

**List of Tables**

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. ....	32
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.....	33
Table 2- 3: Average QM charges and BUFF final charges for the N-terminus methyl. ...	34
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.....	34
Table 2- 5: Average QM charges and BUFF final charges for the C-terminus methyl. ...	34
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. ....	39
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 $\text{Cos}(4q)$ term was needed.....	48
Table 2- 8: A listing of energies at selected conformations of the Gly-Gly-Gly tripeptide. ....	51
Table 2- 9: A listing of energies at selected conformations of the Gly-Ala-Gly tripeptide. ....	51
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.....	51
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.....	54
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.....	56
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.....	58
Table 2- 14: Energy (in kcal/mol) of special points of $\phi/\psi$ scan of N-terminus alanine in alanine tetrapeptide helix. ....	62
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. ....	67
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.....	68
Table 2- 17: A CRMS comparison of BUFF minimized cytochrome C553 (1C75) and the crystal structure. ....	69
Table 2- 18: Extended to helix transition energies of the 4 alanine polypeptide. The endpoints were capped as in Figure 2-26 to neutralize the endpoints.....	71
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.....	73
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. ....	74
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.....	74
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: $E_{\text{torsion}} = A*\text{Cos}(q) + B*\text{Cos}(2q) + C*\text{Cos}(3q)$ . Prior to the torsional fit, all involved torsions but the w torsion were zeroed. The w torsion, $C_{\alpha}\text{-N-C-C}_a$ , was not fit, but was left with a barrier of 10 kcal/mol and a periodicity of 2. $C_{\delta}$ is the side-chain $\delta$ -carbon of proline adjacent to the main chain nitrogen. ....	83
Table 2- 23: The Energy Minima and the Energy Differences of the Minima to the Global Minimum Are Shown with the Conformational $\phi$ and $\psi$ Angles. $\Delta E_{\text{vac}}$ : relative total energy in vacuum. $\Delta E_{\text{sol}}$ : relative solvation energy in water. $\Delta E_{\text{wat}}$ : relative total energy in water.....	87
Table 2- 24: The Relative Energy (kcal/mol) of the $\alpha$ -Helix and $\beta$ -Sheet Conformations to the Global Minimum in Vacuum and Water. All energies are from <i>ab initio</i> calculations (HF, 6-31G** basis) on Gly-X-Gly with a Poisson-Boltzmann description of the solvent. $\alpha$ -Helix is a right-handed $\alpha$ -helix, where (f,y)= (-57, -47). p- $\beta$ -sheet is a parallel $\beta$ -sheet, where (f,y)=(-119, 113). a- $\beta$ -sheet is an antiparallel $\beta$ -sheet, where (f,y)=(-139, 135). $\Delta E_{\text{vac}}$ : relative total energy in vacuum; $\Delta E_{\text{sol}}$ : relative solvation energy in water; $\Delta E_{\text{wat}}$ : relative total energy in water.....	88

## 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 Waals 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.

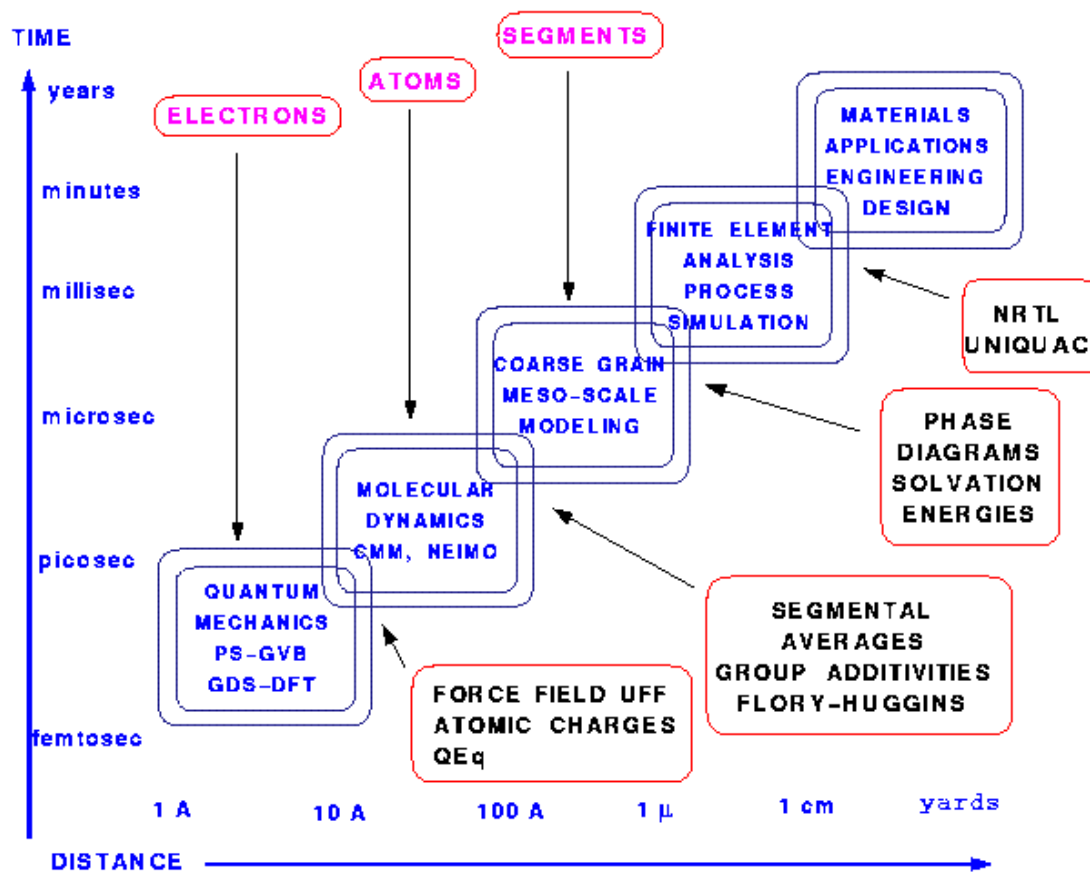
### ***Molecular Modeling***

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.

## The Hierarchy of Materials Modeling



**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  $\Psi$ . 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,  $\Psi_n$ , 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  $\text{H}_2^+$  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,  $\Psi$ , 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,  $\phi_i$ , describes the orbital of a single electron, as shown in equation (1.2).

$$\Psi = f_1 f_2 f_3 \dots f_n \quad (1.2)$$

The total wavefunction must still respect the Pauli principle and be antisymmetric with respect to electron exchange. To construct the individual  $\phi_i$  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.

$$f_i = \sum_k c_{ik} \chi_k \quad (1.3)$$

Here,  $c_{ik}$  are coefficients and  $\chi_k$  is an atomic orbital function. The set of  $\chi$ 's is called a basis set. The problem of solving for  $\Psi$  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,  $\phi_i$ , by using a one-electron Hamiltonian (1.4) containing the interactions with the other electrons.

$$Hf_i = e_i f_i \quad (1.4)$$

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

$$\sum_k c_{ik} \left( \int \mathbf{c}_i H \mathbf{c}_k dv \right) = \epsilon_i \sum_k c_{ik} \left( \int \mathbf{c}_i \mathbf{c}_k dv \right) \quad (1.5)$$

$$\sum_k c_{ik} \left( \int \mathbf{c}_i H \mathbf{c}_k dv - \epsilon_i \int \mathbf{c}_i \mathbf{c}_k dv \right) = 0 \quad (1.6)$$

Now we find the problem that plagues much of quantum mechanics. We can calculate a set of  $\epsilon_{ik}$  by solving equation (1.6) for a given Hamiltonian, but because H in (1.4) depends on all the orbitals  $\phi_i$ , 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  $\epsilon_i$ , and using this temporary set of  $\epsilon_i$  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  $\text{H}_2^-$  molecule. HF calculations arrive at the incorrect dissociation limit for  $\text{H}_2^-$ . 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:

$$c_k = C e^{-\zeta r} Y_{lm} \quad (1.7)$$

where  $Y_{lm}$  is the angular component of the function and  $\zeta$  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).

$$E_{tot} = E_{valence} + E_{nonbond} \quad (1.8)$$

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.

$$E_{valence} = E_{bond} + E_{angle} + E_{torsion} + E_{inversion} \quad (1.9)$$

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  $R$ .

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

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

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.

$$E_{angle} = \frac{1}{2} K_a (\mathbf{q} - \mathbf{q}_0)^2 \quad (1.12)$$

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  $\phi=0^\circ$  or  $\phi=180^\circ$ .

$$E_{torsion} = \sum_{n=1}^6 \frac{1}{2} K_{f,n} (1 - d \cos(n\mathbf{f})) \quad (1.13)$$

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  $n = 3$ .

$$E_{inversion} = \frac{1}{2} K_y \cos[n(\mathbf{y} - \mathbf{y}_o)] \quad (1.14)$$

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

$$E_{inversion} = \frac{1}{2} C (\cos \mathbf{f} - \cos \mathbf{f}_o)^2, \text{ where } K_f = C \sin^2 \mathbf{f}_o \quad (1.15)$$

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.

$$E_{nonbond} = E_{electrostatic} + E_{vdW} + E_{hbond} \quad (1.16)$$

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.

$$E_{elec} = \sum \frac{q_i q_j}{r_{ij}} \quad (1.17)$$

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.”

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

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.

$$E_{\text{exp-6}} = D_0 \left\{ \left[ \left( \frac{6}{\mathbf{x}-6} \right) \exp^{x \left( 1 - \frac{R}{R_0} \right)} \right] - \left[ \left( \frac{6}{\mathbf{x}-6} \right) \left( \frac{R_0}{R} \right)^6 \right] \right\} \quad (1.19)$$

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  $R$ .

$$E_{\text{pureexp}} = D_0 \exp^{g \left( 1 - \frac{R}{R_0} \right)} \quad (1.20)$$

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  $R$ .

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

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.

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

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.

$$E_{Dreidingh-bond} = D_0 \left[ 5 \left( \frac{R_0}{R_{DA}} \right)^2 - 6 \left( \frac{R_0}{R_{DA}} \right)^{10} \right] \cos^2 \mathbf{q}_{DHA} \quad (1.23)$$

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