Force Fields in Molecular Mechanics

Rajarshi Guha (9915607)
and
Rajesh Sardar (9915610)

March 21, 2001

1 Introduction

With the advent of computers chemists have realized the utility of carrying out simulations and studies on various chemical systems using the computer as a tool. Molecular mechanics is one of the computational techniques used by chemists to study molecular systems. Basically molecular mechanics was developed out of a need to describe molecular structures and properties in as *practical* a manner as possible. Here *practical* implies that results should be accurate, available within a reasonable amount of computational time and be applicable to a large variety of systems. As a result of these requirements quantum mechanical first principle methods and semi empirical methods are not always suited (or even feasible). The range of applicability of molecular mechanics (aka force field) techniques include:

- Molecules containing thousands of atoms.
- Organics, oligonucleotides, peptides, and saccharides (metallo-organics and inorganics in some cases).
- Vacuum, implicit, or explicit solvent environments.
- Ground state only.
- Thermodynamic and kinetic properties.

Essentially molecular mechanics is an attempt to provide a reliable recipe for obtaining the potential energy surface for the movement of atoms within molecules.[1] Molecular mechanics is characterized by high computational speed in comparison to first principle or semi empirical methods. This arises from the simple mathematical formulation - mainly simple arithmetic expressions rather than complex integral evaluations. As a result of the speed the technique allows for its use in procedures such as molecular dynamics, conformational energy searching, and docking, that require large numbers of energy evaluations. In the proceeding sections we cover the basis of molecular mechanics - force fields. What they are and how they are designed. We also consider the special case of force fields for metals. Finally we provide a brief comparison of force field and quantum mechanical techniques.

2 What is a Force Field?

Very briefly, a force field is a mathematical function which returns the energy of a system as a function of the conformation of the system. But a better idea may be obtained by considering the situation physically.

Consider a molecule as a collection of atoms held together by elastic forces. (If you want to get even simpler then one could consider a molecule to be a collection of point masses connected by elastic springs). Now the forces can be written in terms of potential energy functions of various structural features such as bond lengths, bond angle, non bonded interactions etc. The force field is the combination of these potential energy terms. Hence force fields are also sometimes referred to as *potentials*. Thus the energy, E, of a molecule in a force field arises from the deviations from the *ideal* structural features and so can be written approximately as

$$E = E_s + E_b + E_\omega + E_{nb} + \cdots$$

Here E is termed as the steric energy. E_s is the energy for bond stretching, E_b is the energy for bond angle bending, E_ω is the torsional energy due to twisting about bonds and E_{nb} is the energy for non bonded interactions. If there are other mechanisms affecting energy such as electrostatic repulsions or hydrogen bonding then these may be included in E by adding appropriate terms into the above expression. It should be kept in mind that E is a measure of the intra-molecular strain relative to a hypothetical situation.

Thus E on its own has no physical meaning. Many different kinds of force-fields have been developed over the years. Some include additional energy terms that describe other kinds of deformations. Some force-fields account for coupling between bending and stretching in adjacent bonds in order to improve the accuracy of the mechanical model.

3 Design of a Force Field

The design of force fields for molecular mechanics is guided by by the following principles:

- Nuclei and electrons are lumped into atom-like particles.
- Atom-like particles are spherical (radii obtained from measurements or theory) and have a net charge (obtained from theory).
- Interactions are based on springs and classical potentials.
- Interactions must be preassigned to specific sets of atoms.
- Interactions determine the spatial distribution of atom-like particles and their energies.

3.1 Stretching & Bending

Considering the idea of a molecule to be a collection of masses connected by springs . Thus by applying Hookes Law we can evaluate the energy required to stretch and bend bonds from their ideal values. Thus E_s and E_b may be expressed as:

$$E_s = \sum_{i=1}^{N} \frac{k_i^s}{2}(l_i - l_i^0)^2$$

$$E_b = \sum_{i<j}^{M} \frac{k_{ij}^b}{2}(\theta_{ij} - \theta_{ij}^0)^2$$

where N is the total number of bonds and M is the total number of bond angles in the molecule. k^s and k^b are the force constants for stretching and bending respectively. l_i and θ_{ij} are the actual bond lengths and bond angles. Finally l_i^0 and θ_{ij}^0 are ideal bond lengths and bond angles. Unique k^s, and l_i^0

and values are assigned to each pair of bonded atoms based on their types (e.g. C-C, C-H, O-C, etc.). Similarly k^b and θ^0_{ij} parameters for angle bending are assigned to each bonded triplet of atoms based on their types (e.g. C-C-C, C-O-C, C-C-H, etc.).

The formulation above is only a first approximation. There are various factors which can be taken into account to improve the accuracy for these terms. These include noting that bond stretching requires more energy than bond bending and so for a molecule being deformed most of the distortion should occur in the bond angles rather than bond lengths. Another point to consider is that Hookes Law overestimates the energy required to achieve large distortions. Another aspect is that as a bond angle gets compressed the two associated bond lengths become longer.

3.2 Torsion

Now we consider the form of the E_ω term. The energy due to torsion is usually expressed in terms of a Fourier series,

$$E_\omega = \sum \frac{1}{2}[V_1(1+\cos\omega) + V_2(1+\cos 2\omega) + V_3(1+\cos 3\omega) + \cdots]$$

where the sum is over all unique sequences of bonded atoms. In general the series is truncated at the third term, V_1, V_2 and V_3 being chosen so that the resultant conformation agree well with experiment for a given group of molecules.

3.3 Non Bonded Interactions

The final term contributing to E is the energy from pairwise non bonded interactions. Such interactions are modeled by London dispersive forces (for the attraction) and Van der Waals forces (for the repulsion). Some of the common potential functions implementing the above are the Lennard Jones and Buckingham potentials,

$$V_{LJ} = \frac{A}{r^{12}} - \frac{B}{r^6}$$

$$V_{Buck} = A' \exp(\frac{B'}{r}) - \frac{C}{r^6}$$

Usually the parameters for the non bonded energy terms are obtained by measuring non bonded contact distances[1] in crystalline hydrocarbons, diamond, graphite and Van der Waals contact data for rare gas atoms. Parameters for other atoms are then obtained by extrapolation or interpolation. One major assumption is that the potential derived from *intermolecular* interactions can accurately reproduce *intramolecular* interactions. In addition the interactions are considered to be pairwise additive.

3.4 Electrostatic Interactions

An important non bonded energy term that is *always* taken into account is the electrostatic interactions. Typically the electrostatic interaction dominates the total energy of a system by a full magnitude. The electrostatic contribution is modeled using a Coulombic potential,

$$E_{Coul} = \sum_{i,j=1} \frac{q_i q_j}{r_{ij}}$$

The electrostatic energy is a function of the charge on the non-bonded atoms, their inter atomic distance, and a molecular dielectric expression that accounts for the decrease of electrostatic interaction due to the environment (such as by solvent or the molecule itself). A linearly varying distance-dependent dielectric (i.e. $1/r$) is sometimes used to account for the increase in environmental bulk as the separation distance between interacting atoms increases. The accuracy of the electrostatic term depends on the correct assignment of charges to individual atoms. There are two main problems in charge assignment:

- Experimental data of charges for molecules at atom resolutions does not exist.
- Molecular mechanics assumes that nuclei are isolated entities. The fact that nuclei share common electrons is not considered. Thus exact localization of electrons by considering atoms as fixed point charges is not correct

Charge assignment methodology differs from force field to force field and are more art than science.

Summarizing we see that the terms describing energy changes from bond length, bond angle and torsions are well understood and can be accurately included in the overall energy expression. The most influential term, the electrostatic term, however is not fully understood. Hence the variation in results from different force fields can be attributed, to a large extent , to the electrostatic term. Of course the nature of the parametrization generating the various force constants and ideal lengths and angles will also affect the applicability of a force field. In general once must consider with which group of molecules or systems a given force field has been parametrized - keeping this fact in mind one may then use the force field on an unknown but similar system.

Generality is still a problem with force fields, though with the development of the Universal Force Field[2] (UFF) an attempt has been made to develop a generalized force field applicable to a large portion of the periodic table and not be restricted to particular groupings of atoms such as proteins, nucleic acids etc.

4 Force Fields for Metals

The basis of the preceding force fields when considering molecules and atoms was pairwise potentials. The simplest example is the Lennard Jones potential. However such potentials are limited in applicability. The most primitive types are applicable only to the noble gas elements where there are no electrons available for bonding and attractions are via weak Van der Waals forces. Of course the more sophisticated force fields take into account more interactions than the simple Lennard Jones or Buckingham potentials and thus model a wider range of real systems. However the systems considered are molecules, large though they may be, yet still isolated molecules. When it comes to modeling metals and semiconductors such pairwise potentials fail.

The effects of many body interactions (ie interactions in metals) can be understood from the following table where experimental data for a few metals are compared with Lennard Jones data.

Property	Cu	Ag	Pt	Au	LJ
E_c/k_bT_m	30	28	33	33	13
E_v/E_c	.33	.36	.26	.25	1
C_{12}/C_{44}	1.5	1.9	3.3	3.7	1

Here E_c/k_bT_m is the ratio between the cohesive energy and melting temperature. This ratio is about 30 in metals and around 10 in two body systems.

This indicates that metals exhibit some extra cohesion with respect to pairwise systems. C_{12}/C_{44} is the ratio between the two elastic constants of a cubic crystal. For a two body system this ration is exactly 1 (and can be shown analytically, the *Cauchy relation*), but deviations in metallic systems are common. E_v/E_c is the ratio between vacancy formation energy and cohesive energy and is around $\frac{1}{4}$ to $\frac{1}{3}$ in metals but around 1 in two body systems. So it is quite clear that two body potential functions are not suited for describing metallic systems.

Progress was made in the development of many body potentials for metals by considering the concept of density or coordination as the key variable. In such a situation the main (physical) point to remember is that bonds become weaker when the local environment becomes more crowded, which is essentially a consequence of the Pauli principle. Thus a plot of cohesive energy as a function of coordination should not decrease linearly as in two body systems but should exhibit a positive curvature. Following this direction a possible form for the attractive part of the potential (the repulsive one still being modeled by a two body potential) can be qualitatively written as

$$E_i \propto \sqrt{\sum_j h_{ij}^2} \propto \sqrt{Z_i}$$

where $h_{ij} = h(r_{ij}) = \langle i|H|j\rangle$ are the overlap integrals between i and its neighbors. This relation is based the tight binding formalism[3] which is basically the MNDO method as applied to solids. In the tight binding formalism the orbitals are localized and these functions vanish beyond a certain cutoff distance. Using this consideration several many atom potential functions have been designed all based on the analytical form:

$$V = \frac{1}{2}\sum_{i,j=1}^{N} \phi(r_{ij}) + \sum_{i=1}^{N} U(n_i)$$

where $\phi(r)$ is a two body part and $U(n)$ is a function giving the energy of the atom as function of *generalized coordination* n. For a given atom n is constructed from a superposition of contributions from neighboring atoms,

$$n_i = \sum_{j=1}^{N} \rho(r_{ij})$$

where $\rho(r)$ is a short ranged, decreasing function of distance. Other schemes that can be grouped under this method include the glue model,[4] the embedded atom method,[5] the Finnis Sinclair potentials[6] and the effective medium theory.[7] Even though the various methods mentioned here share the same analytical form they differ greatly in the way they implement the $\rho(r)$, $\phi(r)$ and $U(r)$ functions.

Both the embedded atom method and effective medium theories are based on density functional theory. That is they consider the total energy to be dependent on the electron density to capture some electronic aspects of bonding and then add empirical pair potentials which are fitted to reproduce experimental geometries, bulk moduli etc. The advantage of these theories is that they overcome the difficulties of applying pair potentials to metallic systems but retain the computational simplicity of pairwise potential techniques; essentially a MD program for metals must evaluate the forces,

$$\vec{F_i} = -\sum_{j \neq i} \left(\phi'(r_{ij}) + [U'(n_i) + U'(n_j)]\rho'(r_{ij})\right) \frac{\vec{r_{ij}}}{r_{ij}}$$

This calculation is only a little more complex than that required for a two body system. Energy and forces can still be obtained using pair wise distances only - ie no angular term or 3 body or 4 body terms are present. Of course the absence of angular forces makes it difficult for these schemes to model metallic systems where covalent effects are important in bonding.

In conclusion we see that force fields for metals are quite a complex matter (and for the case of semiconductors, which have not been discussed here, the design of potentials is a very challenging issue. Examples of semiconductor force fields are the Stillinger Weber[8] potential and the Tersoff potential[9] - which were designed with silicon in mind). Further details regarding metallic force fields may be obtained from the reviews mentioned in the reference.[10][11][12][13][14][15]

5 A Comparison of Force Field Methods and Semi Empirical Methods

In this section we look at a comparison of the applicability of the force field methods and semi empirical methods - the drawbacks and advantages of both techniques.

One of the major drawbacks of force field methods is that many molecules of interest to a chemist are outside the range of molecules for which current force fields are parametrized. Of course there are still a large number of molecules which can be studied using current force fields.In addition it is sometimes possible to guess parameters for unknown bonding situations from chemically similar situations.

5.1 Applicability

Another feature which is a disadvantage for the use of force field methods is the fact that certain problems are essentially electronic in nature. By construction any property related to the electronic structure such as electrical conductivity, optical and magnetic properties are not accessible from force field calculations. In these situations the use of quantum mechanical methods, either first principle or semi empirical methods, are advised. Thus there are a number of problems which would be better treated by such quantum mechanical methods. Some examples include orbital population analysis, evaluation of vibrational frequencies,[16] , electrostatic potentials/densities, charges, following transition states etc. Certain properties like thermodynamic parameters, bulk parameters (for solid systems), diffusive processes, structure and dynamics etc are well evaluated by using force field methods. For these type of problems quantum mechanical techniques are ill suited, mainly due to the extensive electron correlations present in the systems. In such cases a well parametrized force field will yield more accurate results in shorter times. Thus force field approaches can be classified as structural tools.

5.2 System Size

At the same time the study of large systems - large molecules as well as large numbers of molecules/atoms - is carried out more effectively using force field methods. An example can be seen from the study of solid systems - the embedded atom method referred to earlier can handle millions of metal atoms at a time whereas a semi empirical method (such as MNDO) is restricted to 1000 atoms. Polymer dynamics will obviously be handled better by force field methods. At the same time polymer structure can be handled by semi empirical methods, but again the size of the system is limited. Considering various force field methods themselves we see that the sophisticated valence force fields can handle tens of thousands of atoms at a time. Going to simpler force fields (Lennard Jones) one may handle millions of particles at a time

5.3 Timescales

A very important feature of force field techniques is the time scales that can be accessed. Such techniques can explore nano second range time scales. Quantum mechanical techniques are usually limited to the pico second regime. But even then there are very few MD calculations that have been carried out using semi empirical or first principle techniques.

5.4 Accuracy

Accuracy is an important area of comparison. Within the domains for which a given force field has been optimized it can give highly accurate values regarding bond lengths and angles. This is especially true of the force fields for proteins[17][18][19].[20] At the same time it should be kept in mind that high levels of accuracy and wide generality do not go hand in hand. Caution must be taken when applying a force field, parametrized and optimized for a specific system, to a new system. The results from such a task can be misleading if not absolutely wrong. On the other hand semi empirical methods can be accurate or inaccurate depending on the nature of the parameters introduced. In general compared to first principle quantum mechanical techniques semi empirical methods are less accuracies by an order of magnitude. An important point to be noted in force field calculations is that of *over interpreting* results. One single component of the steric energy in a molecule cannot be calculated in absence of other potentials. Thus individual terms like stretching, torsion, bending etc have no *absolute* meaning - all the components must be summed up to get a meaningful value for E.

5.5 Generality

Generality is an important feature of comparison. With the exception of the UFF[2] there are few transferable force fields. In general force fields are best used within the class of compounds for which the parameters had been fitted. Clearly one must have a clear idea of the domain of applicability of the force field in question before using it. Valence force fields are more meaningful for organic and bio molecules since such force fields depend on the concept of well defined bonds. Such a force field would not give very good results when applied to organometallics and would be useless for metals and alloys. For the case of semi empirical methods the simplification of first principles and introduction of system specific parameters limits the generality of such methods.

5.6 Computational Efficiency

Finally we must consider computational efficiency. For small systems (100 atoms) force field methods are nearly a 1000 times faster than first principal methods. If pair wise interactions are truncated then force field calculations can be considered $O(N)$ methods. Of course computational effort also increases with the time span being studied. Force field methods have been implemented on parallel architectures, but it has been seen that there is a significant transfer of data between memory and processors - so that the inter CPU communications and memory transfer times become a deciding factor for the overall efficiency. In the case of semi empirical methods matrix diagonalization is usually the computational bottle neck, such algorithms being $O(N^2)$ in nature.

6 Conclusion

Thus we have presented the basic definitions and techniques of force field calculations. As can be seen force field techniques are not a panacea - there are a large number of systems and problems which are not well suited for study with force field techniques. But even so many important areas of chemistry have been investigated. Proteins and bio molecules have been very fruitfully studied. And in recent times metal, alloy and semiconductor force fields have made significant progress. As can be seen a large range of materials can be modeled and studied by force field methods. The advantages of force field methods - mainly computational simplicity and speed - allow MD runs and Monte Carlo simulations on large systems allowing for significant statistical sampling. Furthermore, in comparison to first principle methods there are many choices for the analytic form of a force field and corresponding parameters. Thus for a given problem one has wide latitude in selecting a specific force field to deal with the problem at hand. Such a free hand is not available when considering attacking a problem with first principle or even semi empirical techniques.

References

1. Boyd, D. B.; Lipkowitz, K. B. *J. Chem. Ed.* **1982,** *59,* 269-274.
2. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992,** *114,* 10035.
3. Carlsson, A. *Solid State Phys.* **1990,** *43,* 1.
4. Ercolessi, F.; Parrinello, M.; Tosatti, E. *Philos. Mag.* **1988,** *A58,* 213.
5. Daw, M. S.; Baskes, M. I. *Phys. Rev.* **1984,** *B 29*, 6443.
6. Finnis, M. W.; Sinclair, J. E. *Philos. Mag.* **1984,** *A 50,* 45.
7. Jacobsen, K. W.; Nrskov, J. K.; Puska, M. J. *Phys. Rev.* **1987,** *B 35,* 7423.
8. Stillinger, F.; Weber, T. A. *Phys. Rev.* **1985,** *B 31,* 5262.
9. Tersoff, J. *Phys. Rev.* **1988,** *B 37,* 6991.
10. Brubaker, G.; Johnson, D. *Coord. Chem. Rev.* **1984,** *53,* 1.
11. Hancock, R. *Progr. Inorg. Chem.* **1989,** *37,* 187.
12. Landis, C.; Root, D.; Cleveland, T. "Molecular mechanics force fields for modeling inorganic and organometallic compounds". In *Reviews in Computational Chemistry*, Vol. 6; Lipowitz, K.; Boyd, D., Eds.; 1995.
13. Hay, B. *Coord. Chem. Rev.* **1993,** *126,* 177-236.
14. Comba, P. *Coord. Chem. Rev.* **1999,** *185-186,* 81-98.
15. Norrby, P. O.; Brandt, P. *Coord. Chem. Rev.* **2001,** *212,* 79-109.
16. Guha, R. “A Theoretical Study of Cyclopropane Carboxylic Acid”, Technical Report, IIT Kharagpur, India, 2001.
17. Cornell, W. D. e. a. *J. Am. Chem. Soc.* **1995,** *115,* 9620.
18. Weiner, *J. Am. Chem. Soc.* **1983,** *106,* 765.
19. Brooks, B. *J. Comp. Chem.* **1983,** *4,* 187.

20. MacKerel, *J. Am. Chem. Soc.* **1995,** 11946.

21. Sutton, A. P.; Chen, J. *Phil. Mag. Lett.* **1990,** *61,* 139.