment designed to test the theory. This has become so common, that it is now rare for a theory to be applied to the real world before being tested by computer simulation. [1]

The calculations described here are both types of computer simulation. In some calculations, computer simulations are used to test peptide conformational energies; others test the accuracy of a new protein forcefield.

Molecular Modeling

The majority of computer simulations performed in chemistry are some form of molecular modeling, usually in the categories of quantum mechanics, molecular mechanics and dynamics, or statistical dynamics. Each of these techniques relies on an approximation of known physical behavior which is then used to numerically calculate and predict the outcome of an experiment.

While chemistry typically conjures up the image of beakers and bottles all bubbling away in some laboratory, computational chemistry is now employed by many synthesis labs. One cannot deny that one of the greatest boons to the field is the ever-increasing speed of computation at an ever decreasing cost. *As the speed of computers continues to increase while the costs decrease, the question faced by a chemist will change from: "Can I do experiment X?" to "Can I do experiment X more cheaply, easily, accurately, and quickly using computational methods rather than traditional bench chemistry?"*[2]. The progress in computer technology may someday progress to the point

where the cost of a calculation can be measured not in CPU hours but in kilowatt hours, but the day of chemical simulations completely replacing lab experiments is still very far off. What computational chemistry can do today and in the near future is help provide insights to the experimentalists in investigating interesting problems, visualizing complex systems, and helping to identify the most promising experimental paths to pursue.

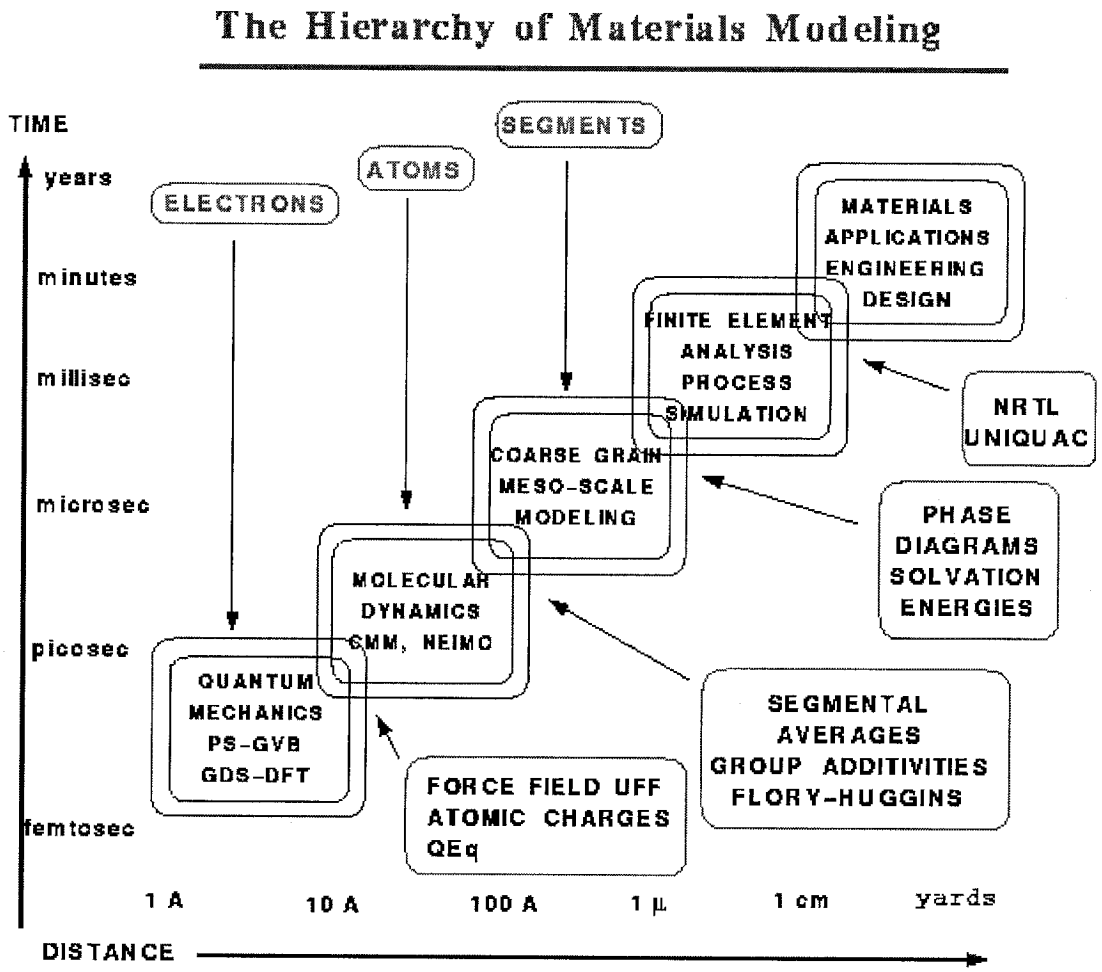

Figure 1- 1. A simulation that investigates properties that occur over long timescales or distances requires broader approximations to be made in order to remain computationally feasible. Biological simulations typically fall within the first two groups. (Figure courtesy of MSC.)

The nature of a chemical system and its properties of interest will dictate which computational tool should be applied. In the grossest sense, this can be summarized by selecting a method appropriate for a given distance and time scale. Figure 1-1 displays the connection between computational methods and increasing distance (basically sample size) and time scale. The lowest box represents methods with the fewest approximations. As we proceed up the hierarchy, successive methods require more approximations to be made in order to complete the computational experiment using a reasonable amount of time and resources. Current biological molecular simulation methods typically fall within the first two boxes of this simulation hierarchy.

Within the area of the first (lowest) box, quantum mechanical (QM) methods are used to calculate the interactions of electrons and nuclei up to the regime of tens of Angstroms and picoseconds. The area of the second box is the domain of molecular mechanics (MM) and molecular dynamics (MD). At this stage in the hierarchy, electrons and nuclei are usually represented by atoms and bonding schemes that behave in a classical dynamical manner.

Further up in the hierarchy lie simulation methods requiring even more gross approximations to maintain computational feasibility for systems operating on time or distance scales greater than 10^{-9} seconds or 100 Å. While many biological systems do fall beyond these limits, most current computational biochemical experiments involve systems that fall below or near this upper limit and methods designed for applications further up in the hierarchy will not be discussed.

Since biological systems are very complex, it is still not feasible to completely include all aspects of a system using quantum mechanics alone. However, because of

this complexity, it is also difficult to obtain accurate experimental data on which to build molecular models of these systems. By using accurate, robust quantum mechanical calculations of small system models, we can obtain reliable data. This data can then be used to build high quality atomistic models to be used in molecular mechanical calculations. By combining the strengths of each of the various simulation techniques, many important problems can be solved. The following sections discuss these simulation techniques in more detail.

Quantum Mechanics

In quantum mechanics, electrons are described by a wavefunction usually denoted Ψ. Any measurable quantity can be found by using an appropriate operator function acting on the wavefunction. One of the most important operators is the Hamiltonian, H, which is used to obtain the energy, E, of the system. This is demonstrated in equation (1.1), the Schrödinger equation.

$$H\Psi = E\Psi \tag{1.1}$$

Solutions to this equation are time independent wavefunctions, Ψ_n, that correspond to a stationary energy, denoted E_n. Allowed wavefunctions must be continuous functions and satisfy the Pauli principle.

The only Schrödinger equation that can be solved exactly is for one electron atoms like the hydrogen atom. Even other one electron problems like H_2^+ can only be solved if one makes the approximation that nuclear and electronic motions can be separated. This particular approximation is called the Born-Oppenheimer approximation and is only one of many further approximations needed in order to study systems of any

significant complexity. However, by making a series of good approximations, a molecular wavefunction, Ψ, can be constructed to sufficient accuracy to allow for calculation of observable properties with an acceptable degree of uncertainty.

After applying the Born-Oppenheimer approximation, we assume that each electron occupies its own molecular orbital. This will allow for the total molecular wavefunction to be expanded so that each function, ϕ_i, describes the orbital of a single electron, as shown in equation (1.2).

$$\Psi = \phi_1\phi_2\phi_3...\phi_n \tag{1.2}$$

The total wavefunction must still respect the Pauli principle and be antisymmetric with respect to electron exchange. To construct the individual ϕ_i orbitals, we can use a linear combination of known atomic orbital functions (1.3), which we could take, for instance, from solutions to the H atom problem.

$$\phi_i = \sum_k c_{ik}\chi_k \tag{1.3}$$

Here, c_{ik} are coefficients and χ_k is an atomic orbital function. The set of χ's is called a basis set. The problem of solving for Ψ becomes the problem of solving for the best set of c_{ik} coefficients in equation (1.3).

The Schrödinger equation (1.1) can also be applied to an individual molecular orbital, ϕ_i, by using a one-electron Hamiltonian (1.4) containing the interactions with the other electrons.

$$H\phi_i = \varepsilon_i\phi_i \tag{1.4}$$

By expanding the molecular orbital into the summation of the individual linear atomic orbitals as in (1.3) multiplying by a basis function χ_i, integrating over all space (1.5), and performing a small amount of algebra, we arrive at equation (1.6).

$$\sum_k c_{ik}\left(\int \chi_i H \chi_k dv\right) = \varepsilon_i \sum_k c_{ik}\left(\int \chi_i \chi_k dv\right) \tag{1.5}$$

$$\sum_k c_{ik}\left(\int \chi_i H \chi_k dv - \varepsilon_i \int \chi_i \chi_k dv\right) = 0 \tag{1.6}$$

Now we find the problem that plagues much of quantum mechanics. We can calculate a set of ε_{ik} by solving equation (1.6) for a given Hamiltonian, but because H in (1.4) depends on all the orbitals ϕ_i, it would seem we need to know the answer before we start to solve the problem. In practice, we can get around this problem by using an initial guess of the coefficients c_{ik}, using them to solve for the eigenvalues ε_i, and using this temporary set of ε_i to solve for new c_{ik} coefficients. We then take the new coefficients and plug them back into (1.4) and repeat the process until the c_{ik} coefficients converge to within a pre-selected limit.

The Hamiltonian operator chosen for much of quantum chemistry is the non relativistic Hartree-Fock (HF) self-consistent field operator. This operator includes a Coulombic term for the interaction of an electron with the average electron field along with an exchange term that has no classical equivalent. It is derived from a summation of terms of electrons with the same spin. While HF calculations are used extensively, they do have limitations. Even with a perfect selection of a complete basis set, a HF calculation will not arrive at the exact solution to the Schrödinger equation. It will instead reach what is called the HF limit.

This HF limit results from two approximations. The first assumption is that relativity does not affect the calculation. This is true for light molecules and most elements involved in biochemistry, but the electrons in the core of heavy atoms often approach the speed of light. HF calculations fail to accommodate the changes that result from core electrons approaching relativistic speeds. The second approximation, and a more drastic one, results from the electron-electron repulsion calculation. Since the electron repulsion of one electron is calculated with regard to the average field of all the other electrons, HF does not take into account the fact that the electrons' motion will be correlated. Simply put, if you have two electrons, they will be more likely to be found on opposite sides of a nuclei than on the same side.

The problems that result from this inexact solution are manifest even in the simple example of the H_2^- molecule. HF calculations arrive at the incorrect dissociation limit for H_2^-. All is not lost, however, because the HF method does perform accurate calculations for molecules near their optimum geometries. The method also does a fairly good job at calculating atomic properties like electrostatic potentials and dipole moments.

Extensions to HF calculations can improve some of the error arising from the assumptions inherent in the calculation, but they come at a computational cost. A frequent resolution is to use HF calculations to obtain quality geometries and then perform single energy calculations with a more rigorous method.

One commonly used, more rigorous method is Moller-Plesset second-order perturbation (MP2). Because MP2 calculations incorporate some of the effects of dynamic electron-electron interactions, conformational energies are calculated with much better results over local changes in bond angles and torsions. MP2 calculations begin

with the HF wavefunction but then perturb this wavefunction to second order to calculate a better energy of the system. One of the benefits of an MP2 calculation is that it is size invariant: the size of the system examined does not have an effect on the quality of the calculated energy.

MP2 calculations are computationally expensive, and, for calculations involving multiple molecules, some errors are introduced during the perturbation calculation which can be partially avoided by using the Local MP2 method. During the MP2 perturbation calculation, the excited electronic states for each pair of electrons is evaluated. Some of these states involve electron-electron interactions over large distances in the molecule. By only considering local excited states for any electron interaction, the cost is greatly reduced.

A discussion of quantum chemical methods is not complete without mention of basis sets. The ideal basis set (set of atomic orbitals) that each molecular orbital is expanded into are atomic orbitals of the form:

$$\chi_k = Ce^{-\varsigma r}Y_{lm} \tag{1.7}$$

where Y_{lm} is the angular component of the function and ζ is the orbital exponent. To ease the computational cost, gaussian functions are often fit to the atomic orbitals and are used instead. In the double zeta basis set, two sets of three gaussian functions are used to approximate (1.7) for each atomic orbital. One of the most common basis sets is denoted by 6-31G. This means that, for a first row atom, six gaussian functions are fit to the core 1s orbital. Each valence orbital is then represented by two functions, one that is a set of 3 gaussians and a second function that is a single gaussian function. Additional polarization functions may be added to the basis set and are indicated by an asterisk. A

6-31G* basis set would indicate all heavy atoms have additional polarization functions added while a 6-31G** basis set indicates additional polarization functions on both the heavy atoms and the hydrogen atoms.

Quantum mechanical calculations herein are usually geometry optimized using the HF 6-31G** basis set. Energy calculations reported here are usually carried out with LMP2/6-31G** calculations, sometimes after further geometry optimizations at the LMP2 level. The computational cost of QM calculations beyond the 10 to 100 atom range is very high. If the simulation of hundreds or thousands of atoms is required, a different approach must be used.

Molecular Mechanics

Quantum mechanics treats atomic nuclei as points and electrons as waves in order to calculate interesting molecular properties. If, however, one approximates atoms as soft spheres bonded to each other with springs, it is possible to model a system using only classical physics. Energies and forces derived from this approximation can be plugged into classical physics formulas to obtain dynamic trajectories or optimized geometries.

Forcefields

At the heart of any molecular mechanics calculation is the forcefield. It is the main set of approximations used to represent the molecular system examined. Once a quality force field is constructed for a system, the application of classical physical principles is enough to derive high quality information about the molecular system

studied. The force field is usually the limiting factor on the accuracy of a molecular mechanics calculation.

The total energy calculated by a force field for a molecular system can be broken down into two terms, a valence term and a nonbond term (1.8).

$$E_{tot} = E_{valence} + E_{nonbond} \tag{1.8}$$

The valence term can be further broken down into bond, angle, torsion, and inversion terms (1.9). Bond, angle, and inversion terms arise directly out of an examination of atomic and molecular orbitals. Torsion terms are not as easily justified using only molecular orbital theory, but are a required element in order for classical physics to correctly describe a molecular system.

$$E_{valence} = E_{bond} + E_{angle} + E_{torsion} + E_{inversion} \tag{1.9}$$

The simplest valence term is a two-body interaction of bonded atoms. The bond term is usually encountered in one of two forms. The simplest and most common is a harmonic bond potential (1.10). In this case, the bond is treated like a classical spring with a spring constant of K_b and an equilibrium length, R_0. This gives excellent results for all bond distances near equilibrium. This expression is also very economical to compute, making it the most commonly used bond term. At distances far from equilibrium, like breaking a chemical bond, the harmonic potential is incorrect. In cases where bond breaking needs to occur, a Morse potential for bonding is used instead (1.11). This allows the bond energy to go to zero for large R.

$$E_{harmonic} = \frac{1}{2} K_b \left(R - R_0\right)^2 \tag{1.10}$$

$$E_{morse} = \frac{1}{2} D_o \left(e^{-\alpha(R-R_0)} - 1 \right)^2 \tag{1.11}$$

The second basic valence term in a forcefield is an angle term. The most common angle term is again a harmonic potential (1.12). In this case, a spring constant is again used in a function that depends on a deviation from the optimal angle.

$$E_{angle} = \frac{1}{2} K_a \left(\theta - \theta_0 \right)^2 \tag{1.12}$$

Torsions are more complex than angle or bond potentials. The torsion potential is typically represented by up to six terms, each of which can have their own periodicity (1.13). The periodicity is determined by n, while d determines whether the torsion has a maximum at ϕ=0° or ϕ=180°.

$$E_{torsion} = \sum_{n=1}^{6} \frac{1}{2} K_{\phi,n} \left(1 - d \cos\left(n\phi \right) \right) \tag{1.13}$$

The most complex of the four common valence terms is the inversion potential. The inversion term is included to make sure that a given atom, i, will remain either planar or non-planar to three other atoms, j, k, and l. Two forms are commonly found. AMBER [3] uses equation (1.14) and insures planar geometries when n = 2 and a tetrahedral geometry when n = 3.

$$E_{inversion} = \frac{1}{2} K_\psi \cos\left[n\left(\psi - \psi_o \right) \right] \tag{1.14}$$

DREIDING [4] uses a simpler harmonic equation (1.15).

$$E_{inversion} = \frac{1}{2} C \left(\cos\phi - \cos\phi_o \right)^2, \quad \text{where } K_\phi = C \sin^2 \phi_0 \tag{1.15}$$

Generic forcefields have shown that quality geometries can be obtained with very simple values for valence spring constants and equilibrium positions [4, 5]. This is not

the case for the nonbond portion of a forcefield. The nonbond portion of a forcefield typically consists of three main parts (1.16): the electrostatic energy of charge-charge interactions, van der Waals interactions, and a special term to represent hydrogen bonding.

$$E_{nonbond} = E_{electrostatic} + E_{vdW} + E_{hbond} \tag{1.16}$$

The electrostatic energy can easily be calculated by evaluating the coulombic interaction between each pair of atoms in the system (1.17). Particularly in biological systems, the electrostatic contribution to the energy can be one of the most important for evaluating intermolecular interactions. This means that a quality force field must also contain a method to arrive at charges that accurately represent the true molecular system.

$$E_{elec} = \sum \frac{q_i q_j}{r_{ij}} \tag{1.17}$$

The van der Waals energy is also a pairwise interaction. There are many functional forms used to describe van der Waals interactions. The simplest one is the Lennard-Jones 6-12 potential (1.18) and it is used in many common forcefields [3, 6]. It requires only two parameters, a D_0 well depth and an equilibrium distance, R_0. It has one main drawback. For R less than R_0, it gives results that tend to be too high in energy. To put it another way, its "inner wall" is too "hard."

$$E_{LJ} = D_0\left[\left(\frac{R_0}{R}\right)^{12} - 2\left(\frac{R_0}{R}\right)^{6}\right] \tag{1.18}$$

Dreiding uses an exponential-6 potential (1.19). This function allows a softer inner wall, but it requires three parameters and for very small R, which are typically

found only in non-physical geometries, computational tricks must be used to prevent the function from becoming attractive again.

$$E_{\exp-6} = D_0 \left\{ \left[\left(\frac{6}{\xi - 6} \right) \exp^{\xi\left(1-\frac{R}{R_0}\right)} \right] - \left[\left(\frac{6}{\xi - 6} \right) \left(\frac{R_0}{R} \right)^6 \right] \right\} \tag{1.19}$$

A pure exponential function is occasionally used and can be thought of as a form of the exponential-6 potential that is repulsive for all R.

$$E_{pure\exp} = D_0 \exp^{\gamma\left(1-\frac{R}{R_0}\right)} \tag{1.20}$$

The Morse function (1.21) also has three parameters and allows a much softer inner wall than the Lennard-Jones 6-12 without the unrealistic features of the exponential-6 form for small R.

$$E_{morse} = D_0 \left\{ \left(\exp\left[\frac{-\gamma}{2} \left(\frac{R}{R_0} - 1 \right) \right] \right)^2 - 2 \left(\exp\left[\frac{-\gamma}{2} \left(\frac{R}{R_0} - 1 \right) \right] \right) \right\} \tag{1.21}$$

Because most forcefields have static charges, there is no ability for polarization to occur on a pair of atoms that might otherwise increase the interaction. As a result, an additional term for hydrogen bonds is often added to a forcefield. Amber [3] uses a Leonard-Jones 10-12 potential (1.22) similar to the 6-12 potential (1.18) used for van der Waals interactions. This 10-12 potential goes to zero much more quickly.

$$E_{Hbond\,12-10} = D_0 \left[5 \left(\frac{R_0}{R} \right)^{12} - 6 \left(\frac{R_0}{R} \right)^{10} \right] \tag{1.22}$$

Dreiding [4] uses a 10-12 potential for hydrogen bonds, but it also incorporates an angle dependence which is based on the angle between the acceptor atom, A, the donor

hydrogen, H, and the heavy atom connected to the hydrogen, D (1.23). This turns off the hydrogen bonding interaction for D-H …A interactions for inappropriate angles.

$$E_{Dreiding\ h-bond} = D_0\left[5\left(\frac{R_0}{R_{DA}}\right)^{12} - 6\left(\frac{R_0}{R_{DA}}\right)^{10}\right]\cos^2\theta_{DHA} \qquad (1.23)$$

Each forcefield function depends on two or more parameters. These parameters are typically chosen to fit or are at least tested against experimental data. Spectroscopic data can be fit well by adjusting valence terms while crystal structures and experiments on small molecular clusters can provide data useful for fitting nonbond parameters. In recent years, high quality quantum mechanics is also providing data with which to fit forcefield parameters.

A typical forcefield will break atom types down into element types and their hybridization. Parameters are then derived for each hybridization of each element of interest. Some forcefields, such as CHARMM [6], AMBER [3], or OPLS [7], are highly parameterized. This means that they have many different atom types, often several atom types for a particular element and hybridization. They use many atom types and all parameters are fit to known data. This often gives good results, but does not easily allow an application to new molecular systems. Since each parameter was derived with some dependence on other parameters, it is not easy to fit a few new parameters to a new system. Other forcefields, such as DREIDING [4] or UFF [5], are more generic. As much as possible, valence and nonbond parameters are generated from a simple metric that depends on only a few experimental numbers, such as electronegativity or atomic size. They often produce results similar to highly parameterized forcefields, but are easily extended to new molecular systems without requiring a new fit. When developing

a forcefield, one must always weigh the benefits of an improved fit by using a more highly parameterized forcefield to the utility of maintaining a much more generic, and therefore more easily extended forcefield.

Pseudoatoms

In order to reduce complexity, some early forcefields did not explicitly include all hydrogen atoms that existed in the system of interest. Each hydrogen removed had its mass added to its connecting atom to create a new implicit hydrogen atom. This implicit hydrogen model essentially removes all hydrogen vibrations within the system and reduces the total number of atom-atom interactions that need to be calculated. An atom in a simulation that is used to represent more than one atom is often called a pseudoatom. Advances in computing power now make the use of an implicit hydrogen model rare. The added cost of an explicit hydrogen model is worth the increased accuracy that it provides.

While implicit hydrogen models are now rarely used, pseudoatoms are still used in many areas of chemical simulation. One such situation is in the area of protein simulation. The regular structure of proteins provides an easy framework to reduce multiple atoms into a single pseudoatom. Every protein contains a sequence of amino acids. Each residue has several backbone atoms that are part of the main chain and at least one atom that comes off of the chain, referred to as the sidechain atoms. The simplest pseudoatom representation of a protein is to reduce all the atoms in each residue to a single pseudoatom. This is usually placed at the C_α coordinate for that residue, and is thus called a C_α model [8, 9]. More complex models can be constructed by adding

pseudoatoms that represent the sidechains as well. Since the first carbon of a sidechain is referred to as the C_β atom, these are called C_β models. C_β details of models vary and can contain a C_α pseudoatom or the main chain atoms can be explicitly present. Occasionally, pseudoatoms are only used for the longest of the amino acid sidechains any may only represent a few of the outer-most atoms.

Pseudoatoms are particularly well suited for coarse-grained searching of conformational space. However, when local geometry and energies are at a high priority, an all atom forcefield is almost always desired.

Minimization Techniques

A summation of all the atomic interactions in a system using a forcefield will give you a numerical value for the energy of the system. However, if there are slight perturbations of just a few atoms into regions disfavored by the forcefield, the total energy of the system will be dominated by those few atoms with "poor" interaction energies. For example, a protein structure may have an excellent geometric conformation for all its atoms except for two hydrogens that are too close to each other. This will result in a extremely high total energy for the system, due to the high energy associated with van der Waals energy evaluated for the pair of hydrogen. An energy minimization procedure can be used to avoid this situation.

Energy minimization is typically performed by perturbing atoms in order to reduce the net force applied to them by the forcefield potentials; known as applying a gradient optimization. Since a minimized structure usually has a decent geometry and

rarely has large forces on any atom, it is preferred to start a molecular dynamics simulation with a minimized structure.

Energy minimization can be performed in Cartesian coordinates by optimizing in $3n$-dimensional space where n is the number of particles in the system. The path chosen is the gradient, ∇, where:

$$\nabla_x = \frac{\partial V}{\partial x} \tag{1.24}$$

Each Cartesian component, x, of the gradient is the derivative of the potential energy of the forcefield with respect to that component. Only interactions involving a particle i contribute to its own gradient, (x_i, y_i, z_i). The $3n$ components of ∇ form a path in $3n$ space. Two points along this pathway are interpolated to find a minimum on the path. The $3n$ components are reexamined at the new minimum. Usually, the gradient is still non-zero so a new path is constructed and a new step of minimization is begun. The path followed at each step can be along the gradient, ∇, but it is more efficient to choose a gradient that is orthogonal to all previous paths. This is referred to as the conjugate gradient minimization procedure and is one of the most popular methods of minimization used [10].

The conjugate gradient minimization method will not always arrive at the lowest possible conformation of the system, the global minimum. In fact, it is extremely unlikely for a conjugate gradient minimization of a large system to arrive at a global minimum. A local minimum is usually found. The energy at a local minimum is lower than the energy of any nearby conformation, but there may be other local minimums that are lower in energy that are distant in conformational space.

Molecular Dynamics

Molecules in the real world are not static. They are constantly fluctuating and changing conformation to respond to external environmental fluctuations. Molecular dynamics is the simulation of these moving molecules and permits the study of time-dependent processes. Two applications of molecular dynamics are particularly important: conformational sampling and the formation of thermodynamic ensembles.

Minimization procedures can take a specific conformation and lower its energy into a local minimum. Since molecular dynamics actually imparts kinetic energy into the system, the system can be excited into a higher kinetic energy state that allows the system to cross over a local barrier. Molecular dynamics, coupled with conjugate gradient minimization, forms a procedure called simulated annealing. Dynamics are performed at a relatively high simulation temperature for a time and are followed with minimization. When repeated for several cycles, simulated annealing can find local conformational minima that are lower in energy than the initial minimum found solely with conjugate gradient methods.

A simple way to create a thermodynamic ensemble is to maintain a constant total energy, volume, and particles to produce a microcanonical ensemble of conformations. Once an ensemble is formed, relative free energies, average densities, and other thermodynamic properties can be calculated.

Molecular dynamics calculations evaluate the forces on a particle and use these forces to determine the particle's acceleration. A particle's initial velocity is usually determined by a random distribution according to the Maxwell-Boltzmann distribution for the given simulation temperature. Once an initial velocity is chosen, it is updated

using the calculated accelerations. Most molecular dynamics methods use Cartesian coordinates, resulting in $3n$ degrees of freedom for systems of n particles. The forces, velocities, and accelerations applied to a specific particle are determined independently for each degree of freedom. The single exception is the common practice of subtracting out translations and rotations that affect the entire system. Since each Cartesian degree of freedom is uncoupled from all others, the force component along the x-axis for a specific particle can be calculated independently. The total force F_x in the x direction is the opposite of the gradient (1.25)

$$F_x = -\frac{\partial V}{\partial x} \tag{1.25}$$

Newton's equation of motion, (1.26), is then used to determine the accelerations where m_i is the mass of particle i.

$$\ddot{x} = -\frac{F_x}{m_i} \tag{1.26}$$

Velocities could ideally be updated from the accelerations by an analytical integration of the equations of motion as in equation (1.27), where $v_x^{t_1} = \dot{x}^{t_1}$ is the x-component of the velocity vector at time t_1.

$$v_x^{t_2} = v_x^{t_1} + \int_{t_1}^{t_2} \ddot{x} dt \tag{1.27}$$

However, an analytical solution for the accelerations would be quite complex for any but the simplest of systems, and we must use a numerical solution instead. Of the many common methods used to perform numerical integration [11], most have been used in molecular dynamics.

One of the most popular numerical solutions is the Verlet algorithm [12, 13]. It has several different formulations [1], one of the most popular being the "leapfrog formulation." The Verlet leapfrog algorithm gets its name from its method of updating the velocities and coordinates at half-timestep intervals, one after the other. Most methods of performing molecular simulations divide the simulation into timesteps, h, which are shorter than the periodicity of the fastest motions in the system. A typical timestep used is one femptosecond (1×10^{-15} s), which is shorter than the period of O-H and N-H bond stretches. In the leapfrog Verlet Algorithm, the velocities at timestep $n+½$ are obtained from the previous velocities and the current accelerations:

$$v_x^{n+\frac{1}{2}} = v_x^{n-\frac{1}{2}} + h\ddot{x}^n \tag{1.28}$$

The new velocities, $v_x^{n+\frac{1}{2}}$, are then used to update the coordinates for timestep $n+1$:

$$x^{n+1} = x^n + hv_x^{n+\frac{1}{2}} \tag{1.29}$$

The new coordinates are then used as input back into equation (1.25) and the dynamics continues on into the next timestep. The simulation can then be continued for a predetermined number of timesteps or until a system property reaches a specified value.

Monte Carlo

First coined by Metropolis and Ulam [14], Monte Carlo methods get their name from the games of chance in the gambling halls of Monaco. The very first computer simulation of a liquid was carried out using the Metropolis Monte Carlo method in the early 1950s [15]. By the end of the decade, Monte Carlo methods were being used for molecular dynamics simulations as well. While the methodology of Monte Carlo

simulations have changed, the basic algorithms used today are much the same as they were in the 1950s.

While molecular dynamics simulations are driven by the physical properties of the system (e.g., coordinates or interatomic forces), Monte Carlo simulations use random numbers to generate a sample population from which properties are then determined. Because of this, Monte Carlo simulations are widely employed in the study of disordered systems like gases and fluids.

The Metropolis Monte Carlo method [15] calculates a molecular property F from a canonical ensemble using equation (1.30).

$$F = \frac{\int F e^{-E/k_b T} dqdp}{\int e^{-E/k_b T} dqdp} \tag{1.30}$$

Here, k_B is the Boltzmann constant, T is the system temperature, and $dqdp$ is integrated over the volume. This integral is typically approximated by producing a large number of sample configurations. Equation (1.31) demonstrates the calculation for a system of N_c sample configurations.

$$F = \frac{\sum_{c=1}^{N_c} F_c e^{-E_c/k_b T}}{\sum_{c=1}^{N_c} e^{-E_c/k_b T}} \tag{1.31}$$

A configuration is generated and then weighted by $\exp(^{-E_c}/_{k_B T})$ to form the canonical ensemble. This leads to inefficiency because many configurations that are generated have high energies and thus have very low weighting factors. The Metropolis version of Monte Carlo avoids this problem. The key is to generate configurations

according to the probability $\exp(^{-E_c}/_{k_BT})$ and weight all generated conformations equally. The simplest way to accomplish this is to perturb the previously generated conformation a small amount to generate a new conformation. This new conformation is only kept if a generated random number n is less than $\exp(^{-E_c}/_{k_BT})$. If so, the new conformation is kept and entered into the ensemble. If not, the new conformation is discarded and a different perturbation is made.

Monte Carlo methods are excellent at coarse grained sampling of conformational space [16] as well as simulating conformational changes which cannot be simulated by molecular dynamics [17]. Monte Carlo methods make a nice complement to molecular dynamics and minimization. For example, coarse-grained Monte Carlo can be used to generate a diverse set of conformations. Molecular dynamics and minimization can then be performed to find the local minima of the conformations.

References

1. Frenkel, D. and B. Smit, *Understanding Molecular Simulation.* 1996, San Diego: Academic Press.
2. Brameld, K.A., *Molecular Modeling of Biological Systems: From Chitinase A to Z-DNA*, in *Chemistry*. 1999, California Institute of Technology: Pasadena.
3. Weiner, S.J., *et al.*, *A New Force-Field For Molecular Mechanical Simulation of Nucleic-Acids and Proteins.* Journal of the American Chemical Society, 1984. **106**(3): p. 765-784.
4. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding - a Generic Force-Field For Molecular Simulations.* Journal of Physical Chemistry, 1990. **94**(26): p. 8897-8909.
5. Rappe, A.K., *et al.*, *Uff, a Full Periodic-Table Force-Field For Molecular Mechanics and Molecular-Dynamics Simulations.* Journal of the American Chemical Society, 1992. **114**(25): p. 10024-10035.
6. Brooks, B.R., *et al.*, *Charmm - a Program For Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
7. Jorgensen, W.L. and J. Tirado-Rives, *Development of the OPLS-AA force field for organic and biomolecular systems.* Abstracts of Papers of the American Chemical Society, 1998. **216**: p. U696-U696.
8. Levitt, M., J. Proc. Natl. Acad. Sci. U.S.A., 1976. **104**: p. 59.
9. Friedrichs, M.S. and P.G. Wolynes, Science, 1989. **246**: p. 371-373.
10. *BIOGRAF Reference Manual*, 1992, Molecular Simulations, Inc.
11. Flannery, B.P., S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes.* 1989, Cambridge: Cambridge University Press.
12. Rahman, A., Phys. Fev., 1964. **136**: p. A405.
13. Verlet, L., Phys. Rev., 1967. **159**: p. 98.
14. Metropolis, N., *The Beginning of the Monte Carlo Method.* Los Alamos Science, 1987. **12**: p. 125-130.
15. Metropolis, N., *et al.*, *Equation of state calculations by fast computing machines.* J. Chem. Phys., 1953. **21**: p. 1087-1092.
16. Debe, D., M. Carlson, and W. Goddard, *The topomer-sampling model of protein folding.* Proc. Nat. Acad. Sci. U.S.A., 1999. **96**(6): p. 2596-2601.
17. Mathiowetz, A., *Dynamic and Stochastic Protein Simulations: From Peptides to Viruses*, in *Chemistry*. 1993, California Institute of Technology: Pasadena.

Chapter 2: Development of a Biological Universal Forcefield (BUFF)

Abstract

A new biological universal forcefield, BUFF, is presented for the simulation of proteins and other biological molecules. Built upon the rule based generic forcefields UFF and Dreiding, the BUFF has additional hydrogen bond and protein backbone torsion terms. A set of charges for common amino acids are also provided. These additional parameters have been fit to *ab initio* quantum mechanical calculations carried out on model systems. Validation studies of peptide trimers demonstrate that the BUFF accurately reproduces the quantum mechanical torsional energies. Several other common, highly parameterized forcefields are also applied to the same tripeptide systems, as well as short α-helical chains and other model systems in order to make a comparison to the BUFF. These studies show that while the BUFF is universal and can be quickly deployed on new systems, such as unnatural amino acids or metal containing systems, it is also at least as accurate as other commonly employed, but highly parameterized, forcefields.

Introduction

The development of forcefields for biological simulations such as protein or nucleic acid macromolecules provides several challenges. The functions of these molecules are often intimately tied with their local structure. It becomes necessary for a forcefield to accurately predict both structural geometries as well as energies. To complicate matters, the electrostatic interactions of many systems are of great interest in fields such as ligand binding and membrane proteins. Current biological forcefields such as Amber94 [1], CHARMM [2], and OPLS-AA [3] attempt to solve this problem by creating specialized parameters for many of the atoms in each amino acid type. Not only are the valence and van der Waals terms optimized, but the electrostatic charge on each atom is optimized along with the other terms. The result of all this parameterization is a loss of generality. Using OPLS-AA as an example: "*If additional parameters are developed by others, it is recommended to use the same procedures, particularly 6-31G* energetics, as a basis for torsional parameters and validation of nonbonded parameters through computations of pure liquid properties and/or free energies of hydration.*[3]" In other words, to study a molecular system with a few unique modifications such as a protein containing an unnatural amino acid, in order to be consistent with the forcefield, a large series of quantum mechanical calculations must first be performed in order to derive the new forcefield charges and other parameters.

Rarely do simulations involve pure nucleic acids or proteins. Typically, systems of interest include modifications such as unnatural amino acids, metals, small ligands

(drug molecules or cofactors), or polymer scaffolds. It is desirable for a forcefield to be general enough to handle such additional molecules self-consistently with those parameters already developed specifically for nucleic acids or proteins. The final goal is parameterization of a biological forcefield (BUFF) from first principles that maintains the flexibility of a generic forcefield. BUFF uses a generic forcefield (UFF [4]) for valence terms along with another general forcefield (Drieding [5]) for nonbond terms. Additional forcefield terms such as hydrogen bonding and charges are added which are specific to biomolecules. A similar, compatible generic forcefield for nucleic acids has already been developed (MSCFF [6, 7]), so the part of this forcefield that covers proteins is presented herein. These new forcefield terms are derived from high level *ab initio* quantum mechanical (QM) calculations of small peptides and other molecular clusters which accurately represent the relevant potential energy surfaces present in a typical protein.

The model clusters and peptides are chosen to create a set of parameters that will correctly reproduce the behavior of the fundamental units of proteins, amino acids. This chapter presents the complete BUFF that includes: a charge scheme for each standard amino acid type developed from a rule-based procedure, a set of high quality hydrogen bond potentials for each hydrogen bonding type found within the standard set of amino acids, and a set of specialized torsions used for glycine and any amino acid type that contains a carbon C_β atom.

A set of validation studies is also performed on tripeptide systems, short helical peptides, and a number of systems containing a metal. Comparisons to several common biological forcefields are made as well.

Methods

Most *ab initio* QM calculations were carried out using the Jaguar 4.0 (or earlier) software package [8]. The Biograf [9] software package was used for most molecular mechanics and dynamics simulations. A listing of forcefield parameters is provided in Appendix A, while atom types and charges are listed in the form of a PDB protein conversion table in Appendix B. When solvation was included, a Poisson-Boltzmann continuum model [10] was used.

The Biological Universal Forcefield (BUFF) uses the following valence energy terms:

$$E_{valence} = E_{bond} + E_{angle} + E_{torsion} + E_{inversion} \qquad (2.1)$$

All valence terms for BUFF are originally derived from UFF [4], with additional torsion parameters used for amino acid backbone torsions. In order to uniquely distinguish these new torsions, a new atom types of C_A was added as the C_α atom in the protein backbone. For all parameters, except for torsions, C_A is the equivalent of C_3.

Because UFF is a rule based forcefield with valence force constants which vary as a result of bond orders determined from electronegativities, it is possible for a parameter involving the same set of atom types to have a slightly different force constant. In order to allow users who may not have access to the UFF forcefield generator to still utilize BUFF, the force constants for common atom types are averaged and reported in Appendix A. This is the parameter set for which all benchmarks are carried out. If the UFF generator is used, similar, but not identical results should be expected.

Van der Waals (vdW) interactions in BUFF use the exponential-6 implementation of the Dreiding [5] forcefield which was derived empirically from small molecule crystal

structures. For the small set of nonstandard elements that do not have van der Waals parameters within the Dreiding forcefield, UFF van der Waals parameters are used. A Morse potential is used for hydrogen bonding, as described in detail in the parameterization section. Standard coulombic potentials are used to find the energy of charge-charge interactions. Charges were derived from the electron density distribution (constrained to reproduce the molecular monopole and dipole moments) calculated from the converged wavefunction of small model systems [11]. This process is described in more detail in the next section.

By using a generic forcefield where charges are not parameterized along with nonbond forces, new atom and molecule types can be modeled without any additional parameterization. When an unnatural amino acid or metal center is involved in a simulation, it is only necessary to derive a set of charges before the simulation can begin. In many highly parameterized forcefields, [1-3], the valence and van der Waals parameters must also be tuned before the simulation can begin. This can present a prohibitive computational cost if simulations are planned for a large number of nonstandard amino acids or metals.

Parameterization

The basic strategy employed to develop BUFF begins with two generic forcefields, UFF [4] and Dreiding [5], which are then tuned to reproduce QM energies for small model systems chosen to span the diverse space of protein molecular interactions. The tuning focuses on developing a universal scheme that is quickly applied to any new

system. Thus, only a few important terms are parameterized, leaving the rest of the parameter set to be generated using a straightforward methodology like UFF.

The parameterization of BUFF involved 4 steps. The first step was to combine the UFF valence terms with Dreiding van der Waals forces as has been discussed. The next step involved deriving a charge model for each amino acid type for common pH levels. Since hydrogen bonding plays a significant role in many protein folds and protein-ligand interactions, the third step involved developing special hydrogen bonding terms. This is particularly important in a forcefield with static point charges. Since the charges are not free to polarize on an atom, hydrogen bonds are poorly reproduced unless treated explicitly. The final step in developing BUFF was to optimize the torsions along the protein backbone to reproduce the HF quantum mechanical energies [12].

The Charge Scheme

Electrostatic interactions have an important role in many biological processes, so it is necessary to have a high quality set of point charges in order to perform accurate biological simulations. A set of charges were derived for each standard amino acid type. Charges were calculated using a model tripeptide system containing the central amino acid of interest, capped at both ends by a methylated glycine as in Figure 3-1. Each model system was minimized at the HF/6-31G** level with solvation, and charges were calculated from the converged wavefunction using an electrostatic potential derived from the electron density distribution [11]. The charges were constrained to reproduce the molecular monopole and dipole moments of the molecule. The net charge on the central residue of interest was set to 0, 1, or –1 by small adjustments to the heavy atoms with

corrections to attached hydrogen to preserve the net dipole of the heavy atom and its hydrogen. The final charges used for the standard amino acid types in H_2O can be found in Appendix B. Additional calculations were performed to create a set of charges for residues in a vacuum or hydrophobic (hexane dielectric) environment, but have not been compiled.

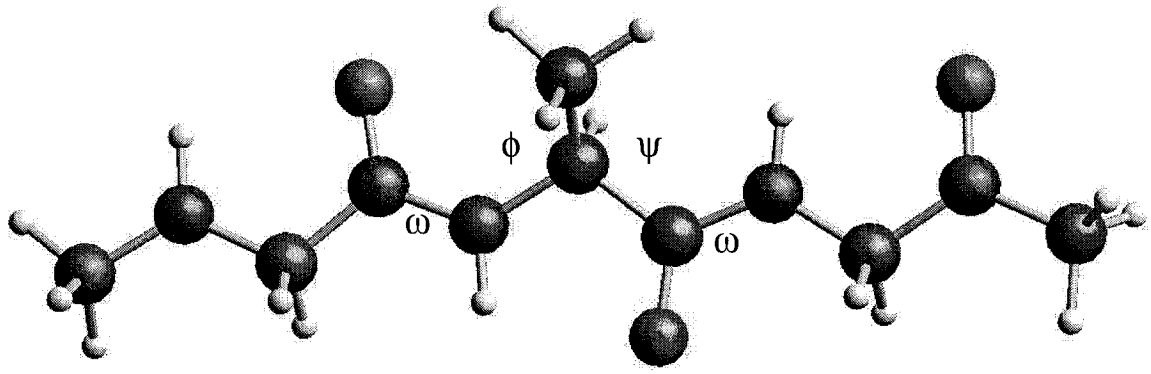

Figure 2- 2. The tripeptide model system for Gly-Ala-Gly calculations. Gly-XXX-Gly tripeptides of this type were used to derive charges for each amino acid type. The central residue was tuned to the appropriate integer charge for each amino acid type. ω is typically planar due to resonance, so its torsion parameters were not optimized. Shown is ϕ= 180° and ψ= 180°.

Since all charges are calculated *ab initio* it is simple to incorporate new amino acid types. A single QM calculation can be performed to obtain new charges. Alternatively, a fast method like charge equilibration [13] could be used for a large system. This is only possible because the charges are not parameterized along with nonbond or valence forces.

The charge scheme described thus far works well for all residues in the protein chain except for residues at the endpoints. Special consideration for residues that begin and end a peptide chain must be made if the simulation is to correctly model end of chain effects. The BUFF charge model has three methods for terminating a peptide chain.

The first method is to leave the peptide chain as a Zwitterion. The N-terminus has a positive charge due to the 3 hydrogen on the starting nitrogen. The C-terminus has a net negative charge due to the extra oxygen. A QM calculation at the HF/6-31G** level was performed on two tetramers. Charges on extended chains of Gly-Ala-Ala-Gly and Ala-Ala-Ala-Ala peptides were calculated using electrostatic potential (ESP) fitting. The net charge on each residue of both tetramers is shown in Table 2-1.

	QM	BUFF			QM	BUFF
GLY1	1.071	1.000		ALA1	1.067	0.970
ALA2	0.006	0.000		ALA2	-0.006	0.000
ALA3	-0.114	0.000		ALA3	-0.024	0.000
GLY4	-0.963	-1.000		ALA4	-1.037	-0.970
Total Charge	0.000	0.000		Total Charge	0.000	0.000

Table 2- 1: HF/6-31G** ESP calculated charges for two tetramers and charges used by BUFF when the Zwitterion endpoint charge model is applied.

The QM calculations show that the first and last residue in the peptide chain are nearly +1 and –1 respectively. Thus, if we adjust the charges only on the first and last residues in a chain, we should have a good approximation of the Zwitterion. An examination of the charges on each atom arrives at the following generic scheme for appropriately adjusting endpoint charges: 1) Extra hydrogen is added to the N-terminus and an additional oxygen is added to the C-terminus at the appropriate charge level for that residue type. 2) An additional 0.16 charge is subtracted from the C-terminus oxygen and 0.32 is added to the N-terminus nitrogen in order to increase the net charge at each endpoint. 3) Finally, since the hydrogen and oxygen charges are different for each residue type, the hydrogen on the N-terminus nitrogen are adjusted to give a net neutral charge to the system. Table 2-1 shows the final charge distribution when this procedure is followed, resulting in only a 0.06 difference between QM and BUFF.

It is sometimes desirable to avoid Zwitterion effects when studying a small peptide system. The second method of capping the termini uses a methyl group at both ends to allow the two termini to have nearly a neutral charge. This was how the tripeptide system used in the charge calculations was capped. (See Figure 2-1.) This provided many tripeptide systems that could be used in developing a methylation charge scheme. In order to preserve the charge scheme for as much of the protein as possible, a scheme for the methyl end groups was derived that only involved the methyl groups themselves.

The charges for the 18 N-termini methyl groups used in the charge fitting are listed in Table 2-2. The charge on all hydrogen and carbon are averaged shown in Table 2-3. The final BUFF charge scheme for methylated N-terminus is to place a charge of +0.110 on the carbon and +0.025 on each hydrogen atom.

	gag	gcg	gfg	ggg	ghg	ghg	gig	glg
C	0.149	0.144	0.103	0.145	0.186	0.141	0.114	0.121
H	0.012	0.017	0.039	0.018	-0.003	0.031	0.039	0.038
H	0.043	0.036	0.056	0.039	0.035	0.045	0.054	0.050
H	0.033	0.042	0.025	0.035	0.025	0.016	0.019	0.020
	gmg	gng	gqg	gsg	gtg	gvg	gwg	gyg
C	0.141	0.013	0.107	0.134	0.123	0.124	0.047	0.098
H	0.035	0.064	0.041	0.037	0.041	0.039	0.051	0.041
H	0.044	0.069	0.051	0.043	0.047	0.048	0.063	0.055
H	0.015	0.049	0.026	0.020	0.022	0.019	0.036	0.026

Table 2- 2: N-terminus methyl charges calculated from HF/6-31G** QM on methylated Gly-XXX-Gly tripeptide systems. Boxes colored in gray indicate charges that deviate from the BUFF methyl charge scheme by greater than 0.05.

Atom	Average			BUFF
C1	0.114		Carbon	0.110
H2	0.036		Hydrogen	0.025
H3	0.050			
H4	0.027			
All H Atoms:	0.038		Total:	0.185

Table 2- 3: Average QM charges and BUFF final charges for the N-terminus methyl.

The charges for 17 C-terminus methyl groups used in the charge fitting are listed in Table 2-4. The charge on all hydrogen and carbon are averaged shown in Table 2-5. The final BUFF charge scheme for methylated C-terminus is to place a charge of –0.590 on the carbon and +0.135 on each hydrogen atom.

	gag	gcg	gfg	ggg	ghg	gig	glg	gmg
C	-0.574	-0.578	-0.566	-0.596	-0.508	-0.584	-0.566	-0.567
H	0.171	0.170	0.165	0.166	0.158	0.164	0.169	0.171
H	0.160	0.168	0.167	0.175	0.131	0.172	0.164	0.161
H	0.162	0.162	0.160	0.168	0.154	0.165	0.159	0.159

	gng	gqg	gsgq	gtg	gvg	gwg	gyg
C	-0.583	-0.576	-0.575	-0.536	-0.572	-0.568	-0.563
H	0.164	0.170	0.160	0.162	0.166	0.166	0.165
H	0.172	0.165	0.172	0.145	0.169	0.166	0.166
H	0.165	0.161	0.162	0.155	0.159	0.159	0.158

Table 2- 4: C-terminus methyl charges calculated from HF/6-31G** QM on methylated Gly-XXX-Gly tripeptide systems. Boxes colored in gray indicate charges that deviate from the BUFF methyl charge scheme by greater than 0.05.

Atom	Average			BUFF
C1	-0.567		Carbon	-0.590
H2	0.166		Hydrogen	0.135
H3	0.164			
H4	0.161			
H Atom Avg:	0.163		Total:	-0.185

Table 2- 5: Average QM charges and BUFF final charges for the C-terminus methyl.

The final BUFF methylated charge scheme results in a small Zwitterion of +0.185 on the N-terminus and –0.185 on the C-terminus. While not a true neutral system at each

endpoint, the charges are significantly smaller than their charged, standard Zwitterion counterparts. One of the benefits of this scheme is that it allows the first and last residues in a chain to preserve their standard charges. If the N and C termini were to be truly neutral, adjustments in the charges of the first and last residue would also have to be made.

A final option for protein chain termini is the "½ glycine model." For some studies of small peptides, it is often advantageous to be able to calculate ϕ and ψ for both the first and the last residues. The N-terminus has an additional methylated C=O group while the C-terminus has a methylated NH. In this case, the glycine equivalent charges are used with the methyl groups having the same charge as a C_α glycine atom. Charges are equilibrated over all additional hydrogen atoms. A minor correction for the final small, non-zero charge can be made if a periodic calculation is desired, but is otherwise not necessary.

Charges have been calculated from QM for all standard residues and a systematic scheme for residues that begin and end protein chains has been provided. This system permits the treatment of new systems without a costly parameterization of a new charge set. Whenever a new system is encountered, *ab initio* charges can be calculated for the system without the need for further parameterization.

Hydrogen bond potentials

Hydrogen bonds play a key role in maintaining structure and specificity of biological systems. In order to improve intermolecular interactions in BUFF, an explicit hydrogen bonding term has been developed for most hydrogen bond types. The common hydrogen bonding terms that have been fit are: sp^3 O, sp^3 N, sp^2 N, and sp^3 S hydrogen donors and sp^3 O, sp^2 O, and sp^2 N acceptors. Initial investigations found that sp^3 S was not a good hydrogen bond acceptor so it has no special hydrogen bonding acceptor term. Donor and acceptor types that were highly charged exhibited different properties than more neutral donors and acceptors, so some additional types were included to differentiate between the neutral and charged forms.

Hydrogen bond terms were derived from LMP2/6-31G** *ab initio* calculations in vacuum of small model systems (Figures 2-2 through 2-8). Each hydrogen bond donor/acceptor pair was geometry optimized at the LMP2/6-31G** level. The two model fragments were then held rigid and were expanded along the hydrogen bonding axis. "Snap bond" energies were calculated at regular intervals using LMP2/6-31G** single point energies. Charges for each of the model fragments were taken from HF/6-31G** ESP calculations of the isolated molecule in a vacuum. These charges were then used with Dreiding van der Waals terms in the BUFF fit. Each dimer interaction analyzed in the *ab initio* calculation was analyzed using BUFF with the van der Waals term between the hydrogen donor atom and the acceptor atom set to zero. The difference between BUFF and the *ab initio* energies was fit to a Morse term for each hydrogen bonding

interaction. A Morse term (equation (2.2)) was used for the functional form. Morse terms have an additional parameter, γ, which allows the curvature of the function to be fit as well as the well depth, D_0, and the equilibrium distance, R_0. This additional parameter allows the potential to have a softer inner wall than is typically found with Lennard-Jones type potentials.

$$E_{morse} = D_0 \left\{ \left(\exp\left[\frac{-\gamma}{2} \left(\frac{R}{R_0} - 1 \right) \right] \right)^2 - 2 \left(\exp\left[\frac{-\gamma}{2} \left(\frac{R}{R_0} - 1 \right) \right] \right) \right\} \qquad (2.2)$$

The fit for each donor/acceptor interaction is shown in Table 2-6. Appendix A contains the final averaged forcefield parameters for each hydrogen bond type.

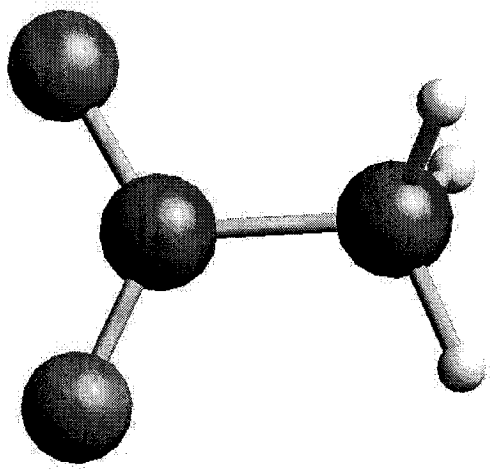

Figure 2- 3: $CH_3CO_2^-$ model fragment used to determine sp^2 O^- hydrogen bonding acceptors.

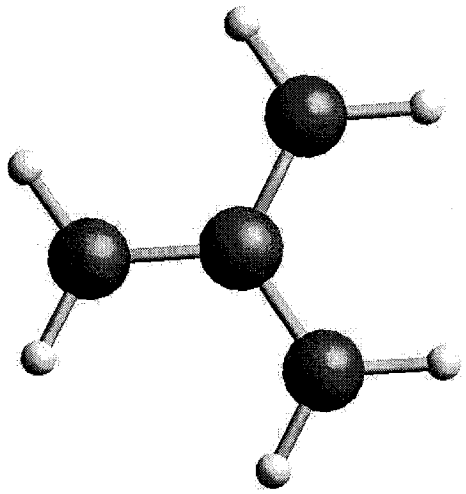

Figure 2- 4: $C(NH)_3^+$ arganine model fragment used to determine sp^2 N^+ hydrogen bonding donors.

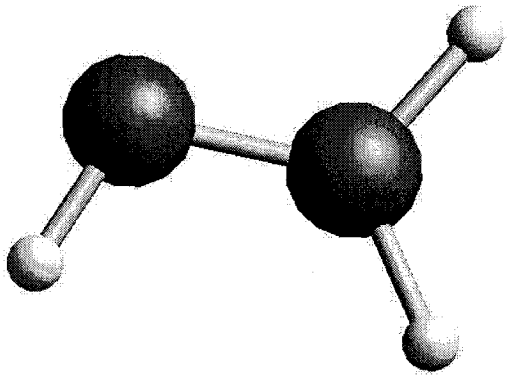

Figure 2- 5: CH_2NH model fragment used to determine sp^2 N hydrogen bonding donors and acceptors.

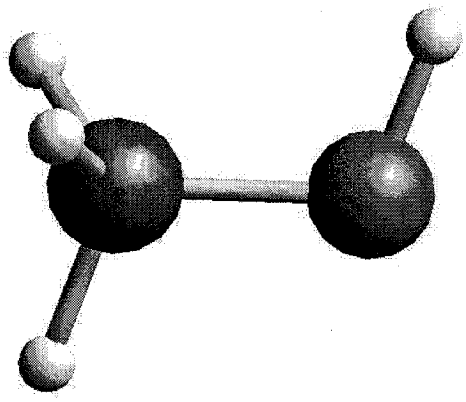

Figure 2- 7: CH_3OH model fragment used to determine sp^3 O hydrogen bonding donors and acceptors.

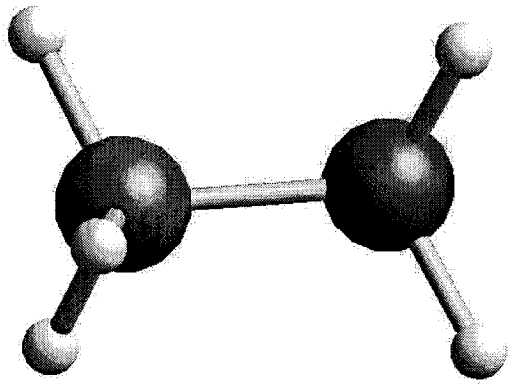

Figure 2- 6: $CH_3NH_3^+$ model fragment used to determine sp^3 N^+ hydrogen bonding donors.

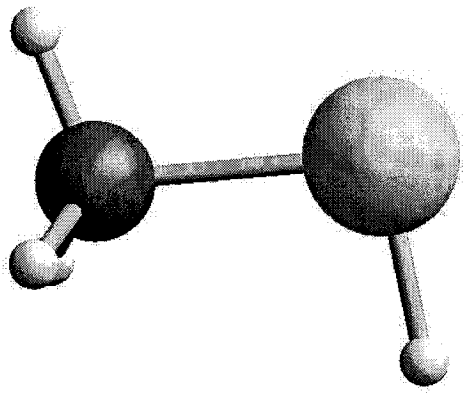

Figure 2- 8: CH_3SH model fragment used to determine sp^3 S hydrogen bonding donors.

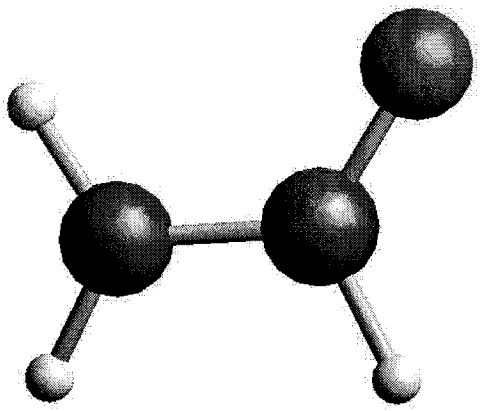

Figure 2- 9: Formamide model fragment used to determine sp^2 O hydrogen bonding acceptors, and sp^2 N hydrogen bonding donors and acceptors.

		sp3 O		sp2 O				sp2 N	
		CH_3OH		formamide		CH_3CO_2-		CH_2NH	
sp3 O	CH_3OH	Ro	OH: 2.0 HH: 3.5	Ro	2.2	Ro	2.29	Ro	2.37
		Do	OH: 1.5 HH: 0.2	Do	0.834	Do	2.900	Do	1.860
		gamma	OH: 10.6 HH: 9.76	gamma	8.86	gamma	6.86	gamma	7.51
sp3 N	CH_3NH_3+	Ro	3.16	Ro	2.00	Ro	2.36	Ro	2.20
		Do	0.100	Do	2.90	Do	3.450	Do	4.250
		gamma	8	gamma	6.60	gamma	5.39	gamma	5.70
sp2 N	formamide	Ro	2.63	Ro	2.58	Ro	2.34	Ro*	3.73
		Do	0.291	Do	0.186	Do	2.41	Do*	1.35
		gamma	6.77	gamma	10.00	gamma	6.27	gamma*	5.27
	CH_2NH	Ro	2.53	Ro	5.90	Ro	2.06	Ro	2.83
		Do	0.852	Do	0.023	Do	6.16	Do	0.446
		gamma	8.54	gamma	8.79	gamma	6	gamma	8.27
	$C(NH_2)_3$+	Ro	2.5	Ro	3.66	Ro	2.09	Ro	3.24
		Do	3.51	Do	0.214	Do	3.68	Do	0.610
		gamma	5.84	gamma	7.95	gamma	6.22	gamma	7.46
sp3 S	CH_3SH	Ro	2.52	Ro	3.07	Ro	1.80	Ro	2.44
		Do	0.760	Do	0.077	Do	8.440	Do	3.310
		gamma	8.26	gamma	10.63	gamma	4.62	Gamma	8.00

Table 2- 6: Morse parameters for hydrogen bonding of the BUFF. Grayed boxes are parameters that use a pure exponential function. The sp^3 O - sp^3 O hydrogen bonding interaction required adjustment of the hydrogen –hydrogen term to accommodate the difference between the C_s type and C_{2v} type interaction symmetries. The parameters for this exponential-6 function are listed alongside the Morse terms for the sp^3 O:::H-sp^3 O interaction.

As can be seen in the following charts (Figures 2-9 through 2-15), a Morse potential can almost always be used to reproduce the QM interaction energies. Each fit focused on fitting the bottom of the potential well, but longer interactions are almost always reproduced as well. Two terms were found to have almost no attractive forces beyond the electrostatic interactions, so they were fit with a pure exponential potential. These two interactions are shown in gray on Table 2-6. Equation (2.3) shows a pure exponential function.

$$E_{pure\,\exp} = D_0 \exp^{\gamma\left(1-\frac{R}{R_0}\right)} \tag{2.3}$$

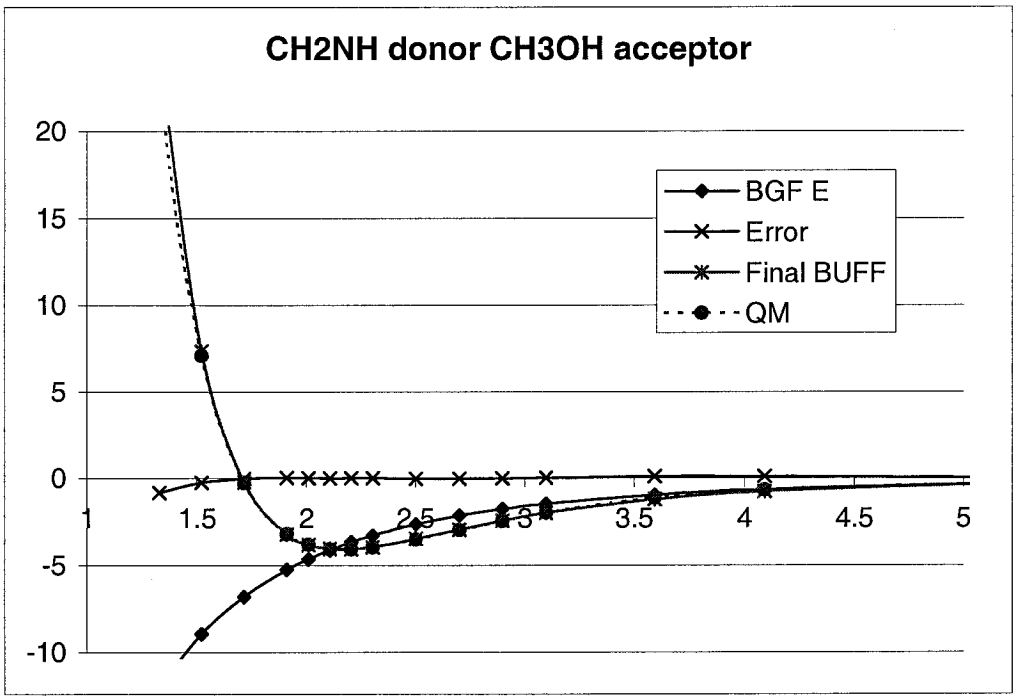

Figure 2- 10: Interaction energies of the CH_2NH - CH_3OH dimer. When the BUFF hydrogen bond term is implemented, the BUFF energies reproduce the LMP2/6-31G** QM energies.

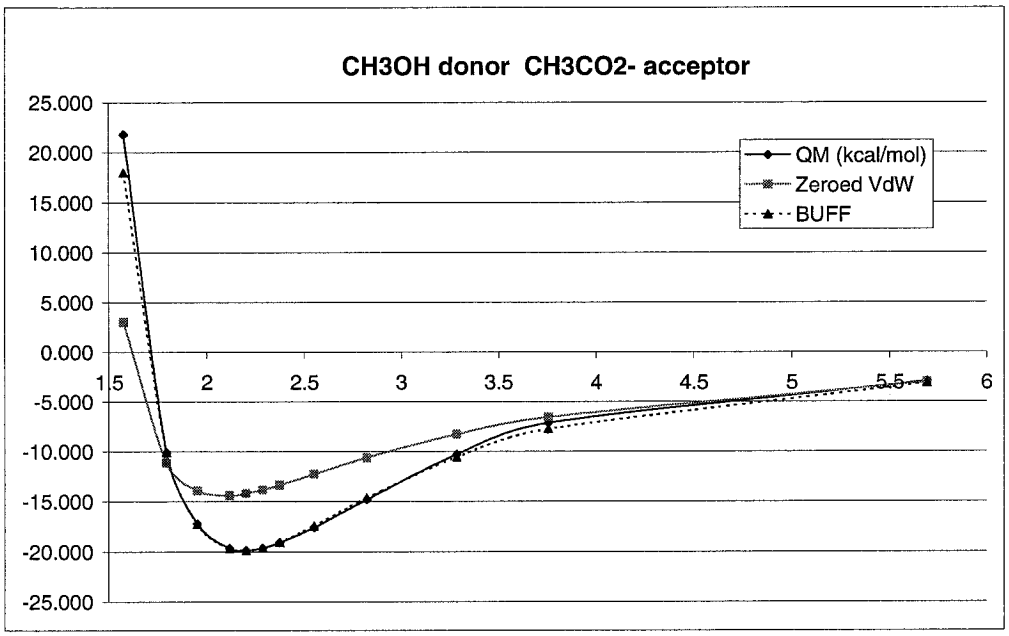

Figure 2- 11: Interaction energies of the CH_3OH - $CH_3O_2^-$ dimer. When the BUFF hydrogen bond term is implemented, the BUFF energies reproduce the LMP2/6-31G** QM energies.

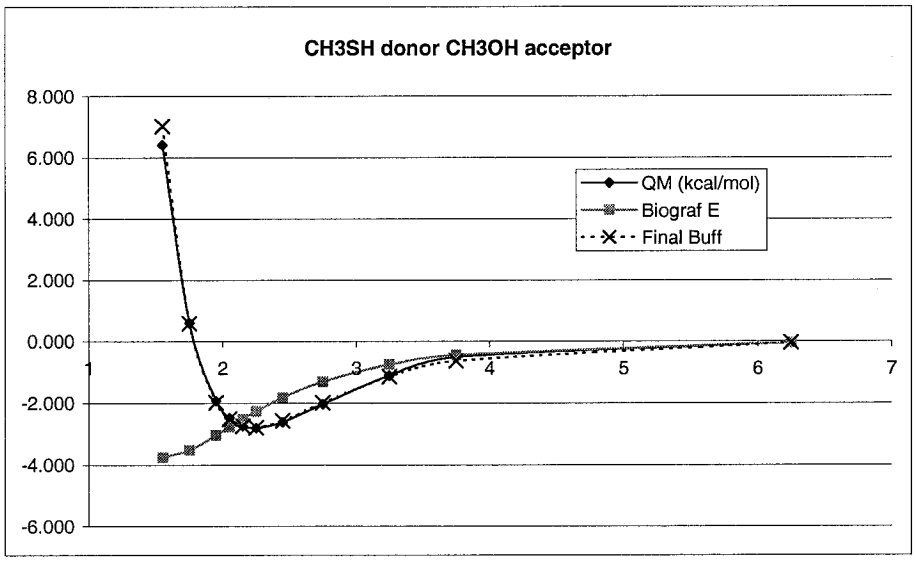

Figure 2- 12: Interaction energies of the CH_3SH - CH_3OH dimer. When the BUFF hydrogen bond term is implemented, the BUFF energies reproduce the LMP2/6-31G** QM energies.

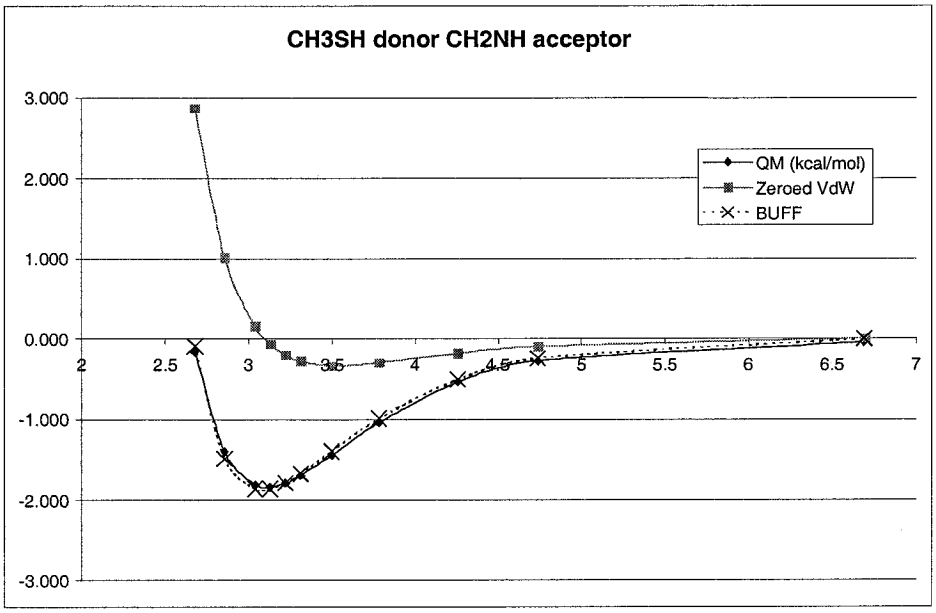

Figure 2- 13: Interaction energies of the CH_3SH – CH_2NH dimer. When the BUFF hydrogen bond term is implemented, the BUFF energies reproduce the LMP2/6-31G** QM energies.

One dimer interaction was problematic. After optimizing the parameters to fit the QM energies of the CH_3OH – CH_3OH interaction, the BUFF was found to minimize to the incorrect structure. The higher energy "box" type structure, Figure 2-13, allowed both hydrogen and oxygen to participate in a hydrogen bond. Since the hydrogen bond term was attractive, the two O:::H interactions of the "box" type structure were lower in energy than the single O:::H interaction found in the standard C_s type structure, Figure 2-14. The QM calculations demonstrated the C_s type structure was actually 2.5 kcal/mol lower in energy than the "box" type structure.

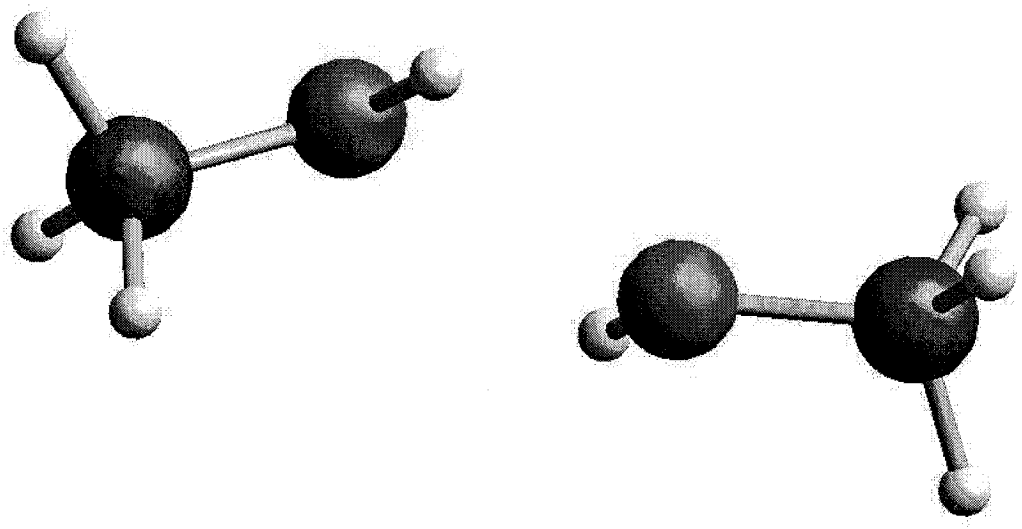

Figure 2- 14: The $CH_3OH - CH_3OH$ "box" type dimer interaction. Each hydrogen/oxygen pair is attempting to hydrogen bond with the other. This structure is 2-3 kcal/mol higher in energy than the low energy structure.

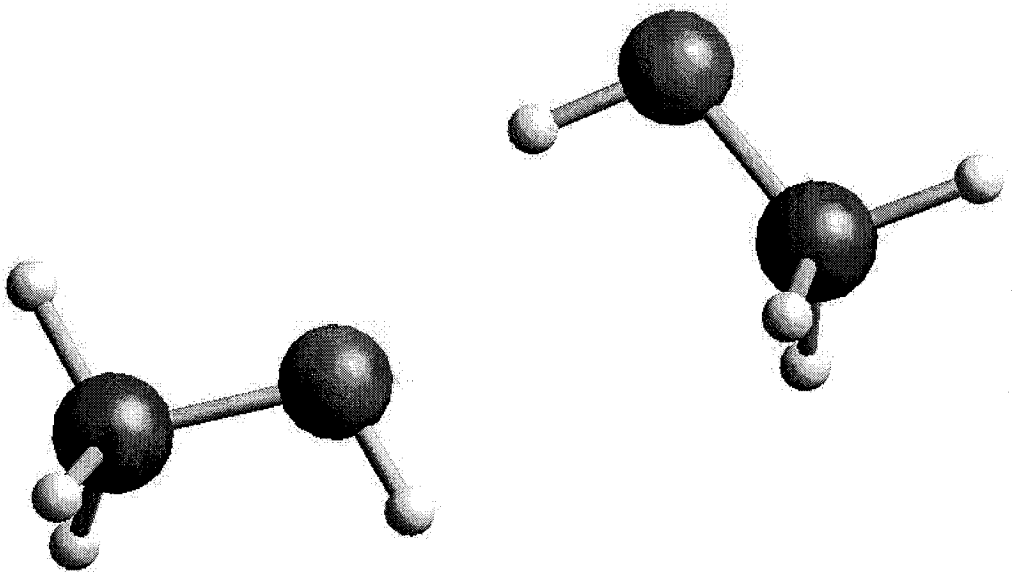

Figure 2- 15: The $CH_3OH - CH_3OH$ C_s type dimer interaction. Each hydrogen/oxygen pair is attempting to hydrogen bond with the other. This is the low energy structure as calculated by LMP2/6-61G** QM. A single hydrogen/oxygen pair is the primary interaction.

There are two solutions to this hydrogen bonding problem. An angle dependence of the hydrogen bond term could be added, as in Dreiding [5]. This would provide an elegant solution to the problem and probably improve other structural qualities of the forcefield as well. However, most popular molecular mechanics programs do not have the capability to incorporate angle-dependent nonbond terms. In order to keep BUFF as widely accessible as possible, a second solution was found. In the "box" state, the two hydrogen atoms are in much closer contact than in the C_s state. By parameterizeing the hydrogen-hydrogen interaction and the hydrogen-oxygen interaction simultaneously, the energies of the two states are correctly reproduced. The exponential-6 parameters for the hydrogen-hydrogen interaction are listed in Table 2-6.

Figures 2-15 and 2-16 show the final results of the parameterization. The C_s dimer interaction reproduces the QM energies. The "box" dimer interaction has approximately the correct well depth as well as the appropriate energies for near interactions. Most importantly, when the "box" form is minimized without constraints, it converts to nearly the correct C_s form. Note should be taken of the energies found for the "box" form at medium (2-3 Å) distances. This is a lower energy than found in the QM calculations. For this reason, the BUFF hydrogen bonding potential for sp^3O-H:::sp^3O should not be used for liquid simulations like methanol. A methanol simulation would be expected to have too high a density. The local hydrogen bond interactions have been optimized in the BUFF at the expense of some longer range inaccuracies. This should not make any difference in most biological simulations since only a few of the nonbond interactions will be hydrogen bonds and the majority of interactions will be Dreiding van der Waals terms.

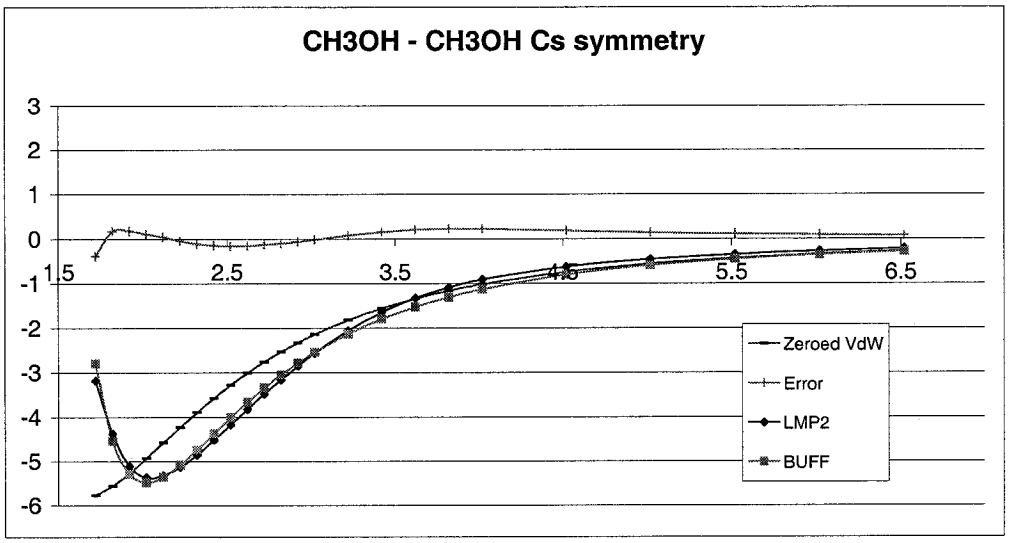

Figure 2- 16: Interaction energies of the C_s form of the CH_3OH - CH_3OH dimer. This structure was found to be the lowest energy dimer using LMP2/6-31G** QM energies. The BUFF hydrogen bond term was parameterized to correctly reproduce this dimer interaction.

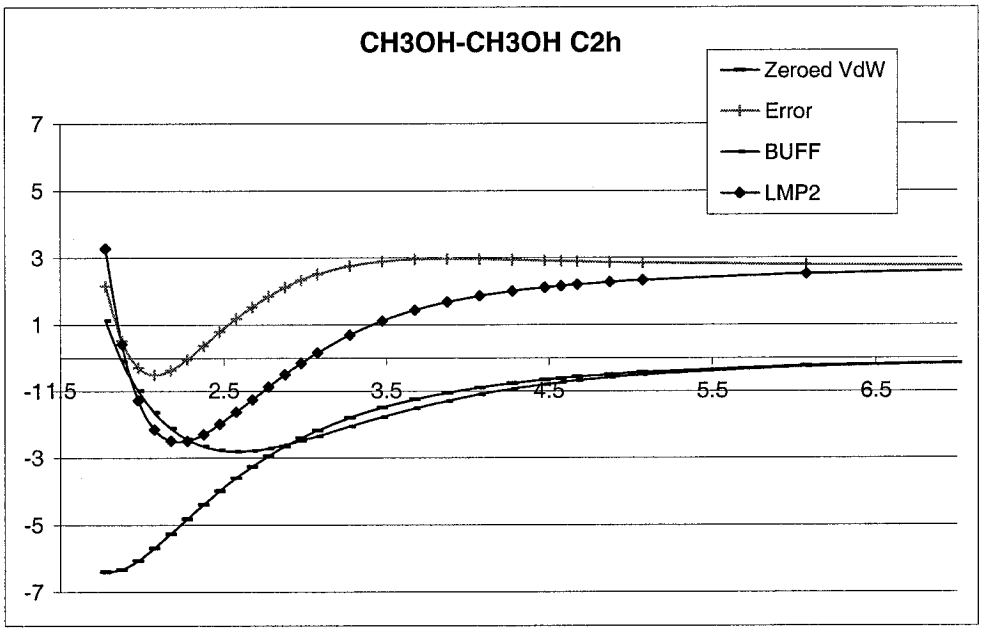

Figure 2- 17: Interaction energies of the C_{2h} "box" form of the CH_3OH - CH_3OH dimer. Since this structure was not the lowest energy dimer found using LMP2/6-31G** QM energies, the BUFF hydrogen bond term was parameterized to only reproduce the dimer interaction near the bottom of its potential well. The correct dimer interaction is found if the structure is minimized with BUFF.

Torsional space

Torsion parameters are one of the most important parameters in a forcefield that is not attempting to reproduce spectroscopic properties. Changes in global structure most frequently occur through changes in torsional conformations. The torsions most responsible for the global tertiary fold of a protein are the ϕ and ψ torsions along the peptide backbone. (See Figure 2-1.) For this reason, the torsion potentials involving the backbone atoms are carefully tuned in BUFF. Parameterization was performed on three tripeptides. Gly-Gly-Gly, Gly-Ala-Gly, and Gly-Pro-Gly were used where torsions developed for the Gly-Ala-Gly system were used for all non-glycine residues.

The backbone ϕ,ψ torsions for each tripeptide had previously been studied at the HF/6-31G** level [12]. The tripeptides were assigned appropriate atom types and charges based on the BUFF charge scheme. Torsions were then parameterized to best reproduce the low lying regions of the middle residue's ϕ,ψ torsions.

First, the Gly-Gly-Gly tripeptide (Figure 2-17) was parameterized to obtain appropriate main chain backbone torsions. The Gly-Gly-Gly torsions were then applied to the Gly-Ala-Gly system (Figure 2-18) and the forcefield was fit to the quantum mechanical potential map using the backbone torsions that include the C_β atom. This set of torsional parameters is intended to be of general use for all non-Gly, non-Pro residues. The Gly-Pro-Gly system (Figure 2-19) included all previous backbone and C_β torsions. It was found to faithfully reproduce the general trends of the QM potential, so no additional proline torsion terms were parameterized. A list of the final torsion terms can be found in Table 2-7.

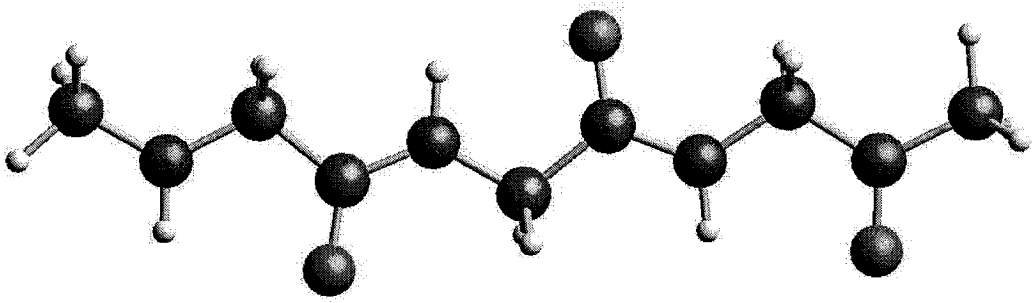

Figure 2- 18: Gly-Gly-Gly tripeptide used in BUFF torsion parameterization.

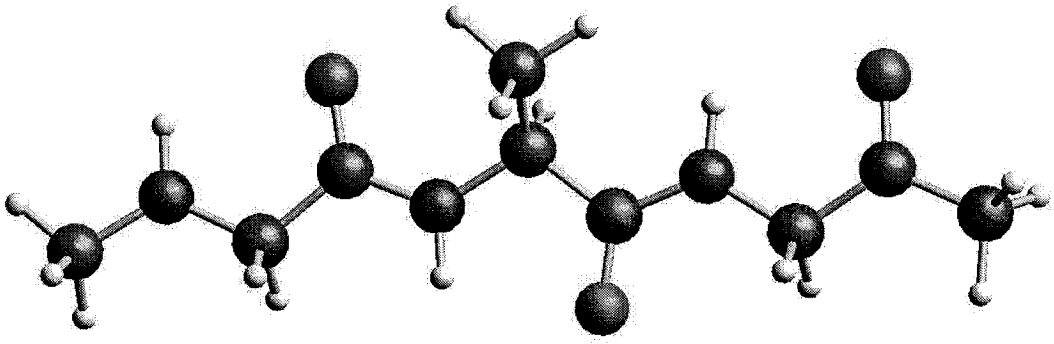

Figure 2- 19: Gly-Ala-Gly tripeptide used in BUFF torsion parameterization.

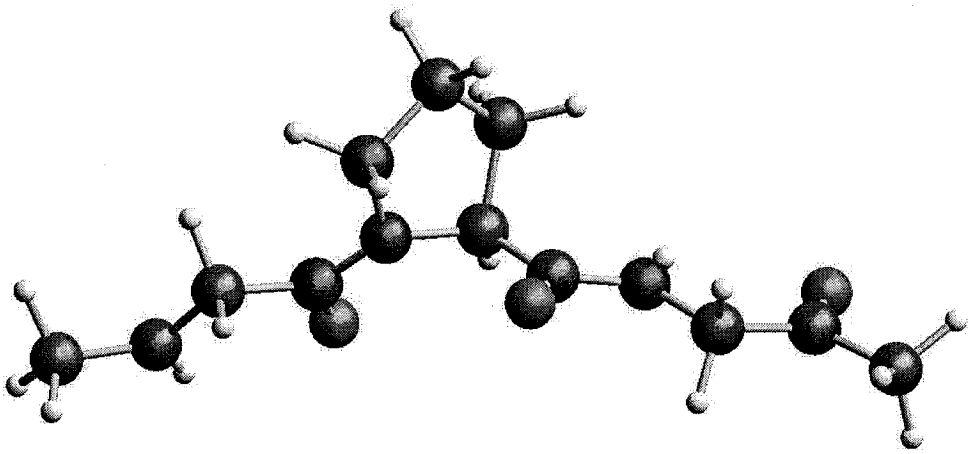

Figure 2- 20: Gly-Pro-Gly tripeptide used in BUFF torsion parameterization. It was found that additional torsions were not required to correctly reproduce the QM studies.

	A * Cos(θ) A in (kcal/mol)	B * Cos(2θ) B in (kcal/mol)	C * Cos(3θ) C in (kcal/mol)	D * Cos(4θ) D in (kcal/mol)
all amino acids				
C-N-C_α-C (ϕ)	0.00	0.00	-0.45	0.00
N-C-C_α-N (ψ)	0.00	-2.50	-0.20	0.00
needed for non-glycine				
C_β-C_α-N-C	-1.00	-1.00	-2.40	-1.50
C_β-C_α-C-N	0.60	0.30	-0.50	0.00

Table 2- 7: Special torsional terms used in the BUFF forcefield that are not found in UFF, but are required to correctly reproduce QM backbone energies. The net function for each potential is a sum of cosine terms. Note that for correct representation of the non-glycine torsions, a Cos(4θ) term was needed.

Each set of torsion parameters were fit in a similar manner. First, all torsions (except for the ω amide torsion) involved directly in the potential energy surface were zeroed out. Then a constrained minimization was performed for each of the 39 HF data points. The difference between the molecular mechanics and the HF energies was fit to a torsional potential. Initial torsional fitting of the tripeptide potential surfaces used a Boltzmann weighting to attempt to fit the lowest energy points of each potential surface. However, it became necessary to make additional adjustments to some of the torsional parameters in order to reproduce the correct relative energies of local minima on each potential surface. The torsion functions are constructed such that the backbone uses two torsional functions, one for ϕ and one for ψ. All other amino acids require the two backbone torsions as well as two new torsions that involve the C_β atom. Proline has an additional special torsional term involving its C_δ atom connected to the main chain nitrogen, but it was determined that parameters for this torsion was not required in order to correctly reproduce the QM potential surface. The HF/6-31G** potential surfaces and

the corresponding BUFF potential surface for each of the three tripeptides are shown in Figures 2-20 through 2-22.

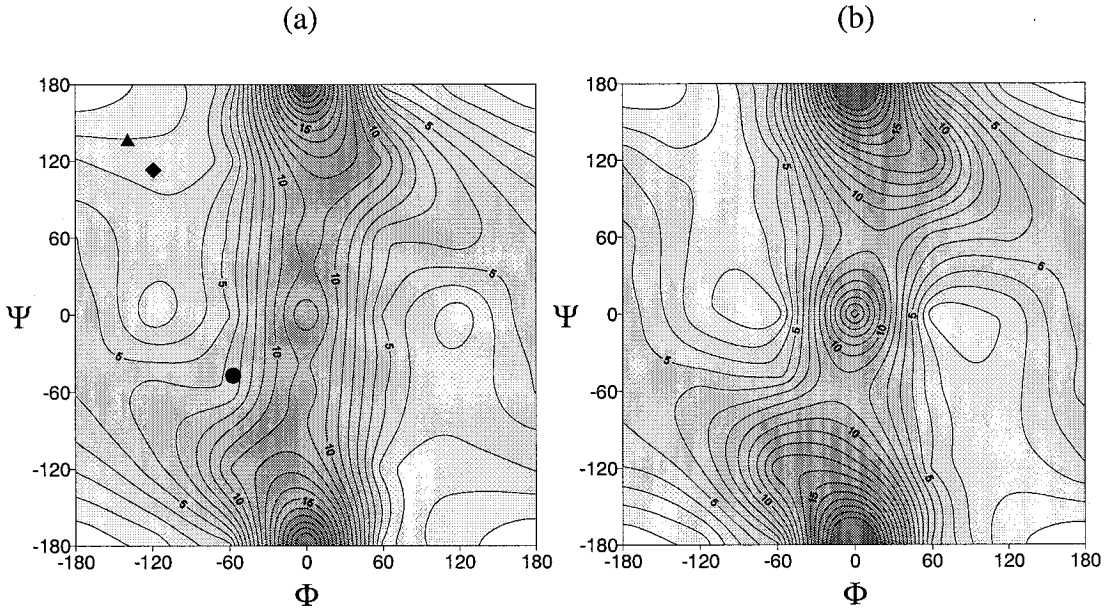

Figure 2- 21: The potential surfaces of the central ϕ,ψ of the Gly-Gly-Gly tripeptide. (a) HF/6-31G** calculated energies. (b) BUFF calculated energies. The contour spacing is 1 kcal/mole. The triangle, diamond, and circle represent ϕ,ψ angles at typical anti-parallel β-sheet, parallel β-sheet, and α-helical conformations respectively. A comparison of special points is listed in Table 2-8.

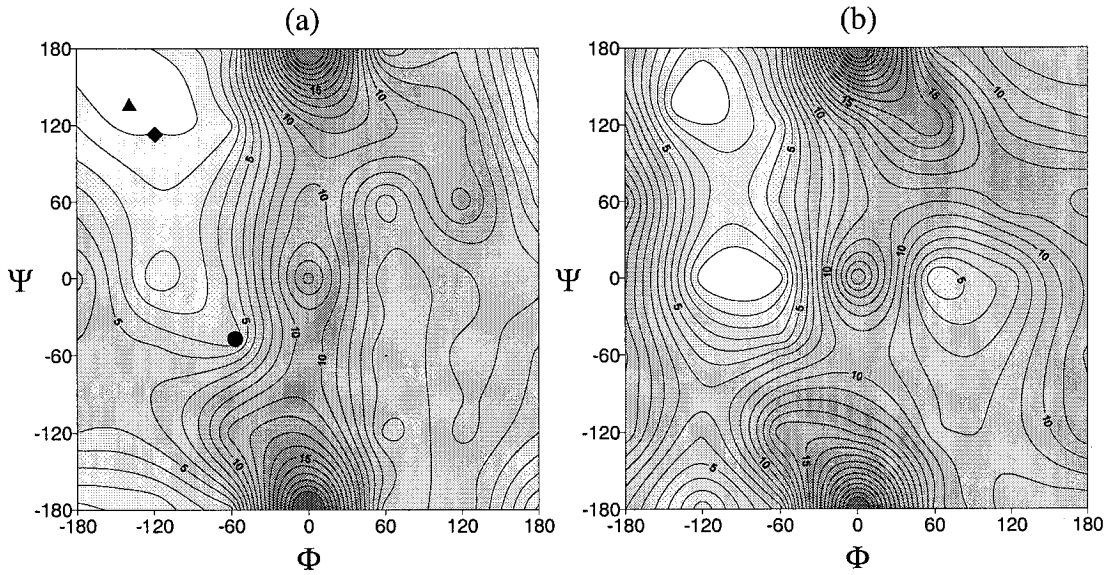

Figure 2- 22: The potential surfaces of the central ϕ,ψ of the Gly-Ala-Gly tripeptide. (a) HF/6-31G** calculated energies. (b) BUFF calculated energies. The contour spacing is 1 kcal/mol. The triangle, diamond, and circle represent ϕ,ψ angles at typical anti-parallel β-sheet, parallel β-sheet, and α-helical conformations respectively. A comparison of special points is listed in Table 2-9.

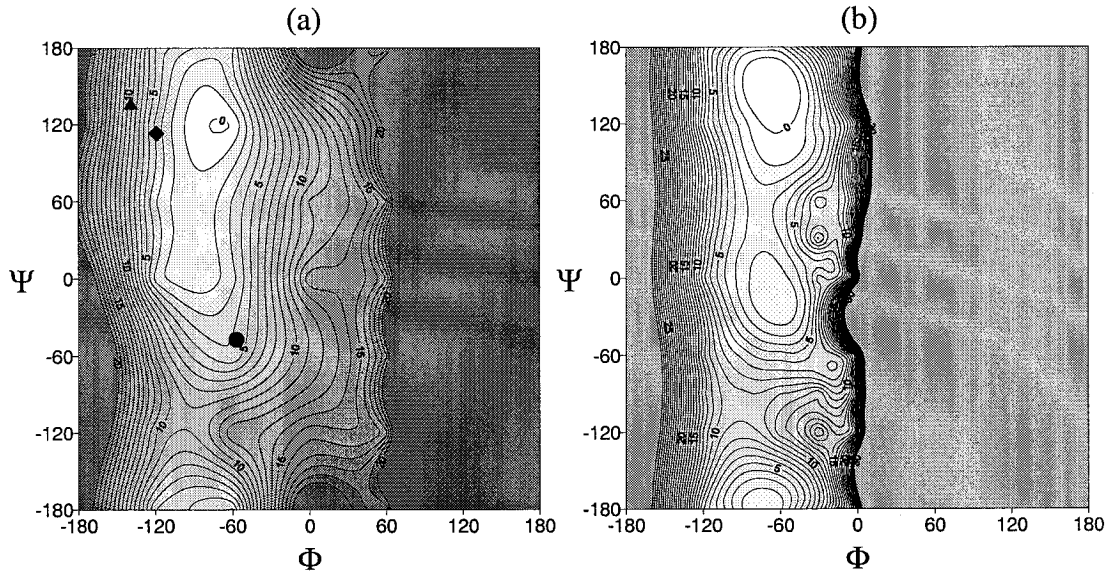

Figure 2- 23: The potential surfaces of the central ϕ,ψ of the Gly-Pro-Gly tripeptide. (a) HF/6-31G** calculated energies. (b) BUFF calculated energies. The contour spacing is 1 kcal/mol. The triangle, diamond, and circle represent ϕ,ψ angles at typical anti-parallel β-sheet, parallel β-sheet, and α-helical conformations respectively. A comparison of special points is listed in Table 2-10.

	(ϕ,ψ)	QM (HF/6-31G**)	BUFF
α-helix	-57,-47	4.60	4.68
parallel β-sheet	-119,113	3.07	2.85
anti-parallel β-sheet	-139,135	0.77	1.68
extended	-180,-180	0.00	0.00
	±60,0	4.89	1.36
	0,0	13.06	15.75

Table 2- 8: A listing of energies at selected conformations of the Gly-Gly-Gly tripeptide.

	(ϕ,ψ)	QM (HF/6-31G**)	BUFF
α-helix	-57,-47	3.27	4.29
parallel β-sheet	-119,113	1.06	1.55
anti-parallel β-sheet	-139,135	0.00	0.00
extended	-180,-180	0.66	7.00
	60,0	5.12	3.80
	-60,0	4.90	0.37
	0,0	13.10	16.30

Table 2- 9: A listing of energies at selected conformations of the Gly-Ala-Gly tripeptide.

(ϕ,ψ)	QM (HF/6-31G**)	BUFF
-60,0	2.59	1.75
-60,120	0.00	0.00

Table 2- 10: A listing of energies at selected conformations of the Gly-Pro-Gly tripeptide. The local minima are correctly ordered with BUFF having a 0.7 kcal/mol error for the higher energy minimum.

The special torsions parameterized for BUFF are easily identified by using a unique atom type for the C_α atom. In this case, the C_α atom has the atom type "C_A" and thus all special, non-UFF torsions are defined with the above torsion terms. Since the special torsion parameters were dependent solely upon the backbone atoms and the C_β atom, they should be universally applicable to most non-natural amino acids that are developed. The only requirement is that they have a C_β atom.

Validation and Comparison Studies

Throughout the comparison studies, calculations were performed using the OPLS-AA [3], Amber [14], and Dreiding [5] forcefields. Unless otherwise mentioned, calculations using OPLS-AA and Amber were performed using their implementation within the Macromodel software package [15]. For all calculations using the BUFF and Dreiding forcefield, the Biograf [9] software package was used. Dreiding was implemented with the exponential-6 van der Waals form. Dreiding charges were derived from a systematic charge equilibration scheme [11] that results in a set of backbone charges that are constant for all amino acid types and a different set of charges for each standard amino acid sidechain.

Gly-XXX-Gly Tripeptides

Quantum mechanical studies of tripeptide systems in various solvents [12] provide a detailed quantum mechanical potential energy surface which can be used as a standard reference for forcefield comparisons. The OPLS-AA, Amber, Dreiding, and BUFF forcefields have been compared to vacuum calculations on the central ϕ,ψ torsions of three tripeptides: Gly-Gly-Gly, Gly-Ala-Gly, and Gly-Pro-Gly (Figures 2-17 through 2-19).

For the glycine and alanine case, the central torsions were constrained with 250 kcal/mol restraints, and the other backbone torsions were kept extended (180,180) with additional 250 kcal/mol constraints. Then complete minimization was performed until an RMS force of less than 0.1 kcal/mol was found. Due to the additional constraints added by the proline ring, not all of the higher energy points were able to be calculated.

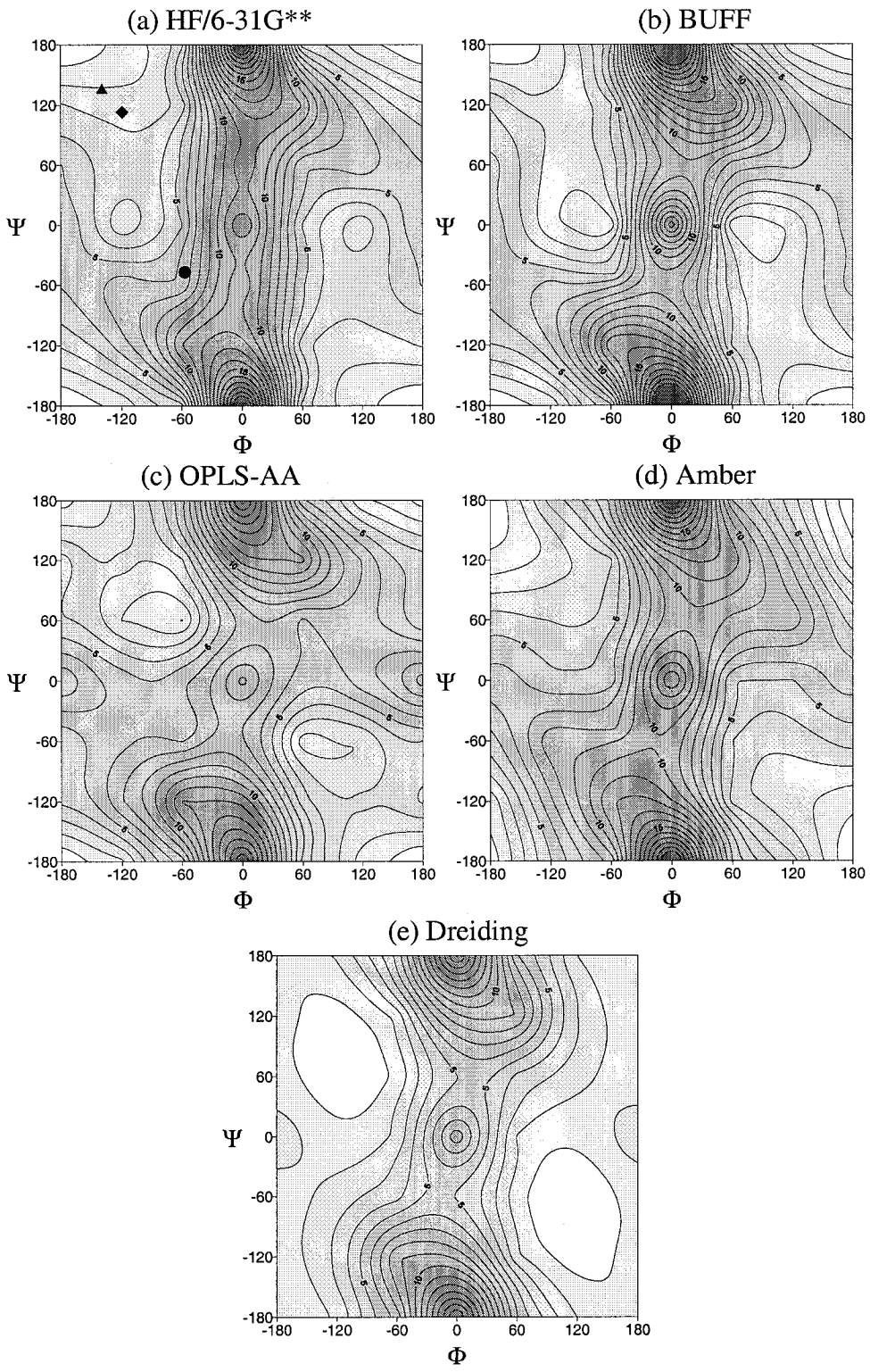

Figure 2- 24: Gly-Gly-Gly Tripeptide. Plots for various forcefields of the central torsion of the Gly-Gly-Gly tripeptide. Contour lines are drawn at 1 kcal/mol intervals.

An examination of the Gly-Gly-Gly torsions is shown in Figure 2-23. While most of the forcefields have the correct global minimum at (180,180), only BUFF results in the correct local minima at (±60,0). Only Dreiding and BUFF have broad low energy areas similar to the QM calculations, but the Dreiding low lying regions are arguably too broad.

Relative energies of the global and local minima, as well as the energies at common special points are shown in Table 2-11. BUFF has the correct ordering and approximately the correct energies for the special points, only differing from the QM by a slightly higher anti-parallel β-sheet region. Amber also accurately reproduces the correct glycine torsion special points, with only the α-helical region slightly high in energy. Dreiding has an incorrect global minimum and fails to order the special points correctly. OPLS-AA gives a correct ordering of special point energies, but gives energies that are, in general, too high. If the transition across $\phi = 0$ is examined, Amber gives approximately correct energies, BUFF is slightly high, while OPLS-AA and Dreiding give a transition across the $\phi = 0$ boundary that is almost half the calculated value.

	(ϕ,ψ)	QM (HF/6-31G**)	BUFF	OPLS	Dreiding	Amber
α-helix	-57,-47	4.60	4.68	5.26	3.40	7.85
parallel β-sheet	-119,113	3.07	2.85	3.20	0.59	2.79
anti-parallel β-sheet	-139,135	0.77	1.68	2.29	1.37	1.57
extended	-180,-180	0.00	0.00	0.00	1.05	0.00
	±60,0	4.89	1.36	5.48	2.26	5.50
	0,0	13.06	15.75	9.32	8.72	14.15

Table 2- 11: Energy (in kcal/mol) of special points within the Gly-Gly-Gly torsion for various forcefields. Constrained minimization was performed at each point. The global minimum for each forcefield is set to 0 kcal/mol.

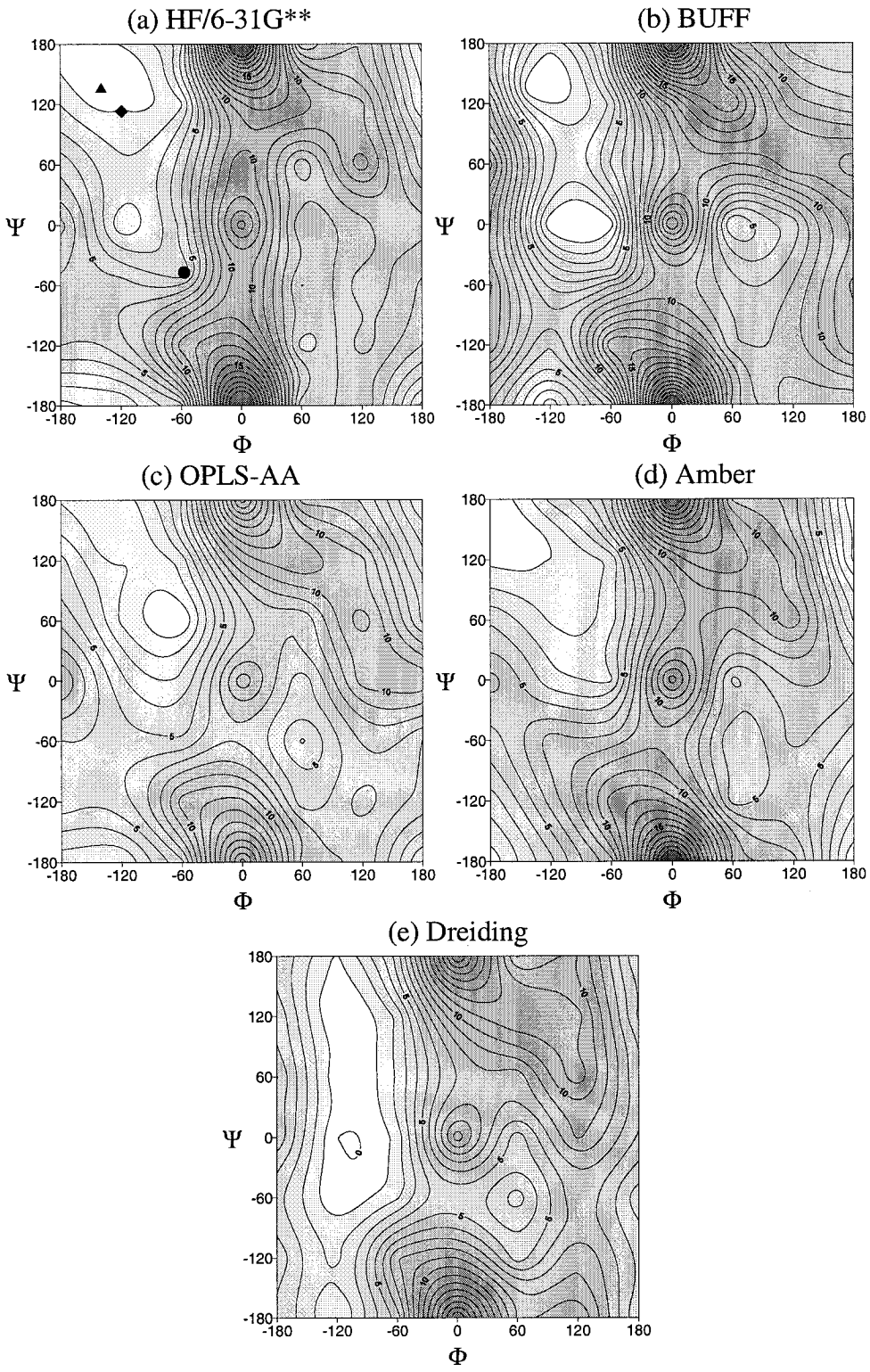

Figure 2- 25: Gly-Ala-Gly Tripeptide. Plots for various forcefields of the central torsion of the Gly-Ala-Gly tripeptide. Contour lines are drawn at 1 kcal/mol intervals.

An examination of the Gly-Ala-Gly torsions is shown in Figure 2-24. Relative energies of the global and local minima, as well as the energies at common special points, are shown in Table 2-12. In this situation, BUFF clearly has a more restrictive potential surface than the QM potential energy surface. The special points give good energies and are ordered correctly, but the torsional space is slightly more confined than that calculated from the QM. The transition across $\phi = 0$ calculated by BUFF is slightly higher than calculated by QM, but it is approximately correct.

OPLS-AA has a global minimum at (-60,60) which is a significant deviation from the calculated QM minimum at (-139,135). Despite this poor global minimum, the ordering of special point energies is reasonable, but OPLS-AA again suffers from a low transition barrier over the $\phi = 0$ barrier. Dreiding gives a broad low energy conformation potential well, but has its global minimum in the wrong place. In addition, Dreiding fails to order the special points correctly. Amber gives an excellent ordering of all energies except for the local minima that should occur around (-120, 0). This local minimum is completely absent resulting in a poor α-helix energy and a global minimum at the fully extended (-180,180) torsion angles.

	(ϕ,ψ)	QM (HF/6-31G**)	BUFF	OPLS	Dreiding	Amber
α-helix	-57,-47	3.27	4.29	4.76	1.63	6.47
parallel β-sheet	-119,113	1.06	1.55	1.92	0.26	1.46
anti-parallel β-sheet	-139,135	0.00	0.00	1.60	1.06	0.23
extended	-180,-180	0.66	7.00	3.51	3.99	0.00
	60,0	5.12	3.80	5.45	4.50	4.76
	-60,0	4.90	0.37	3.22	1.40	2.81
	0,0	13.10	16.30	9.70	10.66	13.65

Table 2- 12: Energy (in kcal/mol) of special points within the Gly-Ala-Gly torsion for various forcefields. The global minimum for each potential energy surface was set to zero. Constrained minimization was performed at each point. The global minimum for OPLS-AA is at (-60,60), for Dreiding it is at (-120,0), and for Amber it is at (-180,180). All other forcefields have a global minimum at –139,135.

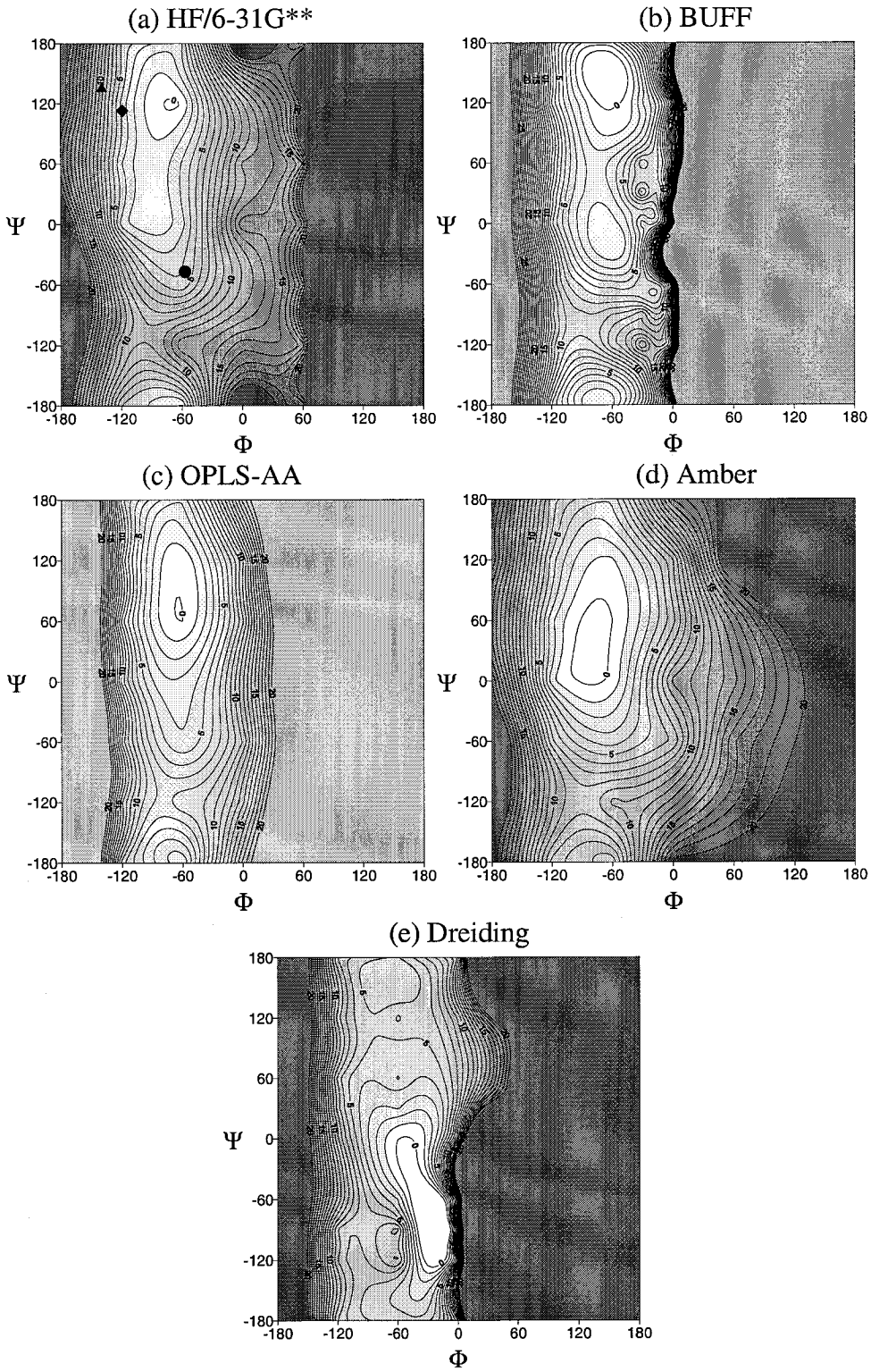

Figure 2- 26: Gly-Pro-Gly Tripeptide. Plots for various forcefields of the central torsion of the Gly-Pro-Gly tripeptide. Contour lines are drawn at 1 kcal/mol intervals.

An examination of the Gly-Pro-Gly torsions is shown in Figure 2-25. All four forcefields clearly show the high energy penalty at positive ϕ angles that results from the proline ring. All the forcefields except for Amber display a slightly smaller low energy torsional space available to the proline than is found in the QM calculation. BUFF and Dreiding both display the narrowest selection of low energy ϕ angles within their potential energy surfaces.

Table 2-13 lists the values of the potential energy surface at the two calculated quantum mechanical minima. Only BUFF succeeds in reproducing both of these local minima correctly. Each of the other forcefields only find one significant low energy minimum. Dreiding and Amber actually order the lower energy (-60,120) state higher than the local minimum at (-60,0). Despite its incorrect global minimum, OLPS-AA returns reasonable energies since its minimum lies between the two QM minima.

(ϕ,ψ)	QM (HF/6-31G**)	BUFF	OPLS	Dreiding	Amber
-60,0	2.59	1.75	3.29	0.00	0.29
-60,120	0.00	0.00	0.23	6.25	0.88

Table 2- 13: Energy (in kcal/mol) of special points within the Gly-Pro-Gly torsion for various forcefields. The global minimum for each potential energy surface was set to zero. Constrained minimization was performed at each point.

Of the forcefields tested on the three tripeptide systems, BUFF was the best at reproducing the correct energies of the low lying minima and other special points. OPLS-AA also performed well and was significantly better than Dreiding or Amber. OPLS-AA gave acceptable results for the low energy states of the system, but tended to err on the side of too much flexibility while BUFF gave a more accurate depiction of the lower energy states but tended to have a more restricted potential energy surface,

particularly in the Gly-Ala-Gly case. BUFF was expected to perform well at this test since the backbone torsions were optimized on the Gly-Gly-Gly and Gly-Ala-Gly case. BUFF is shown to be at least as accurate as OPLS-AA for these systems and has the additional advantage of a rule based universal forcefield as well.

Polyalanine α-Helices

Recent examinations using *ab initio* quantum mechanical calculations (HF/6-31G**) have shown that for a polyalanine α-helix of length greater than or equal to 4, there is a strong preference for a new residue added at the amino or carboxy terminus to adopt an α-helix conformation [16]. QM charges were used in molecular dynamics calculations to determine that the QM effects were dominated by electrostatic dipole-dipole interactions. An examination of the amino terminus (N-terminus) of both a 4 alanine and a 7 alanine peptide α-helix were made with the BUFF and the OPLS-AA forcefield and comparisons were made to the HF/6-31G** QM energies.

For all polyalanine α-helix calculations, the ϕ and ψ torsions of the end residue (N-terminus) were constrained and all other atoms were allowed to relax during the minimization. Figure 2-26 displays the 4 alanine helix N-terminus. As can be seen in the figure, the helix begins and ends with a ½ glycine residue that both neutralizes the endpoints and provides an additional amid environment so that the first and last peptide ω backbone torsions remain planar.

All calculations on α-helix N-terminus residues were performed on 27 points. ψ ranged from 0° to –180° in 60° increments while ψ ranged from –180° to 180°, also in 60° increments. Three additional points were calculated at standard α-helix, parallel, and

anti-parallel β-sheet torsion pairs. The quantum mechanical calculations were performed as single point energies with all atomic coordinates fixed. In order to fairly evaluate the two forcefields, calculations involving BUFF and OPLS-AA were performed with minimization that allowed all parameters but the torsions of interest to relax.

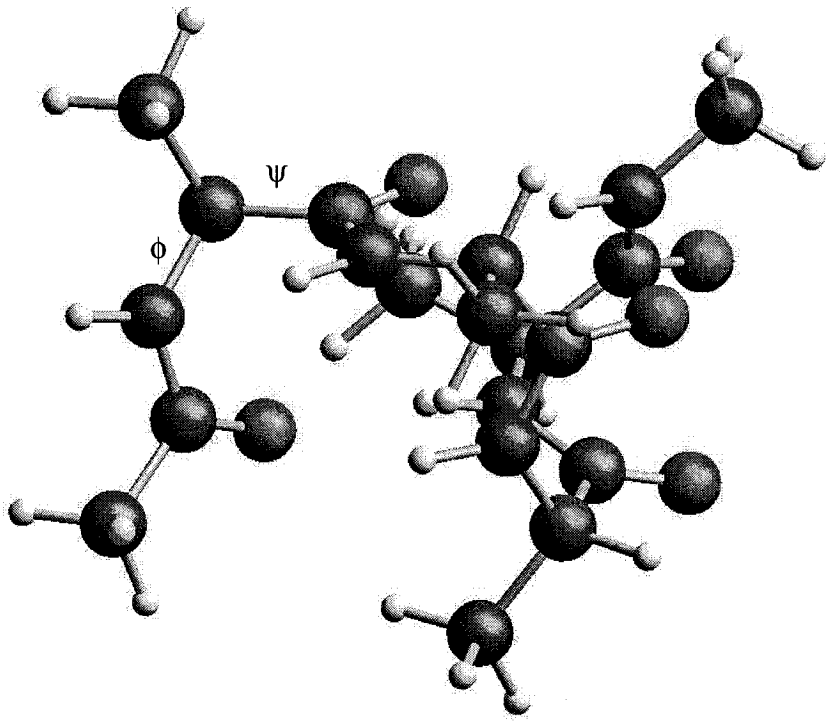

Figure 2- 27: N-terminus of alanine tetrapeptide in helix conformation. ϕ and ψ torsions used to create the potential energy surface are marked. The quantum mechanical energies are from single point calculations at the HF/6-31G** level. All forcefield calculations restrained only the ϕ and ψ torsions.

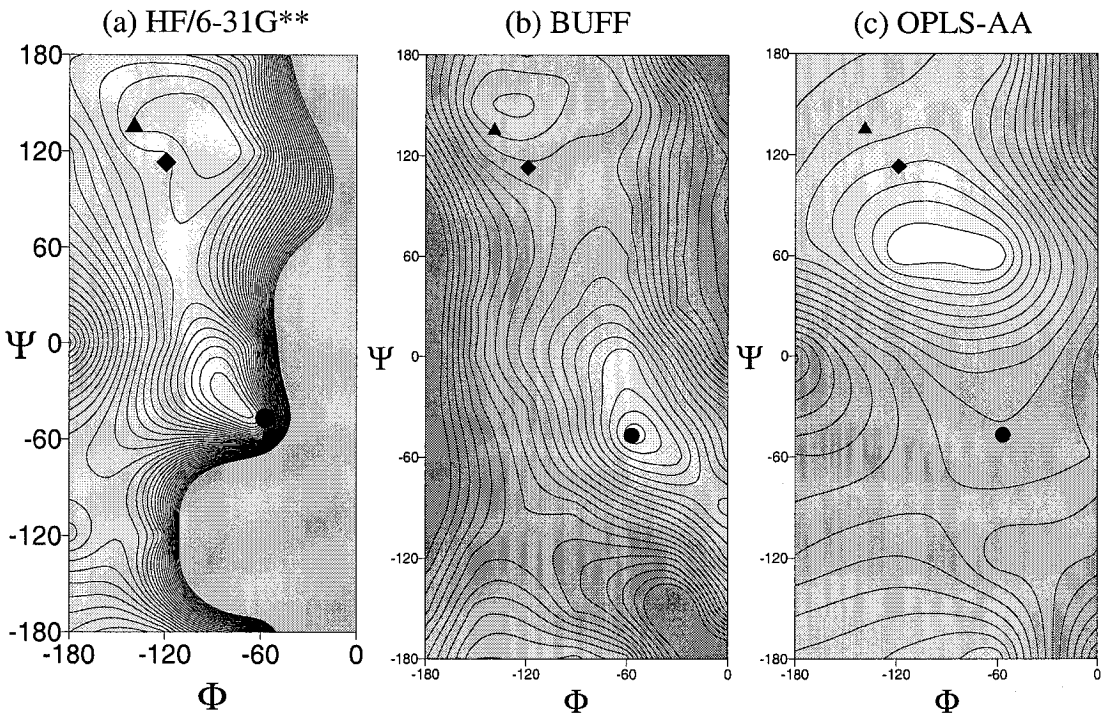

Figure 2- 28: Potential energy surfaces of the alanine helix tetrapeptide N-terminus. The QM calculation was performed on a rigid, idealized helix, while the BUFF and OPLS-AA potential energy surfaces were generated from calculations that only constrained the N-terminus ϕ,ψ angles. Contour lines are plotted at 1 kcal/mol intervals.

Figure 2-27 displays the potential energy surface for N-terminal residue in an alanine tetrapeptide helix. OPLS-AA does a very poor job of reproducing the quantum mechanical trends. Its global minimum is at (–60,60) rather than (-57,-47) as calculated by quantum mechanics. The BUFF, however, does a very good job of reproducing the quantum mechanical energies. Table 2-14 demonstrates how well BUFF reproduces the correct energies and ordering of the QM results.

	(ϕ,ψ)	QM (HF/6-31G**)	BUFF
α-helix	-57,-47	0.00	0.09
parallel β-sheet	-119,113	4.39	4.94
anti-parallel β-sheet	-139,135	2.82	1.71
extended	-180,-180	6.08	4.73

Table 2- 14: Energy (in kcal/mol) of special points of ϕ/ψ scan of N-terminus alanine in alanine tetrapeptide helix.

Similar calculations were performed on the 7 alanine helix N-terminus (Figure 2-28). The results are shown in Figure 2-29. Here, BUFF once again gives an excellent reproduction of the QM potential surface. The local minima and global minimum are well reproduced and have approximately the correct energy spacing. The OPLS-AA forcefield does significantly better on the 7 alanine helix than on the 4 alanine helix. The global minimum is roughly correct, although it does not have a clear local minimum near (-60,120) which is seen in the QM potential surface. The general shape and location of the OPLS-AA low energy surface is correct, but it tends to be wider in shape than found in the QM potential energy surface. This may be attributed, in part, to the QM calculation method. If the helix were allowed to relax in the QM calculation, it might result in a broader lower energy region as well.

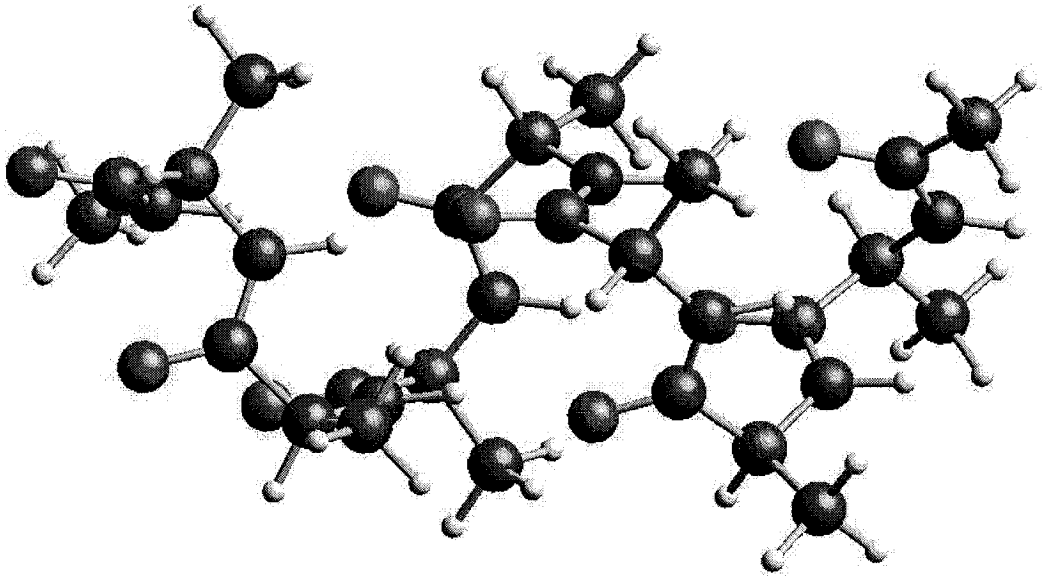

Figure 2- 29: 7mer polyalanine in a helix conformation.

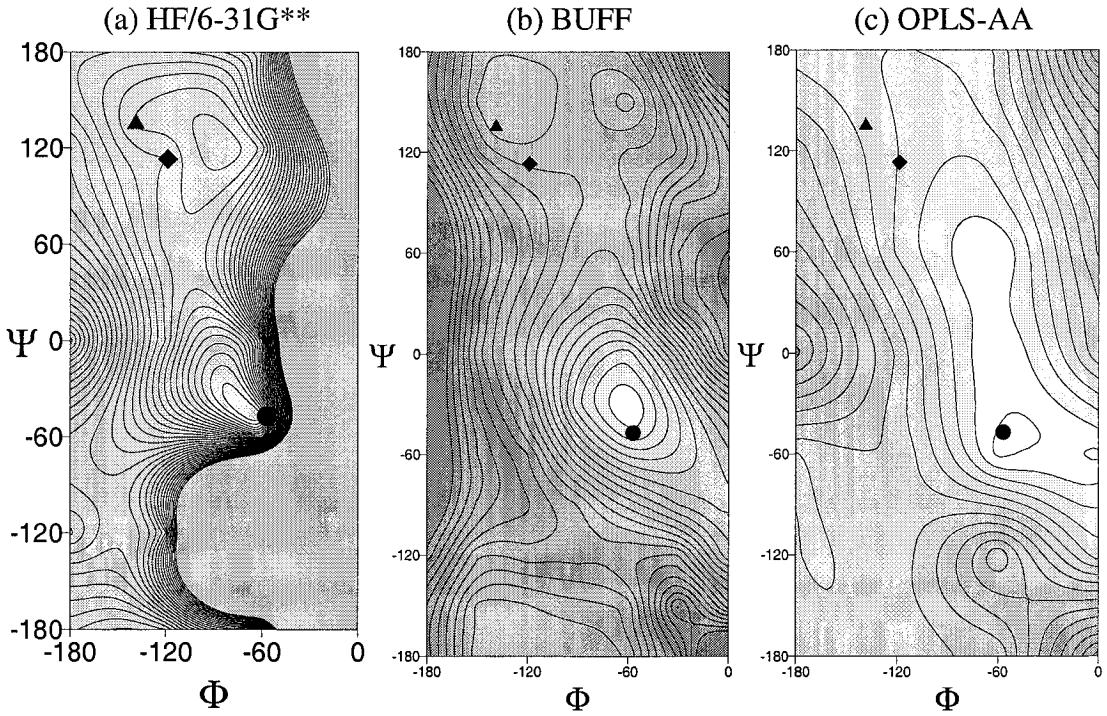

Figure 2- 30: Potential energy surfaces of the 7 alanine helix N-terminus. The QM calculation was performed on a rigid, idealized helix, while the BUFF and OPLS-AA potential energy surfaces were generated from calculations that only constrained the N-terminus ϕ,ψ angles. Contour lines are plotted at 1 kcal/mol intervals.

Figure 2-30 displays a contour plot of the difference in energy between the 7mer and 4mer alanine calculations. Since there may be geometric differences between the minimized state found for the 4mer and 7mer helices, the contours shown for BUFF and OPLS-AA have a much more complex surface. Since the HF calculations on the 4mer and 7mer helix are only single point energy calculations, with no minimization, the difference between the two scans is more easily isolated. A close examination of the forcefield potential energy surface differences shows that they both describe approximately the same trend as the QM calculation. The α-helical regions are slightly more favored in the 7mer helix than in the 4mer helix. This results in a low energy region in the difference plot. This trend can be more clearly seen in Figure 2-31. Here, the BUFF calculation for both the 4 alanine and the 7 alanine are held fixed in order to eliminate effects of changing geometry. The trend that favors the helical conformation as the helix lengthens is clearly shown. The trend is correct, but the overall energies are not correctly reproduced. The BUFF contour plot bears a close resemblance to the QM plot, but the contour lines are drawn at only ½ kcal/mol intervals for the BUFF. Only half of the preference energy is reproduced. This is expected to be a result of the fixed point charges. If the charge within the helix were allowed to relax and polarize along the helical axis, there should be a greater cooperative effect resulting in an even greater favoring of the α-helical conformation with increasing helix length [16].

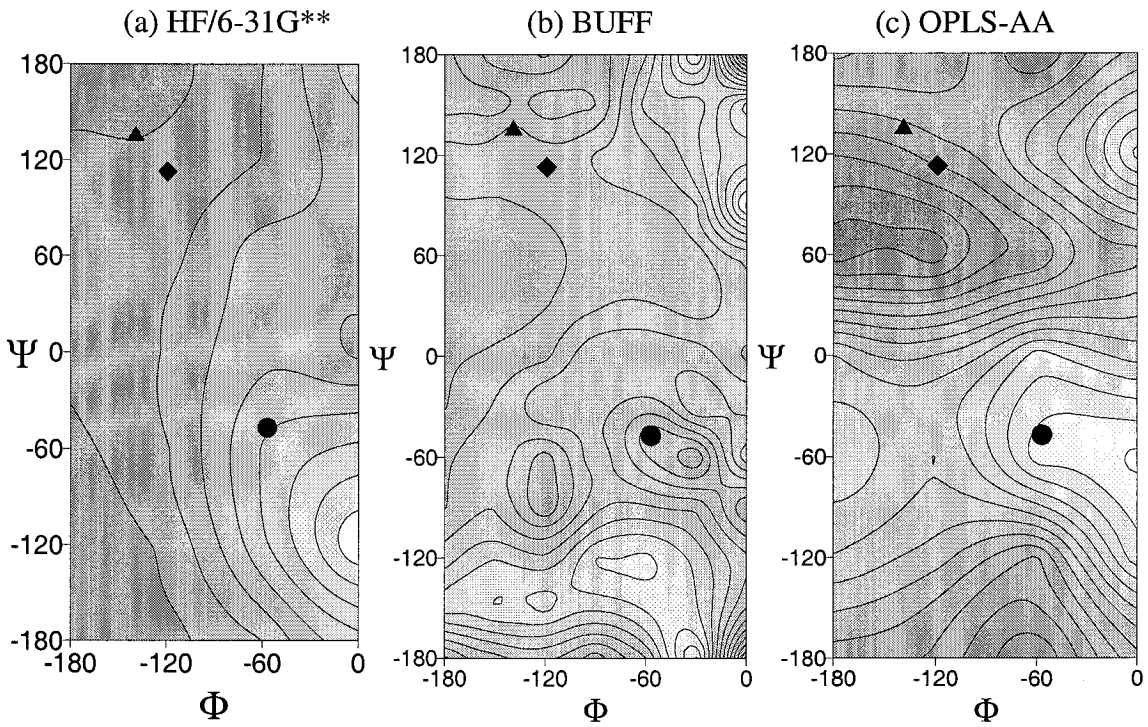

Figure 2- 31: Potential energy surfaces of the difference between the 7 alanine and 4 alanine helix N-terminus. Both the BUFF and OPLS-AA potential energy surfaces are significantly more complex due to the relaxation during minimization. However, both BUFF and OPLS-AA have the correct trend, and the helical conformation is increasingly preferred as the helix length increases. Contour lines are plotted at 1 kcal/mol intervals.

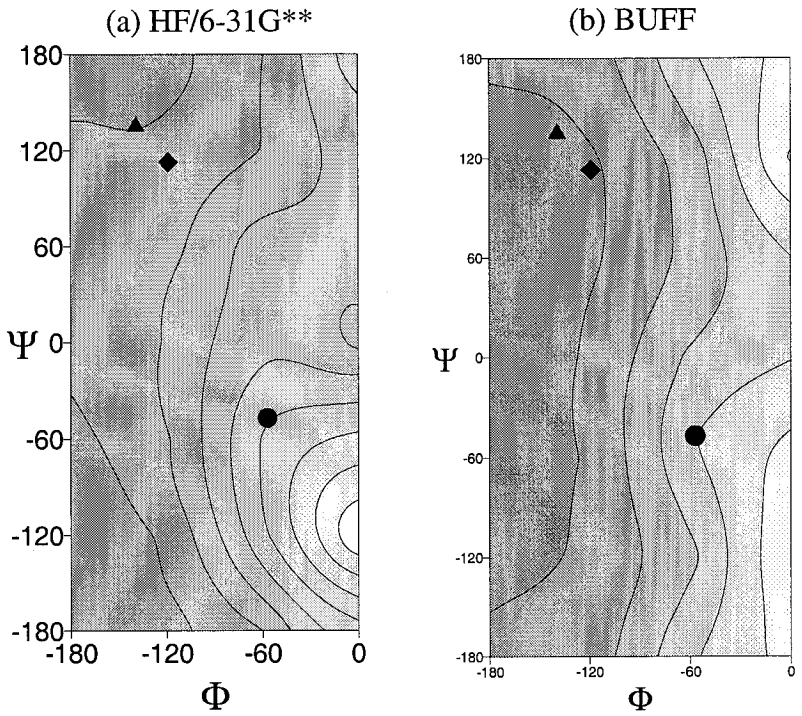

Figure 2- 32: Potential energy surfaces of the difference between the 7 alanine and 4 alanine helix N-terminus. In this BUFF calculation, the helix remains fixed and only single point energies are calculated. This clearly shows that BUFF matches the QM trends. Note that the HF plot has contours at 1 kcal/mol intervals while the BUFF plot has contour lines only at ½ kcal/mol intervals.

Both BUFF and OPLS-AA approximately reproduce the increasing stability of the α-helical conformation as the helix length increases. Neither forcefield results in the correct absolute value of this preference as found by QM calculations, but this limitation is a result of the fixed charge scheme rather than a problem with the forcefields. The BUFF gives an excellent reproduction of the QM potential energy surface for both the 4 alanine and 7 alanine helix N-termini. The OPLS-AA potential energy surface for the 7 alanine helix N-terminus is reasonable, but OPLS-AA gives a poor representation of the potential energy surface of the 4 alanine helix N-terminus.

X-ray crystal structure minimization

Structures examined so far have been short peptides. These validation studies do not involve a very condensed state, and so do not rigorously test the nonbond part of the forcefield. In order to validate the nonbond forces (charges, van der Waals, and hydrogen bonding) in the BUFF, a series of examinations of high quality crystal structures has been performed.

A semi-empirical QM minimization using MOPAC2000 [17] was performed on the 0.83 Å resolution crystal structure of the 46 residue protein crambin (1cbn [18]). This crystal structure was also minimized with Dreiding, Amber94 [1], and BUFF. Biograf [9] was employed for the Dreiding, BUFF, and Amber94 forcefields. Charges for the Dreiding minimization were calculated by performing a charge equilibration calculation [13] on the entire protein. Results of the minimization studies are shown in Table 2-15.

	Mopac 2000	Amber94	Dreiding QEq	BUFF
CRMS to crystal structure	0.10	0.72	0.54	1.00

Table 2- 15: All atom coordinate root mean square (CRMS) structural fits to 0.83 Å resolution 1cbn crystal structure. Structures were minimized, then matched to the original crystal structure to determine the approximate level of perturbation caused by the forcefield.

The Mopac2000 calculation returns an amazingly accurate reproduction of the crystal structure. The semi-empirical QM method required 8 steps of minimization which took approximately 1 hour of CPU time on an SGI origin machine. The high quality of the fit between the crystal structure and the Mopac2000 minimized structure may be due, in part, to the use of *ab initio* methods in the x-ray crystal structure solution of 1cbn.

The Amber94 and Dreiding forcefields minimize the protein to structures that have a CRMS from the crystal that is less than the structure resolution. BUFF minimizes the protein to 1.00 Å CRMS which is only slightly larger than the crystal structure resolution. BUFF does not perform as well as the other two forcefields tested, but BUFF forces do not perturb the crystal structure in a significant way.

In order to demonstrate the versatility of BUFF, minimization of several iron containing heme groups were performed. The BUFF and UFF forcefields were applied to the heme portion of P450 oxidase. The results of a heavy atom CRMS match between the minimized and crystal structures are listed in Table 2-16. BUFF minimization returns a structure that is closer to the crystal structure than a minimization using UFF. Charges for the heme structure were derived from HF/6-31G** Mulliken calculations[19], and were the same in both minimization calculations. The improvement in the BUFF structure over the UFF structure is a result of using Dreiding van der Waals terms. The primary disruption in the structure is a result of the iron settling down into the heme pocket to make a planar center. This forces out the surrounding nitrogen groups and thus, slightly disrupts the rest of the heme. The iron-nitrogen bond lengths stay approximately the same, with the bonds at 1.99 Å in the crystal structure lengthening to 2.02 Å in the BUFF minimization.

	BUFF	UFF
CRMS to P450 Heme	0.77	0.81

Table 2- 16: CRMS values of the heme portion of P450 crystal structure matched to the heme structure, minimized with BUFF and UFF. Charges were derived using Mulliken populations[19] from a HF calculation and are the same in both calculations.

The BUFF was also applied to a minimization calculation of the entire cytochrome C553 (1C75[20]) protein. The original crystal structure is at 0.97 Å resolution. The BUFF minimized structure is 1.39 Å CRMS away from the crystal structure. Figure 2-32 shows a comparison of the heme groups from the crystal and from the minimized structure. A CRMS comparison of heavy atoms shows that the BUFF structure is only 0.68 Å away from the crystal structure. While the BUFF minimization of the entire protein is not as close to the crystal structure as might be preferred, the reproduction of the heme portion is clearly very acceptable.

Cytochrome C553 (1C75) at 0.97 Å resolution	BUFF
All Atom CRMS with 1c75	1.39 Å
Heavy Atom CRMS with only Heme group	0.68 Å

Table 2- 17: A CRMS comparison of BUFF minimized cytochrome C553 (1C75) and the crystal structure.

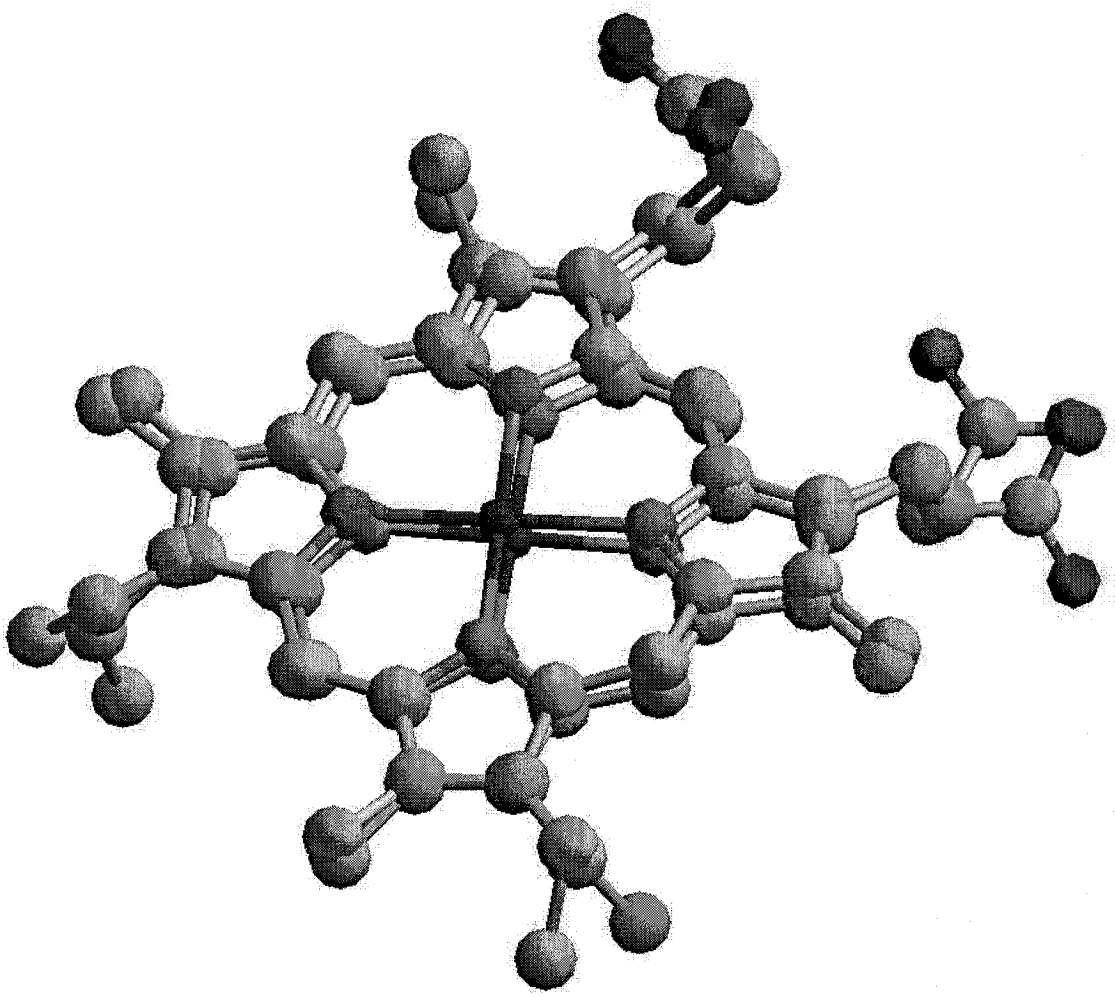

Figure 2- 33: A comparison of the cytochrome C heme group. Minimization of the cytochrome-c structure in BUFF results in a heme group that has a CRMS of only 0.68 Å from the crystal structure heme.

Investigations of the application of BUFF to folded protein crystal structures demonstrate that it is comparable, but not better than other common forcefields. Systems that contain metals are easily investigated with BUFF and are reproduced with acceptable accuracy. The next section addresses possible improvements to the BUFF.

Alanine tetrapeptide helix/sheet folding

As validations of the BUFF were performed, concern arose over the electrostatic energies calculated by the forcefield. A system that would be less complex than an entire

The only contribution to the electrostatic energies are the charge-charge interactions within the peptide. In order to significantly improve these results, the charges used within BUFF would need to be derived from a different source. There are two common ways to calculate charges from quantum mechanics. Electrostatic potential (ESP) fitting [11] uses the molecular dipole of the system to fit the overall charge scheme while Mulliken charges [19] are derived by an analysis of electron densities near each atom in the system. ESP charges were used in creating the BUFF, and are excellent at recreating a molecular dipole. However, since there are many atoms that contribute to the molecular dipole, the charges on interior atoms are not always sufficiently constrained by the ESP method. The charges on the extended tripeptide systems used in deriving charges for BUFF were initially thought to be sufficiently constrained, so a Mulliken charge calculation was also performed on the same tripeptide systems. Table 2-19 shows the charges found for the backbone atoms of selected residues using both ESP and Mulliken charge fitting. On average, the Mulliken charges were found to be 0.1 smaller than the ESP calculated charges. In particular, the C_α carbon was frequently found to have a charge of up to 0.75 in ESP and only around 0.05 in the Mulliken calculation.

ggg	N12	H13	C14	H15	C17
HF Mulliken	-0.76073	0.34095	-0.04232	0.1685	0.7503
LMP2 ESP in h2o	-0.654	0.297	0.001	0.1	0.836
0.07	**0.11**	0.04	0.04	0.07	0.09
gag	N12	H13	C14	H15	C16
HF Mulliken	-0.755	0.34002	0.02111	0.17338	0.78047
LMP2 ESP in h2o	-0.857	0.342	0.487	0.011	0.726
0.09	**0.10**	0.00	**0.47**	**0.16**	0.05
gdg	N12	H13	C14	H15	C16
HF Mulliken	-0.74285	0.32531	0.03426	0.14448	0.83726
LMP2 ESP in h2o	-0.88	0.376	0.13	0.086	0.914
0.11	**0.14**	0.05	0.10	0.06	0.08
geg	N12	H13	C14	H15	C16
HF Mulliken	-0.74382	0.32138	0.03781	0.15572	0.79672
LMP2 ESP in h2o	-0.768	0.321	0.104	0.081	0.871
0.11	0.02	0.00	0.07	0.07	0.07
ghg	N12	H13	C14	H15	C16
HF Mulliken	-0.7481	0.33402	0.05968	0.18376	0.75908
LMP2 ESP in h2o	-0.945	0.359	0.757	-0.044	0.638
0.16	**0.20**	0.02	**0.70**	**0.14**	**0.12**
gkg	N12	H13	C14	H15	C16
HF Mulliken	-0.75571	0.33956	0.05155	0.19961	0.77379
LMP2 ESP in h2o	-0.71	0.323	0.003	0.112	0.872
0.10	0.05	0.02	0.05	0.09	0.10
gmg	N12	H13	C14	H15	C16
HF Mulliken	-0.7521	0.3326	0.05118	0.18452	0.78043
LMP2 ESP in h2o	-0.629	0.291	-0.034	0.129	0.851
0.08	**0.12**	0.04	0.02	0.06	0.07
gng	N12	H13	C14	H15	C16
HF Mulliken	-0.75106	0.34071	0.05507	0.18662	0.81841
LMP2 ESP in h2o	-0.776	0.329	0.259	0.088	0.804
0.09	0.02	0.01	**0.20**	0.10	0.01
grg	N12	H13	C14	H15	C16
HF Mulliken	-0.75622	0.33916	0.05105	0.20164	0.7748
LMP2 ESP in h2o	-0.74	0.334	0.029	0.13	0.877
0.10	0.02	0.01	0.02	0.07	**0.10**
gsg	N12	H13	C14	H15	C16
HF Mulliken	-0.74149	0.33562	0.02856	0.18161	0.79731
LMP2 ESP in h2o	-0.75	0.328	0.189	0.048	0.828
0.10	0.01	0.01	**0.16**	**0.13**	0.03
gtg	N12	H13	C14	H15	C16
HF Mulliken	-0.74236	0.33294	0.04007	0.18444	0.74147
LMP2 ESP in h2o	-0.515	0.314	-0.226	0.142	0.849
0.10	**0.23**	0.02	**0.19**	0.04	**0.11**
gwg	N12	H13	C14	H15	C16
HF Mulliken	-0.74905	0.33242	0.07003	0.17497	0.77644
LMP2 ESP in h2o	-0.928	0.365	0.564	-0.011	0.69
0.09	**0.18**	0.03	**0.49**	**0.16**	0.09
gyg	N12	H13	C14	H15	C16
HF Mulliken	-0.74993	0.33437	0.06988	0.17693	0.7747
LMP2 ESP in h2o	-0.917	0.352	0.648	-0.018	0.617
0.12	**0.17**	0.02	**0.58**	**0.16**	**0.16**

Table 2- 19: A comparison of main chain ESP and Mulliken charges for selected amino acids in Gly-XXX-Gly QM studies. Boxes in gray differ in charge by more than 0.1 and boxed in dark gray differ by more than 0.20. All charges are listed in units of the charge on an electron.

Simulation Method	Extended to Helix ΔE
LMP2/6-31G**	-5.01 kcal/mol
HF/6-31G**	-1.14 kcal/mol
Mopac2000	-12.0 kcal/mol
OPLS-AA	-2.84 kcal/mol
BUFF – ESP Charges	-22.79 kcal/mol
BUFF – Mulliken Charges	-3.41 kcal/mol

Table 2- 20: Extended to helix transition energies of the 4 alanine polypeptide. The endpoints were capped as in Figure 2-26 to neutralize the endpoints. With only an adjustment to the charge scheme, the BUFF calculation is in excellent agreement with the high level QM calculations.

Mulliken charges were taken for glycine and alanine and then a new minimization of the alanine tetrapeptide extended to helix transition was performed. With Mulliken charges, BUFF gives an excellent value of –3.41 kcal/mol, which is significantly better than any of the other methods tested (Table 2-20.)

Mulliken charges were applied to the minimization of crambin (1cbn) as well (Table 2-21.) The CRMS fit improved from 1.00 Å to 0.91 Å. The Mulliken charge scheme should also affect the torsions fit to the Gly-Gly-Gly and Gly-Ala-Gly tripeptides. Presumably, with a complete set of Mulliken charges and refit torsions, the BUFF 1cbn results should improve even further.

	Mopac2000	Amber94	Dreiding QEq	BUFF (ESP)	BUFF (Mulliken)
CRMS	0.10	0.72	0.54	1.00	0.91

Table 2- 21: All atom coordinate root mean square (CRMS) structural fits to 0.83 Å resolution 1cbn crystal structure. Structures were minimized, then matched to the original crystal structure to determine the approximate level of perturbation caused by the forcefield. BUFF calculations were performed with the standard ESP calculated charges, and a set of charges that are approximately what would be derived from HF if Mulliken charges were used as the basis for BUFF.

Conclusion

The Biological Universal Forcefield (BUFF) reproduces the torsional conformation energies of peptides as well as or better than most commonly used forcefields. The BUFF has the additional advantage of being a universal forcefield and can be easily applied to metal systems and unnatural amino acids without further parameterization. The charges calculated for amino acid groups using ESP fitting do not correctly reproduce the energies of folded and extended states like the extended to helix transition of a alanine tetrapeptide. Preliminary indications show that if Mulliken charges are used instead, much more reasonable energies are calculated. The process of fitting tripeptide torsions to correctly reproduce quantum mechanical energies has resulted in excellent torsion energies using ESP charges. It is expected that if Mulliken charges are used and the Gly-Gly-Gly and Gly-Ala-Gly torsion potential surfaces are refit, the potential surfaces will perform as least as well, while the electrostatic nonbond interactions will be greatly improved.

References

1. Cornell, W.D., *et al.*, *A 2nd Generation Force-Field For the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules.* J.A.C.S., 1995. **117**(19): p. 5179-5197.
2. MacKerell, A.D., *et al.*, *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* J. Phys. Chem. B, 1998. **102**(18): p. 3586-3616.
3. Jorgensen, W.L. and Tirado-Rives, J., *Development of the OPLS-AA force field for organic and biomolecular systems.* Abstracts of Papers of the American Chemical Society, 1998. **216**: p. U696-U696.
4. Rappe, A.K., *et al.*, *Uff, a Full Periodic-Table Force-Field For Molecular Mechanics and Molecular-Dynamics Simulations.* J.A.C.S., 1992. **114**(25): p. 10024-10035.
5. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *Dreiding - a Generic Force-Field For Molecular Simulations.* J. Phys. Chem., 1990. **94**(26): p. 8897-8909.
6. Brameld, K.A., *Molecular Modeling of Biological Systems: From Chitinase A to Z-DNA*, in *Chemistry*, 1999. California Institute of Technology: Pasadena.
7. Brameld, K., S. Dasgupta, and W.A.G. III, *Distance Dependent Hydrogen Bond Potentials for Nucleic Acid Base Pairs from ab Initio Quantum Mechanical Calculations.* J. Phys. Chem. B, 1997. **101**: p. 4851-4859.
8. Ringnalda, M.N., *et al.*, *Jaguar*, Schrodinger, Inc.: Portland, OR.
9. *BIOGRAF/POLYGRAF*, 1992. Molecular Simulations, Inc.
10. Collins, J.B., *et al.*, J. Chem. Phys., 1976. **64**: p. 5142.
11. Tannor, D.J., *et al.*, *Accurate First Principles Calculation of Molecular Charge-Distributions and Solvation Energies From Ab-Initio Quantum- Mechanics and Continuum Dielectric Theory.* J.A.C.S., 1994. **116**(26): p. 11875-11882.
12. Park, C., M.J. Carlson, and W.A. Goddard III, *Solvent effects on the secondary structures of proteins.* Journal of Physical Chemistry A, 2000. **104**(11): p. 2498-2503.
13. Rappe, A.K. and W.A.Goddard III, *Charge Equilibration for Molecular Dynamics Simulations.* J. Phys. Chem, 1991. **95**: p. 3358-3363.
14. Weiner, S.J., *et al.*, *A New Force-Field For Molecular Mechanical Simulation of Nucleic-Acids and Proteins.* J.A.C.S., 1984. **106**(3): p. 765-784.
15. Mohamadi, F., *et al.*, *MacroModel- An Integrated Software System for Modeling Organic and Bioorganic Molecules using Molecular Mechanics.* J. Comput. Chem., 1990. **11**: p. 440.
16. Park, C. and W.A.Goddard III, *Stabilization of α-helices by Dipole-Dipole Interactions Within α-helices.* J. Phys. Chem., 2000: p. *Submitted for publication.*
17. Stewart, J.J.P., *MOPAC: A General Molecular Orbital Package.* Quant. Chem. Prog. Exch., 1990. **10**: p. 86.
18. Teeter, M.M., S.M. Roe, and N.H. Heo, *Atomic Resolution (0.83 Angstroms) Crystal Structure of the Hydrophobic Protein Crambin at 130K.* J. Mol. Biol., 1993. **230**: p. 292.
19. Mulliken, R.S., *(Mulliken charges).* J. Chem. Phys., 1955. **23**: p. 1833.

20. Benini, S., *et al.*, *Crystal Structure of Oxidized Bacillus Pasteurii Cytochrome C553 Determined Ab Initio and by MAD Methods: Structure-Function Relationships and Molecular Evolution at 0.97 A Resolution. to be published .*

Chapter 3: An Examination of Solvent Effects on Peptide Torsions

Abstract

An examination of the effect of solvation on the conformational preferences (e.g., α–helix versus β–sheet) of tripeptides using *ab initio* quantum mechanics (Hartree-Fock 6-31G**) with solvation in the Poisson-Boltzmann continuum solvent approximation finds that aqueous solvent preferentially stabilizes the α–helix conformation over β–sheet conformations by 3.5 kcal/mol for Ala, 2.4 kcal/mol for Gly, and 2.0 kcal/mol for Pro. The torsional potential surfaces of the tripeptides, Gly-Ala–Gly, Gly-Gly-Gly, and Gly-Pro-Gly in vacuum, aqueous solvent, and nonpolar solvent conditions were examined. The results were used to demonstrate that simple force-field torsional corrections can be used to accurately reproduce the quantum mechanical potential surfaces.

Introduction

Determining the final folded structure of a protein from a specific sequence remains one of the central challenges to computational biochemistry [1-4]. Beginning with Chou and Fasman, statistical information from the protein databank has been used to predict the secondary structure of a protein from its primary sequence [5,6]. These models have not been able to produce predictions of the desired accuracy.

Good model systems have been developed which demonstrate the stability of α–helices in water [7,8], allowing measurement of the peptide preferences for α-helix conformations. However, model systems have not been as easy to develop to measure stable β-sheet conformations in water [9,10], resulting in little direct experimental evidence for the preferences of residues in β-sheets. Several theoretical studies have been directed at understanding the torsion preferences in secondary structures of a protein [11,12]. Most *ab initio* studies are on dipeptide model systems and have been carried out in a vacuum, ignoring solvent effects [13-16].

The accuracy of forcefield (FF) parameters for the protein main chain is clearly important in investigations seeking to determine a final folded structure of a protein from its primary sequence [1-4]. If the potential curves of a peptide's phi and psi angles are poorly represented by a forcefield, the modeled local protein structure will be distorted and a poor global protein structure may result as well.

Reported herein are *ab initio* quantum mechanical calculations (Hartree-Fock, 6-31G** basis) for the conformational energies of the Gly-Ala–Gly, Gly-Gly-Gly, and Gly-

Pro-Gly tripeptides in water. This is expected to mimic the solvation effects in hydrophilic environments (surface regions). The results show that solvation preferentially stabilizes the α–helix conformations over β–sheet conformations by 2 to 3.5 kcal/mol.

The final quantum mechanical potential surfaces are used to obtain a high quality forcefield using simple torsion terms for each of the peptide trimers. This forcefield is able to reproduce the quantum mechanical potential surfaces of the trimers to a fair degree of accuracy. Particular attention was paid to obtain a quality fit for the low energy regions of the potential surfaces.

Methods

All quantum chemical calculations were at the Hartree-Fock (HF) level using the 6-31G** basis for all atoms. All calculations used the Jaguar quantum chemistry program [17,18]. Solvation was included by solving the Poisson-Boltzmann equations with a realistic molecular surface (van der Waals radius plus solvent radius about each atom) using the Jaguar solvation model (PBF) [19]. ε was assumed to be 80 and R_0 was set at 1.4 Å based on using water as the solvent to mimic hydrophilic environments. The solvent effects were calculated self-consistently. At each iteration the wavefunction is calculated in the field of the solvent and then the charges (based on the electrostatic potential from the HF wavefunction) are used to calculate a new reaction field [19]. This process is repeated until convergence.

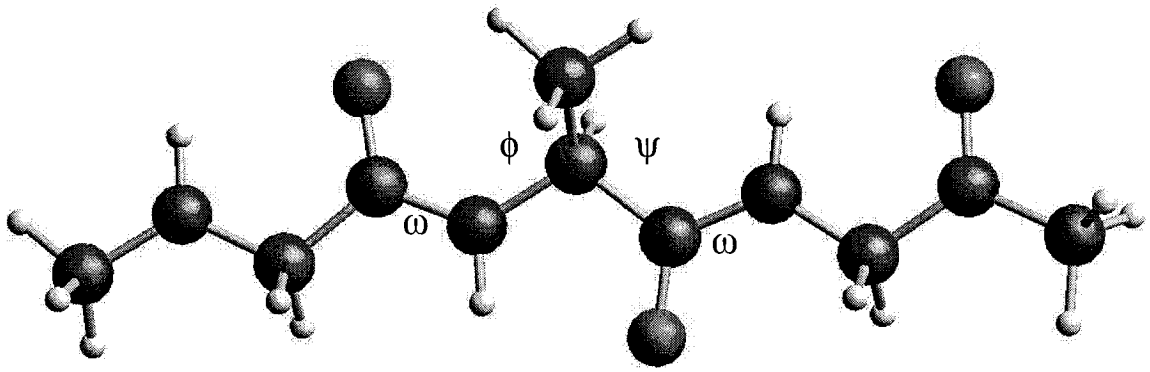

Figure 3- 34. The tripeptide model system for *ab initio* calculations. Both glycines were constrained to have the extended conformation shown for all conformations of the center amino acid. The conformational dihedral angles of the amino acid side chain were optimized for each ϕ and ψ. Shown is ϕ= 180° and ψ= 180°.

To establish the effect of environment on the conformation of amino acids, quantum mechanical calculations were first carried out on the model system (Gly-Ala–Gly) (Figure 3-1) for all ϕ and ψ torsional angles of the center alanine. The two glycines used to provide a proper environment for the central residue were constrained to the extended form (ϕ= 180° and ψ= 180°) for all conformations. The quantum mechanical calculations were carried out for every 60° of the ϕ and ψ torsional angles (36 points) of the center alanine plus three additional conformations corresponding to α–helix (ϕ= -57° and ψ= -47°) and β–sheet (ϕ= -119°and ψ= 113° for parallel and ϕ= -139° and ψ=135° for anti-parallel). The geometry of each of the 39 conformations was fully optimized, with the 3 ϕ,ψ torsions constrained, by quantum mechanical calculations (Hartree-Fock, 6-31G**basis) in vacuum. This leads to the contour maps in Figure 3-2 for the potential energy and solvation energy of alanine. Similar calculations were performed for the Gly-Gly-Gly (Figure 3-3) and Gly-Pro-Gly (Figure 3-4) cases.

A forcefield that reproduces the three potential maps was developed by first fitting the Gly-Gly-Gly potential fit to obtain appropriate backbone ϕ,ψ torsions (Figure 3-5). The Gly-Gly-Gly torsions were then applied to the Gly-Ala-Gly system and the forcefield was fit to the quantum mechanical potential map using the backbone torsions that include the C_β atom (Figure 3-6). This set of torsional parameters is intended to be of general use for all non-Gly, non-Pro residues. The final forcefield fit of Gly-Pro-Gly included all previous backbone and C_β torsions. The potential map was fit using the torsions involved with the backbone and the Nitrogen end of the proline ring (Figure 3-7).

The starting forcefield used the exponential-6 form of DREIDING [20] nonbonds, UFF [21] valence terms, and charges derived from a HF calculation on the extended form of each tripeptide. All torsions (except for the ω amide torsion) involved directly in the potential energy surface were first zeroed out. Then a constrained minimization was performed for each of the 39 HF data points. The difference between the molecular mechanics and the HF energies was fit to a torsional potential. Initial torsional fitting of the tripeptide potential surfaces used a Boltzmann weighting to attempt to fit the lowest energy points of each potential surface. However, it became necessary to make additional adjustments to some of the torsional parameters in order to reproduce the correct relative energies of local minima on each potential surface. The torsion functions are constructed such that glycine uses two torsional functions, one for ϕ and one for ψ. Alanine then requires the two glycine torsions as well as two new torsions that involve the C_β atom. The four special torsions for alanine are intended to be used for all amino acids but glycine, since glycine has no C_β atom. Proline has an additional special

torsional term involving its C_δ atom connected to the main chain nitrogen The final forcefield torsion parameters are shown in Table 2-1.

	A * Cos(θ) *A* in (kcal/mol)	*B* * Cos(2θ) *B* in (kcal/mol)	*C* * Cos(3θ) *C* in (kcal/mol)
all amino acids			
C-N-C_α-C (ϕ)	1.00	-1.00	-1.25
N-C-C_α-N (ψ)	1.50	-2.25	0.00
needed for non-glycine			
C_β-C_α-N-C	-1.70	-1.70	-0.20
C_β-C_α-C-N	1.20	0.60	0.40
proline-specific			
C_δ-N-C_α-C	0.00	0.00	1.50

Table 2- 22. The Force-Field Torsional Cosine Expansion Terms Used in Fit to the Quantum Mechanical Data. The torsion function is a simple cosine sum of the form: $E_{torsion} = A*Cos(\theta) + B*Cos(2\theta) + C*Cos(3\theta)$. Prior to the torsional fit, all involved torsions but the ω torsion were zeroed. The ω torsion, C_α–N-C-C_α, was not fit, but was left with a barrier of 10 kcal/mol and a periodicity of 2. C_δ is the side-chain δ-carbon of proline adjacent to the main chain nitrogen.

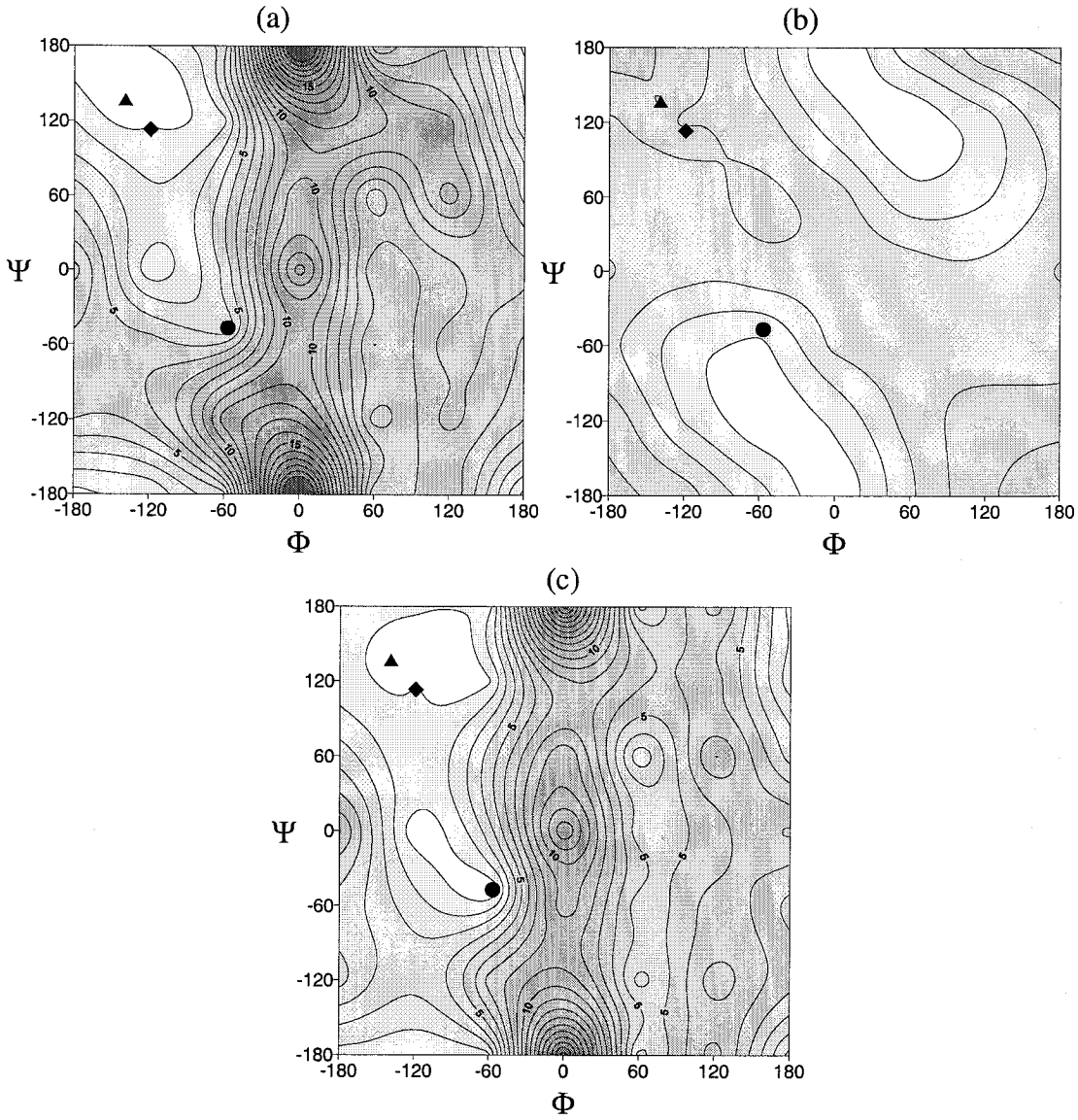

Figure 3- 35. Conformation energies for Gly-Ala–Gly. Each map is based on the energies for 36 pairs of torsional angles (ϕ= 60°, ψ= 60°) plus three additional energies corresponding to the α–helix (ϕ= -57°, ψ= -47°) indicated by solid circle, the parallel β–sheet (ϕ= -119° and ψ=113°) indicated by a solid diamond, and the antiparallel β–sheet (ϕ= -139° and ψ= 135°) indicated by a solid square. The bright region indicates stable conformations and the dark region indicates unstable conformations. The maps show clearly that solvent effects tend to stabilize the α–helical conformation compared to the β–sheet conformation. Contours are spaced at 1.0 kcal/mol intervals. (a) Vacuum HF results, (b) solvation energy for H2O, (c) total energy in H2O.

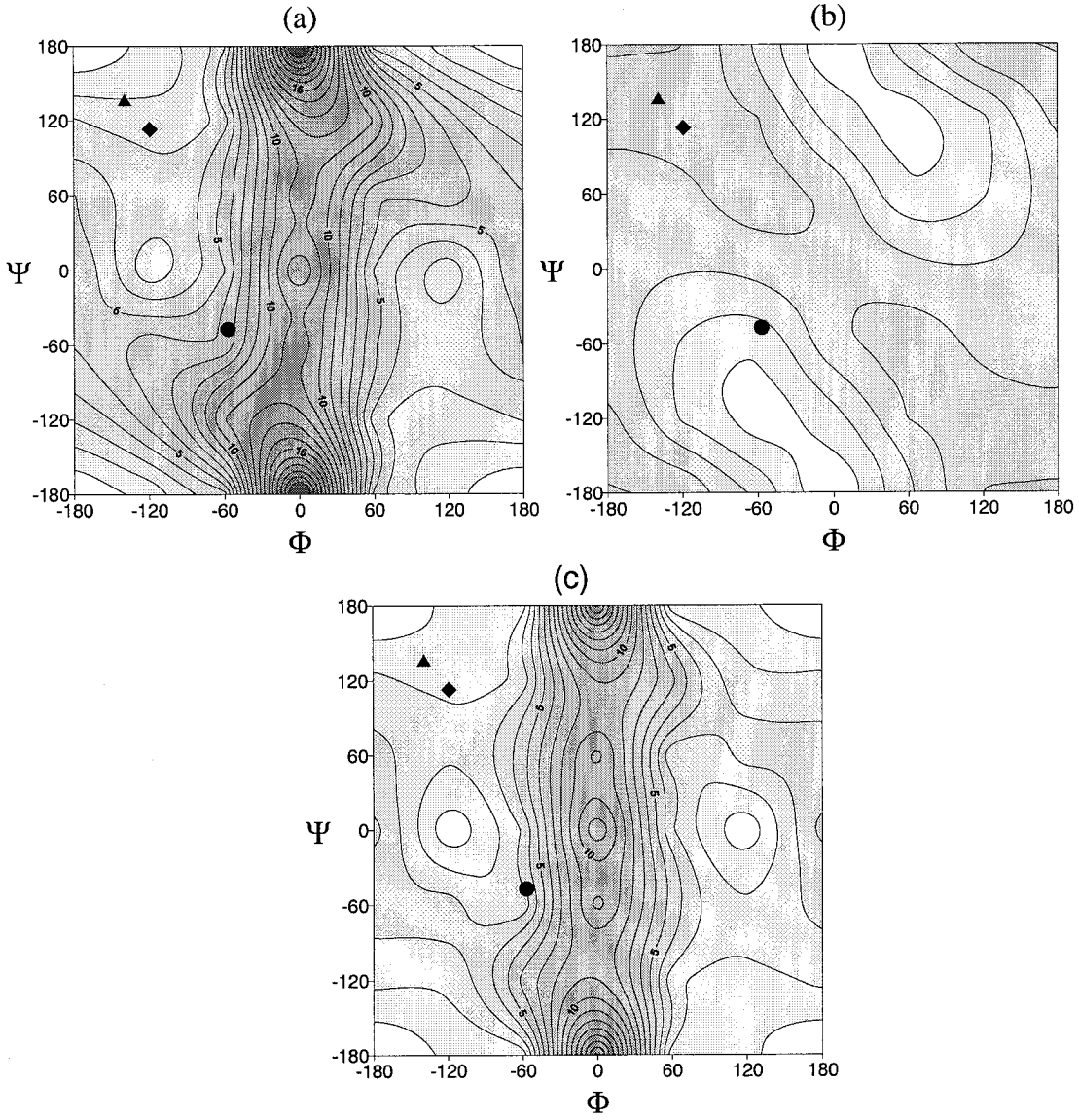

Figure 3- 36. Conformation energies for Gly-Gly-Gly. Contour details are the same as in Figure 3-2. These maps show clearly that solvent effects tend to stabilize the α–helical conformation compared to the β–sheet conformation. (a) Vacuum HF results, (b) solvation energy for H2O, (c) total energy in H2O.

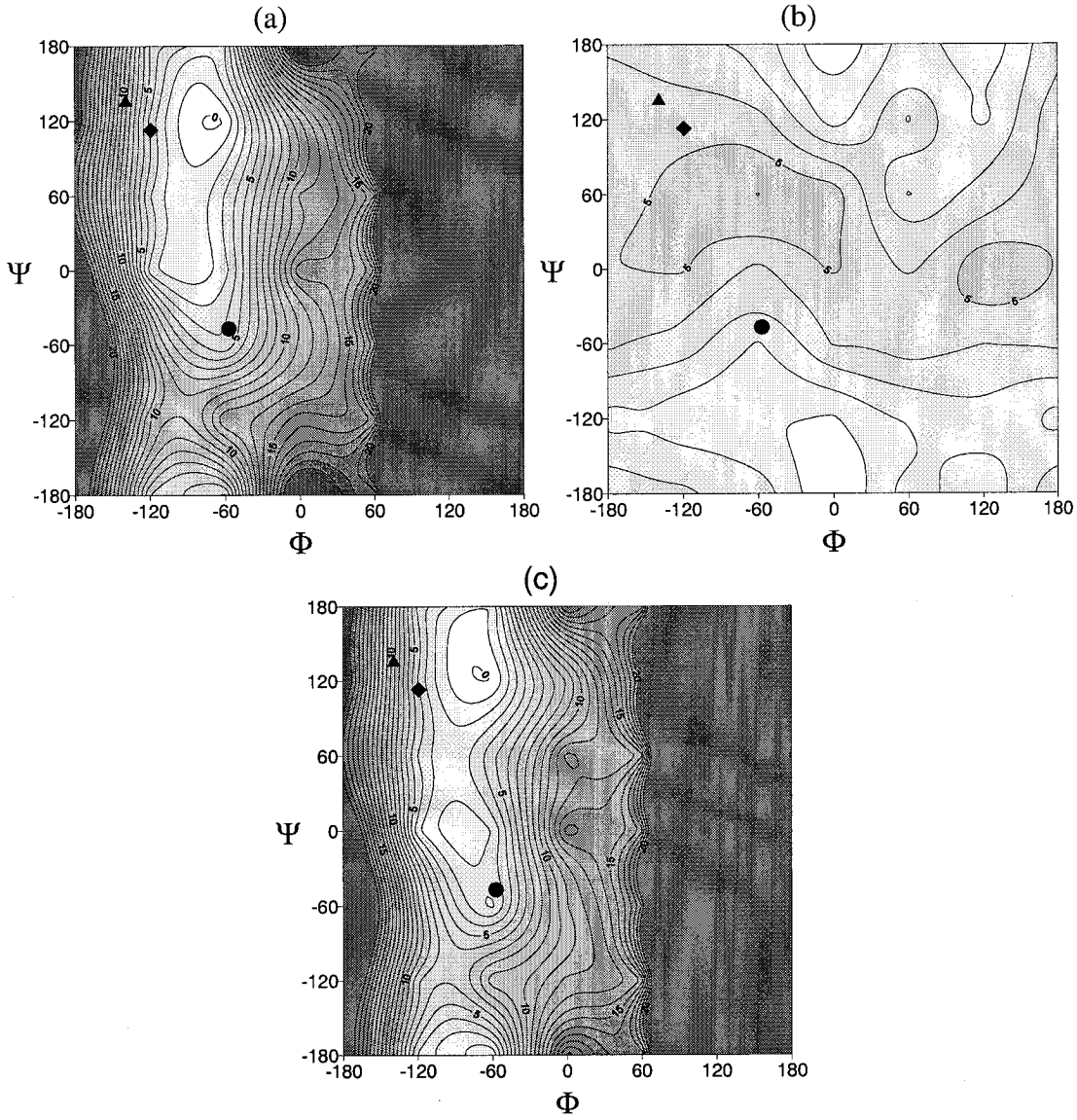

Figure 3- 37. Conformation energies for Gly-Pro-Gly. Contour details are the same as in Figure 3-2. Note that the maps clearly show that solvent effects tend to stabilize the α–helical conformation compared to the β–sheet conformation. (a) Vacuum HF results, (b) solvation energy for H2O, (c) total energy in H2O.

Results

Without solvation, the tripeptide results are similar to those of previous calculations [13,22]. Table 2-2 shows the apparent local minima and Table 2-3 shows the relative energy differences of the α–helix, parallel β–sheet, and antiparallel β–sheet conformations to the global minimum in each case with ϕ and ψ angles used in the current calculations.

Residue					**β-sheet**	**α-helix**	**α-helix**
alanine	(ϕ,ψ)	(-120,0)		(60,-120)	(-138,138)		(60,60)
	ΔE_{vac}	1.343		4.558	0.000		5.234
	(ϕ,ψ)	(0.-180)	(60,120)	(-60,-120)			
	ΔE_{sol}	0.039	0.000	0.238			
	(ϕ,ψ)	(-120,0		(60,-120)	(-120,120)	(-60,-48)	(60,60)
	ΔE_{wat}	0.338		3.592	0.000	0.281	2.138
glycine	(ϕ,ψ)	(±180,-180)	(±120,0)				
	ΔE_{vac}	0.000	2.280				
	(ϕ,ψ)	(0,-180)	(-60,-120)	(60,120)			
	ΔE_{sol}	0.000	0.408	0.408			
	(ϕ,ψ)	(±180,-180)	(±120,0)				
	ΔE_{wat}	0.000	0.322				
proline	(ϕ,ψ)				(-72,120)		
	ΔE_{vac}				0.000		
	(ϕ,ψ)	(0,180)					
	ΔE_{sol}	0.000					
	(ϕ,ψ)				(-72,126)	(-60,-48)	
	ΔE_{wat}				0.000	0.263	

Table 2- 23. The Energy Minima and the Energy Differences of the Minima to the Global Minimum Are Shown with the Conformational ϕ and ψ Angles. ΔE_{vac}: relative total energy in vacuum. ΔE_{sol}: relative solvation energy in water. ΔE_{wat}: relative total energy in water.

Residue	conformation	ΔE_{vac} (kcal/mol)	ΔE_{sol} (kcal/mol)	ΔE_{wat} (kcal/mol)
Alanine	α-helix	3.448	1.227	0.282
	p-β-sheet	1.238	0.483	1.268
	a-β-sheet	0.180	5.057	0.783
Glycine	α-helix	4.603	1.664	1.168
	p-β-sheet	3.070	4.145	2.116
	a-β-sheet	0.767	4.991	0.659
Proline	α-helix	2.492	2.198	0.281
	p-β-sheet	3.861	4.861	4.314
	a-β-sheet	7.166	3.672	6.429

Table 2- 24. The Relative Energy (kcal/mol) of the α-Helix and β–Sheet Conformations to the Global Minimum in Vacuum and Water. All energies are from *ab initio* calculations (HF, 6-31G** basis) on Gly-X-Gly with a Poisson-Boltzmann description of the solvent. α–Helix is a right-handed α–helix, where (ϕ,ψ)= (-57, -47). p-β–sheet is a parallel β–sheet, where (ϕ,ψ) =(-119, 113). a–β–sheet is an antiparallel β–sheet, where (ϕ,ψ) =(-139, 135). ΔE_{vac}: relative total energy in vacuum; ΔE_{sol}: relative solvation energy in water; ΔE_{wat}: relative total energy in water.

The absolute minima of alanine at (ϕ= -138, ψ= 138) and proline at (ϕ= -72, ψ= 120) correspond to a β–sheet while the absolute minimum of glycine at (ϕ= 180, ψ= 180) is the extended conformation. For alanine, the potential surface has a channel pointing toward the α–helix region with local minima at (ϕ= -120, ψ= 0) about 1.343 kcal/mol higher. The actual α–helix conformations of alanine, glycine, and proline are not minima (E= 3.448 kcal/mol, 4.603 kcal/mol, and 2.492 kcal/mol, respectively). In addition, there are relative minima for alanine at (ϕ= -120, ψ= 0) with E= +1.343, at (ϕ=60, ψ= 60) with +5.234 and at (ϕ= 60, ψ= -120) with+4.558 kcal/mol. Glycine has two relative minima at (ϕ= 120,ψ= 0) and (ϕ= -120, ψ= 0), but proline has none.

In water the solvation energies for alanine (22.508 to 27.875 kcal/mol) (Figure 3-2b), glycine (24.113 to 29.212 kcal/mol) (Figure 3-3b), and proline (21.306 to 27.336

kcal/mol) (Figure 3-4b) are large. Though the solvation energies of proline show the strongest preferences only at (ϕ= 0, ψ= 180), those of alanine and glycine show the strongest preferences at (ϕ= 60, ψ= 120), (ϕ= -60, ψ= -120), and (ϕ= 0, ψ= 180). However, the solvation energies with water are strongly biased toward the α–helix conformation for all the three cases. For alanine, this biased solvation effect leads to a minimum at (ϕ= -120, ψ= 120) corresponding to the β–sheet and a second minimum at (ϕ= -60, ψ= -48) corresponding to the α–helix, now only 0.281 kcal/mol higher.

Table 2-3 shows that solvation dramatically changes the relative energy between the α–helix and parallel β–sheet conformations. Thus, ΔE_{vac} (α–helix) - ΔE_{vac} (p-β–sheet), changes

- from +2.210 kcal/mol in vacuum to -0.986 kcal/mol in water for alanine (δ= -3.2 kcal/mol),
- from +1.533 kcal/mol in vacuum to -0.948 kcal/mol in water for glycine (δ= 2.5 kcal/mol), and
- from -1.369 kcal/mol in vacuum to -4.033 kcal/mol in water for proline (δ= 2.7 kcal/mol).

The net aqueous stabilization of α–helix in kcal/mol are Ala(3.2), Gly (2.1), and Pro (2.7). Similar stabilization is observed between the α–helix and antiparallel β–sheet conformation, where ΔE_{vac}(α–helix) - ΔE_{vac}(α–β–sheet) changes

- from +3.268 kcal/mol in vacuum to -0.501 kcal/mol in water for alanine,
- from +3.836 kcal/mol in vacuum to 0.509 kcal/mol in water for glycine, and

- from -4.674 kcal/mol in vacuum to -6.148 kcal/mol in water for proline.

The net solvent stabilization of α–helix in kcal/mol are Ala(3.8), Gly (3.3), and Pro (2.5).

It was possible to fit the QM potential surfaces by adjusting the torsions that involve the main chain atoms in the protein backbone. In the Gly-Gly-Gly tripeptide (Figure 3-5), the location and well depth of all low energy local minima are reproduced by the forcefield parameters. The broad lower energy surfaces are only slightly narrower in the forcefield, with a slightly smaller beta sheet / extended low energy region than is found in the QM potential surface. The high energy regions are, in general, significantly higher than the QM due to the steric interactions of the forcefield, but this should not be a problem since they are effectively unpopulated at typical biological temperatures.

Figure 3-6 compares the QM to the forcefield of the Gly-Ala-Gly tripeptide. Again, the local minima have approximately correct values and locations. The minimum around (-90,0) is lightly lower than the QM but the global minimum is still correct. We find that the relative energies of the low lying regions are in excellent agreement with QM energies. However, the extended / beta sheet region is again slightly smaller than found in the QM. In the positive ϕ region, a local minima occurs horizontally along $\psi=0$ rather than vertically along $\phi = 60$, even though the relative energy of the local minima is approximately correct. Since this is a higher energy region, it is also less populated and should be acceptable for most uses. Finally, the Gly-Pro-Gly (Figure 3-7) forcefield results are shown to be just slightly more constrained by high barrier walls than the sides of the QM well. The local minima are similar in both plots with the surface area of the lowest energy regions being approximately equal. Overall, the forcefield makes an

excellent reproduction of the QM potential energy surfaces. Low energy regions are faithfully reproduced. The high energy regions near $\phi = 0$ tend to be higher in energy in the forcefield than in QM, resulting in slightly narrower low energy regions in the forcefield potential energy surfaces. Since the high-energy regions are effectively unpopulated for typical biological temperatures, they should not adversely affect most calculations.

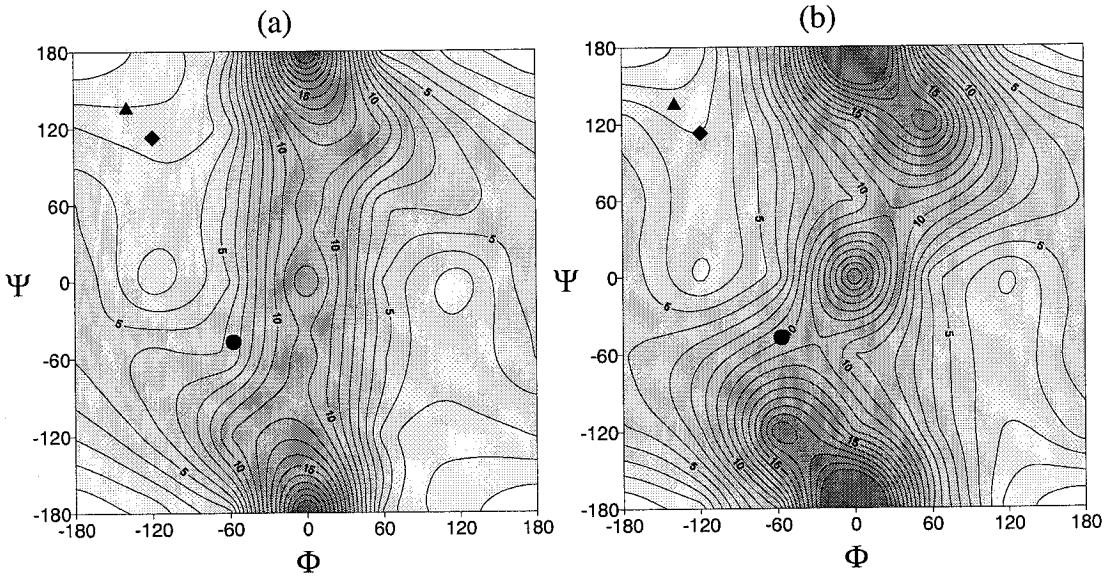

Figure 3- 38. Quantum mechanical (HF/6-31G**) energies (a) and forcefield energies (b) for Gly-Gly-Gly in vacuum. The same 39 data points were used as in Figure 3-2. Contours are spaced at 1.0 kcal/mol intervals.

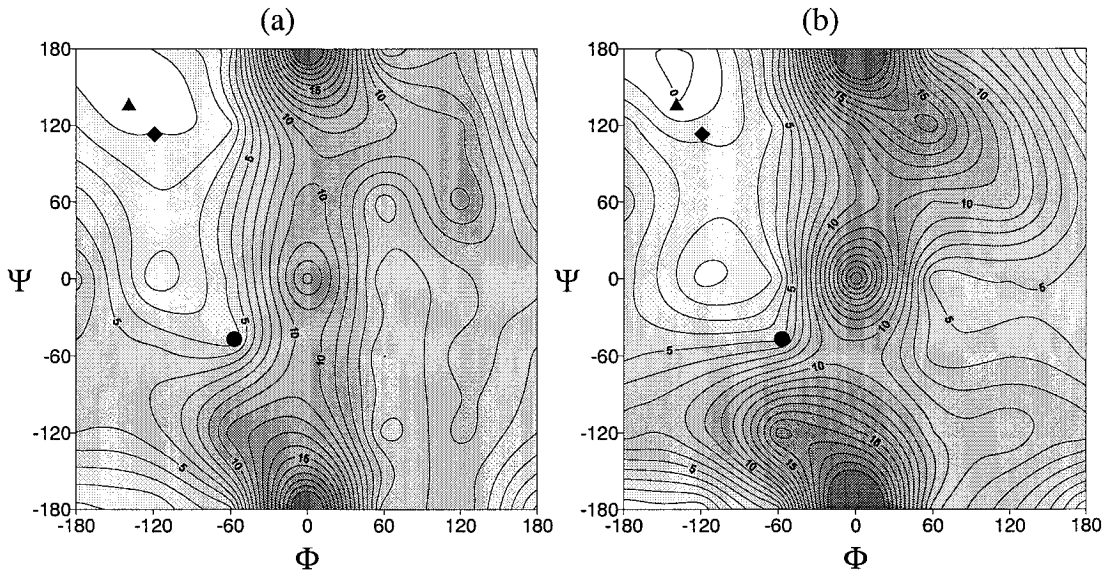

Figure 3- 39. Quantum mechanical (HF/6-31G**) energies (a) and forcefield energies (b) for Gly-Ala-Gly in vacuum. The same 39 data points were used as in Figure 3-2. Contours are spaced at 1.0 kcal/mol intervals.

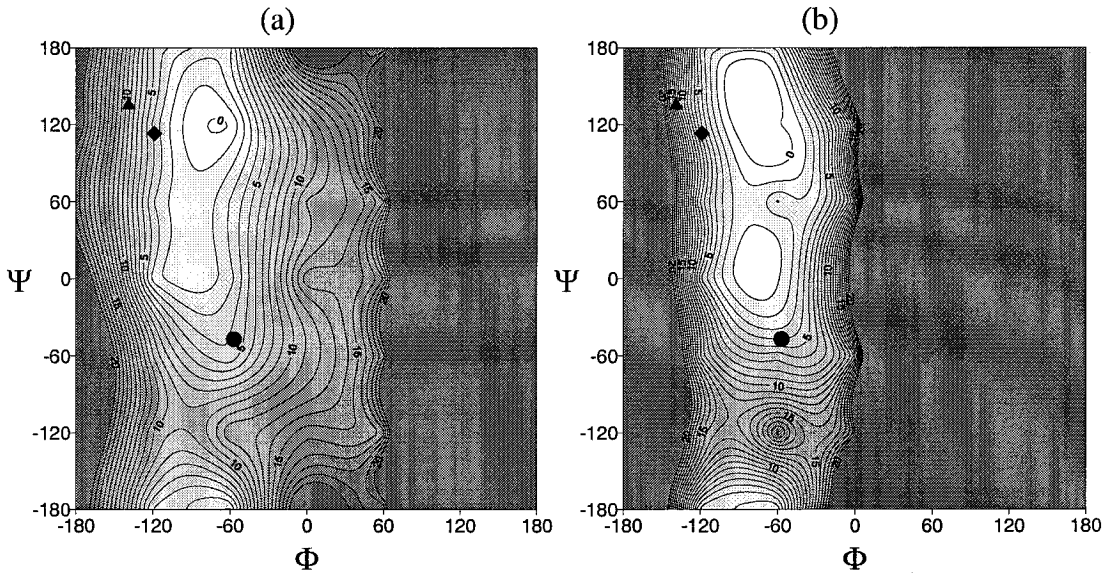

Figure 3- 40. Quantum mechanical (HF/6-31G**) energies (a) and forcefield energies (b) for Gly-Pro-Gly in vacuum. The same 39 data points were used as in Figure 3-2. Contours are spaced at 1.0 kcal/mol intervals.

Discussion

Various results of theoretical calculations [23-25] on model systems for amino acids in vacuum have shown that the right-handed α–helix conformation is not stable (with a few exceptions [26]) while the β–sheet conformation is quite stable. This is not consistent with experiment, and hence it has been proposed that the right-handed α–helix must be stabilized by specific nonbond interactions [13]. It has been suggested that the α–helical conformation is destabilized compared to the β–sheet conformation by the dipole moment interaction between the side chain and the backbone [27].

In contrast, the calculations for water show that this unfavorable dipole moment of the α–helix induces a stronger solvent effect in water, leading to an α–helical conformation nearly as stable as the β–sheet conformation in water (the solvation energy is directly related to the dipole moment of the solute). This strong solvent effect in water for the α–helical conformation agrees with earlier thermodynamic studies, which included the solvent effects on an alanine dipeptide [13]. The HF calculations show that proline's α–helix conformation is more stable (4 to 6 kcal/mol) than its β–sheet conformation and for alanine the α–helical conformation (0.501 kcal/mol), is slightly more stable than the β–sheet conformation (0.986 kcal/mol). For glycine the helical conformation is more stable than parallel β–sheet conformation (0.948 kcal/mol) but less stable than the antiparallel β–sheet conformation (0.509 kcal/mol). This becomes clear for the calculations on the 10 nonpolar amino acids for the α–helix and β–sheet conformations. For all 10 relatively hydrophobic amino acids the β–sheet is more stable than the α–helix conformation in vacuum (hydrophobic environment), but in water the

stability of the α–helix conformation becomes very close to that of the β–sheet conformation due to the strong solvent stabilization of the α–helix [28].

These results support the observation that β–sheets usually occur only inside folded proteins. This is because a protein's interior is usually hydrophobic, favoring the β–sheet conformation. These results are also supported by experiments which show: (i) a transition of polylysine from the α–helix to β–sheet conformations by the addition of anesthetics, and (ii) a transition of polylysine from β–sheet to α–helix occurs by applying pressure [29]. The anesthetics induce a partial dehydration of the polypeptide side chains, creating a more hydrophobic environment favorable for β–sheet conformation for the polypeptide [30]. In contrast, the applied pressure seems to push water near the side chains and makes the environment more hydrophilic [31,32].

These results support the observation that hydrophobic residues have high preferences and polar residues have low preferences for the β–sheet secondary structure [5,6]. Hydrophobic residues are more likely to be inside the protein (in a hydrophobic environment) than are hydrophilic residues while hydrophilic residues have relatively high probabilities to be placed on the exterior of proteins compared to the hydrophobic residues. These results also explain the many α–helix models stable in water, making it easy to study the properties of α–helices, while there are very few β–sheet models stable in water, making it difficult to study β–sheets [13,26]. These conclusions are supported by results that show the presence of a hydrophobic core is essential for the formation of a β–sheet [9,10].

Peptides from the prion protein induce conformational transitions due to addition of acetonitrile and/or salts [33]. The added denaturants make the microenvironment around the peptides more hydrophobic, causing a conformational change in the peptides from α–helix to β–sheet. This observation is consistent with our results, thus providing a possible insight into explaining the Creutzfeldt-Jakob disease, the most common human prion disease [35]. These results also show that for the case of alanine and glycine the α–helical conformation is comparable to the β–sheet conformation in water. For the case of proline the α–helical conformation is much more stable than the β–sheet conformations both in water and in vacuum. This seems to contradict the observation that a proline residue tends to destroy the formation of an α–helix. Proline residues destabilize the α–helix because of the pyrrolidine ring attached to the imide nitrogen. Its presence matters only when the succeeding residue is a proline. The steric interactions of a residue are independent of the nature of the predecessor because only the carbonyl group(C=O) of the preceding residue is involved [36]. This is supported by the observation that proline residues are one of the best residues to initiate an α–helix [37]. The QM results show that an α–helix conformation is stabilized by solvation with water, providing insight into understanding the role of interactions between solvents and proteins in guiding protein folding.

Conclusion

We find that solvents have a significant effect on the conformation of polypeptides. We believe that these effects play an important role in protein folding. We report torsional parameters to use in chemical MD calculations.

References

1) Taylor, W. R.; Thornton, J. M. *J. Mol. Biol.* 1984, **173**, 487.
2) Finkelstein, A. V.; Reva, B. A. *Nature* 1991, **351**, 497.
3) Bowie, J. U.; Luthy, R.; Eisenberg, D. *Science* 1991, **253**, 164.
4) Jones, D. T.; Taylor, W. R.; Thornton, J. M. *Nature* 1992, **358**, 86.
5) Fasman, G. D. *Prediction of Protein Structure and the Principles of Protein Conformation*; Plenum: New York, 1989.
6) Chou, P. Y.; Fasman, G. D. *Biochemistry* 1978, **13**, 222.
7) Poland, D.; Scheraga, H. A. *Theory of Helix-Coil Transitions in Biopolymers*; Academic: New York, 1970.
8) Lockhart, D. J.; Kim, P. S. *Science* 1992, **257**, 947.
9) Diaz, H.; Tsang, K. Y.; Choo, D.; Kelly, J. W. *Tetrahedron* 1993,**49**, 3533.
10) Tsang, K. Y.; Diaz, H.; Graciani, N.; Kelly, J. W. *J. Am. Chem. Soc.* 1994, **166**, 3988.
11) Gould, I. R.; Kollman, P. A. *J. Phys. Chem.* 1992, **96**, 9255.
12) Schafer, L.; Klimkovsky, V. J.; Momany, F. A.; Chuman, H.; vanAlsenoy, C. *Biopolymers* 1984, **23**, 2335.
13) Tobias, D. J.; Brooks, C. L., III. *J. Phys. Chem.* 1992, **96**, 3864.
14) Pettitt, B. M.; Karplus, M. *Chem. Phys. Lett.* 1985, **121**, 194.
15) Pettitt, B. M.; Karplus, M. *J. Phys. Chem.* 1988, **92**, 3994.
16) Anderson, A.; Hermans, J. *Proteins* 1988, **3**, 262.
17) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.;Langlois, J.-M.; Goddard, W. A., III; Donnelly, R. E.; Ringnalda, M. N. *J. Chem. Phys.* 1994, **101**, 4028.
18) Ringnalda, M. N.; Langlois, J.-M.; Greeley, B. H.; Murphy, R. B.;Russo, T. V.; Cortis, C.; Muller, R. P.; Marten, B.; Donnelley, R. E., Jr.;Mainz, D. T.; Wright, J. R.; Pollar, T. W.; Gao, Y.; Won, Y.; Miller, G.H.; Goddard, W. A., III; Friesner, R. A. *PS-GVB, v2.24*; 1995.
19) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Stikoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* 1994, **116**, 11875.
20) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. J. Phys. Chem1990, **94**, 8897.
21) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* 1992, **114**, 10024.
22) McAllister, M. A.; Perzel, A.; Csaszar, P.; Viviani, W.; Rivail, J.-L.;Csizmadia, I. G. *J. Mol. Struct. (Theochem)* 1993, **288**, 161.
23) Klimkowski, V. J.; Schafer, L.; Momany, F. A.; van Alsenoy, C.*J. Mol. Struct.* 1985, **124**, 143.
24) Schafer, L.; van Alsenoy, C.; Klimkowski, V. J.; Scarsdale, J. N. *J. Chem. Phys.* 1982, **76**, 1439.
25) Schafer, J. N.; van Alsenoy, C.; Klimkowski, V. J.; Schafer, L.; Momany, F. A. *J. Am. Chem. Soc.* 1983, **105**, 3438.

26) Perzel, A.; Farkas, O.; Csizmadia, I. G. *J. Am. Chem. Soc.* 1996,**111**, 7809.
27) Perzel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* 1991, **113**, 6256.
28) Park, C.M. *Unpublished data.*
29) Chiou, J.-S.; Tatara, T.; Sawamura, S.; Kaminoh, Y.; Kamaya, H.;Shibata, A.; Udea, I. *Biochim. Biophys. Acta* 1992, **1119**, 211.
30) Lyu, P. C.; Lif, M. I.; Marky, L. A.; Kallenbach, N. R. *Science1990*, **250**, 669.
31) Carrier, D.; Mantsch, H. H.; Wong, P. T. T. *Biochemistry* 1990,**29**, 254.
32) Carrier, D.; Mantsch, H. H.; Wong, P. T. T. *Biopolymers* 1990,**29**, 837.
33) Zhang, H.; Kaneko, K.; Nguyen, J. T.; Livshits, T. L.; Baldwin,M. A.; Cohen, F. E.; James, T. L.; Prusiner S. B. *J. Mol. Biol.* 1995, **250**, 514.
34) Plaxco, K. W.; Morton, C. J.; Grimshaw, S. B.; Jones, J. A.; Pitkeathly, M.; Campbel, I. D.; Dobson, C. M. *J. Biomol. NMR* 1997, **10**,221.
35) Prusiner, S. B. *Science* 1991, **252**, 1515.
36) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry Part I: The Conformation of Biological Macromolecules*; W. H. Freeman and Company: San Francisco, 1980.
37) Richardson, J. S.; Richardson, D. C. *Science* 1988, **240**, 1648.

Appendix A – BUFF Parameters

```
PARAMETER FORMAT (11-89)
  233    5    0    0    0    0    0    0    0    0    0    0    0    0    0    0
UFF parameters for all H amino acid case
*
*  Parameters Current as of April 27, 2000.
*
FORCEFIELD GENFF
DEFAULTS  xxxdataxxx plus comments
LBOND      T         T >> use bond terms
LANGLE     T         T >> use angle terms
ANGX 2 K   F         T >> use true force constants for cosine ang-str cross terms
ANGANGINV  F         T >> use angle-angle inversion terms
LINVERSN   T         T >> use inversion terms
ALL INVER  T         T >> use all possible inversion terms on each center
BNDXANG    F         T >> use bond cross angle terms
ANGXANG    F         T >> use angle cross angle terms
LTORSION   T         T >> use torsion terms
BNDBNDTOR  F         T >> allow coupling of the 1-2 and 3-4 bonds of torsions
ANGANGTOR  F         T >> allow coupling of the 1-2-3 and 2-3-4 angles of torsions
LPITWIST   F         T >> use pi twist terms
TORS SCAL  T         T >> will renormalize torsions (not allow SNGTOR)
ALL TORSN  T         T >> use all possible torsion terms per each central bond
ETOR SCAL     1.0000 exocyclic scaling factor
TORANGSW   F         T >> switch torsion barrier off as angle becomes linear
TORANGR    135. 180. on and off angles for torsion angle switch
UREYBRAD   F         T >> use urey-bradley terms
LNONBOND   T         T >> use nonbond terms
RNB GEOMN  T         T >> use geom mean for nonbond cross terms
NBEXBND    T         T >> exclude 1-2 terms from nonbonds
NBEXANG    T         T >> exclude 1-3 terms from nonbonds
NBEXTOR    F         T >> exclude 1-4 terms from nonbonds
DOALLCOUL  F         T >> do NOT exclude coulomb terms from nonbonds
SCAL NB14     1.0000 factor scale 1-4 nonbonds (1.0 >> full value)
SHRINK CH  F         T >>  allow shrunk CH bonds for
SHRINK FC     1.0000 shrink factor for CH bonds
LCOULMB    T         T >> use Coulomb terms
R*EPS      F         T >> use shielded Coulomb 1/(eps*R**2) instead of 1/(eps*R)
DIELCTRIC     1.0000 Dielectric constant, eps
LHBOND     F         T >> use hb interactions
ATM DEFLT C_3        default atom for FF
MASSZER    F         T >> use zero mass option
POLYENE    F         T >> use polyene option
USRENERGY  F         T >> use user energy expression
*
FFLABEL    ATNO MODIFD MS CHARG HYB BND CPK #IH #LP RES
H_            1            0.00   0   1   8   0   0   F
H___A         1            0.00   0   1   8   0   0   F
H__O3         1            0.00   0   1   8   0   0   F
H__N3         1            0.00   0   1   8   0   0   F
H__NR         1            0.00   0   1   8   0   0   F
H_N3+         1            0.00   0   1   8   0   0   F
H_NR+         1            0.00   0   1   8   0   0   F
H__S3         1            0.00   0   1   8   0   0   F
C_3           6            0.00   3   4   5   0   0   F
C_A           6            0.00   3   4   5   0   0   F
C_R           6            0.00   2   3   5   0   0   T
C_2           6            0.00   2   3   5   0   0   F
N_3           7            0.00   3   3   7   0   0   F
N_R           7            0.00   2   3   7   0   0   T
N_R2          7            0.00   2   3   7   0   0   T
O_3           8            0.00   3   2   2   0   0   F
O_2           8            0.00   2   1   2   0   0   F
O_2m          8            0.00   2   1   2   0   0   F
S_3          16            0.00   3   2   3   0   0   F
```

```
H_F           1    1.0080  0.41   0   0   8   0   0
O_3F          8   15.9994 -0.82   3   0   2   0   2   F
P_3          15            0.00   3   3   3   0   1
Cl           17           -1.00   0   1   4   0   3
Br           35           -1.00   0   1   4   0   3
Na           11            1.00   0  -6   1   0   0
Ca           20            2.00   0  -4   1   0   0
Fe3+2        26            3.00   0  -6   6   0   0
Fe6+2        26            3.00   0  -6   6   0   0
Zn           30            2.00   0  -4   1   0   0
Ru           44            3.00   0  -6   6   0   0
*
ADDED H   HYDROGEN  1IMPLCTH  2IMPLCTH  3IMPLCTH  4IMPLCTH
H_        H_
H___A     H_
C_3       H_
C_A       H_
C_R       H_
C_2       H_
N_3       H_N3+
N_R       H__NR
N_R2      H_NR+
O_3       H__O3
O_2       H___A
S_3       H__S3
H_F       H_F
O_3F      H_F
*
LONE PAIRS
*
VDW AT ITY       RNB      DENB     SCALE
*LJ12-6  1        Re        De  not used      1/R12 fct       1/R6 fct
*exp-6   2        Re        De  exp scal      pre-expon       dispersn  exponent
*morse   3        Re        De  exp scal
*pur exp 5        Re        De  exp scal      pre-expon       not used  exponent
*
* Dreidiing NB  (Directly from paper.  Exp-6)
H_       2   3.19500   0.01520  12.38200    17198.63477       32.33693
H___A    2   3.19500   0.00010  12.00000      113.14890        0.21274
H__O3    2   3.19500   0.00010  12.00000      113.14890        0.21274
H__N3    2   3.19500   0.00010  12.00000      113.14890        0.21274
H__NR    2   3.19500   0.00010  12.00000      113.14890        0.21274
H_N3+    2   3.19500   0.00010  12.00000      113.14890        0.21274
H_NR+    2   3.19500   0.00010  12.00000      113.14890        0.21274
H__S3    2   3.19500   0.00010  12.00000      113.14890        0.21274
C_3      2   3.89830   0.09510  14.03400  1171341.25000      667.51642
C_A      2   3.89830   0.09510  14.03400  1171341.25000      667.51642
C_2      2   3.89830   0.09510  14.03400  1171341.25000      667.51642
C_R      2   3.89830   0.09510  14.03400  1171341.25000      667.51642
N_3      2   3.66210   0.07740  13.84300   450301.56250      373.38098
N_R      2   3.66210   0.07740  13.84300   450301.56250      373.38098
N_R2     2   3.66210   0.07740  13.84300   450301.56250      373.38098
O_3      2   3.40460   0.09570  13.48300   232115.98438      298.08386
O_2      2   3.40460   0.09570  13.48300   232115.98438      298.08386
O_2m     2   3.40460   0.09570  13.48300   232115.98438      298.08386
S_3      2   4.03000   0.34400  12.00000  6312761.00000     2947.26294
P_3      2   4.15000   0.32000  12.00000
Cl       2   3.95030   0.28330  13.86100
Br       2   3.95000   0.37000  12.00000
Na       2   3.14400   0.50000  12.00000
Ca       2   3.47200   0.05000  12.00000
Fe3+2    2   4.54000   0.05500  12.00000
Fe6+2    2   4.54000   0.05500  12.00000
Zn       2   4.54000   0.05500  12.00000
Ru       2   4.54000   0.05500  12.00000
*
* F3C Nonbonds
H_F      1   0.90000   0.01000   12.0000
O_3F     1   3.55320   0.18480   12.0000
```

```
*H_F       6   3.19500   0.00000   0.00000
*O_3F      6   3.40460   0.00000   0.00000
*
NONBOND-OFF
*IIII-JJJJJ
*LJ12-6         1         Re        De  not used      1/R12 fct       1/R6 fct
*exp-6          2         Re        De  exp scal      pre-expon       dispersn  exponent
*morse          3         Re        De  exp scal
*pur exp        5         Re        De  exp scal      pre-expon       not used  exponent
*LJ12-10        7         Re        De  not used      1/R12 fct      1/R10 fct
*
* F3C off-diagonal
*O_3F -H_F       1   3.29800   0.03800  12.93250
O_3F -O_3F      1   3.57237   0.15047   0.00000
*
* SDG Nylon pure repulsive H-bonding term
O_2  -H___A     5   3.01696   0.02800  12.00000
O_2m -H___A     5   3.01696   0.02800  12.00000
O_3  -H___A     5   3.01696   0.02800  12.00000
N_R  -H___A     5   3.01696   0.02800  12.00000
N_R2 -H___A     5   3.01696   0.02800  12.00000
*
*  BUFF H__O3 donors
O_3  -H__O3     3   2.00000   1.50000  10.60000
H__O3-H__O3     2   3.50000   0.20000   9.76000
O_2  -H__O3     3   2.20000   0.83400   8.86000
O_2m -H__O3     3   2.29000   2.90000   6.86000
N_R  -H__O3     3   2.37000   1.86000   7.51000
N_R2 -H__O3     3   2.37000   1.86000   7.51000
*
*  BUFF H_N3+ donors
O_3  -H_N3+     5   3.16000   0.10000   8.00000
O_2  -H_N3+     3   2.00000   2.90000   6.60000
O_2m -H_N3+     5   2.36000   3.45000   5.39000
N_R  -H_N3+     3   2.20000   4.25000   5.70000
N_R2 -H_N3+     3   2.20000   4.25000   5.70000
*
*  BUFF H_NR+  donors
O_3  -H_NR+     3   2.50000   3.51000   5.84000
O_2  -H_NR+     3   3.66000   0.21400   7.95000
O_2m -H_NR+     3   2.09000   3.68000   6.22000
N_R  -H_NR+     3   3.24000   0.61000   7.46000
N_R2 -H_NR+     3   3.24000   0.61000   7.46000
*
*  BUFF H__N3 donors
* Not implemented yet since H__N3 is rare in proteins...
*  H__N3 is usually charged and thus H_N3+
O_3  -H__N3     5   3.01696   0.02800  12.00000
O_2  -H__N3     5   3.01696   0.02800  12.00000
O_2m -H__N3     5   3.01696   0.02800  12.00000
N_R  -H__N3     5   3.01696   0.02800  12.00000
N_R2 -H__N3     5   3.01696   0.02800  12.00000
*
*  BUFF H__NR donors
O_3  -H__NR     3   2.63000   0.29100   6.77000
O_2  -H__NR     3   2.58000   0.18600  10.00000
O_2m -H__NR     3   2.34000   2.41000   6.27000
N_R  -H__NR     3   3.73000   1.35000   5.27000
N_R2 -H__NR     3   3.73000   1.35000   5.27000
*
*  BUFF H__S3 donors
O_3  -H__S3     3   2.52000   0.76000   8.26000
O_2  -H__S3     3   3.07000   0.07700  10.63000
O_2m -H__S3     3   1.80000   8.44000   4.62000
N_R  -H__S3     3   2.44000   3.31000   8.00000
N_R2 -H__S3     3   2.44000   3.31000   8.00000
*
BONDSTRTCH  TYPE
*  morse        2 FORC CNST  BND DIST    BOND E
```

```
*  uff          8        Ke0        Re0  elec dRe       Ren       Ken
*simp harm      1 FORC CNST  BND DIST
*    (Put in using harmonic bonstretch.)
N_R  -H___A     1 1030.9469  1.053000   -0.0096    1.0434 1059.5963    0.0000
N_R  -H__NR     1 1030.9469  1.053000   -0.0096    1.0434 1059.5963    0.0000
N_R  -H_NR+     1 1030.9469  1.053000   -0.0096    1.0434 1059.5963    0.0000
N_R  -C_3       1 1046.4963  1.456000   -0.0059    1.4501 1059.3770    0.0000
N_R  -C_A       1 1046.4963  1.456000   -0.0059    1.4501 1059.3770    0.0000
N_R  -H_        1 1030.9469  1.053000   -0.0096    1.0434 1059.5963    0.0000
N_R2 -H___A     1 1030.9469  1.053000   -0.0096    1.0434 1059.5963    0.0000
N_3  -H___A     1 1028.0154  1.054000   -0.0096    1.0444 1056.5662    0.0000
N_R2 -H__NR     1 1030.9469  1.053000   -0.0096    1.0434 1059.5963    0.0000
N_R2 -H_NR+     1 1030.9469  1.053000   -0.0096    1.0434 1059.5963    0.0000
N_3  -H__N3     1 1028.0154  1.054000   -0.0096    1.0444 1056.5662    0.0000
N_3  -H_N3+     1 1028.0154  1.054000   -0.0096    1.0444 1056.5662    0.0000
C_3  -C_3       1  699.5920  1.514000    0.0000    1.5140  699.5920    0.0000
C_3  -C_A       1  699.5920  1.514000    0.0000    1.5140  699.5920    0.0000
C_3  -C_R       1  739.8881  1.486000    0.0000    1.4860  739.8881    0.0000
C_A  -C_R       1  739.8881  1.486000    0.0000    1.4860  739.8881    0.0000
C_3  -H_        1  659.7507  1.111000   -0.0016    1.1094  662.6080    0.0000
C_A  -H_        1  659.7507  1.111000   -0.0016    1.1094  662.6080    0.0000
C_3  -C_2       1  735.4249  1.489000    0.0000    1.4890  735.4249    0.0000
C_3  -S_3       1  568.4460  1.821000   -0.0073    1.8137  575.2924    0.0000
C_A  -N_3       1 1044.3430  1.457000   -0.0059    1.4511 1057.1967    0.0000
C_3  -N_3       1 1044.3430  1.457000   -0.0059    1.4511 1057.1967    0.0000
C_3  -O_3       1 1030.7742  1.415000   -0.0212    1.3938 1078.4241    0.0000
C_2  -O_3       1 1087.3977  1.390000   -0.0212    1.3938 1078.4241    0.0000
C_2  -O_2       1 1610.4080  1.217000   -0.0204    1.2195 1610.4076    0.0000
C_2  -O_2m      1 1610.4080  1.217000   -0.0204    1.2195 1610.4076    0.0000
C_2  -N_R       1 1284.9920  1.360000   -0.0058    1.3597 1284.9924    0.0000
C_2  -C_2       1  773.7474  1.464000   -0.0058    1.3597 1284.9924    0.0000
C_R  -C_2       1  778.5236  1.461000    0.0000    1.3793  925.3104    0.0000
C_R  -C_R       1  938.6990  1.373000    0.0000    1.3793  925.3104    0.0000
C_R  -H_        1  712.2570  1.083000   -0.0016    1.0814  715.3873    0.0000
C_2  -H_        1  706.3705  1.086000   -0.0016    1.0814  715.3873    0.0000
C_R  -O_2       1 1621.0470  1.217000   -0.0204    1.3426 1206.6206    0.0000
C_R  -O_2m      1 1621.0470  1.217000   -0.0204    1.3426 1206.6206    0.0000
C_2  -N_3       1 1100.0002  1.432000   -0.0204    1.3426 1206.6206    0.0000
*
* Metal bonding (from UFF)
Zn   -N_R       1  326.3616  1.892000    0.0000    0.0000    0.0000    0.0000
Zn   -O_2       1  327.6860  1.827000    0.0000    0.0000    0.0000    0.0000
Zn   -O_3       1  315.1042  1.851000    0.0000    0.0000    0.0000    0.0000
Fe3+2-N_R       1  549.4361  1.955000    0.0000    0.0000    0.0000    0.0000
Fe3+2-S_3       1  343.1248  2.334000    0.0000    0.0000    0.0000    0.0000
Fe6+2-N_R       1  487.8706  2.034000    0.0000    0.0000    0.0000    0.0000
Fe6+2-N_R2      1  487.8706  2.034000    0.0000    0.0000    0.0000    0.0000
Fe6+2-S_3       1  315.9832  2.399000    0.0000    0.0000    0.0000    0.0000
Ru   -N_R       1  557.0706  2.177000    0.0000    0.0000    0.0000    0.0000
*
* CR-NR is backbone type NR
C_R  -N_R       1 1293.1050  1.357000   -0.0058    1.3568 1293.1053    0.0000
* Arg (outter 2 N), His, Trp type NR
C_R  -N_R2      1 1364.3630  1.330000   -0.0058    1.3568 1293.1053    0.0000
*
C_R  -O_3       1 1094.4690  1.387000   -0.0207    1.3663 1144.9427    0.0000
O_3  -H___A     1 1050.0039  1.012000   -0.0217    0.9903 1120.7078    0.0000
S_3  -H___A     1  448.6317  1.418000   -0.0107    1.4073  458.9131    0.0000
O_3  -H__O3     1 1050.0039  1.012000   -0.0217    0.9903 1120.7078    0.0000
S_3  -H__S3     1  448.6317  1.418000   -0.0107    1.4073  458.9131    0.0000
S_3  -S_3       1  503.6175  2.128000    0.0000    2.1280  503.6175    0.0000
*
* F3C bondstretch
O_3F -H_F       1  500.0000  1.000000    0.0000    0.0000    0.0000    0.0000
*
ANGLE-(L-C-R)     TYPE
*simple costhet      1 FORC CNST EQUIL ANG
*
H_   -C_3  -H_      1   75.2779  109.4710    0.0000    0.0000    0.0000    0.0000    0.2233    0.0000    0.0000
H_   -C_A  -H_      1   75.2779  109.4710    0.0000    0.0000    0.0000    0.0000    0.2233    0.0000    0.0000
H_   -C_R  -H_      1   64.1310  120.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000
```

```
H_   -N_R  -H_       1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
H_   -C_2  -H_       1   63.6010  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_2  -C_2  -H_       1  102.2140  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_R  -C_2  -C_3      1  181.9801  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_2  -C_2  -C_3      1  181.4303  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_2  -C_2  -C_2      1  186.1347  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
N_R  -C_2  -C_3      1  268.1890  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
O_3  -C_2  -C_3      1  240.1350  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0825    3.0000   53.5392
C_3  -C_2  -C_3      1  176.9158  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0825    3.0000   53.5392
O_2  -C_2  -C_3      1  262.5995  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0825    3.0000   53.5392
O_2m -C_2  -C_3      1  262.5995  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0825    3.0000   53.5392
O_2  -C_2  -N_R      1  434.1630  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0879    3.0000   75.8849
O_2  -C_2  -O_2      1  399.3190  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0944    3.0000   73.6727
O_2  -C_2  -O_3      1  322.9095  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0944    3.0000   73.6727
O_2m -C_2  -O_2m     1  399.3190  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0944    3.0000   73.6727
N_R  -C_3  -C_2      1  311.5480  109.4710    0.0000    0.0000    0.0000    0.0000    0.0959    0.0000    0.0000
C_2  -C_3  -C_2      1  225.1782  109.4710    0.0000    0.0000    0.0000    0.0000    0.0882    0.0000    0.0000
C_3  -C_3  -C_3      1  214.2065  109.4710    0.0000    0.0000    0.0000    0.0000    0.0882    0.0000    0.0000
C_3  -C_3  -C_A      1  214.2065  109.4710    0.0000    0.0000    0.0000    0.0000    0.0882    0.0000    0.0000
C_2  -C_3  -C_3      1  219.5725  109.4710    0.0000    0.0000    0.0000    0.0000    0.0904    0.0000    0.0000
C_R  -C_3  -C_3      1  220.2246  109.4710    0.0000    0.0000    0.0000    0.0000    0.0907    0.0000    0.0000
C_R  -C_3  -C_A      1  220.2246  109.4710    0.0000    0.0000    0.0000    0.0000    0.0907    0.0000    0.0000
C_R  -C_A  -C_3      1  220.2246  109.4710    0.0000    0.0000    0.0000    0.0000    0.0907    0.0000    0.0000
S_3  -C_3  -C_3      1  224.7200  109.4710    0.0000    0.0000    0.0000    0.0000    0.0650    0.0000    0.0000
S_3  -C_3  -C_A      1  224.7200  109.4710    0.0000    0.0000    0.0000    0.0000    0.0650    0.0000    0.0000
O_3  -C_3  -C_3      1  290.0060  109.4710    0.0000    0.0000    0.0000    0.0000    0.0973    0.0000    0.0000
O_3  -C_3  -C_A      1  290.0060  109.4710    0.0000    0.0000    0.0000    0.0000    0.0973    0.0000    0.0000
N_3  -C_3  -C_3      1  303.2690  109.4710    0.0000    0.0000    0.0000    0.0000    0.0933    0.0000    0.0000
N_3  -C_A  -C_3      1  303.2690  109.4710    0.0000    0.0000    0.0000    0.0000    0.0933    0.0000    0.0000
N_R  -C_3  -C_3      1  303.5660  109.4710    0.0000    0.0000    0.0000    0.0000    0.0934    0.0000    0.0000
N_R  -C_3  -C_A      1  303.5660  109.4710    0.0000    0.0000    0.0000    0.0000    0.0934    0.0000    0.0000
N_3  -C_A  -C_R      1  303.5660  109.4710    0.0000    0.0000    0.0000    0.0000    0.0934    0.0000    0.0000
N_R  -C_A  -C_3      1  303.5660  109.4710    0.0000    0.0000    0.0000    0.0000    0.0934    0.0000    0.0000
N_R  -C_3  -C_R      1  312.5190  109.4710    0.0000    0.0000    0.0000    0.0000    0.0962    0.0000    0.0000
N_R  -C_A  -C_R      1  312.5190  109.4710    0.0000    0.0000    0.0000    0.0000    0.0962    0.0000    0.0000
S_3  -C_3  -H_       1  112.5440  109.4710    0.0000    0.0000    0.0000    0.0000    0.0871    0.0000    0.0000
C_3  -C_3  -H_       1  117.3990  109.4710    0.0000    0.0000    0.0000    0.0000    0.1296    0.0000    0.0000
C_3  -C_A  -H_       1  117.3990  109.4710    0.0000    0.0000    0.0000    0.0000    0.1296    0.0000    0.0000
C_A  -C_3  -H_       1  117.3990  109.4710    0.0000    0.0000    0.0000    0.0000    0.1296    0.0000    0.0000
C_2  -C_3  -H_       1  121.3740  109.4710    0.0000    0.0000    0.0000    0.0000    0.1340    0.0000    0.0000
C_R  -C_3  -H_       1  121.8610  109.4710    0.0000    0.0000    0.0000    0.0000    0.1345    0.0000    0.0000
C_R  -C_A  -H_       1  121.8610  109.4710    0.0000    0.0000    0.0000    0.0000    0.1345    0.0000    0.0000
N_R  -C_3  -H_       1  170.1120  109.4710    0.0000    0.0000    0.0000    0.0000    0.1400    0.0000    0.0000
N_R  -C_A  -H_       1  170.1120  109.4710    0.0000    0.0000    0.0000    0.0000    0.1400    0.0000    0.0000
N_3  -C_3  -H_       1  168.8850  109.4710    0.0000    0.0000    0.0000    0.0000    0.1398    0.0000    0.0000
N_3  -C_A  -H_       1  168.8850  109.4710    0.0000    0.0000    0.0000    0.0000    0.1398    0.0000    0.0000
O_3  -C_3  -H_       1  165.8800  109.4710    0.0000    0.0000    0.0000    0.0000    0.1479    0.0000    0.0000
O_2  -C_R  -C_3      1  281.0080  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0830    3.0000   53.8777
O_2m -C_R  -C_3      1  281.0080  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0830    3.0000   53.8777
O_2  -C_R  -C_A      1  281.0080  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0830    3.0000   53.8777
O_2m -C_R  -C_A      1  281.0080  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0830    3.0000   53.8777
N_R  -C_R  -C_3      1  273.1685  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0777    3.0000   55.8001
N_R  -C_R  -C_A      1  273.1685  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0777    3.0000   55.8001
C_R  -C_R  -C_3      1  199.5395  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0754    3.0000   40.6878
C_A  -C_R  -C_3      1  199.5395  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0754    3.0000   40.6878
C_A  -C_R  -C_A      1  199.5395  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0754    3.0000   40.6878
*(mjc)
O_3  -C_R  -C_R      1  271.5450  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0835    3.0000   54.1926
C_R  -C_R  -C_R      1  226.2168  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0776    3.0000   41.8760
N_R2 -C_R  -C_3      1  273.1685  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0777    3.0000   55.8001
N_R2 -C_R  -C_R      1  214.9725  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0801    3.0000   57.4616
N_R2 -C_R  -H_       1  162.3270  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1186    3.0000   31.7273
N_R2 -C_R  -N_R      1  436.6740  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0826    3.0000   78.8955
N_R2 -C_R  -N_R2     1  436.6740  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0826    3.0000   78.8955
N_R  -C_R  -C_R      1  214.9725  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0801    3.0000   57.4616
N_R  -C_R  -H_       1  162.3270  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1186    3.0000   31.7273
C_R  -C_R  -H_       1  115.6673  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1140    3.0000   22.9257
O_2  -C_R  -H_       1  140.6971  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1140    3.0000   22.9257
O_2m -C_R  -H_       1  140.6971  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1140    3.0000   22.9257
N_R  -C_R  -N_R      1  436.6740  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0826    3.0000   78.8955
O_2  -C_R  -N_R      1  436.9620  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0885    3.0000   76.3751
O_2  -C_R  -O_2      1  401.9570  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0950    3.0000   74.1603
O_2m -C_R  -N_R      1  436.9620  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0885    3.0000   76.3751
O_2m -C_R  -O_2m     1  401.9570  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0950    3.0000   74.1603
C_3  -N_3  -C_3      1  260.8690  106.7000    0.0000    0.0000    0.0000    0.0000    0.1574    0.0000    0.0000
C_A  -N_3  -C_3      1  260.8690  106.7000    0.0000    0.0000    0.0000    0.0000    0.1574    0.0000    0.0000
C_3  -N_3  -H___A    1  144.7980  106.7000    0.0000    0.0000    0.0000    0.0000    0.1574    0.0000    0.0000
C_A  -N_3  -H_N3+    1  144.7980  106.7000    0.0000    0.0000    0.0000    0.0000    0.1574    0.0000    0.0000
H___A-N_3  -H___A    1   97.1150  106.7000    0.0000    0.0000    0.0000    0.0000    0.2804    0.0000    0.0000
C_3  -N_3  -H_N3+    1  144.7980  106.7000    0.0000    0.0000    0.0000    0.0000    0.1574    0.0000    0.0000
C_3  -N_3  -H__N3    1  144.7980  106.7000    0.0000    0.0000    0.0000    0.0000    0.1574    0.0000    0.0000
H__N3-N_3  -H__N3    1   97.1150  106.7000    0.0000    0.0000    0.0000    0.0000    0.2804    0.0000    0.0000
H___A-N_R  -H___A    1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
H__NR-N_R  -H__NR    1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
H___A-N_R2 -H___A    1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
H__NR-N_R2 -H__NR    1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
C_3  -N_R  -C_3      1  191.5490  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0779    3.0000   42.0488
C_A  -N_R  -C_3      1  191.5490  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0779    3.0000   42.0488
C_A  -N_R  -C_A      1  191.5490  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0779    3.0000   42.0488
C_R  -N_R  -C_3      1  210.9740  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0802    3.0000   43.2802
C_R  -N_R  -C_A      1  210.9740  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0802    3.0000   43.2802
C_2  -N_R  -C_3      1  210.3520  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0800    3.0000   43.1465
C_R  -N_R  -C_R      1  246.6940  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0826    3.0000   44.5711
```

```
C_3  -N_R  -H_       1  108.4940  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1178    3.0000   23.6902
C_2  -N_R  -H___A    1  122.6710  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1218    3.0000   24.4966
C_3  -N_R  -H___A    1  108.4940  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1178    3.0000   23.6902
C_R  -N_R  -H___A    1  128.4833  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1223    3.0000   24.5955
C_2  -N_R  -H__NR    1  122.6710  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1218    3.0000   24.4966
C_3  -N_R  -H__NR    1  108.4940  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1178    3.0000   23.6902
C_A  -N_R  -H__NR    1  108.4940  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1178    3.0000   23.6902
C_R  -N_R  -H__NR    1  128.4833  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1223    3.0000   24.5955
C_R  -N_R2 -C_R      1  246.6940  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0826    3.0000   44.5711
C_R  -N_R2 -H___A    1  128.4833  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1223    3.0000   24.5955
C_3  -O_3  -H___A    1  174.2860  104.5100    0.0000    0.0000    0.0000    0.0000    0.1830    0.0000    0.0000
C_R  -O_3  -H___A    1  181.7360  104.5100    0.0000    0.0000    0.0000    0.0000    0.1908    0.0000    0.0000
C_2  -O_3  -H__O3    1  180.9220  104.5100    0.0000    0.0000    0.0000    0.0000    0.1908    0.0000    0.0000
C_R  -N_R2 -H__NR    1  128.4833  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1223    3.0000   24.5955
C_3  -O_3  -H__O3    1  174.2860  104.5100    0.0000    0.0000    0.0000    0.0000    0.1830    0.0000    0.0000
C_R  -O_3  -H__O3    1  181.7360  104.5100    0.0000    0.0000    0.0000    0.0000    0.1908    0.0000    0.0000
C_3  -S_3  -C_3      1  201.9560   92.1000    0.0000    0.0000    0.0000    0.0000    0.0822    0.0000    0.0000
S_3  -S_3  -C_3      1  217.9600   92.1000    0.0000    0.0000    0.0000    0.0000    0.0632    0.0000    0.0000
C_3  -S_3  -H___A    1  102.0450   92.1000    0.0000    0.0000    0.0000    0.0000    0.1111    0.0000    0.0000
C_3  -S_3  -H__S3    1  102.0450   92.1000    0.0000    0.0000    0.0000    0.0000    0.1111    0.0000    0.0000
C_2  -N_R  -H_NR+   1  122.6710  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1218    3.0000   24.4966
C_3  -N_R  -H_NR+   1  108.4940  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1178    3.0000   23.6902
C_R  -N_R  -H_NR+   1  128.4833  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1223    3.0000   24.5955
C_R  -N_R2 -H_NR+   1  128.4833  120.0000    0.0000    0.0000    0.0000   -1.0000    0.1223    3.0000   24.5955
H_N3+-N_3  -H_N3+   1   97.1150  106.7000    0.0000    0.0000    0.0000    0.0000    0.2804    0.0000    0.0000
H___A-N_R  -H___A   1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
H_NR+-N_R  -H_NR+   1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
H___A-N_R2 -H___A   1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
H_NR+-N_R2 -H_NR+   1   71.3950  120.0000    0.0000    0.0000    0.0000   -1.0000    0.2060    3.0000   15.4366
Zn   -N_R  -C_R      1   86.4530  120.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
Zn   -O_2  -C_2      1  100.4620  120.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
Zn   -O_3F -H_F      1   64.1760  104.5100    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
O_3  -Zn   -N_R      1  187.7090  109.4710    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
N_R  -Zn   -N_R      1  197.4920  109.4710    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
O_2  -Zn   -O_3      1  182.2370  109.4710    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
O_2  -Zn   -N_R      1  191.2200  109.4710    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
O_2  -Zn   -O_2      1  185.8360  109.4710    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
* Stuff for Heme group
N_R  -Fe3+2-N_R      1  188.3380  109.4710    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
N_R  -Fe6+2-N_R      1  289.5630   90.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
N_R2 -Fe6+2-N_R      1  289.5630   90.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
C_2  -N_R  -C_2      1  291.3030  111.3000    0.0000    0.0000    0.0000   -1.0000    0.0800    3.0000   43.1465
N_R  -C_2  -C_2      1  279.4170  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
N_R  -C_2  -C_R      1  280.3450  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_2  -C_2  -C_R      1  186.7077  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_2  -C_R  -C_2      1  187.2837  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_2  -C_R  -H_       1  102.9250  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
Fe3+2-N_R  -C_2      1  198.2260  111.3000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
Fe3+2-S_3  -C_3      1  171.0890   92.1000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
S_3  -Fe3+2-N_R      1  148.3640  109.4710    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
S_3  -Fe6+2-N_R      1  217.2056   90.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
S_3  -Fe6+2-N_R2     1  217.2056   90.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
Fe6+2-N_R  -C_2      1  151.0240  120.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
Fe6+2-N_R2 -C_R      1  137.5332  120.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
Fe6+2-S_3  -C_3      1  199.5525   92.1000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
*
* P450 Ru Linker params.
N_R  -Ru   -N_R      1  238.6410   90.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
Ru   -N_R  -C_R      1  188.2300  120.0000    0.0000    0.0000    0.0000    0.0000    0.0000    3.0000   15.4366
N_3  -C_2  -C_3      1  249.3046  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
O_2  -C_2  -N_3      1  240.9266  120.0000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_2  -N_3  -C_3      1  267.6580  106.7000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
C_2  -N_3  -H__N3    1  147.3000  106.7000    0.0000    0.0000    0.0000   -1.0000    0.0773    3.0000   55.4570
*
* F3C angle
H_F  -O_3F -H_F     21  120.0000  109.4700    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000
*
TORSION                  CASE   BARRIER    PERIOD CISMIN(1)     ANGANG    BNDTOR MPHI  B-B
POLY
*must have angang etc on last one
* Taken from UFF and placed in Cos expansion form.
TORSION FOURIER
*LLLL-CCCCC-CCCCC-RRRRR CASE      v0       v1       v2       v3       v4       v5       v6
v7      v8      v9     v10     v11     v12
X    -N_R  -C_3  -X        1  1.0000  0.0000  0.0000  1.0000
X    -C_3  -C_3  -X        1  1.0595  0.0000  0.0000  1.0595
X    -C_3  -C_R  -X        1  0.5000  0.0000  0.0000  0.0000  0.0000  0.0000 -0.5000
X    -N_R  -C_R  -X        1 14.1644  0.0000-14.1644
X    -N_R2 -C_R  -X        1 14.1644  0.0000-14.1644
X    -S_3  -C_3  -X        1  0.5064  0.0000  0.0000  0.5064
X    -N_R  -C_2  -X        1 12.1810  0.0000-12.1810
X    -C_2  -C_3  -X        1  0.5000  0.0000  0.0000  0.0000  0.0000  0.0000 -0.5000
X    -C_R  -C_R  -X        1 14.2069  0.0000-14.2069
X    -C_2  -C_R  -X        8  5.0000  0.0000 -5.0000
X    -C_2  -C_2  -X        8  5.0000  0.0000 -5.0000
X    -C_2  -O_3  -X        8  5.0000  0.0000 -5.0000
```

```
X    -N_3  -C_3  -X        1  0.4882  0.0000  0.0000  0.4882
X    -O_3  -C_3  -X        1  0.0976  0.0000  0.0000  0.0976
X    -O_3  -C_R  -X        1  5.0000  0.0000 -5.0000
X    -S_3  -S_3  -X        1  0.2420  0.0000  0.2420
X    -Zn   -N_R  -X        1  0.5000  0.0000  0.0000  0.0000  0.0000  0.0000 -0.5000
X    -Zn   -O_2  -X        1  0.5000  0.0000  0.0000  0.0000  0.0000  0.0000 -0.5000
X    -Ru   -N_R  -X       41  0.0000
X    -N_3  -C_2  -X        1  1.0000  0.0000  0.0000  1.0000
* Fe - N_R torsion zero rather than 1 since N_R is N_2 actually.
X    -Fe3+2-N_R  -X       44  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
X    -Fe3+2-S_3  -X        1  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
X    -Fe6+2-N_R  -X       44  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
X    -Fe6+2-N_R2 -X       44  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
X    -Fe6+2-S_3  -X        1  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
*
*  The following zeroed to make sure phi/psi/proline torsions work out correctly..
X    -C_3  -C_A  -X        1  0.0000  0.0000  0.0000  0.0000
X    -N_3  -C_A  -X        1  0.0000  0.0000  0.0000  0.0000
X    -N_R  -C_A  -X        1  0.0000  0.0000  0.0000  0.0000
X    -C_A  -C_R  -X        1  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
*
* Modified torsions as per CMP/MJC/WAG torsion paper of tripeptides (12/10/99)
* *** Multiplied by 6 to get correct barriers... ***
*phi (CNCC)  psi (NCCN)  (CCCN is Cbeta torsion (psi + 120), CCNC is Cb (phi -120) )
*
*PHI*
C_R  -N_R  -C_A  -C_R      0  0.0000  0.0000  0.0000 -2.7000  0.0000  0.0000  0.0000
*PSI*
N_R  -C_R  -C_A  -N_R      0  0.0000  0.0000-15.0000 -1.2000  0.0000  0.0000  0.0000
*(added for Nterminal N_3)
N_R  -C_R  -C_A  -N_3      0  0.0000  0.0000-15.0000 -1.2000  0.0000  0.0000  0.0000
*
*
*"C-b phi"
C_R  -N_R  -C_A  -C_3      0  0.0000 -6.0000 -6.0000-14.4000 -9.0000
*"C-beta psi"
N_R  -C_R  -C_A  -C_3      0  0.0000  3.6000  1.8000 -3.0000
*
*
INVERSION (CENT AT 1ST) TYPE  FRC CNST  EQU ANGL         D         E         F
*CCCC-JJJJJ-KKKKK-LLLLL  TYPE=1 FOR CHARMM,TYPE=2 FOR SPECTROSCOPIC, TYPE=3 FOR AMBER
C_R  -X    -X    -X        2    6.0000    0.0000
C_R  -O_2  -X    -X        2   50.0000    0.0000
C_R  -X    -X    -O_2      2   50.0000    0.0000
C_2  -O_2  -X    -X        2   50.0000    0.0000
C_2  -X    -X    -O_2      2   50.0000    0.0000
C_3  -X    -X    -X        2    0.0000    0.0000
C_2  -X    -X    -X        2    6.0000    0.0000
C_R  -O_2m -X    -X        2   50.0000    0.0000
C_R  -X    -X    -O_2m     2   50.0000    0.0000
C_2  -O_2m -X    -X        2   50.0000    0.0000
C_2  -X    -X    -O_2m     2   50.0000    0.0000
N_R  -X    -X    -X        2    6.0000    0.0000
N_R2 -X    -X    -X        2    6.0000    0.0000
N_3  -X    -X    -X        2    0.0000   61.2230
S_3  -X    -X    -X        2    0.0000    0.0000
*
END OF DATA
```

```
 ARG #H       H__NR     1  0   0.3340
 ARG HN       H__NR     1  0   0.3340
 ARG  CA      C_A       4  0   0.0290
 ARG HCA      H_        1  0   0.1300
 ARG  HA      H_        1  0   0.1300
 ARG  C       C_R       3  0   0.8770
 ARG  O       O_2       1  2  -0.6990
 ARG  OXT     O_2       1  2  -0.6990
 ARG  CB      C_3       4  0  -0.2210
 ARG HCB      H_        1  0   0.0850
 ARG #HB      H_        1  0   0.0850
 ARG HB       H_        1  0   0.0850
 ARG  CG      C_3       4  0   0.1720
 ARG HCG      H_        1  0   0.0360
 ARG #HG      H_        1  0   0.0360
 ARG #DG      H_        1  0   0.0360
 ARG HG1      H_        1  0   0.0360
 ARG HG2      H_        1  0   0.0360
 ARG  CD      C_3       4  0  -0.0750
 ARG HCD      H_        1  0   0.1070
 ARG #HD      H_        1  0   0.1070
 ARG #DD      H_        1  0   0.1070
 ARG HD1      H_        1  0   0.1070
 ARG HD2      H_        1  0   0.1070
 ARG  NE      N_R       2  1  -0.6240
 ARG HNE      H__NR     1  0   0.3820
 ARG  HE      H__NR     1  0   0.3820
 ARG #HE      H__NR     1  0   0.3820
 ARG #DE      H__NR     1  0   0.3820
 ARG  CZ      C_R       3  0   1.1000
 ARG  NH1     N_R2      3  0  -1.0930
 ARG HNH1     H_NR+     1  0   0.5160
 ARG HN11     H_NR+     1  0   0.5160
 ARG HN12     H_NR+     1  0   0.5160
 ARG HH11     H_NR+     1  0   0.5160
 ARG  NH2     N_R2      3  0  -1.0930
 ARG HNH2     H_NR+     1  0   0.5160
 ARG HN21     H_NR+     1  0   0.5160
 ARG HN22     H_NR+     1  0   0.5160
 ARG HH21     H_NR+     1  0   0.5160
 ARG HH22     H_NR+     1  0   0.5160
 ARG #HH1     H_NR+     1  0   0.5160
 ARG #DH1     H_NR+     1  0   0.5160
 ARG #HH2     H_NR+     1  0   0.5160
 ARG #DH2     H_NR+     1  0   0.5160
* *ARG done*
 ASN  N       N_R       2  1  -0.7760
 ASN HN       H__NR     1  0   0.3290
 ASN  H       H__NR     1  0   0.3290
 ASN #H       H__NR     1  0   0.3290
 ASN  CA      C_A       4  0   0.2590
 ASN HCA      H_        1  0   0.0880
 ASN  HA      H_        1  0   0.0880
 ASN  C       C_R       3  0   0.8040
 ASN  O       O_2       1  2  -0.6580
```

```
 ASN  OXT     O_2       1  2  -0.6580
 ASN  CB      C_3       4  0  -0.4180
 ASN HCB      H_        1  0   0.1440
 ASN #HB      H_        1  0   0.1440
 ASN HB       H_        1  0   0.1440
 ASN  CG      C_R       3  0   0.9380
 ASN  OD1     O_2       1  2  -0.7280
 ASN  AD1     O_2       1  2  -0.7280
 ASN  ND2     N_R       2  1  -1.0200
 ASN  AD2     N_R       2  1  -1.0200
 ASN HND1     H__NR     1  0   0.4470
 ASN HND2     H__NR     1  0   0.4470
 ASN HD21     H__NR     1  0   0.4470
 ASN HD22     H__NR     1  0   0.4470
 ASN HAD2     H__NR     1  0   0.4470
 ASN #HD2     H__NR     1  0   0.4470
 ASN #DD2     H__NR     1  0   0.4470
* *ASN done*
 ASP  N       N_R       2  1  -0.8800
 ASP HN       H__NR     1  0   0.3760
 ASP  CA      C_A       4  0   0.1300
 ASP HCA      H_        1  0   0.0860
 ASP  HA      H_        1  0   0.0860
 ASP  C       C_R       3  0   0.9140
 ASP  O       O_2       1  2  -0.7290
 ASP  OXT     O_2       1  2  -0.7290
 ASP  CB      C_3       4  0  -0.0960
 ASP HCB      H_        1  0   0.0330
 ASP #HB      H_        1  0   0.0330
 ASP HB       H_        1  0   0.0330
 ASP  H       H__NR     1  0   0.3760
 ASP #H       H__NR     1  0   0.3760
 ASP  CG      C_R       3  0   0.9150
 ASP  OD1     O_2m      1  2  -0.8910
 ASP  OD2     O_2m      1  2  -0.8910
* *ASP done using h2o optimized geometry for charges*
 CYS  N       N_R       2  1  -0.9160
 CYS HN       H__NR     1  0   0.3860
 CYS  CA      C_A       4  0   0.4020
 CYS HCA      H_        1  0   0.0700
 CYS  HA      H_        1  0   0.0700
 CYS  C       C_R       3  0   0.8120
 CYS  O       O_2       1  2  -0.6360
 CYS  OXT     O_2       1  2  -0.6360
 CYS  CB      C_3       4  0  -0.4080
 CYS HCB      H_        1  0   0.1990
 CYS #HB      H_        1  0   0.1990
 CYS HB       H_        1  0   0.1990
 CYS  H       H__NR     1  0   0.3860
 CYS #H       H__NR     1  0   0.3860
 CYS  SG      S_3       2  2  -0.1080
 CYS  S1      S_3       2  2  -0.1080
* (for no disulfide)
* CYS  SG      S_3       2  2  -0.3850
* (If you have no disulfide, SG needs to be changed to -0.385)
```

```
 CYS HSG      H__S3     1  0   0.2770
 CYS HG       H__S3     1  0   0.2770
 CYS  HG      H__S3     1  0   0.2770
 CYS  DG      H__S3     1  0   0.2770
*
*  If no disulfide, SG needs to be -0.3850
*  *CYS done*
 GLU  N       N_R       2  1  -0.7680
 GLU HN       H__NR     1  0   0.3210
 GLU  CA      C_A       4  0   0.1040
 GLU HCA      H_        1  0   0.0810
 GLU  HA      H_        1  0   0.0810
 GLU  C       C_R       3  0   0.8710
 GLU  O       O_2       1  2  -0.6800
 GLU  OXT     O_2       1  2  -0.6800
 GLU  CB      C_3       4  0  -0.0250
 GLU HCB      H_        1  0   0.0260
 GLU #HB      H_        1  0   0.0260
 GLU HB       H_        1  0   0.0260
 GLU  H       H__NR     1  0   0.3210
 GLU #H       H__NR     1  0   0.3210
 GLU  CG      C_3       4  0  -0.1470
 GLU HCG      H_        1  0   0.0470
 GLU #HG      H_        1  0   0.0470
 GLU #DG      H_        1  0   0.0470
 GLU HG1      H_        1  0   0.0470
 GLU HG2      H_        1  0   0.0470
 GLU  CD      C_R       3  0   0.9510
 GLU  OE1     O_2m      1  2  -0.9270
 GLU  OE2     O_2m      1  2  -0.9270
* *GLU       done*
 GLN  N       N_R       2  1  -0.7240
 GLN HN       H__NR     1  0   0.3160
 GLN  CA      C_A       4  0   0.1480
 GLN HCA      H_        1  0   0.0770
 GLN  HA      H_        1  0   0.0770
 GLN  C       C_R       3  0   0.7920
 GLN  O       O_2       1  2  -0.6690
 GLN  OXT     O_2       1  2  -0.6690
 GLN  CB      C_3       4  0  -0.0060
 GLN HCB      H_        1  0   0.0320
 GLN #HB      H_        1  0   0.0320
 GLN HB       H_        1  0   0.0320
 GLN  H       H__NR     1  0   0.3160
 GLN #H       H__NR     1  0   0.3160
 GLN  CG      C_3       4  0  -0.2460
 GLN HCG      H_        1  0   0.1000
 GLN #HG      H_        1  0   0.1000
 GLN #DG      H_        1  0   0.1000
 GLN HG1      H_        1  0   0.1000
 GLN HG2      H_        1  0   0.1000
 GLN  CD      C_R       3  0   0.9340
 GLN  OE1     O_2       1  2  -0.7410
 GLN  AE1     O_2       1  2  -0.7410
 GLN  NE2     N_R       2  1  -1.1230
```

```
 GLN  AE2     N_R       2  1  -1.1230
 GLN HNE1     H__NR     1  0   0.4890
 GLN HNE2     H__NR     1  0   0.4890
 GLN HE21     H__NR     1  0   0.4890
 GLN HE22     H__NR     1  0   0.4890
 GLN HAE2     H__NR     1  0   0.4890
 GLN #HE2     H__NR     1  0   0.4890
 GLN #DE2     H__NR     1  0   0.4890
* *GLN done*
 GLY  N       N_R       2  1  -0.6540
 GLY HN       H__NR     1  0   0.2970
 GLY  CA      C_A       4  0   0.0010
 GLY HCA      H_        1  0   0.1000
 GLY  HA      H_        1  0   0.1000
 GLY  C       C_R       3  0   0.8360
 GLY  O       O_2       1  2  -0.6800
 GLY  OXT     O_2       1  2  -0.6800
 GLY #HA      H_        1  0   0.1000
 GLY HA1      H_        1  0   0.1000
 GLY HA2      H_        1  0   0.1000
 GLY  H       H__NR     1  0   0.2970
 GLY #H       H__NR     1  0   0.2970
* *GLY done*
 HIS  N       N_R       2  1  -0.9440
 HIS HN       H__NR     1  0   0.3590
 HIS  CA      C_A       4  0   0.7570
 HIS HCA      H_        1  0  -0.0440
 HIS  HA      H_        1  0  -0.0440
 HIS  C       C_R       3  0   0.6380
 HIS  O       O_2       1  2  -0.6240
 HIS  OXT     O_2       1  2  -0.6240
 HIS  CB      C_3       4  0  -0.5280
 HIS HCB      H_        1  0   0.1630
 HIS #HB      H_        1  0   0.1630
 HIS HB       H_        1  0   0.1630
 HIS  H       H__NR     1  0   0.3590
 HIS #H       H__NR     1  0   0.3590
 HIS  CG      C_R       3  0   0.0790
 HIS  ND1     N_R2      3  0  -0.3440
 HIS  AD1     N_R2      3  0  -0.3440
 HIS HND1     H__NR     1  0   0.3240
 HIS HD1      H__NR     1  0   0.3240
 HIS HAD1     H__NR     1  0   0.3240
 HIS  HD1     H__NR     1  0   0.3240
 HIS  DD1     H__NR     1  0   0.3240
 HIS  CD2     C_R       3  0   0.2180
 HIS  AD2     C_R       3  0   0.2180
 HIS HCD2     H_        1  0   0.0910
 HIS  HD2     H_        1  0   0.0910
 HIS  DD2     H_        1  0   0.0910
 HIS  CE1     C_R       3  0   0.3930
 HIS  AE1     C_R       3  0   0.3930
 HIS HCE1     H_        1  0   0.1080
 HIS  HE1     H_        1  0   0.1080
 HIS  DE1     H_        1  0   0.1080
```

```
 HIS  NE2    N_R2     2  1  -0.8090
 HIS  AE2    N_R2     2  1  -0.8090
*  *HIS done*  -> assume "HSD" is deprotonated
*           and "HSP" is protonated form.
*  HIS has no proton on outter most nitrogen...
*
 HSD  N      N_R      2  1  -1.1790
 HSD HN      H__NR    1  0   0.4120
 HSD  CA     C_A      4  0   1.0990
 HSD HCA     H_       1  0  -0.0990
 HSD  HA     H_       1  0  -0.0990
 HSD  C      C_R      3  0   0.5540
 HSD  O      O_2      1  2  -0.6300
 HSD  OXT    O_2      1  2  -0.6300
 HSD  CB     C_3      4  0  -0.8520
 HSD HCB     H_       1  0   0.1950
 HSD #HB     H_       1  0   0.1950
 HSD HB      H_       1  0   0.1950
 HSD  H      H__NR    1  0   0.4120
 HSD #H      H__NR    1  0   0.4120
 HSD  CG     C_R      3  0   0.2900
 HSD  ND1    N_R2     2  1  -0.6270
 HSD  AD1    N_R2     2  1  -0.6270
 HSD  CD2    C_R      3  0   0.0580
 HSD  AD2    C_R      3  0   0.0580
 HSD HCD2    H_       1  0   0.0810
 HSD HD2     H_       1  0   0.0810
 HSD HAD2    H_       1  0   0.0810
 HSD  CE1    C_R      3  0   0.4070
 HSD  AE1    C_R      3  0   0.4070
 HSD HCE1    H_       1  0   0.0320
 HSD HE1     H_       1  0   0.0320
 HSD HAE1    H_       1  0   0.0320
 HSD  NE2    N_R2     3  0  -0.9360
 HSD  AE2    N_R2     3  0  -0.9360
*  *HSD Done.*  Assumed HSD had charge -1, no HN hydrogens.*
 HSP  N      N_R      2  1  -0.8980
 HSP HN      H__NR    1  0   0.4080
 HSP  H      H__NR    1  0   0.4080
 HSP #H      H__NR    1  0   0.4080
 HSP  CA     C_A      4  0   0.2970
 HSP HCA     H_       1  0   0.1300
 HSP  HA     H_       1  0   0.1300
 HSP  C      C_R      3  0   0.8650
 HSP  O      O_2      1  2  -0.6810
 HSP  OXT    O_2      1  2  -0.6810
 HSP  CB     C_3      4  0  -0.6540
 HSP HCB     H_       1  0   0.2150
 HSP #HB     H_       1  0   0.2150
 HSP HB      H_       1  0   0.2150
 HSP  CG     C_R      3  0   0.4130
 HSP  ND1    N_R2     3  0  -0.2410
 HSP  AD1    N_R2     3  0  -0.2410
 HSP HND1    H_NR+    1  0   0.3710
 HSP HD1     H_NR+    1  0   0.3710
```

```
 HSP HAD1     H_NR+     1  0    0.3710
 HSP  CD2     C_R       3  0   -0.2710
 HSP  AD2     C_R       3  0   -0.2710
 HSP HCD2     H_        1  0    0.2640
 HSP HD2      H_        1  0    0.2640
 HSP HAD2     H_        1  0    0.2640
 HSP  CE1     C_R       3  0    0.1280
 HSP  AE1     C_R       3  0    0.1280
 HSP HCE1     H_        1  0    0.2440
 HSP HE1      H_        1  0    0.2440
 HSP HAE1     H_        1  0    0.2440
 HSP  NE2     N_R2      3  0   -0.2380
 HSP  AE2     N_R2      3  0   -0.2380
 HSP HNE2     H_NR+     1  0    0.4330
 HSP HE2      H_NR+     1  0    0.4330
 HSP HAE2     H_NR+     1  0    0.4330
*  *HSP done*  Used +1 total net charge, 2 NH hydrogens
*     *Used H2O solvent optimized geometry to get charges*
*              (H__NR in HSP may need special plus hydrogen terms...)
*
 ILE  N       N_R       2  1   -0.6820
 ILE HN       H__NR     1  0    0.2830
 ILE  H       H__NR     1  0    0.2830
 ILE #H       H__NR     1  0    0.2830
 ILE  CA      C_A       4  0    0.0380
 ILE HCA      H_        1  0    0.1090
 ILE  HA      H_        1  0    0.1090
 ILE  C       C_R       3  0    0.8600
 ILE  O       O_2       1  2   -0.7080
 ILE  OXT     O_2       1  2   -0.7080
 ILE  CB      C_3       4  0    0.1600
 ILE HCB      H_        1  0    0.0080
 ILE #HB      H_        1  0    0.0080
 ILE HB       H_        1  0    0.0080
 ILE  CG1     C_3       4  0    0.0500
 ILE HCG1     H_        1  0    0.0290
 ILE #HG1     H_        1  0    0.0290
 ILE #DG1     H_        1  0    0.0290
 ILE HG1      H_        1  0    0.0290
 ILE  CG2     C_3       4  0   -0.3880
 ILE HCG2     H_        1  0    0.1000
 ILE #HG2     H_        1  0    0.1000
 ILE #DG2     H_        1  0    0.1000
 ILE HG2      H_        1  0    0.1000
 ILE  CD1     C_3       4  0   -0.3640
 ILE HCD1     H_        1  0    0.0920
 ILE #HD1     H_        1  0    0.0920
 ILE #DD1     H_        1  0    0.0920
 ILE HD1      H_        1  0    0.0920
 ILE HD2      H_        1  0    0.0920
 ILE HD3      H_        1  0    0.0920
*  *ILE done* *H's need to be double checked.*
*  *removed HD4 through HD6 since unspecified.*
 LEU  N       N_R       2  1   -0.7070
 LEU HN       H__NR     1  0    0.2970
```

```
 LEU  H      H__NR    1  0   0.2970
 LEU #H      H__NR    1  0   0.2970
 LEU  CA     C_A      4  0   0.3160
 LEU HCA     H_       1  0   0.0530
 LEU  HA     H_       1  0   0.0530
 LEU  C      C_R      3  0   0.6760
 LEU  O      O_2      1  2  -0.6510
 LEU  OXT    O_2      1  2  -0.6510
 LEU  CB     C_3      4  0  -0.3170
 LEU HCB     H_       1  0   0.0800
 LEU #HB     H_       1  0   0.0800
 LEU HB      H_       1  0   0.0800
 LEU  CG     C_3      4  0   0.5390
 LEU HCG     H_       1  0  -0.0480
 LEU  HG     H_       1  0  -0.0480
 LEU  DG     H_       1  0  -0.0480
 LEU  CD1    C_3      4  0  -0.5160
 LEU HCD1    H_       1  0   0.1190
 LEU #HD1    H_       1  0   0.1190
 LEU #DD1    H_       1  0   0.1190
 LEU  CD2    C_3      4  0  -0.5160
 LEU HCD2    H_       1  0   0.1190
 LEU #HD2    H_       1  0   0.1190
 LEU #DD2    H_       1  0   0.1190
 LEU HD1     H_       1  0   0.1190
 LEU HD2     H_       1  0   0.1190
 LEU HD3     H_       1  0   0.1190
 LEU HD4     H_       1  0   0.1190
 LEU HD5     H_       1  0   0.1190
 LEU HD6     H_       1  0   0.1190
* *LEU done*
 LYS  N      N_R      2  1  -0.7090
 LYS HN      H__NR    1  0   0.3230
 LYS  CA     C_A      4  0   0.0030
 LYS HCA     H_       1  0   0.1120
 LYS  HA     H_       1  0   0.1120
 LYS  C      C_R      3  0   0.8720
 LYS  O      O_2      1  2  -0.6920
 LYS  OXT    O_2      1  2  -0.6920
 LYS  CB     C_3      4  0  -0.0380
 LYS HCB     H_       1  0   0.0400
 LYS #HB     H_       1  0   0.0400
 LYS HB      H_       1  0   0.0400
 LYS  H      H__NR    1  0   0.3230
 LYS #H      H__NR    1  0   0.3230
 LYS  CG     C_3      4  0   0.0360
 LYS HCG     H_       1  0   0.0200
 LYS #HG     H_       1  0   0.0200
 LYS #DG     H_       1  0   0.0200
 LYS HG1     H_       1  0   0.0200
 LYS HG2     H_       1  0   0.0200
 LYS  CD     C_3      4  0  -0.1640
 LYS HCD     H_       1  0   0.0680
 LYS #HD     H_       1  0   0.0680
 LYS #DD     H_       1  0   0.0680
```

```
 LYS HD1     H_        1  0   0.0680
 LYS HD2     H_        1  0   0.0680
 LYS  CE     C_3       4  0   0.2910
 LYS HCE     H_        1  0   0.0340
 LYS #HE     H_        1  0   0.0340
 LYS #DE     H_        1  0   0.0340
 LYS HE1     H_        1  0   0.0340
 LYS HE2     H_        1  0   0.0340
 LYS  NZ     N_3       4  0  -0.4920
 LYS HNZ     H_N3+     1  0   0.3780
 LYS HZ1     H_N3+     1  0   0.3780
 LYS HZ2     H_N3+     1  0   0.3780
 LYS HZ3     H_N3+     1  0   0.3780
 LYS #HZ     H_N3+     1  0   0.3780
 LYS #DZ     H_N3+     1  0   0.3780
 LYS HNZ1    H_N3+     1  0   0.3780
 LYS HNZ2    H_N3+     1  0   0.3780
 LYS HNZ3    H_N3+     1  0   0.3780
* May want special plus H__NR term for LYS.
*
 MET  N      N_R       2  1  -0.6300
 MET HN      H__NR     1  0   0.2910
 MET  CA     C_A       4  0  -0.0340
 MET HCA     H_        1  0   0.1290
 MET  HA     H_        1  0   0.1290
 MET  C      C_R       3  0   0.8510
 MET  O      O_2       1  2  -0.6890
 MET  OXT    O_2       1  2  -0.6890
 MET  CB     C_3       4  0  -0.0620
 MET HCB     H_        1  0   0.0990
 MET #HB     H_        1  0   0.0990
 MET HB      H_        1  0   0.0990
 MET  H      H__NR     1  0   0.2910
 MET #H      H__NR     1  0   0.2910
 MET  CG     C_3       4  0  -0.3000
 MET HCG     H_        1  0   0.2120
 MET #HG     H_        1  0   0.2120
 MET #DG     H_        1  0   0.2120
 MET HG1     H_        1  0   0.2120
 MET HG2     H_        1  0   0.2120
 MET  SD     S_3       2  0  -0.3290
 MET  CE     C_3       4  0  -0.2810
 MET HCE     H_        1  0   0.1440
 MET #HE     H_        1  0   0.1440
 MET #DE     H_        1  0   0.1440
 MET HE1     H_        1  0   0.1440
 MET HE2     H_        1  0   0.1440
 MET HE3     H_        1  0   0.1440
* *MET done*
 PHE  N      N_R       2  1  -0.9340
 PHE HN      H__NR     1  0   0.3470
 PHE  CA     C_A       4  0   0.7780
 PHE HCA     H_        1  0  -0.0410
 PHE  HA     H_        1  0  -0.0410
 PHE  C      C_R       3  0   0.5610
```

```
 PHE  O      O_2      1  2  -0.6260
 PHE  OXT    O_2      1  2  -0.6260
 PHE  CB     C_3      4  0  -0.5280
 PHE HCB     H_       1  0   0.1410
 PHE #HB     H_       1  0   0.1410
 PHE HB      H_       1  0   0.1410
 PHE  H      H__NR    1  0   0.3470
 PHE #H      H__NR    1  0   0.3470
 PHE  CG     C_R      3  0   0.2540
 PHE  CD1    C_R      3  0  -0.2010
 PHE  CD2    C_R      3  0  -0.2010
 PHE HCD1    H_       1  0   0.1470
 PHE  HD1    H_       1  0   0.1470
 PHE  DD1    H_       1  0   0.1470
 PHE HCD2    H_       1  0   0.1470
 PHE  HD2    H_       1  0   0.1470
 PHE  DD2    H_       1  0   0.1470
 PHE  CE1    C_R      3  0  -0.1470
 PHE  CE2    C_R      3  0  -0.1470
 PHE HCE1    H_       1  0   0.1560
 PHE  HE1    H_       1  0   0.1560
 PHE  DE1    H_       1  0   0.1560
 PHE HCE2    H_       1  0   0.1560
 PHE  HE2    H_       1  0   0.1560
 PHE  DE2    H_       1  0   0.1560
 PHE  CZ     C_R      3  0  -0.1560
 PHE HCZ     H_       1  0   0.1530
 PHE  HZ     H_       1  0   0.1530
 PHE  DZ     H_       1  0   0.1530
* *PHE done*
 PRO  N      N_R      2  1  -0.2750
 PRO  CA     C_A      4  0  -0.1030
 PRO HCA     H_       1  0   0.1130
 PRO  HA     H_       1  0   0.1130
 PRO  C      C_R      3  0   1.0140
 PRO  O      O_2      1  2  -0.7850
 PRO  OXT    O_2      1  2  -0.7850
 PRO  CB     C_3      4  0  -0.3090
 PRO HCB     H_       1  0   0.0970
 PRO #HB     H_       1  0   0.0970
 PRO HB      H_       1  0   0.0970
 PRO  CG     C_3      4  0   0.1520
 PRO HCG     H_       1  0  -0.0010
 PRO #HG     H_       1  0  -0.0010
 PRO #DG     H_       1  0  -0.0010
 PRO HG1     H_       1  0  -0.0010
 PRO HG2     H_       1  0  -0.0010
 PRO  CD     C_3      4  0  -0.0950
 PRO HCD     H_       1  0   0.0480
 PRO #HD     H_       1  0   0.0480
 PRO #DD     H_       1  0   0.0480
 PRO HD1     H_       1  0   0.0480
 PRO HD2     H_       1  0   0.0480
*  *PRO Done*
 SER  N      N_R      2  1  -0.7490
```

```
 SER HN       H__NR     1  0   0.3280
 SER  CA      C_A       4  0   0.1890
 SER HCA      H_        1  0   0.0480
 SER  HA      H_        1  0   0.0480
 SER  C       C_R       3  0   0.8280
 SER  O       O_2       1  2  -0.6790
 SER  OXT     O_2       1  2  -0.6790
 SER  CB      C_3       4  0   0.2960
 SER HCB      H_        1  0   0.0060
 SER #HB      H_        1  0   0.0060
 SER HB       H_        1  0   0.0060
 SER  H       H__NR     1  0   0.3280
 SER #H       H__NR     1  0   0.3280
 SER  OG      O_3       2  2  -0.7640
 SER HOG      H__O3     1  0   0.4910
 SER HG       H__O3     1  0   0.4910
 SER  HG      H__O3     1  0   0.4910
* *SER done*
 THR  N       N_R       2  1  -0.5460
 THR HN       H__NR     1  0   0.2820
 THR  CA      C_A       4  0  -0.1010
 THR HCA      H_        1  0   0.1350
 THR  HA      H_        1  0   0.1350
 THR  C       C_R       3  0   0.7760
 THR  O       O_2       1  2  -0.6860
 THR  OXT     O_2       1  2  -0.6860
 THR  CB      C_3       4  0   0.5580
 THR HCB      H_        1  0  -0.0300
 THR #HB      H_        1  0  -0.0300
 THR HB       H_        1  0  -0.0300
 THR  H       H__NR     1  0   0.2820
 THR #H       H__NR     1  0   0.2820
 THR  OG1     O_3       2  2  -0.7770
 THR HOG1     H__O3     1  0   0.4520
 THR HG1      H__O3     1  0   0.4520
 THR HOG      H__O3     1  0   0.4520
 THR #HG1     H__O3     1  0   0.4520
 THR #DG1     H__O3     1  0   0.4520
 THR HG1      H__O3     1  0   0.4520
 THR  CG2     C_3       4  0  -0.3840
 THR HCG2     H_        1  0   0.1070
 THR #HG2     H_        1  0   0.1070
 THR #DG2     H_        1  0   0.1070
*  *THR Done*
 TRP  N       N_R       2  1  -0.9290
 TRP HN       H__NR     1  0   0.3650
 TRP  CA      C_A       4  0   0.5640
 TRP HCA      H_        1  0  -0.0110
 TRP  HA      H_        1  0  -0.0110
 TRP  C       C_R       3  0   0.6900
 TRP  O       O_2       1  2  -0.6580
 TRP  OXT     O_2       1  2  -0.6580
 TRP  CB      C_3       4  0  -0.3000
 TRP HCB      H_        1  0   0.1110
 TRP #HB      H_        1  0   0.1110
```

```
 TRP HB       H_        1  0   0.1110
 TRP  H       H__NR     1  0   0.3650
 TRP #H       H__NR     1  0   0.3650
 TRP  CG      C_R       3  0  -0.1420
 TRP  CD1     C_R       3  0  -0.0290
 TRP HCD1     H_        1  0   0.1960
 TRP #HD1     H_        1  0   0.1960
 TRP #DD1     H_        1  0   0.1960
 TRP  CD2     C_R       3  0   0.0760
 TRP  NE1     N_R2      3  0  -0.5950
 TRP HNE1     H__NR     1  0   0.4550
 TRP HNE      H__NR     1  0   0.4550
 TRP HE1      H__NR     1  0   0.4550
 TRP #HE1     H__NR     1  0   0.4550
 TRP #DE1     H__NR     1  0   0.4550
 TRP  CE2     C_R       3  0   0.3410
 TRP  CE3     C_R       3  0  -0.1400
 TRP HCE3     H_        1  0   0.1110
 TRP #HE3     H_        1  0   0.1110
 TRP #DE3     H_        1  0   0.1110
 TRP  CZ2     C_R       3  0  -0.4010
 TRP HCZ2     H_        1  0   0.2120
 TRP #HZ2     H_        1  0   0.2120
 TRP #DZ2     H_        1  0   0.2120
 TRP HZ1      H_        1  0   0.2120
 TRP HZ2      H_        1  0   0.2120
 TRP  CZ3     C_R       3  0  -0.2600
 TRP #HZ3     H_        1  0   0.1660
 TRP HCZ3     H_        1  0   0.1660
 TRP #DZ3     H_        1  0   0.1660
 TRP  CH2     C_R       3  0  -0.0790
 TRP HCH2     H_        1  0   0.1460
 TRP #HH2     H_        1  0   0.1460
 TRP #DH2     H_        1  0   0.1460
 TRP HH       H_        1  0   0.1460
* *TRP Done*
 TYR  N       N_R       2  1  -0.9160
 TYR HN       H__NR     1  0   0.3520
 TYR  CA      C_A       4  0   0.6480
 TYR HCA      H_        1  0  -0.0180
 TYR  HA      H_        1  0  -0.0180
 TYR  C       C_R       3  0   0.6170
 TYR  O       O_2       1  2  -0.6380
 TYR  OXT     O_2       1  2  -0.6380
 TYR  CB      C_3       4  0  -0.3860
 TYR HCB      H_        1  0   0.1130
 TYR #HB      H_        1  0   0.1130
 TYR HB       H_        1  0   0.1130
 TYR  H       H__NR     1  0   0.3520
 TYR #H       H__NR     1  0   0.3520
 TYR  CG      C_R       3  0   0.0740
 TYR  CD1     C_R       3  0  -0.0800
 TYR HCD1     H_        1  0   0.1500
 TYR  HD1     H_        1  0   0.1500
 TYR  DD1     H_        1  0   0.1500
```

```
 TYR  CD2     C_R       3  0  -0.0800
 TYR HCD2     H_        1  0   0.1500
 TYR  HD2     H_        1  0   0.1500
 TYR  DD2     H_        1  0   0.1500
 TYR  CE1     C_R       3  0  -0.4520
 TYR HCE1     H_        1  0   0.2240
 TYR  HE1     H_        1  0   0.2240
 TYR  DE1     H_        1  0   0.2240
 TYR  CE2     C_R       3  0  -0.4520
 TYR HCE2     H_        1  0   0.2240
 TYR  HE2     H_        1  0   0.2240
 TYR  DE2     H_        1  0   0.2240
 TYR  CZ      C_R       3  0   0.5700
 TYR  OH      O_3       2  2  -0.7230
 TYR HOH      H__O3     1  0   0.5100
 TYR HH       H__O3     1  0   0.5100
 TYR  HH      H__O3     1  0   0.5100
*  *TYR Done*
 VAL  N       N_R       2  1  -0.8090
 VAL HN       H__NR     1  0   0.3500
 VAL  CA      C_A       4  0   0.1180
 VAL HCA      H_        1  0   0.0780
 VAL  HA      H_        1  0   0.0780
 VAL  C       C_R       3  0   0.8330
 VAL  O       O_2       1  2  -0.6910
 VAL  OXT     O_2       1  2  -0.6910
 VAL  CB      C_3       4  0   0.4030
 VAL HCB      H_        1  0   0.0060
 VAL #HB      H_        1  0   0.0060
 VAL HB       H_        1  0   0.0060
 VAL  H       H__NR     1  0   0.3500
 VAL #H       H__NR     1  0   0.3500
 VAL  CG1     C_3       4  0  -0.5130
 VAL HCG1     H_        1  0   0.1230
 VAL #HG1     H_        1  0   0.1230
 VAL #DG1     H_        1  0   0.1230
 VAL  CG2     C_3       4  0  -0.5130
 VAL HCG2     H_        1  0   0.1230
 VAL #HG2     H_        1  0   0.1230
 VAL #DG2     H_        1  0   0.1230
 VAL HG1      H_        1  0   0.1230
 VAL HG2      H_        1  0   0.1230
 VAL HG3      H_        1  0   0.1230
 VAL HG4      H_        1  0   0.1230
 VAL HG5      H_        1  0   0.1230
 VAL HG6      H_        1  0   0.1230
* *VAL Done*
C backbone and first side chain carbon
C for UNKnown residues from crystallographic
C studies
*  Taken from ALA charges
 UNK  N       N_R       2  1  -0.8570
 UNK HN       H__NR     1  0   0.3420
 UNK  CA      C_A       4  0   0.4870
 UNK HCA      H_        1  0   0.0110
```

```
 UNK  HA     H_       1  0   0.0110
 UNK  C      C_R      3  0   0.7260
 UNK  O      O_2      1  2  -0.6500
 UNK  OXT    O_2      1  2  -0.6500
 UNK  CB     C_3      4  0  -0.3950
 UNK  H      H__NR    1  0   0.3420
 UNK #H      H__NR    1  0   0.3420
 UNK HCB     H_       1  0   0.1120
 UNK #HB     H_       1  0   0.1120
 UNK HB      H_       1  0   0.1120
*
C  Methylated amino terminus
 CBX  N      N_3      3  0   0.0000
 CBX HN      H__N3    1  0   0.0000
 CBX H       H__N3    1  0   0.0000
 CBX  CA     C_A      4  0   0.0000
*
C  N-methyl aminine
* (Charges from Mulliken HF/631G**)
 NME  N      N_R      2  1  -0.9287
 NME HN      H_       1  0   0.0834
 NME H       H_       1  0   0.0834
 NME  CA     C_3      4  0   0.0136
 NME  C      C_3      4  0   0.0136
 NME HC1     H_       1  0  -0.0561
 NME HC2     H_       1  0  -0.0561
 NME HC3     H_       1  0  -0.0561
*
C  Conversion for ACE
*  (Charges from Mullikan HF/631G**.)
 ACE  CH3    C_3      4  0  -0.3977
 ACE  C3     C_3      4  0  -0.3977
 ACE  C      C_2      3  0   0.7119
 ACE  O      O_2      1  2  -0.7600
 ACE  C1     C_2      3  0   0.7119
 ACE  C2     C_3      3  0  -0.3977
 ACE  O1     O_2      1  2  -0.7600
 ACE  CG     C_2      3  0   0.7119
 ACE  OD1    O_2      1  2  -0.7600
 ACE  OD2    O_2      1  2  -0.7600
 ACE  CB     C_3      3  0  -0.3977
 ACE HC2     H_       1  0   0.0686
 ACE HCB     H_       1  0   0.0686
*
C  Formyl amino terminus
 FRM  C      C_R      3  0   0.0000
 FRM  O      O_2      1  2   0.0000
*
C  N terminus
 NTE  HT1    H_N3+    1  0   0.0000
 NTE  HT2    H_N3+    1  0   0.0000
 NTE  HT3    H_N3+    1  0   0.0000
 NTR  HT1    H_N3+    1  0   0.0000
 NTR  HT2    H_N3+    1  0   0.0000
 NTR  HT3    H_N3+    1  0   0.0000
```

```
*
C  C terminus
 CTE  OT2     O_2m      1  2  -0.0000
 CTR  OT2     O_2m      1  2  -0.0000
 CTE  OXT     O_2m      1  2  -0.0000
 ***  OXT     O_2m      1  2  -0.0000
*
C  Sulfate Ion
*  mjc - 4/9/00 HF/631G** calculation in h2o
 SO4  S       S_3       2  2   1.4320
 SO4  O1      O_2       1  2  -0.8580
 SO4  O2      O_2       1  2  -0.8580
 SO4  O3      O_2       1  2  -0.8580
 SO4  O4      O_2       1  2  -0.8580
*
*
C  Water
 HOH  OH2     O_3F      2  2  -0.8200
 HOH  O       O_3F      2  2  -0.8200
 HOH  H1      H_F       1  0   0.4100
 HOH  H2      H_F       1  0   0.4100
 HOH HO       H_F       1  0   0.4100
 WAT  OH2     O_3F      2  2  -0.8200
 WAT  O       O_3F      2  2  -0.8200
 WAT  H1      H_F       1  0   0.4100
 WAT  H2      H_F       1  0   0.4100
 WAT HO       H_F       1  0   0.4100
 WAT HOH2     H_F       1  0   0.4100
 DOD  OD2     O_3F      2  2  -0.8200
 DOD  O       O_3F      2  2  -0.8200
 DOD  D1      H_F       1  0   0.4100
 DOD  D2      H_F       1  0   0.4100
 DOD DO       H_F       1  0   0.4100
 H2O  OH2     O_3F      2  2  -0.8200
 H2O  O       O_3F      2  2  -0.8200
 H2O  H1      H_F       1  0   0.4100
 H2O  H2      H_F       1  0   0.4100
 H2O HO       H_F       1  0   0.4100
 OH2  OH2     O_3F      2  2  -0.8200
 OH2  O       O_3F      2  2  -0.8200
 OH2  H1      H_F       1  0   0.4100
 OH2  H2      H_F       1  0   0.4100
 OH2 HO       H_F       1  0   0.4100
*
*
* The following Atom types & charges are
*     Dreidii defaults....
*
*
C  Copper (put in as zinc for now)
 *** CU       Zn       -4  0   2.0000
 ***  CU      Zn       -4  0   2.0000
 *** Cu       Zn       -4  0   2.0000
 ***  Cu      Zn       -4  0   2.0000
*
```

```
C  Zinc
 *** ZN       Zn       -4  0   2.0000
 ***  ZN      Zn       -4  0   2.0000
 *** Zn       Zn       -4  0   2.0000
 ***  Zn      Zn       -4  0   2.0000
*
C  Calcium
 *** CAL      Ca       -4  0   2.0000
 ***  CAL     Ca       -4  0   2.0000
 *** CA       Ca       -4  0   2.0000
 ***  CA      Ca       -4  0   2.0000
 *** Ca       Ca       -4  0   2.0000
 ***  Ca      Ca       -4  0   2.0000
*
C  Barium (as Ca)
 *** BA       Ca       -4  0   2.0000
 ***  BA      Ca       -4  0   2.0000
 *** Ba       Ca       -4  0   2.0000
 ***  Ba      Ca       -4  0   2.0000
*
C  Sr (as Ca)
 *** SR       Ca       -4  0   2.0000
 ***  SR      Ca       -4  0   2.0000
 *** Sr       Ca       -4  0   2.0000
 ***  Sr      Ca       -4  0   2.0000
*
C  Sodium
 *** NA       Na       -6  0   1.0000
 ***  NA      Na       -6  0   1.0000
 *** Na       Na       -6  0   1.0000
 ***  Na      Na       -6  0   1.0000
*
C  Iron
 *** FE       Fe       -6  0   3.0000
 ***  FE      Fe       -6  0   3.0000
 *** Fe       Fe       -6  0   3.0000
 ***  Fe      Fe       -6  0   3.0000
*
C  Chlorine
! ***  CL     CL_B     -6  0  -1.0000
! *** CL       CL_B     -6  0  -1.0000
C  Titanium
 *** TI       Ti       -6  0   4.0000
 ***  TI      Ti       -6  0   4.0000
 *** Ti       Ti       -6  0   4.0000
 ***  Ti      Ti       -6  0   4.0000
*
C  La (as Ti)
 *** LA       Ti       -6  0   4.0000
 ***  LA      Ti       -6  0   4.0000
 *** La       Ti       -6  0   4.0000
 ***  La      Ti       -6  0   4.0000
*
C  Ruthenium
 *** RU       Ru       -6  0   3.0000
```

```
 ***  RU      Ru       -6  0   3.0000
 *** Ru       Ru       -6  0   3.0000
 ***  Ru      Ru       -6  0   3.0000
*
C  Yittrium (as Ru)
 *** Y_       Ru       -6  0   3.0000
 ***  Y_      Ru       -6  0   3.0000
*
*
C  Bromine conversion to Dreiding atom type
  BR BR       Br        1  0
*
```