

**Developing the Polarizable Protein Force Field:
Focus on Amino Acid Residues**

by

Qina Sa

A Thesis

Submitted to the Faculty

of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Chemistry and biochemistry

Aug 2011

APPROVED:

Professor George A. Kaminski, Research Advisor

Abstract

Polarizable force field has been successfully used in molecular modeling for years, especially in biological and protein simulations. In this research thesis, development of a new polarizable force field —POSSIM (POLarizable SIMulations with Second order Interaction Model) involving electrostatic polarization is described and parameters for several protein residues were produced.

In this research thesis, the POSSIM force field was extended to the side chains of the following residues: lysine, glutamic acid, protonated histidine, phenylalanine and tryptophan. This work involved producing parameters for methyl ammonium, acetate ion, imidazolium cation, benzene and pyrrole molecules. The parameters fitting procedure starts from the molecular complex with dipolar electrostatic probes of a many-body system to produce polarizabilities, compute the energies, then charges and Lennard-Jones parameters are produced by fitting to gas-phase dimerization calculations, followed by the torsional parameters fitting and end up with the pure liquid simulations. In all the cases, three-body energies, dimerization energies and distances agree well to the accurate quantum mechanical results. The final parameters obtained assured the error of less than 2% in the heat of vaporization and average volume results compared with the available experimental data.

Unlike the quantum mechanical calculations, the polarizable force field

computations require a relatively small amount of computational resources. Moreover, compared to fixed-charges empirical force fields, polarizable force fields are much more accurate in a number of energy calculations. In the following chapters, the results obtained with this particular polarizable force field are discussed.

Acknowledgment

I would like to express my gratitude to my adviser, Professor George A. Kaminski, who is kindly supporting me as his research assistant and helping me with the research. Professor Kaminski is always very patient with my endless questions and always ready to help. I am so thankful to be his student.

I want to thank all my lab mates. Dr. Tim Click, a previous postdoc of our lab, although he has gone to Kansas for new job. Dr. Sergei Y. Ponomarev, the postdoc of our lab who helped a lot about my project. Zhen Chen, an undergrad student who worked in our lab last summer and winter. Zhen helped me with the computer programming part. I would also like to thank Xinbi Li, Ity Sharma, another two graduate students in our lab.

It is great to be lab mates and classmates with both of them. Dr. Haijun Yang, the new postdoc in our lab who is very helpful.

I have to gratefully appreciate the love and support from my family and Friends, and all the other people who ever helped me.

Table of Contents

Abstract	i
Acknowledgment	iii
Table of Contents	iv
1. Introduction	1
References	12
2. Background	14
2.1 Introduction	14
2.2 From Fixed-Charges Force Field to Polarizable Force Field	14
2.3 Parameterization of the Polarizable Force Field	24
References	31
3. Results and Discussion	32
3.1 Introduction	32
3.2 Electrostatic parameterization in many-body system	33
3.3 Gas-Phase Dimerization Energies and Distances	35
3.4 Torsional parameters fitting	39
3.5 Pure Liquid Simulations-The Final Tuning of the Force Field	41
4. Summary	44
List of tables and Figures	46

Chapter 1 Introduction

Computational chemistry has been widely and successfully used in chemistry and biochemistry research^{[1][2]}. It permits to predict and explain both structures and energetics of biomolecules. The main issue in computational chemistry methods is usually calculating energies of the molecular systems. Even if a project is focused on structure prediction, this is usually done by finding the potential energy minimum. Quantum mechanics (QM) has been an approach to energy calculations for decades^[3]. However, application of QM to protein simulations is problematic for a number of reasons. First, QM calculations need prohibitively large computational resources when dealing with large size systems. Second, QM calculations can produce different results when using different levels of theory, and the deviation can be very noticeable for both large and small systems.

Empirical force fields are used to overcome the above difficulties. A force field is a group of functions and parameters which describe molecular system energetics. They are derived from quantum mechanical calculations and experimental data. Many traditional force fields used for macromolecular systems use Hamiltonians (total energy expressions) which contain several functionally common parts: valence bond stretching, valence angle bending, torsional energies, Coulomb and Lennard-Jones terms for non-bonded interactions. Eq.(a) is common to such force fields including AMBER^[4], CHARMM^{[5][6]}, and OPLS^[7], among others. The standard forms of these force fields

use fixed Coulomb charges^[8].

$$U(\vec{R}) = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedral} \times K_\chi(1+\cos(n\chi-\delta)) + \sum_{impropers} K_{imp}(\varphi-\varphi_0)^2 + \sum_{nonbond} (\epsilon_{ij}[(\frac{Rmin_{ij}}{r_{ij}})^{12} - (\frac{Rmin_{ij}}{r_{ij}})^6]) + \frac{q_i q_j}{\epsilon r_{ij}} \dots\dots\dots(a)$$

There are also some additional or alternative terms beyond Eq.(a). A number of force field including higher order terms, for example, to treat the bond and valence terms^{[9][10][11][12][13][14]}, while others involved the alternative use of a Morse function for bonds^{[15][16]}. For non bonded portion of potential energy, the alternate form involves both treatment of both van der Waals and electrostatic interactions. For example, a buffered LJ 14-7 term is used in MMFF force field, the exponential term used by Buckingham potential for repulsion treatment^[17]. However, although this kind of alternative can treat repulsive wall more accurately^[18], the harmonic internal terms in earlier force fields suggest that for biomolecular simulations near room temperature, the 12-6 LJ terms appears to be adequate.

Such fixed-charges force fields have achieved success in some subjects. For example, a Monte Carlo simulation of efavirenz(Sustiva) bound to HIV-1 reverse transcriptase(HIV-RT) calculated by MC/FEP using MCPRO provided confidence in the predicted structure for the efavirenz/HIV-RT complex(Figure 1)^{[22][24]}. Another successful case is the folding of U(1-17)T9D heptadecapeptide. The backbone RMSD between the simulated structure by MCPRO, OPLS-AA force field, and the NMR

structure is 2.5Å(Figure 2)^{[22][23]}.

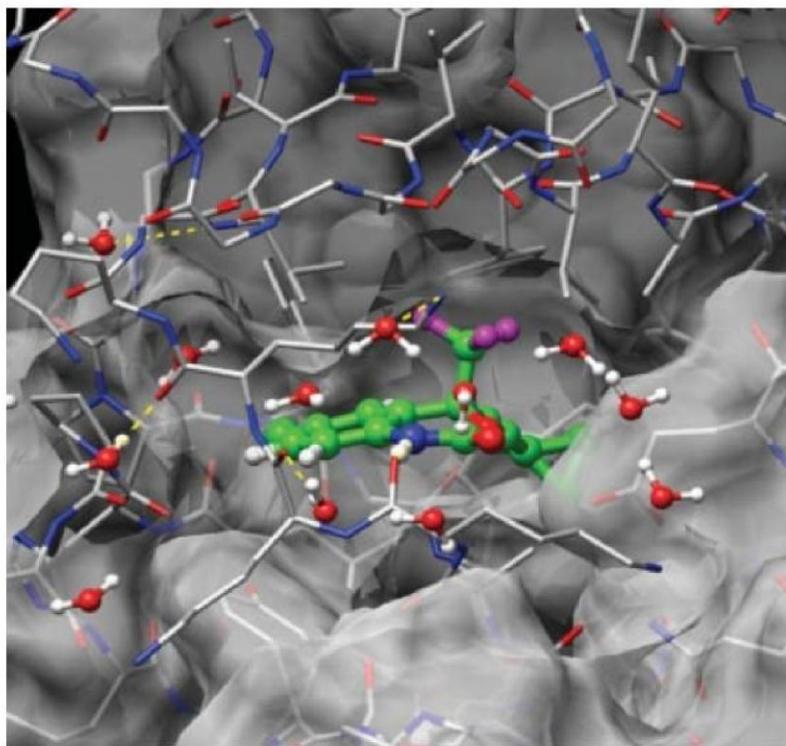


Figure 1. A configuration from a Monte Carlo simulation of efavirenz(Sustiva) bound to HIV-1 reverse transcriptase (HIV-RT).

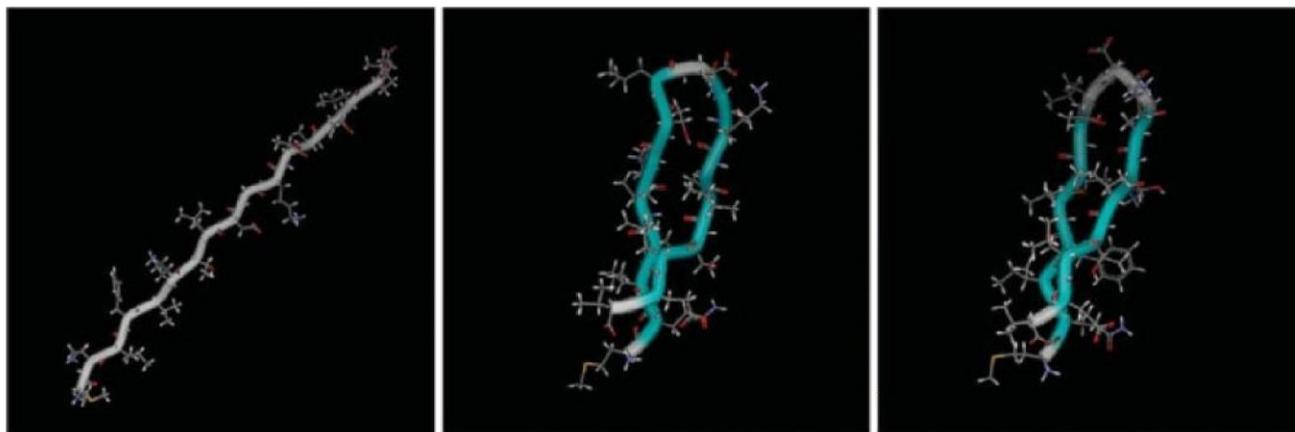


Figure 2. Folding of the U(1-17)T9D heptadecapeptide using MCPRO, the OPLS-AA force field, GB/SA hydration, and the CRA backbone sampling algorithm. Left: start of the MC run. Middle: after 24 million MC steps. Right: NMR structure 1e0q. The backbone RMSD is 2.5Å between the computed and NMR structures.

However, all the fixed-charges force fields described above have a very significant shortcoming. These models cannot describe redistribution of atomic partial charges. Therefore, they are not able to reproduce the response to changes in the electrostatic environment. Here is an example of showing that the absence of polarization can be fatal to the calculations. The inhibitor SCH66336 (4-{2-[4-(3,10-dibromo-8-chloro-6,11-dihydro-5H-benzo[5,6]cyclohepta[1,2-B]pyridin-11-yl)piperidin-1-yl]-2-oxoethyl} piperidine-1-carboxamide) is known to form a stable binding with the protein farnesyl transferase (Figure 3). Calculations of this complex's formation energy were carried out with the OPLS-AA and the PFF. The binding energy results from non-polarizable OPLS-AA and PFF are +55.84 kcal/mol and -28.04kcal/mol, respectively. The negative value of energy from PFF agrees with the existence of the stable protein-inhibitor complex. It can be concluded that explicit polarization is essential in

researching (at least in some cases) farnesyl transferase inhibition. This is why it has been widely acknowledged that including polarization into a force field calculation is very often desirable or even crucial. There are several ways to deal with this problem in terms of the functional form. Some examples include fluctuating charge, Drude oscillator and induced dipole models. The induced point dipole method is probably the most traditional one due to its intuitive physical basis.

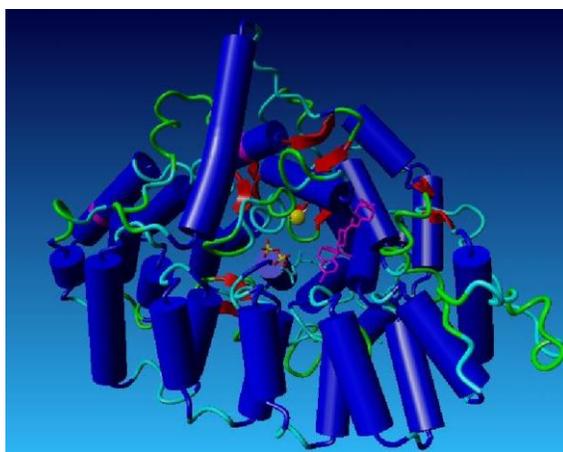


Figure 3. SCH66336 (magenta) in complex with farnesyl transferase, PDB 1O5M

The explicit inclusion of polarizability is showing next significant improvement in the non bonded interactions of force field^{[19][20][21]}. Briefly, a term that describes the energy determined by polarization of charge distribution is included into the potential energy function.

In this research thesis, we are presenting a polarizable force field based on induced point dipoles which is designed for organic and protein simulations. In order to avoid the usual problem of large CPU time and memory consumption in polarizable

calculations, we employ an approximation described in the background part.

References

1. Kollman, P. Chem. ReV. 1993, 93, 2395.
2. Gunsteren, W. F. v.; Berendsen, H. J. C. Angew. Chem., Int. Ed. Engl. 1990, 29, 992
3. He, X.; Mei, Y.; Xiang, Y.; Zhang, D.W.; Zhang, J.Z.H. —Quantum computational analysis for drug resistance of HIV-1 reverse transcriptase to nevirapine through point mutations , *Proteins*, 61, 423-432, **2005**.
4. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179
5. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
6. MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In Encyclopedia of Computational Chemistry; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F., III; Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998, p. 271, vol. 1.
7. Jorgensen, W. L.; Tirado-Rives, J. J Am Chem Soc 1988, 110, 1657.
8. Mackwell, A. D., Journal of Computational Chem, Vol 25, No.13. 2004.
9. Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J.-R.; Hwang, M.-J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P.; Thacher, T. S.; Yan, L.; Ni, X.; Hagler, A. T. J Comp Chem 2001, 22, 1782.
10. Lii, J.-L.; Allinger, N. L. J Comp Chem 1991, 12, 186.
11. Sun, H. J Phys Chem B 1998, 102, 7338.
12. Derreumaux, P.; Vergoten, G. J Chem Phys 1995, 102, 8586.
13. Halgren, T. A. J Comput Chem 1996, 17, 490.
14. Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S. Biopolymers 2003, 68, 383.
15. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. J Phys Chem 1990, 94, 8897. Rappe, A. K.; Colwell, C. J.; Goddard, W. A., III; Skiff, W. M. J Am Chem Soc 1992, 114, 10024.
16. Halgren, T. A. J Comput Chem 1996, 17, 520.
17. Halgren, T. A. J Am Chem Soc 1992, 114, 7827.
18. Halgren, T. A.; Damm, W. Curr Opin Struct Biol 2001, 11, 236.
19. Ponder, J. W.; Case, D. A. Adv Protein Chem 2003, 66, 27.
20. Kaminski, G.A.; Friesner, R.A.; Zhou, R.H. —A Computationally Inexpensive Modification of Point Dipole Electrostatic Polarization Model for Molecular Simulations , *J. Comput. Chem.*, 24, 267-276, 2003.
21. William L. Jorgensen, Julian Tirado-Rives. —Molecular Modeling of Organic and Biomolecular Systems Using *BOSS* and *MCPRO*” J Comput Chem 26: 1689– 1700, 2005
22. Ulmschneider, J. P.; Jorgensen, W. L. J Am Chem Soc 2004, 126, 1849.

24. Rizzo, R. C.; Wang, D.-P.; Tirado-Rives, J.; Jorgensen, W. L. *J Am Chem Soc* 2000, 122, 12898.

Chapter 2 Background

2.1 Introduction

This chapter includes the necessary theoretical background of PFF(polarizable force field). First of all, the basic theories of OPLS-AA force field were introduced, and followed by the transition from empirical force field to polarize force field. The last part of the background explains the parameterization procedure for producing parameters in POSSIM software package.

2.2 From Fixed-Charges Force Field to Polarizable Force Field

2.2.1 OPLS-AA fixed-charges force field

OPLS-AA fixed-charges force field developed by Prof. W.L.Jorgensen group^[1] is the basic foundation of the polarizable force field introduced in this research thesis. Thus, a brief introduction about OPLS-AA is necessary and helpful.

In OPLS-AA force field, the total energy of the system E_{tot} is calculated as following equation:

$$E_{tot}=E_{electrostatic}+E_{vdw}+E_{stretch}+E_{bind}+E_{torsion} \dots\dots\dots(1)$$

In this equation, $E_{\text{electrostatic}}$ is the electrostatic interaction, which includes the dipole-dipole, dipole-charge, charge-charge contributions. E_{vdw} is the nonelectrostatic part of the nonbonded inter- and intramolecular energy; E_{stretch} is the harmonic bond stretching energy and the E_{bend} is the angle bending energy; E_{torsion} is the Fourier expansion for the torsional energy(Figure 4).

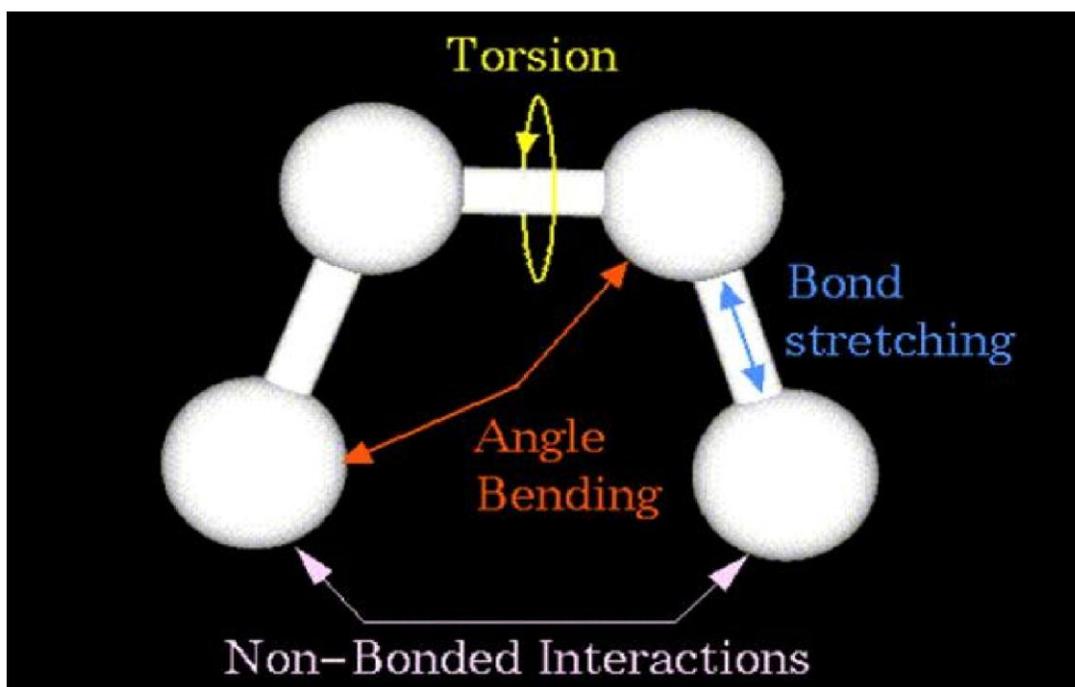


Figure 4. The energy contributions in molecular structure.

The non-bonded part is computed as a sum of both the Coulomb and Lennard-Jones contributions for intra- and intermolecular interactions:

$$E_{nb} = \sum_i^{onA} \sum_j^{onB} [q_i q_j e^2 / r_{ij} + 4\varepsilon_{ij} (\sigma_{ij}^{12} / r_{ij}^{12} - \sigma_{ij}^6 / r_{ij}^6)] f_{ij} \dots \dots \dots (2)$$

As shown in Equation(2), geometric combining rules for the Lennard-Jones coefficients are employed: $\sigma_{ij} = (\sigma_{ii} \sigma_{jj})^{1/2}$ and $\varepsilon_{ij} = (\varepsilon_{ii} \varepsilon_{jj})^{1/2}$. The summation runs over all the pairs of atoms $i < j$ on molecules A and B or A and A for the intra molecular interactions. Moreover, in the latter case, the coefficient f_{ij} is equal to 0.0 for any $i-j$ pairs connected by a valence bond (1-2 pairs) or a valence bond angle (1-3 pairs). $f_{ij} = 0.5$ for 1,4-interactions (atoms separated by exactly 3 bonds) and $f_{ij} = 1.0$ for all the other cases.

The bond stretching energy is computed as follows:

$$E_{bond} = \sum_{bonds} K_r (r - r_{eq})^2 \dots \dots \dots (3)$$

The angle bend is computed as follows:

$$E_{angle} = \sum_{angles} K_{\theta} (\theta - \theta_{eq})^2 \dots \dots \dots (4)$$

Here the subscripts eq are used to denote the equilibrium values of the bond length r and angle θ .

And the torsional energy is obtained as follows:

$$E_{torsion} = \sum_i \frac{V_1^i}{2}[1 + \cos(f_i)] + \frac{V_2^i}{2}[1 - \cos(2f_i)] + \frac{V_3^i}{2}[1 + \cos(3f_i)] + \frac{V_4^i}{2}[1 - \cos(4f_i)] \dots \dots \dots (5)$$

This is the summation of the all the dihedral angles i with values f_i .

2.2.2 Polarizable Force Field

Now let's consider the polarizable force field. When calculating the interaction between two particles in a two-body system, the equation $E(1-2) = E(1) + E(2) - E(1-2)$ is used (Figure 5-a). But in a many-body system, the way calculating the interaction energies will be different. For example, if there is a third particle (Figure 5-b), it will affect the interaction between particle 1 and 2 so that the interaction between 1 and 2 will be different with the previous simple isolated two-body system showed in Figure 5-a. This —different here is the so-called —polarizable. The existence of third particle makes particle 1 and 2 polarizable and the interaction between them is changed.

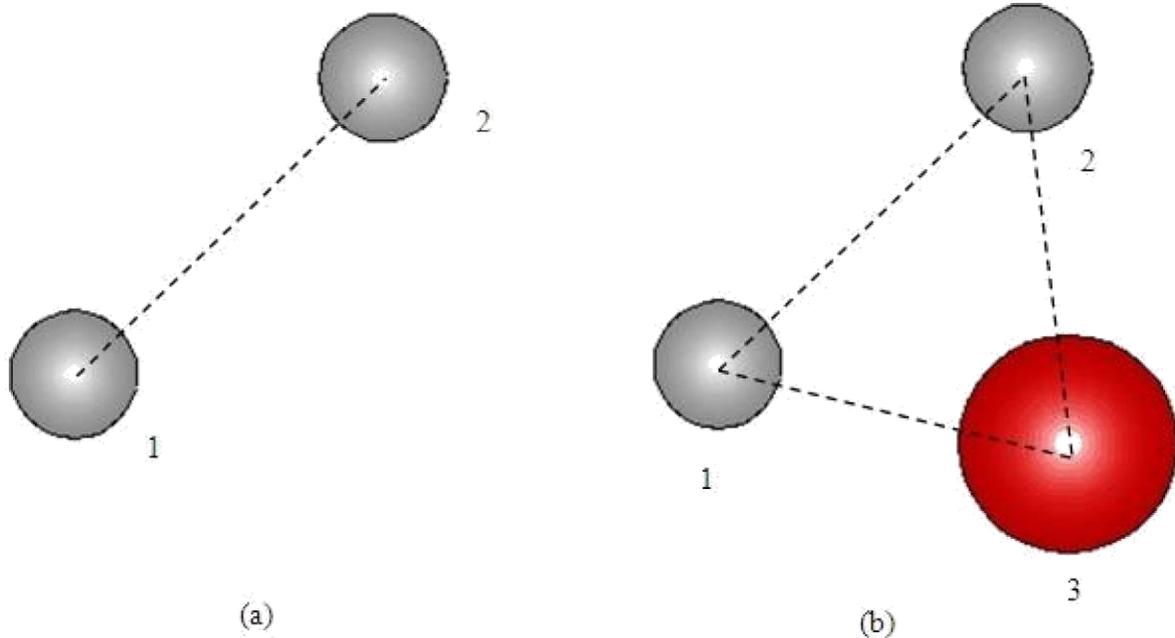


Figure 5. (a)The interaction Energy in two-body system is calculated by $E(1-2)=E(12)-E(1)-E(2)$.

(b)The existence of third particle makes particle1 and 2 polarizable.

For the polarizable force field, an additional term E_{pol} was added to $E_{tot}[2]$.

The electrostatic polarization is taken into account of the induced point dipole moment.

Following component is added to the E_{tot} :

$$E_{pol} = -\frac{1}{2} \sum_i \mu_i E_i^0 \dots\dots\dots(6)$$

In equation(6), μ_i represents the induced dipole moment on the i th polarizable site. E_i^0 is the electrostatic field produced by permanent charges only in the absence of the induced dipoles. The induced dipole moment depends on the total electrostatic field (produced by both the permanent charges and other dipoles) as shown in Equation (7).

$$\mu_i = \alpha_i E_i^{tot} \dots\dots\dots(7)$$

Here, α_i is the polarizability of the i th site. The total field E_i^{tot} is computed as follows:

$$E_i^{tot} = E_i^0 + \sum_{i \neq j} T_{ij} \mu_j \dots\dots\dots(8)$$

where

$$T_{ij} = \frac{1}{R_{ij}^3} \left(\frac{3R_{ij}R_j}{R_{ij}^2} - I \right) \dots\dots\dots(9)$$

is the dipole-dipole interaction tensor, and \mathbf{I} is the unit tensor. Thus,

$$\mu_i = \alpha_i E_i^0 + \alpha_i \sum_{j \neq 1} T_{ij} \mu_j \dots\dots\dots(10)$$

or with $A = \alpha^{-1}(\mathbf{I} - \alpha T)$,

$$A\mu = E^0 \dots\dots\dots(11)$$

And a system of linear equations is supposed to be solved in order to determine the values of induced dipole moments μ_i and substituted into the Equation (2).

To avoid any unphysical growth of the induced dipoles at close distances to

each other and to the permanent electrostatic charges, a cutoff procedure is involved, which is for the small interatomic distances R_{ij} [3]. In this program, the —perceived of—effective distance is modified if it is under a cutoff R_{cut} .

$$R_{ij}^{cut} = R_i^{cut} + R_j^{cut} \dots\dots\dots(12)$$

Where,

$$R_{ij} = (1 - x^2 + x^3)R_{ij}^{cut}, \text{ where, } x = R_{ij}/R_{ij}^{cut} \dots\dots\dots(13)$$

At this point, the effective distance w is used in polarization energy calculation can never be zero.

Using direct matrix inversion with elimination method, or iteratively, the system in Equation (11) can be solved. First, the left side of the Equation (10) is calculated by substituting an initial guess for μ into the right side, then the cycle is repeated until the desired level of self-consistency is obtained. The latter technique or the extended Lagrangian method is widely used in the computational application of molecular systems, because they are less demanding in terms of the computational resources[4].

However, even with those methods employed, the CPU-time is still huge when dealing with large size biomolecular systems. To avoid this problem, only a relatively small part of the large system is treated as polarizable atomic sites. The other atoms are treated in the other way, which is, no polarizability, as a dielectric continuum medium.

Moreover, most of the liquid-phase simulations with explicitly included atomic

polarizabilities are performed with molecular dynamics rather than the Monte Carlo technique due to the following two facts: First, in spite of its general computational simplicity, Monte Carlo with explicit polarization requires to solve Eq.(7) every time when even only one molecule is moved in the system. Second, the configurations number of an average Monte Carlo computation is greater than that of a molecular dynamics computation. As a consequence, for polarizable systems, utilization of Monte Carlo becomes unpractical.

In order to reduce the computational resources to apply the dipole polarization system, we proposed an approximation described as follows.

Equation (14) provides the iterative procedure that usually is employed in solving Equation (11).

$$\mu_i^I = \alpha_i E_i^0 \dots\dots\dots(14a)$$

$$\mu_i^{II} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \mu_j^I = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j E_j^0 \dots\dots\dots(14b)$$

$$\mu_i^{III} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \mu_j^{II} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j E_j^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j \sum_{k \neq j} T_{jk} \alpha_k E_k^0 \dots\dots\dots(14c)$$

Now let us start to consider the first- and the second-order approximations^[5] (Equations 14a and 14b, respectively). Equation (6) is still used to calculate the energy.

The first-order approximation is utilized by other researchers and found it has a good level of computational efficiency but the improvement of accuracy is limited,

comparing to pairwise-additive non-polarizable force fields^[6].

In this POSSIM software, we use a higher level of theory, the second-order approximation (Equation 14b), which not only retains a greater part of the dipole-dipole interactions than the first-order approximation but also reduces the computational resource comparing to the full-scale polarization model. In fact, the time needed to find the dipole moment vector (μ) will be equal to that for just one interaction in the full scale point dipole method if Equation (14b) is replaced by equation(7). Thus, the computational cost is dramatically decreased by this approximation.

Another issue that should be pointed out here is the usage of Monte Carlo simulations. As we mentioned above, the proposed approximation can significantly reduce the computing resource. In consequence, this allows the utilization of Monte Carlo simulations. The Metropolis sampling^[7] was involved when carrying out the MC runs. When generating the trial configurations, an attempt of either moving one molecules or changing the volume is made at each step and the following value is calculated:

$$C = [E(\text{new}) - E(\text{old})] / RT \quad \dots\dots\dots(15)$$

where E(old) is the energy of the last accepted configuration, E(new) is the energy of the new attempted configuration. R is the universal gas constant, and T is the temperature in K. When there is a new random attempt volume move, the following value is calculated:

$$C = [E(\text{new}) - E(\text{old})] / RT - N_{\text{mol}} \log[V(\text{new}) / V(\text{old})] \quad \dots\dots\dots(16)$$

Where N_{mol} is the total number of molecules. If the random number $\leq \exp(-C)$, the attempt is accepted; if the random number $> \exp(-C)$, reject the attempt moving.

2.3 Parameterization of the Polarizable Force Field.

From eq.(1), we know that the total energy is formed by several parts, $E_{\text{electrostatic}}$, E_{vdw} , E_{stretch} , E_{bind} and E_{torsion} . So the procedure of producing parameters of force field follows eq.(1) and includes the following stages, the parameters of different energies are fitted^[2]:

First, fitting the electrostatic polarizabilities.

Second, fitting the permanent electrostatic charges.

Third, fitting the Lennard-Jones parameters.

Fourth, fitting the torsional parameters.

Final step, Liquid simulations and the fine-tuning of the force field.

All the procedures were performed from POSSIM (POLarizable Simulations with Second Order Interaction Model).

2.3.1 Fitting the electrostatic polarizabilities.

The electrostatic polarizabilities fitting were finished in 3-body model. Two dipolar probes were applied to the target molecules as the electrostatic perturbation, consisting of two opposite charges of magnitude $0.78e$, 0.58 \AA apart (for a dipole moment of 2.17 D similar to that of nonpolarizable models for liquid water such as SPC/E11),

and located at the position that hydrogen bonds to the molecule were formed. The perturbations formed were the same as with the previous generation polarizable force field (PFF0)^{[8][9]}. Density functional theory (DFT) with the B3LYP method and cc-pVTZ(-f)^[11] basis set was used to calculate the change in the electrostatic potential at a set of grid points outside the van der Waals surface of the molecules and the energy of the perturbed system for each perturbation. All calculations were finished by Jaguar electrostatic code^[10]. The minimum of the deviations of the three-body energy produced in PFF was obtained by fitting the polarizability alpha in the three-body system. Figure 6 shows the two- and three-body energies of a molecule with dipolar probes. The 3-body energies are calculated by Equation(17):

$$E(3\text{body}) = E(1+2+3) - E(1+2) - E(1+3) - E(2+3) + E(1) + E(2) + E(3)\dots\dots(17)$$

In equation 17, the central molecule for which the parameters are being produced in PFF is marked as 1, while the two dipolar probes are denoted as 2 and 3. In non-polarized fixed-charges force field, the three-body energy equals to 0, because no many-body interactions are involved. From previous work it has been revealed that the three-body energy is independent of the permanent charges but depends upon the polarizability values^[5]. Therefore, in our force field the very first parameter was fitted is the polarizabilities of atoms in the molecule.

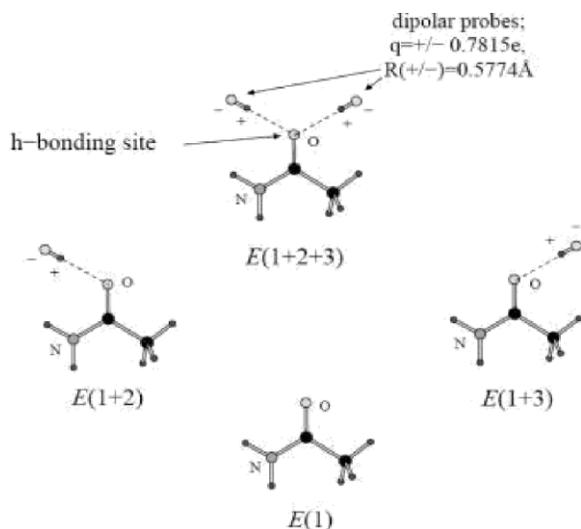


Figure 6. Calculating two- and three-body energies of a molecule with dipolar probes.

2.3.2 Fitting the permanent electrostatic charges.

We have used the same quantum mechanical systems as above, except that two-body energies were employed as the fitting target:

$$E(2\text{body}) = E(1 + 2) - E(1) - E(2) \dots \dots \dots (18)$$

The polarizability values obtained from previous step are constants in this step. In the two body system, the charges were adjusted to get the minimum deviations comparing to ab initio energies. The Rcut of all the cases are set to 0.8 Å, including dipoles and charges of the molecules.

2.3.3 Fitting the Lennard-Jones parameters.

Accurate binding energies are obtained from JAGUAR calculations of hydrogen-bonding interaction energies and distances. Here, the geometries were first optimized in LMP2/6-31** and cc-pVTZ(-f) level and then calculate the binding energies in cc-pVTZ(-f) and cc-pVQZ level.

In this step, we tried to reproduce the dimerization energies and the geometries as close as possible to the quantum mechanical calculations. The accurate quantum mechanical dimerization energies and the optimized geometries can be obtained from JAGUAR by using MP2 optimizations with a cc-pVTZ(-f) basis set. Then the geometries were utilized in our polarized force field to reproduce the dimerization energies and geometries^[12]. The binding energy calculations in JAGUAR consist of the LMP2 binding energy of cc-pVTZ(-f) basis set(Eccpvtz) and the LMP2 binding energy of cc-pVQZ(-g) basis set (Eccpvqz).

Following is the form of binding energy calculation^[13]:

$$E_{\text{bind}} = C1 \text{ Eccpvtz} + C2 \text{ Eccpvqz} \dots\dots\dots (19a)$$

$$C1 = a1/(a1 - a2); C2 = -a2/(a1-a2) \dots\dots\dots (19b)$$

$$a1 = \exp(-2.7); a2 = \exp(-1.8) \dots\dots\dots (19c)$$

In the calculation procedure, the Hartree-Fock energies are further corrected for basis

set superposition error by the counterpoise methods. E_{bind} is the target hydrogen-bonded energy, and the hydrogen-bond distances were obtained from the LMP2/cc-pVTZ(-f) optimizations.

2.3.4 Fitting the torsional parameters.

Finally, the torsional parameters defined in eq (5) were adjusted in single molecule models. The major dihedral angles were fixed at different values to obtain different torsional energies for the molecule. Torsional parameters V were adjusted to minimize the error of torsional energies comparing to QM. The relative quantum mechanical torsional energies were calculated from the LMP2/cc-pVTZ(-f) basis set.

2.3.5 Liquid simulations and fine-tuning of the force field.

The last step of the parameterization is the pure liquid simulations run in POSSIM, where all the parameters obtained in previous step were final tuned. Each liquid model was a cubic cell with periodic boundary conditions and formed by 216 molecules. All the simulations were performed in NPT ensembles (constant temperature, pressure and the number of the molecules), where the temperature was set at 25 °C, the pressure was 1 atm. The calculations were achieved by the Monte Carlo technique, and the heat of vaporization was calculated by the equation below:

$$H(\text{vap})= E(\text{gas}) - E(\text{liq}) + RT \quad \dots\dots\dots(20)$$

$$\text{av.}V= V/216$$

All of the calculations were finished by at least 1×10^6 Monte Carlo configurations of averaging were followed by more than 5×10^6 configurations of averaging for the thermodynamic properties.

In order not to slow down the computation, there are only 216 particles in one liquid model. As a result, some of the molecules are near the edge of the sample, that is near its surface. To avoid surface effects in our computations, the system size has to be extremely large to ensure that the surface has only a small influence on the bulk properties, but this system would be a huge cost of computing resources. Surface effects can be avoided for all system sizes if periodic boundary condition is applied^[14].

In periodic boundary conditions, the cubical simulation box is replicated throughout space to form an infinite lattice. In the simulation procedure, when a molecule moves in the central box, its periodic image in every one of the other boxes moves exactly the same way with same orientations. Thus, when a molecule leaves the central box, one of its images will enter through the opposite face. Overall, there are no walls at the boundary of the central box, and the system has no surface. The central box simply forms a convenient coordinate system for measuring locations of the **N** molecules.

Figure 7 shows the three-dimensional version of such a periodic model. As a particle moves through a boundary, all its corresponding images move across their corresponding boundaries. The number of particles in the central box is constant.

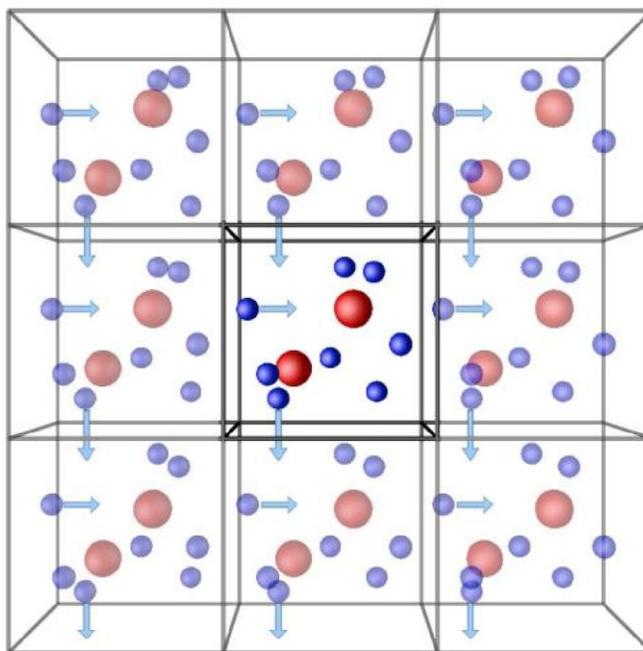


Figure 7. Periodic boundary conditions. The central box is outlined by a thicker line.

References

1. Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J. —Development and testing of the OPLS-AA all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc.*, *118*, 11225-11236, **1996**.
2. Kaminski, G.A., Ponomarev, S. Y., Liu, A. B. *J. Chem. Theory Comput.* **2009**, *5*, 2935-2943
3. (a) Prevost, M.; van Belle, D.; Lippens, G.; Wodak, S. *Mol. Phys.*, **1990**, *71*, 587-603; (b) Jorgensen, W.L. *BOSS, Version 3.6*; Yale University: New Haven, CT, 1995; (c) Stern, H.A.; Kaminski, G.A.; Banks, J.L.; Zhou, R.H.; Berne, B.J.; Friesner, R.A. *J. Phys. Chem. B*, **1999**, *103*, 4730-4737.
4. Halgren, T.A.; Damm, W. *Curr. Opin. Struc. Biol.*, **2001**, *11*, 236-242.
5. Kaminski, G.A.; Friesner, R.A.; Zhou, R.H. —A Computationally Inexpensive Modification of Point Dipole Electrostatic Polarization Model for Molecular Simulations, *J. Comput. Chem.*, *24*, 267-276, **2003**.
6. (a) Roux, B. *Chem. Phys. Lett.* **1993**, *212*, 231-240; (b) Straatsma, T.P.; McCammon, J.A. *Mol. Simul.*, *5*, 181-192, **1990**. (c) Straatsma, T.P.; McCammon, *Chem. Phys. Lett.*, *167*, 252-254, **1990**. (d) Straatsma, T.P.; McCammon, *Chem. Phys. Lett.*, *177*, 433-440, **1991**.
7. Owicki, J. C., *Computer modeling of matter*, chapter 14, 1978
8. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. J. *Comput. Chem.* 2002, *23*, 1515–1531.
9. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* 2004, *108*, 621–627.
10. (a) Jaguar, v3.5; Schro ¨dinger, Inc.: Portland, OR, 1998. (b) Jaguar, v4.2; Schro ¨dinger, Inc.: Portland, OR, 2000.
11. (a) Becke, A. D. *Phys. Rev. A* 1988, *38*, 3098–3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* 1988, *37*, 785–789.
12. Dunning, T. H. *J. Chem. Phys.* 1989, *90*, 1007–1023.
13. Kaminski, G. A.; Maple, J. R.; Murphy, R. B.; Braden, D; Friesner, R. A. *J. Chem. Theory Comput.* 2005, *1*, 248–254.
14. <http://polymer.bu.edu/Wasser/robert/work/node6.html>

Chapter 3 Results and Discussion

3.1 Introduction

In the present work, we explored the electrostatic, gas and liquid properties for several amino acids residues (Figure 8) in POSSIM software package. The polarizable force field parameters of the following molecules were obtained: methyl ammonium, acetate ion, imidazolium cation, benzene and pyrrole, which are the residues of lysine, glutamic acid, protonated histidine, phenylalanine and tryptophan, respectively.

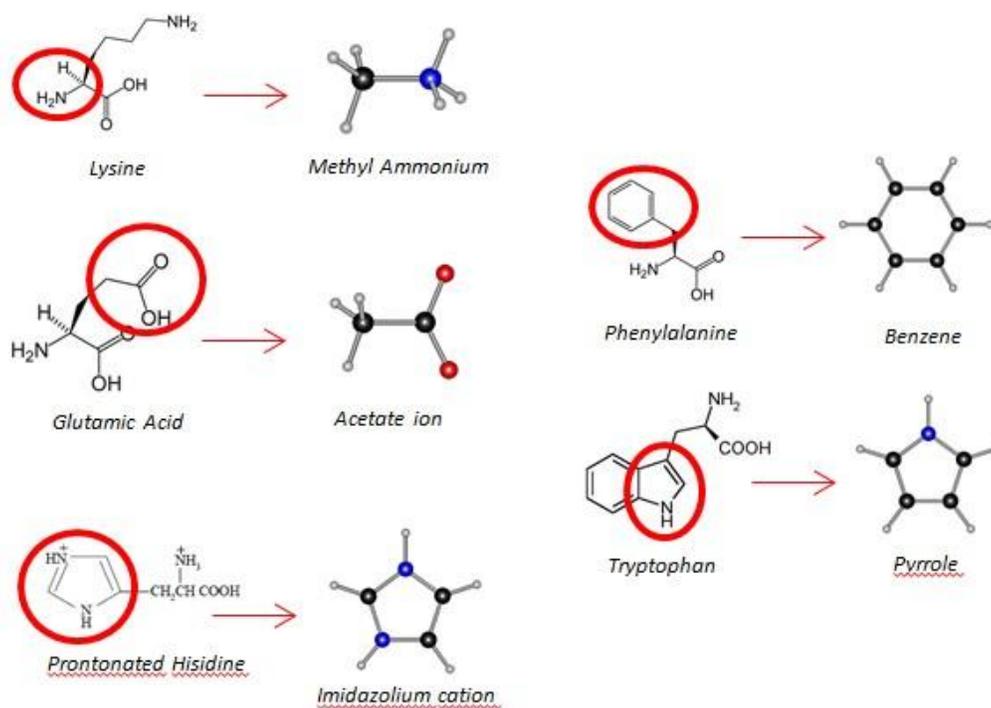


Figure 8. Chemical structures of the amino acids with the residues circled.

3.2 Electrostatic parameterization in a many-body system.

The first step of POSSIM force field parameters production is fitting the polarizabilities in three-body system in which the dipolar probes are involved. After the parameterization was done in the first three steps (electrostatic part fitting, gas-phase Lennard-Jones parameters fitting and torsional parameters fitting), fine tuning and some adjustments are needed in pure liquid simulations. Table 1 presents the three-body energies resulted from final parameters comparing to QM calculations, showing in rms deviations.

The polarizabilities of following atoms were fitted: carbon in benzene, nitrogens in pyrrole, nitrogens in imidazolium cation, -COO- group in acetate acid, -NH₃ group in methyl ammonium.

It is demonstrated that in all cases, a very small rms deviation no greater than 0.22 kcal/mol can be obtained by fitting the polarizabilities for particular atoms. This is concluding that POSSIM is adequate enough to reproduce an explicit many-body characteristic such as three-body energies.

Table 1. Three-Body Energy Deviations from Quantum Mechanical Data(kcal/mol)

molecules	error(kcal/mol)
C ₆ H ₆	0.22
C ₄ H ₅ N	0.18
C ₃ H ₅ N ₂ ⁺	0.12
CH ₃ COO ⁻	0.12
CH ₃ NH ₃ ⁺	0.27

The charges of some particular atoms which are mentioned above were adjusted in two body system in order to get closer two-body energies comparing to QM. We pointed out the atoms whose polarizabilities were fitted in many-body model in Figure 9.

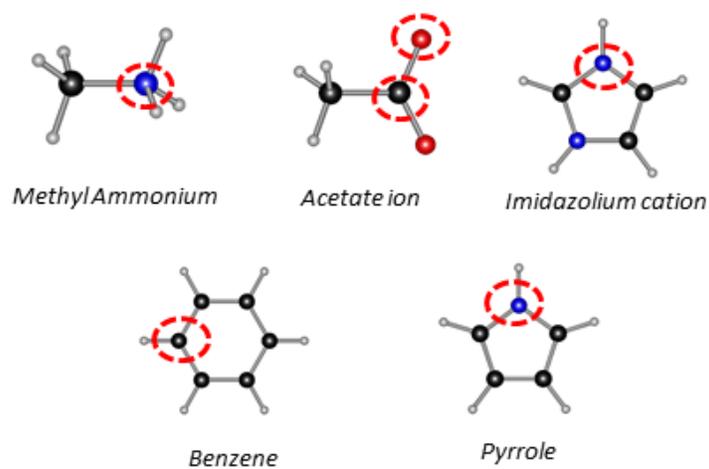


Figure 9. The polarizabilities of circled atoms were fitted in 3 -body system.

It should be noted that we did the parameterization of polarizabilities in three-body system before the two-body system. The reason for this procedure is that the charges of atoms actually do not affect the three-body while it impacts two-body energies, so the three-body fitting came out before the two-body fitting.

3.3 Gas-Phase Dimerization Energies and Distances.

The next step of developing the POSSIM PFF is fitting the Lennard-Jones parameters in dimer models to reproduce the gas-phase dimerization energies and geometries. This was done with high accuracy quantum mechanical results for hydrogen bonding interactions as objectives. Table 2 and 3 are showing the comparison between both the POSSIM results and quantum mechanical results, including dimerization energies and distances of different dimer models. The QM results were carried out by the calculation of LMP2/cc-PVTZ(-f) and LMP2/cc-pVQZ basis sets.

The Lennard-Jones parameters fitted in this step are the sigma and epsilon used in eq.2. The geometries of dimers were first optimized in LMP2/6-31G** level, then followed by LMP2/cc-PVTZ(-f) optimizations. In the case of benzene, there are two dimer geometries. In both of them, the O atom from the water molecule serves as the electron donor and the H atom of benzene serves as the electron acceptor. And the Lennard-Jones parameters of carbon atoms and H atoms are fitted in POSSIM. In imidazolium, there is one type of dimer, where oxygen in water is electron donor while the hydrogen of water molecule is acceptor, the other one is the hydrogen connected to nitrogen is electron acceptor and the oxygen in water is donor. The epsilon and sigma of two nitrogens and the hydrogen which is binding to one of the nitrogen were adjusted in this case. In pyrrole, only one dimer is built, where the hydrogen which is connected to

nitrogen points to the oxygen atom of water, the parameters of nitrogen were fitted in POSSIM. In CH₃COO⁻, two H atoms in water molecule are pointing to the two O atoms of CH₃COO⁻. And the H atom connecting to N in CH₃NH₃⁺ is the electron acceptor and the O in water is donator. All of the dimer models were shown in Figure 10.

The dimerization energies and distances between the molecules are calculated in POSSIM and shown in Table 2 and 3. The results from POSSIM are compared with those of quantum mechanical calculations. An average error no great than 0.5 Å of distances between the heavy atoms were obtained, while the error of dimerization energies were acceptable.

Table 2. Gas-Phase Dimerization Energies and distances(kcal/mol, Å)

Dimer models	Energy(QM) kcal/mol	Energy(POSSIM) kcal/mol	error	Distance(QM)Å	Distance(POSSIM) Å	error
C6H6-H2O-1	-2.97	-3.33	0.36	2.59	2.66	0.07
C6H6-H2O-2	-1.21	-0.96	0.25	3.47	3.85	0.38
C4H5N-H2O	-4.13	-3.73	0.40	1.93	1.59	0.34
C3H5N ⁺ ₂ ⁻ H2O	-18.14	-17.92	0.24	2.75	2.75	0.00
CH ₃ COO ⁻ H2O	-20.21	-20.17	0.04	2.83	2.92	0.09
CH ₃ NH ₃ ⁺ H2O	-19.63	-20.10	0.47	2.69	2.81	0.12

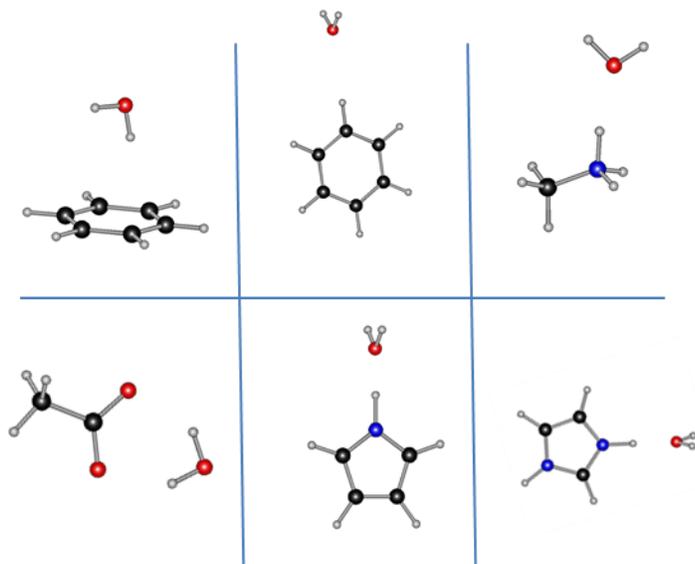


Figure 10. The Gas-phase dimers of benzene, pyrrole, methyl ammonium, acetate ion, imidazolium cation with water.

3.4 Torsional parameters fittings

In benzene, we fixed the dihedral angles of C-C-C-C at 0° , 15° and H-C-C-C at 150° , 180° . For imidazolium cation and pyrrole, the dihedral angle was H-N-C-C and the angle values are 150° and 180° . It is H-C-C-O fixed at 150° and 180° in CH_3COO^- and H-C-N-H at 120° and 150° in CH_3NH_3^+ . All of the dihedral angles are shown in Figure 11.

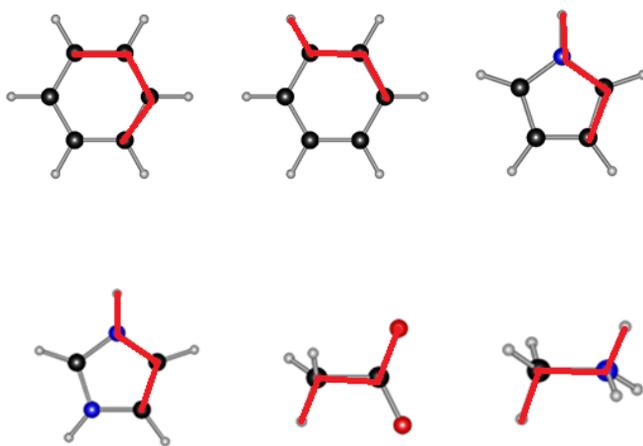


Figure 11. Torsional angles were shown in red lines.

Table 3. Torsional Energy(kcal/mol)

Dihedral angles	errors
C6H6, C-C-C-C	0.09
C6H6, H-C-C-C	0.0
C4H5N, H-N-C-C	0.42
C3H5N2 ⁺ , H-N-C-C	0.30
CH3COO ⁻ , H-C-C-O	0.14
CH3NH3 ⁺ , H-C-N-H	0.0006

We calculated the differences between two different dihedral angles and compared them to the QM running results. For example, in the benzene model, two dihedral types C-C-C-C and H-C-C-C were involved. First, C-C-C-C dihedral was fixed at 0° then fixed at 15°; the energy difference under the two situations were obtained, which are 1.74 kcal/mol in QM and 1.65 kcal/mol in POSSIM force field. From Table 4, we can see that the errors between QM and POSSIM are very small and in a reasonable range. It is revealed that the torsional fitting is done and we can move to the final step— pure liquid simulations.

3.5 Pure Liquid Simulations-The Fine Tuning of the Force Field.

In pure liquid simulations, the heat of vaporization and molecular density are calculated by equation 20. The results of liquid state heat of vaporization and average volume are shown in Tables 5 and 6. The POSSIM and QM results are very close to each other and successfully obtained average errors less than 2%. All the liquid models are 216 molecules cubic cell and at 25° C room temperature. As mentioned before, all of the calculations were finished by at least 1×10^6 Monte Carlo configurations of averaging were followed by more than 5×10^6 configurations of averaging for the thermodynamic properties. Figure 12 is showing one of the liquid models, which is formed by 216 benzene molecules.

Overall, the models presented here are all performing well in pure liquid simulations. So it is clearly revealed that the POSSIM program involving second-order approximation is an absolutely good one for reproducing high accuracy gas-phase and liquid properties.

Table 4. Liquid state Heats of Vaporization(in kcal/mol, at 25° C)

	$\Delta H_{\text{vap}}(\text{exp.})$	$\Delta H_{\text{vap}}(\text{POSSIM})$	$V(\text{exp.})$	$V(\text{POSSIM})$
C6H6 liquid	8.09	7.85	148.40	149.28
C4H5N liquid	10.43	10.80	115.75	113.63

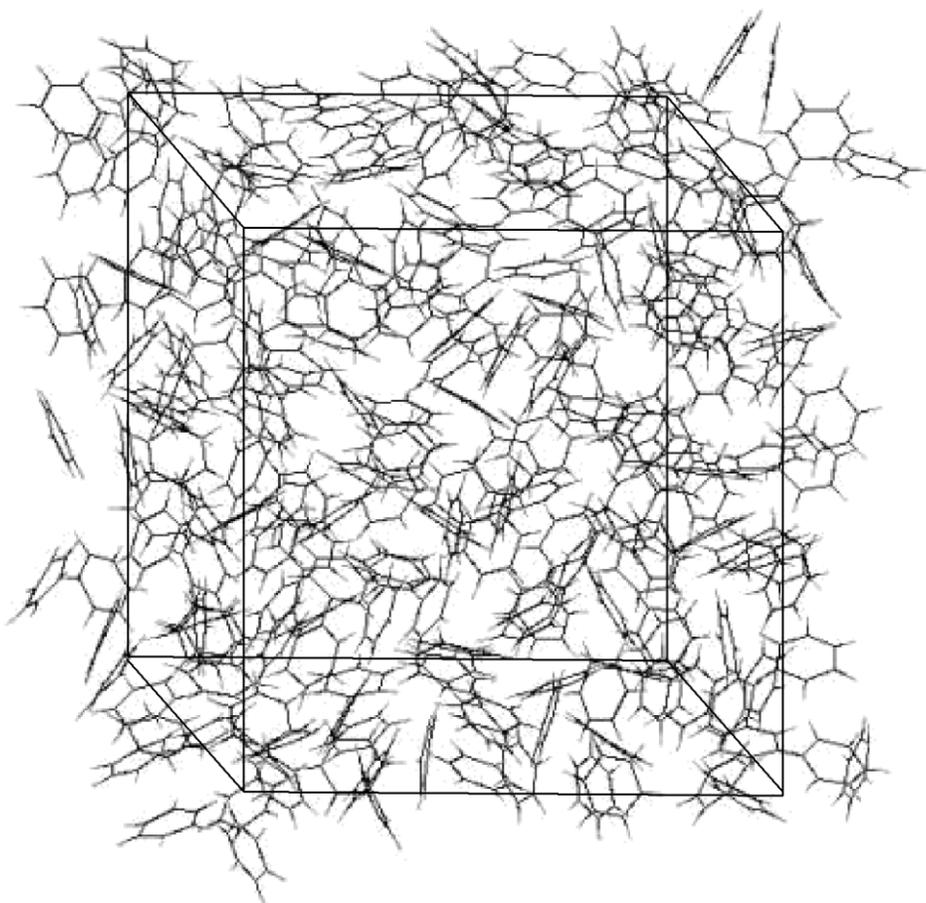


Figure 12. A liquid model formed by 216 benzene molecule

Chapter 4 Summary and Conclusion

A complete parameter fitting procedure was carried out by POSSIM software package, which takes electrostatic polarizability into account and allows a reduction of the computing time and resources. This proposed method was applied to several small molecules that are important in chemical reactions and protein simulations, benzene, imidazole and pyrrole.

The explicit treatment of electrostatic polarization in empirical force field provides a more efficient way to make polarizable calculations affordable and accurate. The fast polarization technique proposed by my research advisor and being developed in our lab has yielded a very good level of accuracy when compared to high-level quantum mechanical calculations.

Before each stage of parameter fitting except pure liquid simulation, all the systems were simulated and the geometries were optimized by JAGUAR quantum mechanical calculations. The results from JAGUAR software were set as target and comparison for POSSIM simulations. We obtained acceptable errors in gas-phase dimerization and torsional parameters fitting, and less than 0.22 kcal/mol error in the property of three-body system. Finally, an average error of no more than 2% in the heat of vaporization and volume in pure liquid simulations were received.

Moreover, the present method significantly reduces the total CPU time, so the liquid state Monte Carlo calculations were able to be carried out. Each liquid system was finished by at least 1×10^6 configurations for equilibrium and 5×10^6 for averaging as shown in Figure 13.

The results presented in this research report are not the end to our POSSIM development. We are looking forward to lowering the errors by adjusting the parameters. Several other molecules like water, methane, ethane, methanol, have been successfully tested in POSSIM force field and published. The ultimate goal of this parameter development project is to create a complete fast polarizable POSSIM force field for proteins and protein-ligand systems.

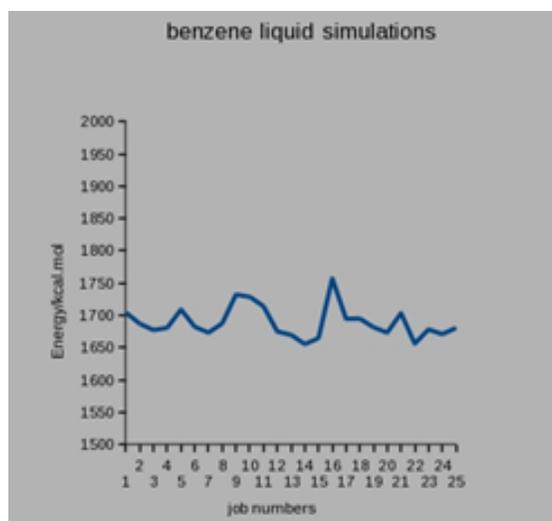


Figure 13. The last 25 jobs of liquid simulation of Benzene. The fluctuating blue lines represent the equilibrium of simulation is successfully done.

List of tables and figures

Figure 1. A configuration from a Monte Carlo simulation of efavirenz(Sustiva) bound to HIV-1 reverse transcriptase (HIV-RT).

Figure 2. Folding of the U(1-17)T9D heptadecapeptide using MCPRO, the OPLS-AA force field, GB/SA hydration, and the CRA backbone sampling algorithm.

Figure 3. SCH66336 (magenta) in complex with farnesyl transferase

Figure 4. The energy contributions in molecular structure

Figure 5. (a)The interaction Energy in two-body system is calculated by $E(1-2)=E(12)-E(1)-E(2)$. (b) The existence of third particle makes particle1 and 2 polarizable.

Figure 6. Calculating two- and three-body energies of a molecule with dipolar probes. **Figure 7.** Periodic boundary conditions.

Figure 8. Chemical structures of the four amino acids with the residues circled

Figure 9. The polarizabilities of circled atoms were fitted in 3-body system

Figure 10. The Gas-phase dimers.

Figure 11. Torsional angles were shown in red lines.

Figure 12. A liquid model formed by 216 benzene molecules.

Figure 13. The last 25 jobs of liquid simulation of Benzene and Imidzole.

Table 1.Three-Body Energy Deviations from Quantum Mechanical
Data(kcal/mol)

Table 2. Gas-Phase Dimerization Energies and distances

Table 3. Torsional Energy

Table 4. Liquid state Heats of Vaporization