Developing and validating Fuzzy-Border continuum solvation model with POlarizable Simulations Second order Interaction Model (POSSIM) force field for proteins

by

Ity Sharma

A Dissertation

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Chemistry

2015

ACKNOWLEDGEMENTS

The completion of this dissertation has been a long journey and could not have been possible without the help and support of my family and friends.

First and foremost I would like to express my deepest gratitude to my advisor **Professor George A. Kaminski** for providing me an opportunity to work in his research group and making my Ph.D. experience very productive. I would like to thank him for not only mentoring and encouraging me to pursue my goals but also allowing me to grow as an independent thinker.

Many thanks to the members of the Kaminski research group, both past and present, for their years of friendship, help and technical support; **Qina Sa, Dr. Xinbi Li, John Peter Cvitkovic, Dr. Sergei Y. Ponomarev, Dr. Timothy Click** and **Dr. Haijun Yang**. They immensely supported towards my intellectual development. I am also thankful to **Zhen Chen, Zijian Xia, Daniel Sigalovsky** and **Anetta Goldsher**, for all their support during their undergraduate research in Dr. Kaminski group. I am also thankful to **Siamak M Najafi** for making sure that my computer and software ran smoothly.

I would like to thank my PhD dissertation committee, **Prof James P. Dittami, Prof John C. MacDonald and Prof Erkan Tuzel** for their time, valuable suggestions and guidance. Special thanks is reserved for late **Prof Robert E Connors** for his enormous support, valuable suggestions and guidance during my first year seminar, second year qualifier and as my TA advisor.

I am especially grateful to **Prof Arne Gericke** for all his help as the head of the department and **Prof Kristin K. Wobbe** for her valuable suggestions during very critical time of my graduate studies.

I am also very thankful to **Prof Uma Kumar**, **Prof Glazer**, **Prof Wen-Chao Lai**, **Prof Drew Brodeur, Prof Destin Heilman and Prof Sal Triolo** for all they have taught me to become a better TA. Thanks also go out to **Mary**, **Rebecca** and **Paula** for their tremendous help during my TA duties. Special thanks to **Ann Mondor** for always being there no matter the task or circumstance.

I owe a special thanks to **Ms. Shobha Reddy, Dr. Varnitha Reddy** and **Dr. Sarva Lakshmi** for their friendship and support throughout my graduate studies at WPI.

I would like to acknowledge the **Chemistry and Biochemistry department** at **Worcester Polytechnic Institute** for the financial support and assistantship during my course of graduate studies.

I am also thankful to **WPI** for the financial support I recieved from the **Backlin Fund** for the Fall Semester 2015.

The PhD dissertation, which I completed at **Worcester Polytechnic Institute**, was indeed one long journey.
My time at the **University of Connecticut** was brief but forever memorable. The friendships made with **Dupinderjeet Kaur Mann, Kalpanie Bandara, Narendran Gummudipundi D and Sang-Yong Ju** in the beautiful campus of **UConn** will last forever. **UConn** offered me the opportunity to work with an outstanding professor, **Fotios Papadimitrakopoulos** who taught me the value of hard work.

I will always be very grateful to the **Department of Chemistry, Guru Nanak Dev University (GNDU)**, **Amritsar, India** from where this quest for knowledge began. I am thankful to my professors at **GNDU** for their encouragement and guidance. I am also grateful to my friends in **GNDU** for their contribution to my personal and professional development.

No acknowledgments would be complete without my family.

My sincere most gratitude goes out to my parents **Kiran Sharma** and **Janak Raj Sharma**, without whom none of my success would have been possible. They always believed in me, provided me the best education and made numerous sacrifices to make my life better. I lost my mother a month before joining the graduate program in WPI but I've always felt her around me especially when the days were tough and nights were long.

I would also like to thank my sister **Upasana Dhillon** and brother in law **Tanveer Dhillon,** for their love, support, and understanding. Similar thanks are reserved for my sister **Nidhi Sharma**

and her husband **Amit K. Sharma** for their love and understanding. I am also grateful to my younger brother **Abhishek Sharma** for his support and encouragement. **Upasana** and **Nidhi** are not only my sisters but my best friends and their unconditional love has encouraged and motivated me at every step during my graduate studies.

I am also grateful to my **mother in law (Kanta Sharma)** and **father in law (Jaswinder P. Sharma)** as well as my brother in law **(Amit K Sharma)** and his wife **Neerja** for their love, support and best wishes.

My final, and most heartfelt, acknowledgement goes to my husband and friend **Vaneet Kumar Sharma** who has supported me in whatever I intended to do and without whom I would have not made it this far. He is source of my courage to take on challenges that have come in the way. The understanding and love we have shared together have helped us in enduring and surviving the experience of graduate school while being parents. Our daughter, **Sanjana Sharma** is probably the person I owe my success to. She has always been my greatest fan and supporter, reassuring me whenever in doubt. No matter how things look, her smile has always driven me closer to my goals.

Hopefully this thesis would be a roadmap for Sanjana and my nephews Ishaan and Saman so that they understand the value of persistence and education.

TABLE OF CONTENTS

ABSTRACT

The accurate, fast and low cost computational tools are indispensable for studying the structure and dynamics of biological macromolecules in aqueous solution. The goal of this thesis is development and validation of continuum Fuzzy-Border (FB) solvation model to work with the Polarizable Simulations Second-order Interaction Model (POSSIM) force field for proteins developed by Professor G A Kaminski. The implicit FB model has advantages over the popularly used Poisson Boltzmann (PB) solvation model. The FB continuum model attenuates the noise and convergence issues commonly present in numerical treatments of the PB model by employing fixed position cubic grid to compute interactions. It also uses either second or first-order approximation for the solvent polarization which is similar to the second-order explicit polarization applied in POSSIM force field.

The FB model was first developed and parameterized with nonpolarizable OPLS-AA force field for small molecules which are not only important in themselves but also building blocks of proteins and peptide side chains. The hydration parameters are fitted to reproduce the experimental or quantum mechanical hydration energies of the molecules with the overall average unsigned error of ca. 0.076kcal/mol. It was further validated by computing the absolute pK_a values of 11 substituted phenols with the average unsigned error of 0.41pH units in comparison with the quantum mechanical error of 0.38pH units for this set of molecules. There was a good transferability of hydration parameters and the results were produced only with fitting of the specific atoms to the hydration energy and pK_a targets. This clearly demonstrates the numerical and physical basis of the model is good enough and with proper fitting can reproduce the acidity constants for other systems as well.

After the successful development of FB model with the fixed charges OPLS-AA force field, it was expanded to permit simulations with Polarizable Simulations Second-order Interaction Model (POSSIM) force field. The hydration parameters of the small molecules representing analogues of protein side chains were fitted to their solvation energies at 298.15K with an average error of ca.0.136kcal/mol. Second, the resulting parameters were used to reproduce the pK_a values of the reference systems and the carboxylic (Asp7, Glu10, Glu19, Asp27 and Glu43) and basic residues (Lys13, Lys29, Lys34, His52 and Lys55) of the turkey ovomucoid third domain (OMTKY3) protein. The overall average unsigned error in the pK_a values of the acid residues was found to be 0.37pH units and the basic residues was 0.38 pH units compared to 0.58pH units and 0.72 pH units calculated previously using polarizable force field (PFF) and Poisson Boltzmann formalism (PBF) continuum solvation model. These results are produced with fitting of specific atoms of the reference systems and carboxylic and basic residues of the OMTKY3 protein. Since FB model has produced improved pK_a shifts of carboxylic residues and basic protein residues in OMTKY3 protein compared to PBF/PFF, it suggests the methodology of first-order FB continuum solvation model works well in such calculations. In this study the importance of explicit treatment of the electrostatic polarization in calculating pK_a of both acid and basic protein residues is also emphasized. Moreover, the presented results demonstrate not only the consistently good degree of accuracy of protein pK_a calculations with the second-degree POSSIM approximation of the polarizable calculations and the first-order approximation used in the Fuzzy-Border model for the continuum solvation energy, but also a high degree of transferability of both the POSSIM and continuum solvent Fuzzy Border parameters. Therefore, the FB model of solvation combined with the POSSIM force field can be successfully applied to study the protein and protein-ligand systems in water.

ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
AMBER	Assisted Model Building and Energy Refinement
AMOEBA	Atomic Multipole Optimized Energetics for Biomolecular Simulation
AM1	Austin Model 1
CFF	Consistent Force Field
CHARMM	Chemistry at HARvard Macromolecular Mechanics
CM1	Compact Model 1
COSMO	Conductor–like Screening Model
CVFF	Consistent Valence Force Field
DAPY	Diarylpyrimidine
DATA	Diaryltriazine
DFT	Density Functional Theory
DO	Drude Oscillator
DPCM	Dielectric Polarizable Continuum Model
ECEPP	Empirical Conformational Energy Program for Peptides
EQ	Electronegativity Equalization
FQ	Fluctuating Charge
FTIs	Farnesyl Transferase Inhibitors
GAFF	Generalized Amber Force Field
GB	Generalized-Born
GROMOS	GROningen MOlecular Simulation
IEF	Integral Equation Formulation
IPD	Induced Point Dipole
LJ	Lennard-Jones
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics

MMFF	Merck Molecular Force Field
MO	Molecular orbital methods
MPa	MegaPascals
NEMO	Non-Empirical Molecular Orbital
NMR	Nuclear Magnetic Resonance
NNRTIs	Non-Nucleoside Reverse Transcriptase Inhibitors
NRTIs	Nucleoside Reverse Transcriptase Inhibitors
OPLS	Optimized Potentials for Liquid Simulations
PB	Poisson-Boltzmann
PCM	Polarizable Continuum Models
PFF	Polarizable Force Field
POSSIM	Polarizable Simulations Second-order Interaction Model
QM	Quantum Mechanics
RISM	Reference Interaction Site Model
Rnase Sa	Ribonuclease Sa
RT	Reverse Transcriptase
SASA	Solvent Accessible Surface Area
SDFF	Spectroscopically Determined Force Field
SIBFA	Sum of Interactions Between Fragments Ab initio computed
TIP4P-FQ	Four-site Transferable Intermolecular Potential Model-Fluctuating Charge
UFF	Universal Force Field
WHO	World Health Organization

Chapter 1

Introduction: Force Fields and Solvation Models in Biomolecular Simulations

1.1 Computational Chemistry

Computational chemistry has become increasingly significant in study of structure and function of biological macromolecules as well as organic molecules. It is a major tool in investigating areas such as folding and conformational changes of proteins[1], protein-protein interaction[2], structure-based drug design[3], computing binding free energy of ligands[4] and modeling enzyme mechanisms.[5]

Currently two main methods are used to evaluate energy in theoretical chemistry[6]:

- **Quantum Mechanics (QM)**
- **Molecular Mechanics (MM)**

Both these computational methods calculate potential energy, the difference being in their approach. Quantum mechanics (QM) calculates the potential energy based on the information of electronic structures and the results can be described as the solutions of the Schrödinger equation. In the case of Molecular mechanics (MM) electrons are not considered explicitly in the molecule although there are some exceptions. The atom and its electrons are treated as a single unit represented by potential energy functions or force fields.

Quantum mechanics (QM) calculations can be further divided in two categories[7]:

- **Ab initio**
- **Semi-empirical**

Ab initio methods use the Schrödinger equation with approximations to calculate total energy of the system. Such a calculation is based on quantum mechanics only and no experimental data is used. In the case of semiempirical methods the potential energy is calculated using experimental parameters as well as Schrödinger equation.

Quantum mechanics (QM) is generally regarded as the most accurate for potential energy calculations and has been the most popular approach to calculate energy.

However, there are limitations to its applications;

F**irstly**, QM calculations are computer intensive. They need large computational resources and longer time when dealing with larger systems and thus their area of applications is limited.

Secondly, QM calculations can produce different results when using different levels of theory and this deviation is evident for both small and large systems.

Thus in order to have results which are reasonably accurate, complete within a reasonable computational time and applicable to different molecular systems, molecular mechanics (MM) calculations are performed. Molecular mechanics calculates the potential energy using parameters derived from experimental data or *ab inito* calculation using *force fields.*

1.2 Force field

A Force field constitutes a set of analytical potential energy functions derived from classical mechanics. These potential energy functions are used to calculate the energy of a molecular system using parameters derived from the experimental or quantum mechanical techniques such

as ab initio, DFT (Density Functional Theory). **This combined set of potential energy functions and their parameters is known as a force field.**[8]

The general equation used to calculate energy in a force field consists of the bonded and the non-bonded interactions. The bonded energy interactions is a sum of intramolecular bonds, angles and torsion terms and the nonbonded consist of both intermolecular and intramolecular van der Waal and electrostatic Coulomb interaction terms. **(Equation (1a) (1b), Figure 1)**.

$$E_{total} = E_{bonded} + E_{nonbonded} \quad \textbf{(1a)}$$

$$E_{total} = E_{bond} + E_{angle} + E_{dihedral} + E_{van\,der\,Waals} + E_{Coulombic} \quad \textbf{(1b)}$$

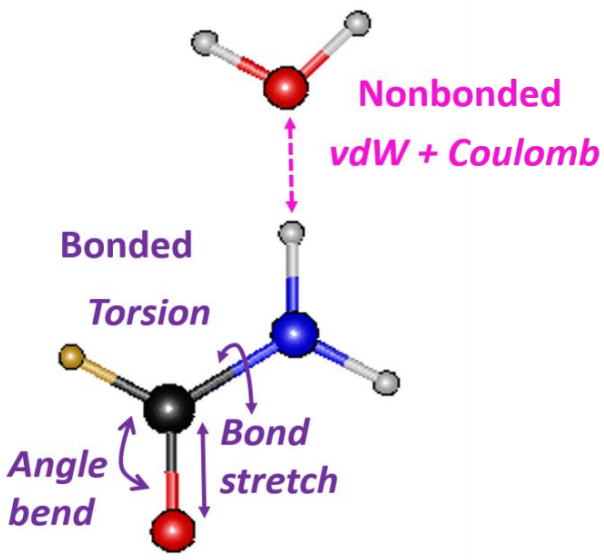


Figure 1: Bonded and non-bonded interactions in a molecule

Force fields can be generic or specific depending on their implementation. Force fields such as universal force field (UFF)[9] and generalized Amber force field (GAFF)[10] are of general applicability but recently developed force fields are more specialized to organic, inorganic or

biological molecules. They are more specifically designed for either organic molecules such as sugars or lipids or bio macromolecules such as proteins and nucleic acids.

1.2.1 *Types of Force field*

Force fields can be grouped under three major classes:

- **Class I force fields**
- **Class II force fields**
- **Class III force fields**

Class I force fields: Force fields such as **AMBER**[11,12], **CHARMM**[13], **OPLS**[14], **MMFF**[15], **GROMOS**[16], **ECEPP**[17] have been successfully applied to address many problems. The functional form of this type of force fields represents minimum forces to describe the molecular structure.

The total energy for bonded interactions consists of harmonic terms for bond stretching, angle bending and a Fourier series for each dihedral angle as shown in **equation (2)**.

$$E_{bonded} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedral} \frac{V_n}{2}[1 + \cos(n\varphi) - \gamma] \quad \textbf{(2)}$$

In **equation (2)**, r_{eq} and θ_{eq} represent the equilibrium values for the bond lengths and the angles between the atoms respectively. In the dihedral term, n represents the multiplicity and γ is the

dihedral phase angle. K_r, K_θ and V_n are the force constants for bonds, angles and the dihedral terms respectively.

The nonbonded energy term is given as sum of van der Waals and the Coulomb electrostatic interactions between the atoms separated by more than two bonds in both the intramolecular and intermolecular atom pairs (**equation (3)**). The van der Waals interactions are evaluated by Lennard-Jones (LJ) formalism. The LJ interaction energy term contains the short range repulsive and long range attractive term. For the repulsive term, the energy varies as a function of r^{-12} whereas for the attractive case it is proportional to r^{-6} as in London dispersion energy between the two atoms with polarizabilities a ($-a^2/r^6$).

$$E_{nonbonded} = \sum_{van\ der\ Waals}^{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] + \sum_{electrostatic}^{i<j} \frac{q_i q_j}{\varepsilon R_{ij}} \qquad (3)$$

The constant A and B in the first term in **equation (3)** are the van der Waals coefficients describing interactions for same atom types when the well depth (ε_i) and atomic radii (R_i) are known **(equation (4) and equation (5))**.

$$\mathbf{A} = \varepsilon_{ij} R_{ij}^{*12}, \mathbf{B} = 2\varepsilon_{ij} R_{ij}^{*6} \qquad (4)$$

$$\varepsilon_{ij} = \left(\varepsilon_i \varepsilon_j\right)^{\frac{1}{2}}, R_{ij}^{*} = R_i^{*} + R_j^{*} \qquad (5)$$

The electrostatic and van der Waals interactions between the atoms separated by three or more bonds or 1-4 nonbonded interactions are treated separately and their magnitude is often scaled

down. This formalism is typical but not the only possible one for class I force fields. Some force fields of this type have additional energy terms or small variations in the functional forms.

The **equations (2) to (5)** contain several constants or parameters that are produced to reproduce the experimental or the quantum mechanically obtained conformational energies and geometries, binding energies, vibrational frequencies, heats of formation, and other properties characterizing the condensed or gas phase.[18]

The electrostatic interactions in **Class I** force field use pairwise additive potential energy functions in terms of fixed charges, usually centered on atoms. This results in lack of accuracy in calculation of potential energy functions in some cases like treating molecules in environments of different dielectrics. The accuracy of many force fields have been increased by reparameterizing the current set of parameters or parameterizing the complete force field after adding new functional forms as in highly successful OPLS (Optimized Potentials for Liquid Simulations) force field where several improvements have been incorporated in the last thirty years resulting in faster and more accurate liquid simulations for large organic molecules and biomolecules such as proteins. These modifications have resulted in improved predictions of thermodynamic properties from the liquid state such as heat of vaporization, and free energy of hydration.

The limitations due to the use of fixed charges and pairwise additive approximation has led to the development of more improved force fields.

Class II force fields: Class II force fields have more complex functional form and include terms in addition to **equations (2)-(5)**. These are higher-order stretch bend valence terms to treat anharmonicity as well as cross terms between stretch and bend valence and/or bend and dihedral angles.[19] Some also include a Morse function that allows for bond breaking in empirical force fields and a cosine angle term for nonlinear angle.[11, 20]

The nonbonded electrostatic interactions between the point charges are represented by Coulomb formalism for most of the force fields. These point charges are mostly located on the nuclei except in case of MM3 force field where the electrostatic interactions are evaluated as point dipoles on the chemical bonds[21]. In standard force fields van der Waal interactions are proportional to the distance, R, between the two interacting points and varies as 12-6 (R^{-12} to R^{-6}) in standard Lennard-Jones interaction energy term. Other alternatives to this term include 9-6 term, buffered 14-7 used in MMFF force field or exponential Buckingham potential[22] that are more realistic and expensive to compute. The accuracy of class II force fields increase with addition of these terms and is particularly useful for reproducing conformational energies and equilibria, molecular structures and molecular vibrations. **Examples of class II force field are CFF,[22] CVFF,[23] MMFF,[24] MM3/MM4[21,25] and UFF.[11]**

Though Class II force fields have been validated over a number of decades and are found to be robust for treating structure and energetics as well in reproducing properties of biological systems, they typically do not perform well in many condensed phase simulations.[26] In these simulations the high dielectric medium such as water polarizes the charge distribution of the solute. The molecule of water itself in the gas phase carries a dipole moment of 1.85 Debye and

its polarization increases by about one Debye in bulk water. Therefore, the explicit treatment of electrostatic polarization interactions is critical for such simulations. Examples of modeling of protein folding where the section of amino acids form a hydrophobic core and must be transferred from its water environment to the interior of a protein with a different dielectric, folding of RNA in divalent ions media or folding of membrane proteins in a lipid environment all underline the need of explicit polarization effects.[27]

In both **Class I** and **Class II** force fields the polarization is included implicitly in the averaged manner. It is either included in Lennard-Jones interactions or by assignment of the fixed enhanced partial charges, q_i, to the atoms. These partial charges are produced through quantum mechanical methods, which overestimate the values of charge. These methods treating polarization in an effective manner limits the accuracy of nonpolarizable or **Class I** and **Class II** force fields as the polarization can vary significantly in a biomolecular system extending from the polar environment at the protein surface to the non-polar interior of the protein. The response of charge distribution to the changing dielectric environment can only be accounted for by incorporating explicit polarization effects.

In the Class I and Class II force fields, molecules do not respond to the changes in the electrostatic environment (temperature, pressure, pH, ion concentration and type of solvent) contrary to the real molecular systems which get perturbed due to the presence of a charged body, thus disturbing its geometry and energetics. Therefore, a major focus is in the development of the force field to treat electrostatic polarization of charge distribution by the environment of different dielectrics.

One of the first attempts to treat polarization was undertaken by Warshel and Lewitt in the study of lysozyme reaction in 1976.[28] Polarization has also shown to significantly affect intermolecular interactions in the gas-phase environment. Caldwell and Kollman[29] published development of a model to study aromatic-cation interactions by including polarization explicitly in additive force fields. Importance of polarization in molecular modeling was further presented by Rick, Stuart and Berne in their study of the hydration of the chloride ion in a small water droplet.[30] The chloride ion preferred to remain buried in the center of the droplet using the nonpolarizable OPLS/AA force field whereas with the polarizable water model TIP4P-FQ clearly showed preference of the ion to remain on the surface, hence depicting the entropic effect consistent with experimental evidence.

Thus published reports similar to presented above emphasized the need to include explicitly many-body induced polarization leading to the development of the class III polarizable force fields.

Class III force fields: Class III force fields are the most recent area of computational research that incorporate explicit polarizability term in the total energy, thus allowing the tuning of charge distribution to the changing dielectric environment. This polarizability or redistribution of charge in response to the changing electric field in molecular simulations is non-additive. The non-additivity arises from the different electron polarization of two atoms in the presence of one or more bonded atoms.

Examples of Class III force fields are either the ones which have included the polarization term since their inception such as AMOEBA,[31] SIBFA,[32] SDFF,[33] NEMO,[34] POSSIM[35] or are the counterparts of the existing standard class I force fields for example AMBER *ff02, ff02EP*,[36] CHARMM,[37] PIPF-CHARMM,[38] OPLS/PFF,[39] OPLS-AAP/OPLS-CM1AP,[40] and GROMOS.[41]

The specific examples illustrating the importance of explicit electrostatic polarizability are evident in the following examples. First is the calculation of accurate binding energies of E-selectin forming a complex with a calcium ion and Sialyl LewisX (Slx). It is known that the surfaces of cancer cells are found to be rich in these sugars. Selectins on the surface of the platelets bind to the Slx carrying cancer cells into the circulatory system. The X-Ray structure of E-selectin-Slx complex reveals a stable complex with two hydrogen bonds between one of the saccharide monomers in Slx, fucose and Ca^{++} ion, figure 2(a). The energy of formation of the E-selectin-Slx complex was found to be thermodynamically unstable +14.52kcal/mol with the OPLS-AA force field compared to -17.93kcal/mol calculated with the PFF or Polarizable Force Field.[39(b), 39(d)]

In another example the formation energy of stable complex between protein farnesyl transferase and inhibitor SCH66336 (4-{2-[4-(3,10-dibromo-8-chloro-6,11-dihydro-5H-benzo[5,6]cyclohepta[1,2-B]pyridin-11-yl)piperdin-1-yl]-2-oxoethyl} piperidine-1-carboxamide)[43] was studied, figure 2(b). The complex formation energy of this stable complex computed with OPLS-AA force field was +55.84kcal/mol while the polarizable force field

(PFF)[39(d)] calculated a value of -28.04kcal/mol in agreement with the stable protein-inhibitor complex.

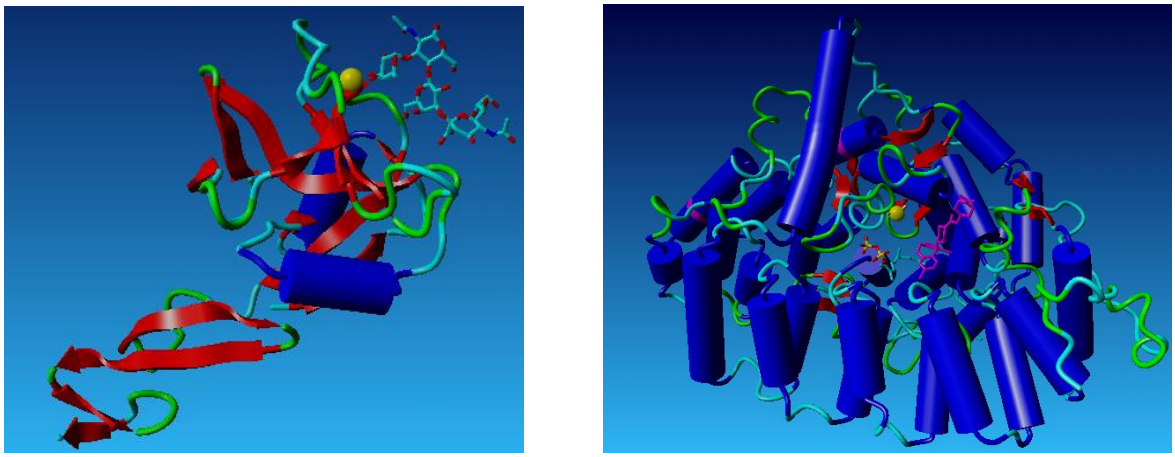
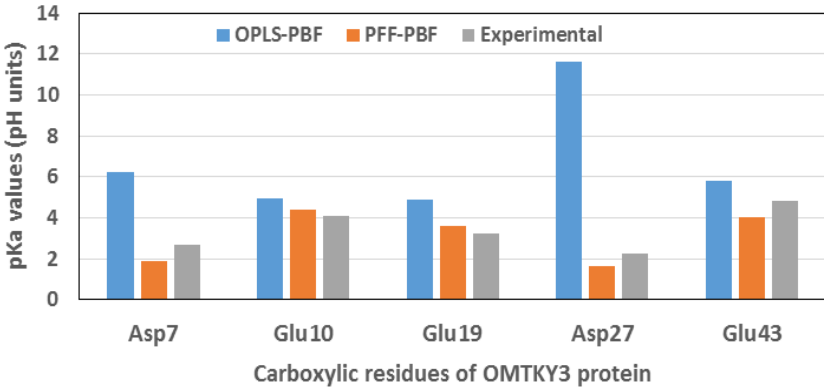
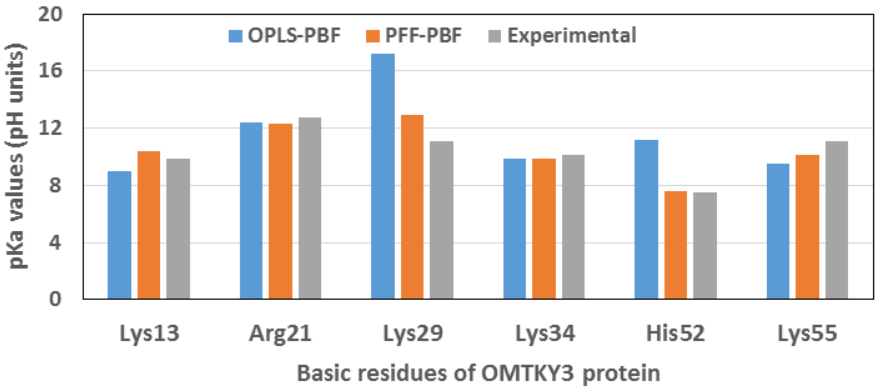


Figure 2: Complex of E-selectin and Ca^{++} with Sialyl Lewis X, PDB 1G1T (a) and SCH66336 (magenta) in complex with farnesyl transferase, PDB 1O5M (b).

It is imperative to include the electrostatic polarization explicitly in calculations such as pK_a and ion binding that involves strong electrostatic interactions as shown by our group. Figure 3 show the calculation of pK_a values of carboxylic [42(a)] and basic [42(b)] OMTKY3 protein residues with nonpolarizable OPLS-AA and polarizable PFF force fields. The accurate pK_a determination of ca. 0.6 and 0.7 pH units was achieved with a polarizable force field.



(a)



(b)

Figure 3: pK_a values of carboxylic (a) and basic (b) OMTKY3 residues[42]

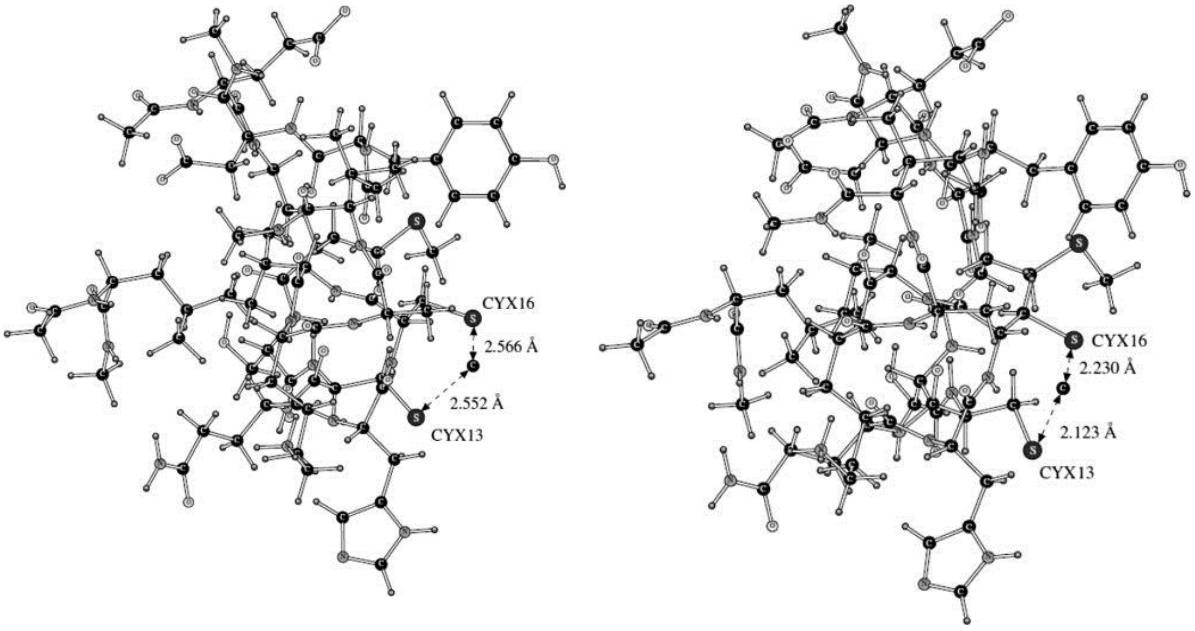
Our group has also shown increased accuracy in simulations of ion interactions with small molecules and proteins with a polarizable force field. The Cu(I) complexes with benzene show improved geometry and energy with a polarizable force field (PFF) than the fixed charge OPLS force field shown in the table 1.[43] Parameters for copper (I) were refitted with both OPLS and PFF force fields for copper-water gas phase complexes but was observed to work well for the copper complex with benzene. The error in the hydration energy of the copper ion with the PFF was also found to be only 1.8% compared to the OPLS error of 22%.

.

Table1: Energy and distances in complex of Cu(I) with benzene molecule [43]

Model systems	Energy, kcal/mol	Cu^{+}…C(benzene) distance, Å
OPLS	-14.0	2.77
OPLS, refitted for TIP3P	-25.2	2.14
OPLS, refitted for TIP4P	-26.0	2.11
PFF	-54.4	2.30
Reference [44]	-56.9 to -61.3	2.31

Ion binding calculations were also extended to a Cu^{+} complex with bacillus subtilis CopZ protein. The binding energy with non-polarizable OPLS an incorrect value of +9.98 kcal/mol while with the polarizable PFF force field was -33.05kcal/mol. The PFF force field also predicts correct Cu^{+}…S^{-} distances within the accuracy of 0.06Å compared to the ca. 0.4Å error with the fixed charge model from the experimental results, figure 4.[44]



(a) (b)

Figure 4: A fragment of Copz protein -copper (I) complex as simulated with OPLS (a) and PFF (b) force fields [44]

The above examples emphasize the importance of explicit electrostatic polarization interaction in studying many protein-ligand interactions. Polarization has proven to be significant particularly in computing acidity constants of small molecules and proteins, dimerization energies (aromatic cation interactions), binding energy (such as ion binding with small molecules and protein and sugar protein complexes) and in energetics and/or directionality of formation of hydrogen bonds.

Although the force fields with explicit polarization yield accurate results in many simulations, those force fields require 3 to 10 times greater computing time depending on the system than their additive analogs. This issue has been partially addressed by massive progress made in computer technologies and advancements in programming such as the particle-mesh Ewald (PME) [45] method for accurate and fast calculation of electrostatic energy. As mentioned above the challenge is to accurately evaluate many body interactions in a reasonably time efficient manner. Several methods have been proposed to incorporate electronic polarization in molecular simulations, including fluctuating charges (FQ) as well as Drude oscillator and induced dipole models.

1.2.2 Methods to include electronic polarization effects in force fields

Currently, the basic methods proposed to include the electronic polarization effects in force fields are fluctuating charge (FQ), Drude oscillator (DO) and induced point dipole (IPD) models.[46,48]

- **Fluctuating charge (FQ) model [46]**

This model is also known as the electronegativity equalization (EQ) model as it allow flow of charges between the atoms to equalize their instantaneous electronegativity. This approach involves assigning fictitious masses to the fluctuating charges (FQs) and treating them as additional degrees of freedom in the equations of the motion. The resulting equations of motions are solved more efficiently in molecular dynamics (MD) simulations than the Monte Carlo (MC) simulations. In the MD simulations, these equations are solved using the extended Lagrangian method[47] at the associated computational cost slightly higher than required for the fixed atomic charges of pairwise additive force fields. The parameters in the FQ model used to determine charge and response in polarization can be obtained empirically or fit to reproduce the two-body, three-body quantum chemical energies of water dimers and trimers.

The FQ model has been used to include polarization in the universal force field (UFF)[10], PFF[39], and CHARMM[13] force fields. One disadvantage of this model is the confinement of the polarizability in the molecular plane whereas experimentally it is found to be nearly isotropic.

- **Drude oscillator (DO) method [48]**

DO models also known as shell models are commonly used in the simulations of solid-state ionic materials and many other systems as well. The electronic polarization is incorporated in this model by representing an atom or ion as a two particle system. The two particles are the core and a shell linked with a harmonic spring and associated with certain fixed charges. This core and shell together is known as a Drude particle. The electronic polarization is linked to the response in the relative displacement of the charges to the external electric field. This approach to add

electrostatic polarizability has been incorporated in CHARMM[37] and GROMOS[41] molecular modeling packages.

- **Inducible point dipole model**

This is the most widely used method to treat molecular polarizability and its applicability varies from atomic to molecular systems such as noble gases to water to proteins. This approach has been used in many force fields such as OPLS/PFF[39], AMOEBA[31] and AMBER ff02, ff02EP[36]. Many new water models being developed employ this method to incorporate electronic polarization. According to this model, a point dipole, or PD, is induced at each contributing center in response to the total electric field, E. Hence the total energy, E_{total}, includes an additional energy term, E_{pol} **(Equation 6)**.

$$\boldsymbol{E_{total} = E_{bonded} + E_{nonbonded} + E_{pol}} \qquad \textbf{(6)}$$

The polarization resulting from the dipolar interactions between the permanent partial charges and the induced dipoles is incorporated in the E_{pol} energy term. The explicit polarization energy is then calculated using the formula given in **equation (7)**.

$$\boldsymbol{E_{pol} = -\frac{1}{2}\sum_i \mu_i E_i^0 = -\frac{1}{2}\sum_i \alpha_i E_i^{(0)} E_i} \qquad \textbf{(7)}$$

In **equation (7)**, α_i represents isotropic point polarizability of atom i. $\boldsymbol{E_i^0}$ denote the electrostatic field created on atom *i* in response to the partial charges. $\boldsymbol{E_i}$ is the electrostatic field due to the atomic charges (or higher order multipoles) and induced dipoles.

This total electric field is a result of both the permanent atomic charges as well as the induced dipoles and is determined self-consistently via an iterative procedure that minimizes the polarization energy or by means of the extended Lagrangian method.[47]

Polarizable Simulations Second-Order Interaction Model (POSSIM) Force Field

The Kaminski group has developed the Polarizable Simulations Second-Order Interaction Model (POSSIM)[35] force field using the inducible point dipole (IPD) method for protein simulations. This method is combined with the fast **second-order approximation** to decrease the computational time by about an order of magnitude without any loss of accuracy. It has also eliminated the problem of polarization catastrophe associated with the polarizable force fields. The second-order technique used for polarizable simulations forms the basis of the polarizable POSSIM force field. The parameters used in POSSIM force field also show good transferability, thus reducing the number of parameters fitted for biomolecular simulations. This also proves the correct physical basis of the model and permits it to predict the physical properties of a molecule in different environments. The polarizable POSSIM force field and software package is particularly targeted for use in biomolecular simulations.

Since most of the biomolecular processes are likely to occur in aqueous solution, theoretical study of such processes requires adequate representation of water. There are different methods to treat solvent and each has its own advantages and disadvantages. The choice of a particular solvation model in simulations depends on the requirements of the problem and size of the solute. Some solvation models have higher accuracy while others have high computational cost.

Thus, developing a computational model for water that is both reasonably accurate and fast is an ever evolving and ongoing quest.

Our group also is developing an implicit solvation model named as the Fuzzy-Border continuum solvation model[49] that is intended to work with both the OPLS and mainly Polarizable Simulations Second-Order Interaction Model (POSSIM) force fields for simulations targeted especially for proteins. The following sections give an overview of solvation models used in biomolecular simulations.

1.3 Solvation Models

There are different approaches for representing solvent in biomolecular simulations particularly in understanding structure and function of biomolecules with increased accuracy and efficiency. These solvation models range from very expensive and accurate representation of solvent to the less expensive continuous isotropic structureless medium representing averaged properties of water and other solvents.

Broadly, there are two main methods to study solvation at the molecular level - explicit and the implicit solvation models

- Explicit solvation model
- Implicit solvation model

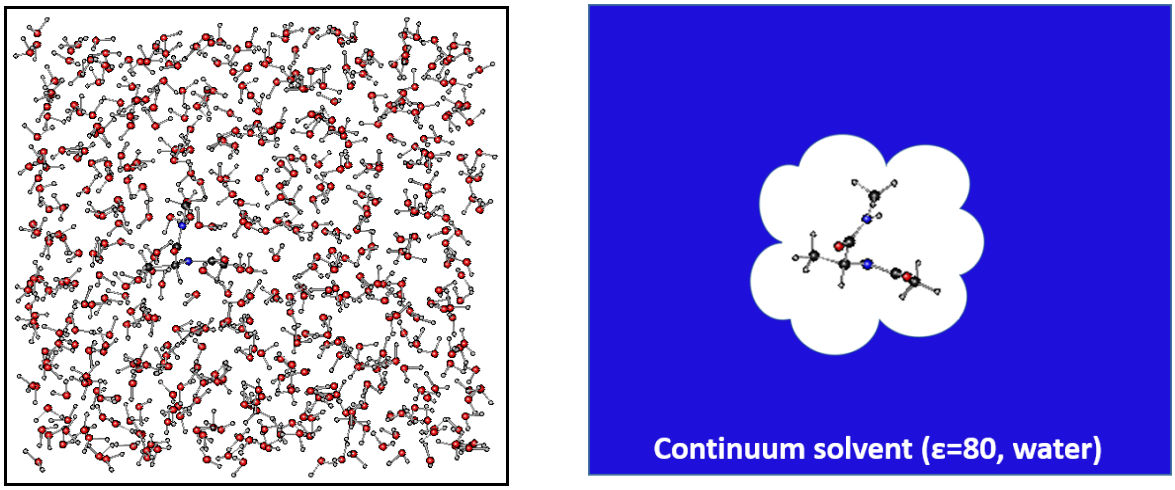
The **explicit solvation model** provides the most detailed and realistic approach to treat solution around the molecules by including all the degrees of the freedom of the solvent molecules.[50] The explicit solvent environment takes into account all interactions and is known to accurately simulate the interactions between the solutes, water, ions and formation of hydrogen bonds (**Figure 5a**).

However, such simulations increases the system size by an order of magnitude compared to the solute alone and are carried out at huge computational expense. Although there is significant advancement in the computational power, these calculations are still not feasible for many applications. It demands long simulation time in calculating water-water interactions as each water molecule is represented by at least three charges. The interactions between the solvent surrounding the solute also requires averaging several times in order to make the results with respect to solute structure and dynamics meaningful.

Due to these limitations of explicit solvation models, the **implicit solvation models** have become more popular (**Figure 5b**). The implicit solvation model represents solvent as a dielectric continuum with the solute-solvent interactions described in the spirit of a mean-field approach as a function of solute configuration.[51]

Recent years has seen much progress in the continuum or the implicit solvation models for biomolecular simulations owing to their fast nature and reasonably accuracy in comparison to the explicit solvation model. The implicit solvation model based on the experimental dielectric constant treat the electrostatic long-range forces accurately and thus is known to work better than

many explicit solvation models. Also, the continuum solvation models works well with polarizable solutes whereas many explicit solvation models that neglect the solute electronic polarization owing to its computational cost.



(a) (b)

Figure 5: (a) Explicit and (b) Implicit solvation models used in biomolecular simulations

Although implicit simulations offer fast treatment of complex systems, it is not suitable for modeling reactions in biomolecular systems. Such systems are simulated with the combined quantum mechanical/molecular mechanics (QM/MM) methods. In the QM/MM method, the system is divided in two parts. The solute and the nearby solvent molecules are treated quantum mechanically for high level description whereas the remaining solvent is modeled using a molecular mechanics force field. Hybrid QM/MM methods rely on use of efficient level of QM theory for solute interactions, MM force field or explicit water for the solvent and partial charges of solute for solute-solvent interactions. The simplified version of fully polarizable QM/MM was first used by Warshal and Lewitt in 1976.[28] Some other examples of the combined QM/MM approach are AM1/OPLS/CM1[52], AM1/TIP3P.[53]

The continuum models can be used within the quantum mechanics (QM) or molecular mechanics (MM) framework. Continuum models such as Polarizable Continuum Models (PCM) are used in QM to model solvent effects. These models use Poisson-Boltzmann (PB) model or Generalized-Born (GB) formalism to calculate the electrostatic potential of the system. Some of the examples of PCM models[54] are original dielectric PCM or D-PCM, the integral equation formulation (IEF-PCM), and conductor-like screening model (COSMO)[55] and SMx models.[56]

1.3.1 *Solvation free energy*

The solvation free energy is most important component of free energy calculations in biomolecules. The implicit solvation model takes into account average influence of solvent by directly computing the solvation free energy. The solvation free energy is defined as the change in free energy associated with the transfer of solute in a fixed configuration from vacuum to the solvent[57] shown in **figure 6**.

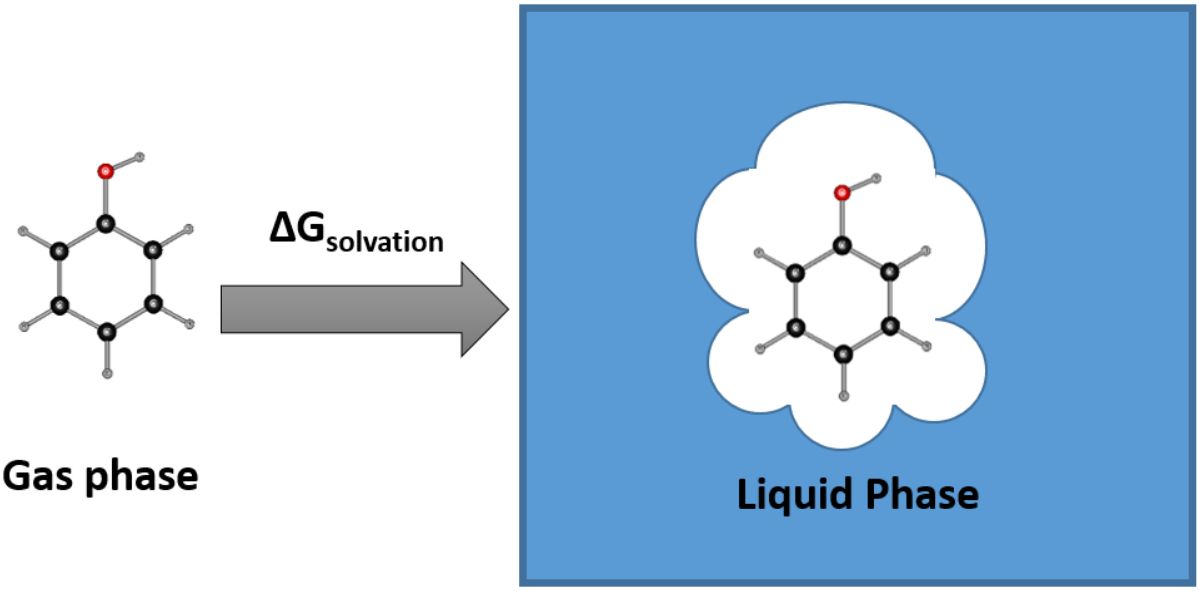


Figure 6: Schematic view of solvation free energy

The free energy of solvation broadly constitutes nonpolar and electrostatic forces between the solute and the solvent (**Equation 8**).[58]

$$\Delta G_{solvation} = \Delta G_{nonpolar} + \Delta G_{electrostatic} \quad \textbf{(8)}$$

The biological processes in water are mainly dominated by inter and intramolecular electrostatic interactions because of their long range nature and the fact that proteins and nucleic acids are charged molecules. Electrostatic interactions substantially affect the structure and dynamics of the biomolecules and are also crucial for stability of macromolecules and their interactions with ions, solvent and other molecules.

The nonpolar component of the total solvation energy arises from the energy penalty for creating a cavity against the solvent pressure, van der Waals interactions with the solvent and for the entropy associated with the reorganization of the solvent around the solute molecule. The nonpolar contribution to the solvation free energy is significant whereever hydrophobic interactions play a key role.[59] Examples of this can be seen in structure and function of proteins in water[60] and ligand binding to proteins.[61] Hydrophobic interactions also play a key role in hydration of hydrophobic molecular assemblies resulting in formation of micelles and phospho-lipid membranes and their mechanism of interaction with plasma and membrane bound proteins.[62]

The most accurate description of a solvent model requires calculation of all the interactions between the solute and the solvent and then averaging these over many solvent configurations.

The huge computational requirements for such calculations have been alleviated by faster theoretical methods such as implicit solvation models and huge advancements in computational power.

1.3.2 *Types of Implicit Solvation Models*

There are different types of implicit solvation models targeted for evaluating the solute-solvent interactions with varying speed and accuracy as shown in **figure 7**. The electrostatic contribution to the solvation free energy is computed using the approaches based on Poisson-Boltzmann (PB) equation, Generalized-Born (GB) formalisms and dielectric screening functions. The nonpolar component is usually modeled as proportional to solvent-accessible surface areas (SASA). The electrostatic PB and GB models are also combined with the nonpolar models such as solvent accessible surface area (SASA) for achieving accuracy in total solvation energy particularly in case of biomolecular simulations.

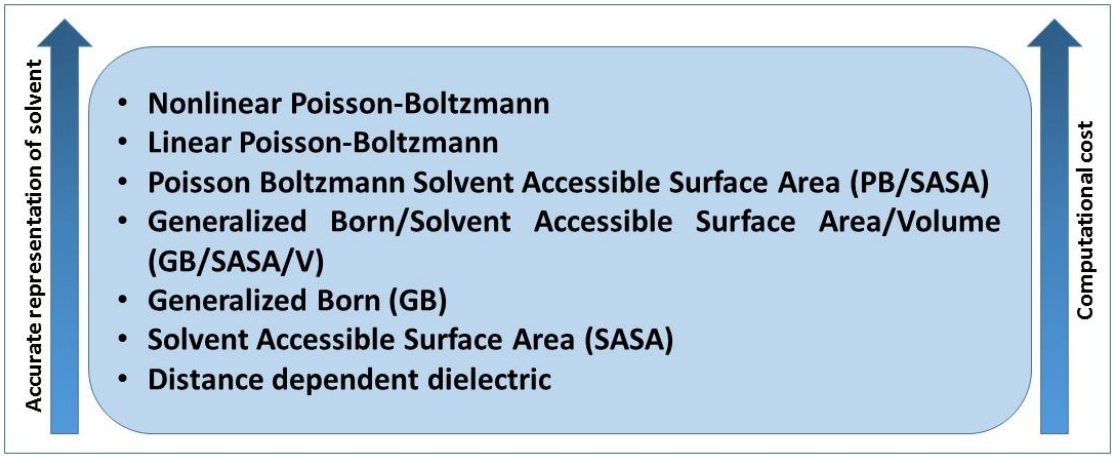


Figure 7: Implicit Solvation models

The basic laws and theories used to compute electrostatic interactions[63] in implicit models are

- **Coulombs law**
- **Poisson Boltzmann equation**
- **Born equation**

The non-electrostatic contribution to solvation free energy is usually modeled as a linear function of solvent-accessible surface area.

Coulomb Equation

The calculation of the electrostatic potential at every point in space in a given distribution of charges is the most difficult problem in classical electrostatic theory. The electrostatic potential $\phi(r)$ at a specific position in space for a point charge in a homogeneous medium such as vacuum can be evaluated using Coulomb's law.

$$\varphi(r) = \frac{1}{4\pi\varepsilon_0\varepsilon}\sum_i \frac{q_i}{r_i} \qquad \textbf{(9)}$$

In the **equation 9** φ is the electrostatic potential, q_i is the charge and r_i is the distance from the point charge i. ε_0 and ε designates the dielectric constant of vacuum and medium respectively.

This can be used to evaluate the total electrostatic energy of complex biomolecular systems like protein of N point charges immersed in the solvent

$$\Delta G_{electrostatic} = \frac{332kcal/mol}{\varepsilon_r}\sum_{i=1}^{N}\sum_{\substack{j=1,\\ i<j}}^{N}\frac{q_i q_j}{r_{ij}} \qquad \textbf{(10)}$$

In **equation (10)** $\Delta G_{electrostatic}$ represent the change in electrostatic interaction energy at room temperature in kcal/mol to the energy of charges placed at infinite separation. q_i and q_j are the point charges and r_{ij} is the distance in Å between the point charges. ε_r designates dielectric constant of the medium with respect to the vacuum. This equation is frequently used to calculate electrostatic forces in microscopic modeling of proteins.

Poisson-Boltzmann Equation (Poisson-Boltzmann Model)

Poisson Equation

In case of complex protein-solvent systems, the evaluation of the electrostatic energy of vast number of point charges can be time demanding process. The explicit representation and reorientation of all the point charges in these systems are approximated as dielectric constant in continuum solvation models.

Poisson equation relates the electrostatic potential φ to the total charge density, ρ (**Equation 11**).

$$\nabla \cdot \boldsymbol{\varepsilon}(\boldsymbol{r})\nabla\boldsymbol{\varphi}(\boldsymbol{r}) + 4\pi\boldsymbol{\rho}(\boldsymbol{r}) = 0 \qquad \textbf{(11a)}$$

ρ(r) or the charge density represent the distribution of charges in the system, *ε(r)* is the dielectric constant that includes effects such as induced dipole and/or relaxation of charges that are not explicitly modeled.

Poisson-Boltzmann Equation (11(a)) for a set of point charges placed in a cavity with dielectric constant, *ε(r)*, can be written as surface integral formulation including the induced polarization charge as shown in **equation 11(b)** [64]

$$\varphi(r) = \sum_k \frac{q_k}{|r - r_k|} + \int_s \frac{\sigma(R) d^2 R}{|r - R|} \qquad \textbf{(11b)}$$

In **equation 11(b)** q_k is the charge and r_k is coordinate of atom k. $\sigma(R)$ represents the induced polarization charge density on the dielectric boundary at point R, where R is the vector of integration over the surface of the molecule.

Boltzmann distribution of ions

The evaluation of charge density, $\rho(r)$ in the Poisson equation is a straightforward process if all the positions of the charges are known such as the C=O bonds in backbone of proteins and the dipoles on side chains that can reorient only in certain allowed geometries within the small conformational changes in the protein. But there are ions in solution such as Na^+, Cl^-, K^+, Mg^{2+} which constantly change their position under the influence of local electrostatic potential and the surrounding water solvent. The probability distribution function known as Boltzmann function is used to describe the positions of mobile ions in a solution:

$$n(r) = Ne^{-\frac{q\varphi(r)}{kT}} \qquad \textbf{(12)}$$

In **equation (12)** $n(r)$ is the concentration of the positive or negative ions in the solution. $\varphi(r)$ is the mean potential at a particular location r in the solution. N is the bulk concentration of the ions, k is the Boltzmann's constant (1.38×10^{-23}J/K) and q is the charge of ion considered. The charge density of the mobile ions can be calculated from the concentration of ions in the solution.

$$\rho_{ion}(r) = qn(r) \qquad \textbf{(13)}$$

These **equations (12) and (13)** account for all the mobile ions in the system and combined with the Poisson equation forms the **Poisson Boltzmann equation (PBE)** used for modeling the electrostatic interactions in the continuum solvation models (**Equation 14**).

$$\nabla \cdot \varepsilon(r)\nabla\varphi(r) - \varepsilon(r)K(r)^2 sinh(\varphi(r)) + \frac{4\pi\rho(r)}{kT} = 0 \qquad \textbf{(14)}$$

In the PBE **equation (14)**, *K* is the Debye-Huckel inverse length parameter dependent on the ionic strength, *I*, of the solution according to the **equation 15**

$$K^2 = \frac{8\pi N_A e^2 I}{1000kT} \qquad \textbf{(15)}$$

N_A, is the Avogadro's number and *e*, *k*, and *T* represent the electronic charge, Boltzmann constant and temperature respectively.

The ionic strength of the solution affects the electrostatic attractions/repulsions in the protein-solvent solutions and changing the ionic strength between the charges can result in the value of quantity being calculated.

Equation (14) is the non-linearized form of PBE equation and the linear form of Poisson-Boltzmann Equation can be written by assuming $sinh\varphi(r) \sim \varphi(r)$:

$$\nabla \cdot \boldsymbol{\varepsilon}(\mathbf{r})\nabla\boldsymbol{\varphi}(\mathbf{r}) - \boldsymbol{\varepsilon}(\mathbf{r})\mathbf{K}(r)^2\boldsymbol{\varphi}(\mathbf{r}) + \frac{4\pi\boldsymbol{\rho}(\mathbf{r})}{kT} = 0 \qquad \textbf{(16)}$$

This **equation (16)** combined with the nonpolar component that accounts for the van der Waals solute-solvent interactions and the entropy penalty for the cavity formation of solute together forms the total solvation energy.

Although PBE equation gives the most accurate treatment of electrostatic interactions, the high cost involved in solving this equation has limited its applications in many areas such as molecular dynamics (MD) simulations. There are methods suggested to overcome above limitations by not optimizing the forces due to the solvent at every simulation step or the solutions to Poisson equation for similar conformations in subsequent time steps.

Poisson Boltzmann Solvation Model

The Poisson-Boltzmann model based on Poisson Boltzmann Equation relates the electrostatic potential of a complex molecule to the charge density, ionic strength and the dielectric constants.[65] PBE is the most rigorous theoretical method for formulating and computing the electrostatic solute-solvent and solvent-solvent interactions of the total free energy of solvation.

PB solvation model involves explicit representation of solute in a cavity with atomic coordinates including the corresponding atomic radii and partial charges on each atom. The solute is placed in a cavity of low dielectric constant embedded in a continuum solvent of high dielectric constant. The solute and the solvent boundary are obtained by rolling the probe of the size of solvent over the van der Waals surface as shown in **figure 8**. The electrostatic potentials are then calculated with PB equation using iterative procedures for quick solutions.

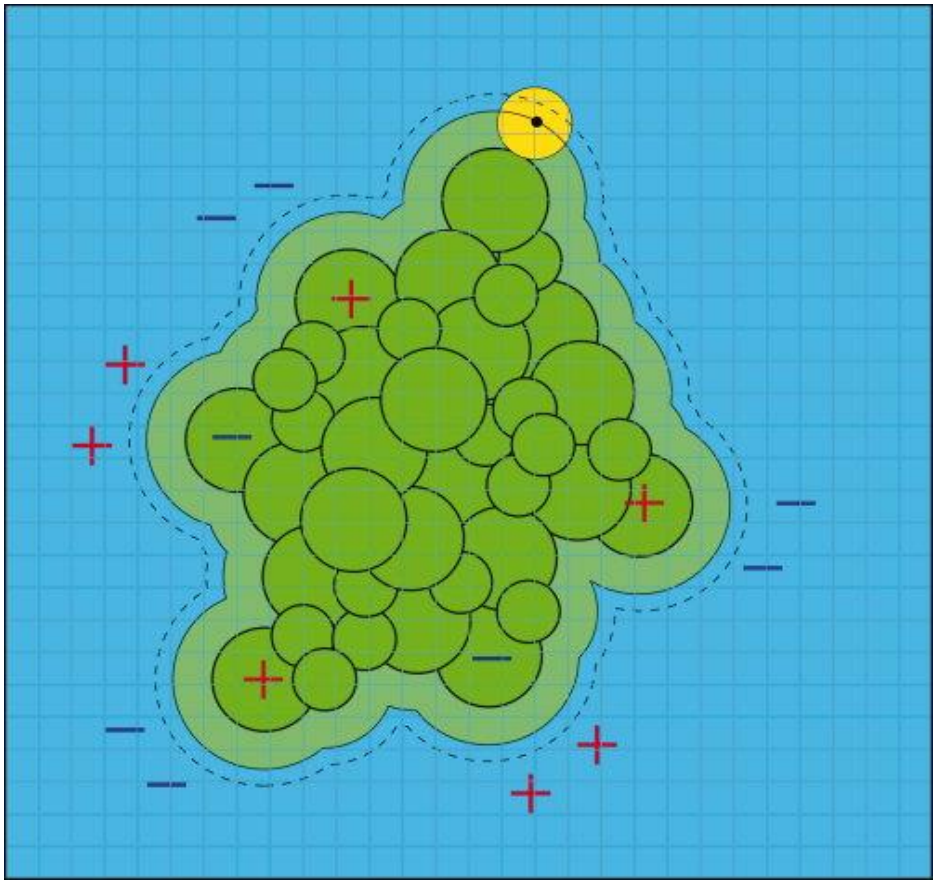


Figure 8: Schematic representation of the Poisson Boltzmann model of a molecule. The atoms in the molecule are represented by green spheres with partial charges and van der Waals radii. The high dielectric constant solvent is depicted in blue. The PB equation is solved on a three dimensional grid depicted in gray. The black line contour is obtained by rolling sphere with radius of water molecule shown in yellow on the van der Waals surface of the molecule. The boundary of ion-accessible volume is denoted by dashed line contour.[66]

The higher dielectric constant of the solvent in the PBE equation includes the induced, permanent dipoles and the orientation of the solvent around the solute. The dielectric constant of the solute, mainly in case of protein, has lower dielectric constant in the range of $\varepsilon \sim 2$-20. Its value varies depending on the type of protein and the simulation method used in PBE model. The lower dielectric constant value, $\varepsilon = 2$, is used if only electronic polarizability of the protein is considered where as higher value (~ 20) can be used to account for the polarizability and charges reorganization.[67, 68] The dielectric constant of solute particularly in case of protein is a

nontransferable parameter. It is required to calculate the electrostatic interactions such as charge-charge interactions or charge-solvation which depends on the shape and the exact location of the charges in the solute. The accuracy of the biomolecular applications require separate parameterization of macromolecule dielectric constant depending on the applications used.

PB model is a physically simple method to compute the electrostatic component but its numerical solutions for complex shapes and charge distributions are associated with high computational cost and do not scale well with increase in the size of the system. Typically, this differential integral equation is solved using finite-difference method (FDM) [69] in molecular mechanics simulations. In this method, molecular charges and dielectric are discretized on the grid and the Poisson-Boltzmann equation is solved and recast in a finite difference form. There are several problems associated with the discretization procedure such as the grid must be fine enough to represent accurately solute-solvent interactions and not merge opposite charges on the same node. Also, the free energy will depend on the grid spacing and the relative position of charges on the grid. Since the algorithms for solving Poisson-Boltzmann equation using finite difference methods is still computationally demanding many advancements such as multigrid methods are applied in biochemistry for faster simulations. Other methods that avoid the discretization problems are boundary element and finite element methods (FEM). The boundary element approach is less popular in molecular mechanics and is mostly applied in quantum calculations for small organic molecules.[70]

Although several methods have been devised to solve PB equation but it is still not feasible to solve it for molecular dynamics (MD) or Monte Carlo (MC) simulations where large

conformational sampling is required. The earliest attempts of wide applicability of PB equation to dynamic simulations demanded high computational effort and thus had limited scope. It was observed that simulation cost per-step with FD method even with 1-Å grid spacing was higher than with the explicit water simulation although the latter took longer time to equilibrate whereas with PB model, water is always equilibrated. This limits the practical applications of PB equation in MD simulations of biological molecules.

Though there is a continuous progress in numerical methods to solve PB equation but high computational effort and complexity in case of macromolecules has led to the development of approximations to the PB equation through methods such as Generalized Born (GB) model,[71] Dielectric Screening model,[72] Induced Multipole Solvent models[73] and others. These approximate models are widely used to treat solvation but none of them model the solvation effects as accurately as the PB equation especially in case of desolvation of charged groups occurring often in protein dynamics.

The approximations to Poisson-Boltzmann equation such as Born equation and its modifications are used for more faster simulations of complex systems. The Born equation is the simplest case of calculating electrostatic solvation free energy of a charged ion from gas phase to the solution.

Born Equation (Generalized Born (GB) model) [71]

Born Equation

Born equation illustrates the electrostatic free energy in transferring a spherical charged ion with radius α from a medium of dielectric constant ε_i to a medium of dielectric constant ε_o **(Equation 17).**

$$\Delta G_{Born} = -\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\frac{q^2}{2\alpha} \qquad \textbf{(17)}$$

Born equation was first derived by setting the dielectric constant $\varepsilon_i = 1$ as in the case of vacuum. Born equation can be used with Coulomb's law to calculate the free energy change in moving a point charge between the two homogeneous media (**Figure 9).**

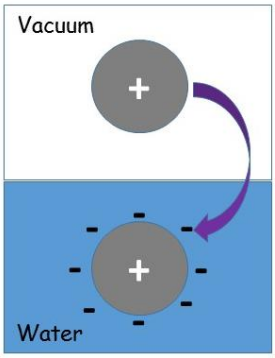


Figure 9: Schematic illustration of Born equation (ion). Spherical ion of radius α transferred from vacuum to water. The reaction field due to surface charges produced as a result of induced polarization in water stabilizes the ion.

Born equation is derived from classical electrostatics theory according to which the total electrostatic energy in the dielectric media is given by the **equation (18) and equation (19)**

$$G = \frac{1}{8\pi}\int E(r).D(r)d^3r \qquad \textbf{(18)}$$

$$\mathrm{D(r)} = \epsilon\, E(r) \qquad \textbf{(19)}$$

E and *D* in the **equation (18) and (19)** represent the electric field and electric field displacement respectively. ε is the dielectric constant of the medium.

Gauss law is used to obtain *E* and *D*.

$$\int_s \mathrm{D(r)} \,.\, \boldsymbol{n(r)} \boldsymbol{d^2 r} = \int_V 4\pi \boldsymbol{\rho} \boldsymbol{d^3 r}$$

or **(20)**

$$\int_s \boldsymbol{E(r)} \,.\, \boldsymbol{n(r)} \boldsymbol{d^2 r} = \int_V 4\pi \frac{\boldsymbol{\rho}}{\boldsymbol{\epsilon}} \boldsymbol{d^3 r}$$

The left integral in the **equation (20)** depict the area integral over the surface whereas the right integral is the volume integrated over the whole space enclosed by the surface. Here ***n(r)*** represent the normal of the surface and ρ is the free charge density.

The electric field and electric displacement inside and outside of the uniformly charged spherical shell with dielectric ε_i inside and outside can be given

$$\boldsymbol{E_{in}} = 0 \qquad \boldsymbol{D_{in}} = 0$$

$$\boldsymbol{E_{out}} = \frac{\boldsymbol{q}}{\boldsymbol{\epsilon_i r^3}} \boldsymbol{r} \qquad \boldsymbol{D_{out}} = \frac{\boldsymbol{q}}{\boldsymbol{r^3}} \boldsymbol{r} \qquad \textbf{(21)}$$

In the **equation (21)**, *q* represents the total charge of the sphere and the center of the coordinate is set at the center of the sphere. The total electrostatic energy of the system can now be calculated using the above equations as

$$G_1 = \frac{q^2}{2\epsilon_i \alpha} \qquad \textbf{(22)}$$

In the **equation (22)**, α represents the radius of the sphere. Similarly, the total electrostatic energy for a system of uniformly charged sphere with dielectric constant inside and outside as ε_i and ε_o respectively can be written as

$$G_2 = \frac{q^2}{2\epsilon_o \alpha} \qquad \textbf{(22)}$$

The energy difference between the two systems is evaluated as **Born equation**.

However if there is more than one charge, an approximation to the Born equation known as Generalized Born Equation (GBE) is used.

Generalized Born (GB) model) [71]

The pairwise GB model is based on the same dielectric continuum solvent model as PBE. Generalized Born model have been widely used to calculate the ligand binding free energies, in conformational analysis of proteins and in drug designing. It is one of the most efficient approximations of the solution of Poisson-Boltzmann equation for a charge in the centre of an ideal spherical solute of radius α and dielectric constant ε_i for the interior and the ε_0 for the exterior solvent. This model is extension of Born model which evaluates the change in free energy in moving a point charge from vacuum to spherical cavity of the solvent.

The generalization of Born model to solutes of different cavity shape and simulating the solutes as a collection of small spheres of atoms of charges q_i and radius α_i or point charges placed in the center of the spheres with the inner dielectric constant of the sphere as ε_i forms the GB formalism.

The electrostatic interactions between the point charges are calculated as a sum of Coulomb interactions in vacuum and the self-energies of the spheres. The self-energy can be decomposed into the total electrostatic energy of spheres placed in medium with dielectric constant ε_i, and the electrostatic solvation energy. In case of real solutes the Still and coworkers used pairwise sum over interacting point charges approximation to calculate solvent induced reaction field energy known as Generalized Born (GB) equation (**Equation 24-26**)

$$\Delta G_{GB} = -\frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-D}}} \qquad \textbf{(24)}$$

$$\alpha_{ij} = \sqrt{\alpha_i \alpha_j} \qquad \textbf{(25)}$$

$$D = \frac{r_{ij}^2}{\left(2\alpha_{ij}\right)^2} \qquad \textbf{(26)}$$

In **equation (27)** ε_i and ε_0 are dielectric constant of interior and exterior medium, r_{ij} is the distance between the atoms *i* and *j*, and α_i is the generalized Born radius of atom *i*.

The estimation of effective Born solvation radius, $\alpha_{i,}$ or the distance between charge and the protein-solvent boundary is central to accurate determination of electrostatic solvation free

energy. This is adjustable parameter and can be calculated using solvation free energy from Poisson-Boltzmann equation or less expensive alternating methods. Although PB solvation model is the most accurate representation of continuum solvation, GB methods provide potentials for faster simulations of larger systems. GB models are also combined with the surface area and referred to as GBSA models to estimate the hydrophobic contributions to the solvation free energy as well. These models are particularly useful in many ligand docking programs.

Implicit Solvation Models based on solvent-accessible surface area

The continuum solvent accessible surface area (SASA) solvation models are based on assumption that interactions between the solute and the solvent are proportional to the surface area. It computes the nonpolar contribution of total free energy of solvation. This model was first parameterized by Eisenberg and McLachlan[74] to compute free energy of transfer of amino acids between octanol and water. The solvent-accessible surface area as defined by Lee and Richards and others is the area moved by center of water molecule of radius 1.4Å around the group without any unobstructed contact with the group. The SASA based continuum model was also parameterized by Ooi[75] et al to compute thermodynamic solvation parameters for seven classes of groups occurring in peptides by fitting to the experimental free energy of solvation of small aliphatic and aromatic molecules.

The non-electrostatic contributions to the total free energy of solvation are usually given as a linear function of the solvent-accessible surface area according to the **equation (27)** [58(e), 61(d), 76]

$$\Delta G_{np} = \gamma A + b \qquad \textbf{(27)}$$

ΔG_{np} represents the nonpolar free energy and A is the solvent accessible surface area in the **equation 27**. The proportionality constant γ or the surface tension is the contribution to the solvation energy per unit surface area obtained by fitting to the experimental data. Another constant b represents the free energy of hydration for a point solute.

Although surface area models have worked well based on theoretical and experimental observations of transfer free energies of small chain alkanes from oil to water and vacuum to water in being related linearly to surface area, there are discrepancies in this model. Some of these include the wide range of surface tension proportionality constant corresponding to the definition of solute surface area[77] (van der Waal, molecular or solvent accessible surface area), parameterization of the model to the different experimental data as well as the application of the model to small organic molecule solvation and complex molecules and binding.

On the careful analysis of nonpolar contribution to the solvation energy in case of small and complex molecules such as proteins, this has been decomposed into energy penalty for cavity formation due to excluded volume effects and van der Waal dispersion forces between the solute and solvent **(Equation 28)**.[57]

$$\Delta G_{np} = \Delta G_{cavity} + \Delta G_{van\ der\ Waal} \qquad \textbf{(28)}$$

It has also been shown that ΔG_{vdW} or free energy change for establishing attractive interactions between the solute and solvent for a set of alkanes of similar size is a function of the solute composition and not its surface area. This explains not only the small hydration energy of cyclic

alkanes in comparison to the linear alkanes but also the requirement of two surface tension parameters of alkanes to reproduce their hydration free energies and conformational equilibria.

The applications of these models have been limited due to the high cost of calculating accurate solvent accessible surface areas. Some of these limitations have been circumvented by approximating solvent accessible surface areas using fast methods[78], or by extensive parameterization of atomic solvation parameters.[79]

Implicit Solvation Model Based on Dielectric Screening Functions

The simplest is the distance-dependent dielectric model where dielectric effect is accounted due to both the solute atoms and the surrounding water molecules. This model directly evaluates the electrostatic field due to the dipoles induced by polarizing the protein atoms and orienting the surrounding water molecules. The model was used to study the factors affecting the stabilization of carbonium ion in reaction of lysozyme. This is the most convenient and low cost solvation model; however, it is the crude representation of the solvation effects. Recent advancements in this approach is the EEF1[80] (effective energy function1) model that modulates the dielectric screening as a function of surface excluded volume and screening functions adjustment to the distance of charge site from the surface. The other advancements include the consideration of relative atomic position within the solvent and using dielectric screening functions without defining the specific solute solvent boundary.[81]

The main goal of this work is parameterizing and validating Fuzzy-Border (FB) continuum model of solvation to work with POlarizable Simulations Second-order Interaction Model

(POSSIM) force field for proteins. There are two distinct features of implicit FB model. Firstly, the formalism of FB model is an approximation to Poisson-Boltzmann methodology truncating the self-consistent equations and thus leading to better convergence. Second, the use of fixed three dimensional grid results in reduction of noise in the solvation energy calculations. Both the convergence and less noise in continuum model simulations are achieved without any loss in accuracy of the simulations.

The rest of the thesis is divided as follows. The **second chapter** introduces the underlying theory of first-order Fuzzy-Border (FB) continuum solvation model. It also presents the parameterization and acidity constant calculations of substituted alcohols with FB model and fixed charge OPLS-AA force field. The parameterization of FB model with polarizable POSSIM force field followed by computing pKa values of protein residues is discussed in chapter 3. The **fourth chapter** is the study of impact of high pressure on the backbone conformational equlibria of N-acetyl-L-alanine-N'-methylamide in aqueous solution with Polarizable Simulations Second-order Interaction Model (POSSIM) and fixed charge OPLS-AA force field. The **fifth chapter** discusses the future directions.

References

1. (a) Daggett, V., Protein folding-simulation. Chemical reviews 2006, 106 (5), 1898-1916. (b) Elcock, A. H., Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. 2006.

2. (a) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil; CR, Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. British journal of pharmacology 2008, 153 (S1), S7-S26. (b) McGuffee, S. R.; Elcock, A. H., Atomically detailed simulations of concentrated protein solutions: the effects of salt, pH, point mutations, and protein concentration in simulations of 1000-molecule systems. Journal of the American Chemical Society 2006, 128 (37), 12098-12110. (c) Ritchie, D. W., Recent progress and future directions in protein-protein docking. Current Protein and Peptide Science 2008, 9 (1), 1-15.

3. Taft, C. A.; Da Silva, V. B.; da Silva, C. H. T. d., Current topics in computer-aided drug design. Journal of pharmaceutical sciences 2008, 97 (3), 1089-1098.

4. Gilson, M. K.; Zhou, H.-X., Calculation of protein-ligand binding affinities*. Annu. Rev. Biophys. Biomol. Struct. 2007, 36, 21-42.

5. Beckstein, O.; Biggin, P. C.; Bond, P.; Bright, J. N.; Domene, C.; Grottesi, A.; Holyoake, J.; Sansom, M. S., Ion channel gating: insights via molecular simulations. FEBS letters 2003, 555 (1), 85-90.

6. Jensen, F., Introduction to computational chemistry. John Wiley & Sons: 2013.

7. Young, D., Computational chemistry: a practical guide for applying techniques to real world problems. John Wiley & Sons: 2004.

8. Jorgensen, W. L.; Tirado-Rives, J., Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proceedings of the National Academy of Sciences of the United States of America 2005, 102 (19), 6665-6670.

9. Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard Iii, W.; Skiff, W., UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. Journal of the American Chemical Society 1992, 114 (25), 10024-10035.

10. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. Journal of computational chemistry 2004, 25 (9), 1157-1174.

11. (a) 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., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society 1995, 117 (19), 5179-5197. (b) Wang, J.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry 2000, 21 (12), 1049-1074.

12. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. Journal of computational chemistry 2003, 24 (16), 1999-2012.

13. MacKerell, A. D.; Brooks, B.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M., CHARMM: the energy function and its parameterization. Encyclopedia of computational chemistry 1998.

14. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. Journal of the American Chemical Society 1996, 118 (45), 11225-11236.

15. Halgren, Thomas A. "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94." Journal of computational chemistry 17.5-6 (**1996**): 490-519.

16. Schuler, L. D.; Daura, X.; Van Gunsteren, W. F., An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. Journal of Computational Chemistry 2001, 22 (11), 1205-1218.

17. Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A., Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP. Macromolecules 1977, 10 (1), 1-9.

18. Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J., Polarization effects in molecular mechanical force fields. Journal of Physics: Condensed Matter 2009, 21 (33), 333102.

19. (a) Lii, J. H.; Allinger, N. L., The MM3 force field for amides, polypeptides and proteins. Journal of computational chemistry 1991, 12 (2), 186-199. (b) Ewig, C. S.; Berry, R.; Dinur, U.; Hill, J. R.; Hwang, M. J.; Li, H.; Liang, C.; Maple, J.; Peng, Z.; Stockfisch, T. P., Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds. Journal of computational chemistry 2001, 22 (15), 1782-1800. (c) Sun, H., COMPASS: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. The Journal of Physical Chemistry B 1998, 102 (38), 7338-7364. (d) Derreumaux, P.; Vergoten, G., A new spectroscopic molecular mechanics force field. Parameters for proteins. The Journal of chemical physics 1995, 102 (21), 8586-8605. (e) Halgren, T. A., Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. Journal of computational chemistry 1996, 17 (5-6), 490-519. (f) Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S., Potential energy functions: from consistent force fields to spectroscopically determined polarizable force fields. Biopolymers 2003, 68 (3), 383-394.

20. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., DREIDING: a generic force field for molecular simulations. Journal of Physical Chemistry 1990, 94 (26), 8897-8909.

21. Allinger, N. L.; Yuh, Y. H.; Lii, J. H., Molecular mechanics. The MM3 force field for hydrocarbons. 1. Journal of the American Chemical Society 1989, 111 (23), 8551-8566.

22. Buckingham, A.; Fowler, P., A model for the geometries of van der Waals complexes. Canadian journal of chemistry 1985, 63 (7), 2018-2025.

23. Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T., Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins: Structure, Function, and Bioinformatics 1988, 4 (1), 31-47.

24. Halgren, T. A., MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. Journal of Computational Chemistry 1999, 20 (7), 730-748.

25. Allinger, N. L.; Chen, K. H.; Lii, J. H.; Durkin, K. A., Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds. Journal of computational chemistry 2003, 24 (12), 1447-1472.

26. Halgren, T. A.; Damm, W., Polarizable force fields. Current opinion in structural biology 2001, 11 (2), 236-242.

27. Rick, S. W.; Stuart, S. J., Potentials and algorithms for incorporating polarizability in computer simulations. Reviews in computational chemistry 2002, 18, 89-146.

28. Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. Journal of molecular biology 1976, 103 (2), 227-249.

29. Caldwell, J. W.; Kollman, P. A., Cation-. pi. Interactions: Nonadditive Effects Are Critical in Their Accurate Representation. Journal of the American Chemical Society 1995, 117 (14), 4177-4178.

30. Stuart, S. J.; Berne, B., Effects of polarizability on the hydration of the chloride ion. The Journal of Physical Chemistry 1996, 100 (29), 11934-11943.

31. (a) Ren, P.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. The Journal of Physical Chemistry B 2003, 107 (24), 5933-5947. (b) Ren, P.; Ponder, J. W., Temperature and pressure dependence of the

AMOEBA water model. The Journal of Physical Chemistry B 2004, 108 (35), 13427-13437.

32. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P., Anisotropic, polarizable molecular mechanics studies of inter-and intramolecular interactions and ligand-macromolecule complexes. A bottom-up strategy. Journal of chemical theory and computation 2007, 3 (6), 1960-1986.

33. Palmo, K.; Mannfors, B.; Mirkin, N. G.; Krimm, S., Potential energy functions: from consistent force fields to spectroscopically determined polarizable force fields. Biopolymers 2003, 68 (3), 383-394.

34. Hermida-Ramón, J. M.; Brdarski, S.; Karlström, G.; Berg, U., Inter-and intramolecular potential for the N-formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. Journal of computational chemistry 2003, 24 (2), 161-176.

35. Kaminski, G. A.; Ponomarev, S. Y.; Liu, A. B., Polarizable Simulations with Second-Order Interaction Model□ Force Field and Software for Fast Polarizable Calculations: Parameters for Small Model Systems and Free Energy Calculations. Journal of chemical theory and computation 2009, 5 (11), 2935-2943.

36. (a) Cieplak, P.; Caldwell, J.; Kollman, P., Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. Journal of Computational Chemistry 2001, 22 (10), 1048-1057. (b) Wang, Z. X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y., Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. Journal of computational chemistry 2006, 27 (6), 781-790.

37. (a) Lamoureux, G.; Roux, B. t., Modeling induced polarization with classical drude oscillators: Theory and molecular dynamics simulation algorithm. The Journal of Chemical Physics 2003, 119 (6), 3025-3039. (b) Patel, S.; Brooks, C. L., CHARMM

fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. Journal of computational chemistry 2004, 25 (1), 1-16. (c) Patel, S.; Mackerell, A. D.; Brooks, C. L., CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. Journal of computational chemistry 2004, 25 (12), 1504-1514.

38. Xie, W.; Pu, J.; MacKerell, A. D.; Gao, J., Development of a polarizable intermolecular potential function (PIPF) for liquid amides and alkanes. Journal of chemical theory and computation 2007, 3 (6), 1878-1889.

39. (a) Friesner, R. A., Modeling polarization in proteins and protein–ligand complexes: Methods and preliminary results. Advances in protein chemistry 2005, 72, 79-104. (b) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A., Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. The Journal of Physical Chemistry A 2004, 108 (4), 621-627. (c) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A., Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. Journal of computational chemistry 2002, 23 (16), 1515-1531. (d) Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A., A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. Journal of Chemical Theory and Computation 2005, 1 (4), 694-715.

40. Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N., Polarization effects for hydrogen-bonded complexes of substituted phenols with water and chloride ion. Journal of chemical theory and computation 2007, 3 (6), 1987-1992.

41. Geerke, D. P.; van Gunsteren, W. F., On the calculation of atomic forces in classical simulation using the charge-on-spring method to explicitly treat electronic polarization. Journal of Chemical Theory and Computation 2007, 3 (6), 2128-2137.

42. (a) MacDermaid, C. M.; Kaminski, G. A., Electrostatic polarization is crucial for reproducing pKa shifts of carboxylic residues in turkey ovomucoid third domain. The

Journal of Physical Chemistry B 2007, 111 (30), 9036-9044. (b) Click, T. H.; Kaminski, G. A., Reproducing Basic p K a Values for Turkey Ovomucoid Third Domain Using a Polarizable Force Field. The Journal of Physical Chemistry B 2009, 113 (22), 7844-7850.

43. Ponomarev, S. Y.; Click, T. H.; Kaminski, G. A., Electrostatic polarization is crucial in reproducing Cu (I) interaction energies and hydration. The journal of physical chemistry B 2011, 115 (33), 10079-10085.

44. Click, T. H.; Ponomarev, S. Y.; Kaminski, G. A., Importance of electrostatic polarizability in calculating cysteine acidity constants and copper (I) binding energy of Bacillus subtilis CopZ. Journal of computational chemistry 2012, 33 (11), 1142-1151.

45. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. The Journal of chemical physics 1993, 98 (12), 10089-10092.

46. Rick, S. W.; Stuart, S. J.; Berne, B. J., Dynamical fluctuating charge force fields: Application to liquid water. The Journal of chemical physics 1994, 101 (7), 6141-6156.

47. Van Belle, D.; Froeyen, M.; Lippens, G.; Wodak, S. J., Molecular dynamics simulation of polarizable water by an extended Lagrangian method. Molecular physics 1992, 77 (2), 239-255.

48. Rick, S. W.; Stuart, S. J., Potentials and algorithms for incorporating polarizability in computer simulations. Reviews in computational chemistry 2002, 18, 89-146.

49. Sharma, I.; Kaminski, G. A., Calculating pKa values for substituted phenols and hydration energies for other compounds with the first-order fuzzy-border continuum solvation model. Journal of computational chemistry 2012, 33 (30), 2388-2399.

50. Levy, R. M.; Gallicchio, E., Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. Annual review of physical chemistry 1998, 49 (1), 531-567.

51. Roux, B.; Simonson, T., Implicit solvent models. Biophysical chemistry 1999, 78 (1), 1-20.

52. Kaminski, G. A.; Jorgensen, W. L., A quantum mechanical and molecular mechanical method based on CM1A charges: applications to solvent effects on organic equilibria and reactions. The Journal of Physical Chemistry B 1998, 102 (10), 1787-1796.

53. Gao, J.; Xia, X., A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. Science 1992, 258 (5082), 631-635.

54. Barone, V.; Cossi, M.; Tomasi, J., A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. The Journal of chemical physics 1997, 107 (8), 3210-3221.

55. Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C., Refinement and parametrization of COSMO-RS. The Journal of Physical Chemistry A 1998, 102 (26), 5074-5085.

56. Cramer, C. J.; Truhlar, D. G., SMx continuum models for condensed phases. Trends and Perspectives in Modern Computational Science 2006, 6, 112-140.

57. Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K., On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. Journal of the American Chemical Society 2003, 125 (31), 9523-9530.

58. (a) Honig, B.; Nicholls, A., Classical electrostatics in biology and chemistry. Science 1995, 268 (5214), 1144-1149. (b) Gilson, M. K.; Honig, B., Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins: Structure, Function, and Bioinformatics 1988, 4 (1), 7-18. (c) Cramer, C. J.; Truhlar, D. G., Implicit solvation models: equilibria, structure, spectra, and dynamics. Chemical Reviews 1999, 99 (8), 2161-2200. (d) Cramer, C. J.; Truhlar, D. G., An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. Science 1992, 256 (5054), 213-217. (e) Sitkoff, D.; Sharp, K. A.; Honig, B., Accurate calculation of hydration free energies using macroscopic solvent models. The Journal of Physical Chemistry 1994, 98 (7),

1978-1988. (f) Luo, R.; Moult, J.; Gilson, M. K., Dielectric screening treatment of electrostatic solvation. The Journal of Physical Chemistry B 1997, 101 (51), 11226-11236.

59. Ben-Naim, A., Hydrophobic interaction and structural changes in the solvent. Biopolymers 1975, 14 (7), 1337-1355.

60. (a) Honig, B.; Yang, A.-S., Free energy balance in protein folding. Advances in protein chemistry 1994, 46, 27-58. (b) Dill, K. A., Dominant forces in protein folding. Biochemistry 1990, 29 (31), 7133-7155. (c) Kauzmann, W., OF PROTEIN DENATURATION1. Advances in protein chemistry 1959, 14, 1. (d) Privalov, P. L.; Makhatadze, G. I., Contribution of hydration to protein folding thermodynamics: II. The entropy and Gibbs energy of hydration. Journal of molecular biology 1993, 232 (2), 660-679.

61. (a) Sturtevant, J. M., Heat capacity and entropy changes in processes involving proteins. Proceedings of the National Academy of Sciences 1977, 74 (6), 2236-2240. (b) Williams, D. H.; Searle, M. S.; Mackay, J. P.; Gerhard, U.; Maplestone, R. A., Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics. Proceedings of the National Academy of Sciences 1993, 90 (4), 1172-1178. (c) Froloff, N.; Windemuth, A.; Honig, B., On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. Protein Science 1997, 6 (6), 1293-1301. (d) Lee, M. R.; Duan, Y.; Kollman, P. A., Use of MM-PB/SA in estimating the free energies of proteins: Application to native, intermediates, and unfolded villin headpiece. Proteins: Structure, Function, and Bioinformatics 2000, 39 (4), 309-316.

62. Tanford, C., The Hydrophobic Effect: Formation of Micelles and Biological Membranes 2d Ed. J. Wiley.: 1980.

63. Kukic, P.; Nielsen, J. E., Electrostatics in proteins and protein-ligand complexes. Future medicinal chemistry 2010, 2 (4), 647-666.

64. Ghosh, A.; Rapp, C. S.; Friesner, R. A., Generalized Born model based on a surface integral formulation. The Journal of Physical Chemistry B 1998, 102 (52), 10983-10990.

65. Baker, N. A., Biomolecular applications of Poisson-Boltzmann methods. Reviews in computational chemistry 2005, 21, 349.

66. Grochowski, P.; Trylska, J., Continuum molecular electrostatics, salt effects, and counterion binding—a review of the Poisson–Boltzmann theory and its modifications. Biopolymers 2008, 89 (2), 93-113.

67. Tynan-Connolly, B. M.; Nielsen, J. E., Redesigning protein pKa values. Protein science 2007, 16 (2), 239-249.

68. Baran, K. L.; Chimenti, M. S.; Schlessman, J. L.; Fitch, C. A.; Herbst, K. J.; Garcia-Moreno, B. E., Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease. Journal of molecular biology 2008, 379 (5), 1045-1062.

69. (a) Warwicker, J.; Watson, H., Calculation of the electric potential in the active site cleft due to α-helix dipoles. Journal of molecular biology 1982, 157 (4), 671-679. (b) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B., Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. Proteins: Structure, Function, and Bioinformatics 1986, 1 (1), 47-59. (c) Davis, M. E.; McCammon, J. A., Solving the finite difference linearized Poisson-Boltzmann equation: A comparison of relaxation and conjugate gradient methods. Journal of computational chemistry 1989, 10 (3), 386-391. (d) Nicholls, A.; Sharp, K. A.; Honig, B., Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Structure, Function, and Bioinformatics 1991, 11 (4), 281-296.

70. Tomasi, J.; Persico, M., Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. Chemical Reviews 1994, 94 (7), 2027-2094.

71. (a) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M., Solvent models for protein–ligand binding: Comparison of implicit solvent Poisson and surface generalized Born models with explicit solvent simulations. Journal of Computational Chemistry 2001, 22 (6), 591-607. (b) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical treatment of solvation for molecular mechanics and dynamics. Journal of the American Chemical Society 1990, 112 (16), 6127-6129.

72. (a) Luo, R.; Moult, J.; Gilson, M. K., Dielectric screening treatment of electrostatic solvation. The Journal of Physical Chemistry B 1997, 101 (51), 11226-11236. (b) Mehler, E.; Eichele, G., Electrostatic effects in water-accessible regions of proteins. Biochemistry 1984, 23 (17), 3887-3891.

73. (a) Davis, M. E., The inducible multipole solvation model: A new model for solvation effects on solute electrostatics. The Journal of chemical physics 1994, 100 (7), 5149-5159. (b) David, L.; Field, M. J., Adapting the inducible multipole solvation model for use in molecular dynamics simulations. Chemical physics letters 1995, 245 (4), 371-376.

74. Eisenberg, D.; McLachlan, A. D., Solvation energy in protein folding and binding. 1986.

75. Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A., Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proceedings of the National Academy of Sciences 1987, 84 (10), 3086-3090.

76. (a) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B., New model for calculation of solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. The Journal of Physical Chemistry 1996, 100 (28), 11775-11788. (b) Hünenberger, P. H.; Helms, V.; Narayana, N.; Taylor, S. S.; McCammon, J. A., Determinants of ligand binding to cAMP-dependent protein kinase. Biochemistry 1999, 38 (8), 2358-2366. (c) Simonson, T.; Bruenger, A. T., Solvation free energies estimated from macroscopic continuum theory: an accuracy assessment. The Journal of Physical Chemistry 1994, 98 (17), 4683-4694. (d) Rapp, C. S.; Friesner, R. A.,

Prediction of loop geometries using a generalized born model of solvation effects. Proteins: Structure, Function, and Bioinformatics 1999, 35 (2), 173-183. (e) Fogolari, F.; Esposito, G.; Viglino, P.; Molinari, H., Molecular mechanics and dynamics of biomolecules using a solvent continuum model. Journal of computational chemistry 2001, 22 (15), 1830-1842. (f) Pellegrini, E.; Field, M. J., A generalized-born solvation model for macromolecular hybrid-potential calculations. The Journal of Physical Chemistry A 2002, 106 (7), 1316-1326. (g) Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Manuel, F.; Rinaldi, D.; Orozco, M.; Luque, F. J., Electrostatic component of solvation: comparison of SCRF continuum models. Journal of computational chemistry 2003, 24 (3), 284-297.

77. (a) Hermann, R. B., Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. The Journal of Physical Chemistry 1972, 76 (19), 2754-2759. (b) Chothia, C., Hydrophobic bonding and accessible surface area in proteins. Nature 1974, 248 (5446), 338-339. (c) Reynolds, J. A.; Gilbert, D. B.; Tanford, C., Empirical correlation between hydrophobic free energy and aqueous cavity surface area. Proceedings of the National Academy of Sciences 1974, 71 (8), 2925-2927.

78. (a) Ferrara, P.; Apostolakis, J.; Caflisch, A., Evaluation of a fast implicit solvent model for molecular dynamics simulations. Proteins: Structure, Function, and Bioinformatics 2002, 46 (1), 24-33. (b) Guvench, O.; Brooks, C. L., Efficient approximate all-atom solvent accessible surface area method parameterized for folded and denatured protein conformations. Journal of computational chemistry 2004, 25 (8), 1005-1014. (c) Weiser, J.; Shenkin, P. S.; Still, W. C., Approximate solvent-accessible surface areas from tetrahedrally directed neighbor densities. Biopolymers 1999, 50 (4), 373-380.

79. Hou, T.; Qiao, X.; Zhang, W.; Xu, X., Empirical aqueous solvation models based on accessible surface areas with implicit electrostatics. The Journal of Physical Chemistry B 2002, 106 (43), 11295-11304.

80. Lazaridis, T.; Karplus, M., Effective energy function for proteins in solution. Proteins: Structure, Function, and Bioinformatics 1999, 35 (2), 133-152.

81. (a) Haberthür, U.; Majeux, N.; Werner, P.; Caflisch, A., Efficient evaluation of the effective dielectric function of a macromolecule in aqueous solution. Journal of computational chemistry 2003, 24 (15), 1936-1949. (b) Hassan, S. A.; Mehler, E. L.; Zhang, D.; Weinstein, H., Molecular dynamics simulations of peptides and proteins with a continuum electrostatic model based on screened Coulomb potentials. Proteins: Structure, Function, and Bioinformatics 2003, 51 (1), 109-125.

CHAPTER 2

Developing and parameterizing first-order Fuzzy-Border (FB) continuum solvation model with OPLS-AA force field and calculating hydration energies of small molecules and pK_a values of substituted phenols

Part of the material covered in this chapter was also published in the following journal:

"Calculating pKa values for substituted phenols and hydration energies for other compounds with the first-order fuzzy-border continuum solvation model."
Sharma, Ity and George A. Kaminski, *Journal of computational chemistry* 33.30 (2012): 2388-2399.

Portion of this work was also presented at the following meeting:

"Developing first-order continuum Fuzzy-Border solvation model and applying for calculating pKa values of substituted phenols and hydration energies of small molecules."

Ity Sharma, George A Kaminski, American Chemical Society (ACS), 244th ACS National Meeting & Exposition - August 19-23, 2012, Philadelphia , USA, 2012

2.1 Introduction

Since many organic, physical and almost all biophysical and biochemical processes take place in water, an accurate and efficient representation of the solvent is crucial in computer simulations of such systems. Some properties simply cannot be addressed without having an adequate water model for fundamental reasons. Calculating pK_a values is one example of such simulations. Accurate assessment of acidity constants is both an important and a challenging task. Success of many applications depend on robustness of these calculations, including those in pharmaceutical industry and related to protein-ligand binding in general, as protein structure and function depend on the states of ionizable residues. At the same time, the problem of calculating pK_a values correctly is not trivial as it involves finding a careful balance between two components of almost equally large magnitude, the bond breaking (gas-phase) deprotonation energy and the energy of the ionic hydration. Therefore, utilizing a good model for the surrounding water is crucial in pK_a calculations.

There are two main ways to simulate aqueous environment around a solute. One of them consists of representing water molecules explicitly. While this method has proved to be successful in a number of studies, it suffers from one significant drawback. Even a small molecular system has to be surrounded by hundreds of water molecules, and this number grows rapidly as the system protein size increases. There are also other issues which may preclude use of an explicit solvation model for particular tasks, for example, the need to approximate an essentially infinitely large bulk solvent with a finite number of solvent molecules.

This is why continuum solvation models have proven to be of great help in assessing solvation

free energy. Our lab is just one of those utilizing the benefits of such models, and we have been able to do so in a number of projects, including those which involve calculating acidity constant values for both proteins and smaller molecular systems.[1-3]

The two most common continuum solvent models in the surface formulation are the Poisson-Boltzmann and the Generalized Born models. In both the models, the whole space is divided into the part occupied by the solute and the rest of the space which is taken up by the solvent. The interface between these two parts of the space is assumed to have a certain charge density. The electrostatic component of the solute-solvent interactions is assumed to be the most important one. The Poisson-Boltzmann continuum solvation technique is based on the Poisson-Boltzmann equation which relates the electrostatic potential to the integral containing this surface density over the interface. In turn, this field affects the surface density itself, as any two different points of the surface polarize each other, and the resulting formalism leads to a self-consistent equation not entirely dissimilar from the regular electrostatic polarization equation.[4] While the above method permits the exact answer to the problem of finding the electrostatic potential of the solvent as interacting with the solute, and thus finding the exact value of the solvation energy, a more approximate Generalized Born formalism can also be used.[4]

The Poisson-Boltzmann approach relies on solving the corresponding equations on a numerical grid, and some versions of Generalized Born model use a grid to compute effective Born radii. This can and does lead to two kinds of problems. First, even a small movement of a solute can lead to generation of a new grid, in which a point is added, omitted or significantly shifted, depending on the specific solute geometry and the rules applied in grid building. This leads to a

noise in the solvation energy which can be as high as several kcal/mol, which presents obvious problems in energy minimizations and thermodynamic sampling, especially molecular dynamics, because of the non-smoothness of the solvation energy as a function of the solute coordinates. The second problem arises from the fact that convergence of the self-consistent Poisson-Boltzmann problem is not always achieved automatically on a numerical grid. Methods have been suggested to address the above issues. For example, the numeric Poisson-Boltzmann solution can be sought on a fixed equally spaced grid which is built independently of the solute coordinates.[5] Smoothing, antialiasing and careful choosing of the grid parameters can also be employed to ensure convergence.[5,6]

The Fuzzy-Border (FB) continuum solvation methodology combines the following two features. First, the above mentioned solutions are followed making use of a fixed-position equally spaced three dimensional grid with position of the nodes not dependent on the solute coordinates. Second, an approximation to the full-scale Poisson-Boltzmann procedure is used which is similar to fast second-order polarization technique for solutes and explicit solutions and pure liquids developed and tested in our lab.[7,8a] It truncates the self-consistent procedure of solving the complete Poisson-Boltzmann equation at either the first or the second iteration. While this method increases the computational speed as compared to the full-scale Poisson-Boltzmann formalism, its main advantage in the case of applying it to a continuum solvent is in reducing the self-consistent problem to an analytical one. The resulting model, while being an approximation as compared to the true Poisson-Boltzmann technique, is still functionally closer to reproducing the true many-body nature of the solvation energy than the Generalized Born method.

It is also important to mention that FB method is distinct enough from the other continuum

solvation techniques available from other research groups. Unlike the PCM models,[9] it is geared toward use with empirical force fields and not quantum mechanics. It contains no conductor-like screening of COSMO[10] and is not based on the Generalized Born methodology of the SMx models.[11] Moreover, while there have been successful attempts in noise reduction in Poisson-Boltzmann calculations by using a fixed grid or smoothing functions,[12] as well as in Generalized Born models,[13] it has not been combined with the iteration truncation implemented in the FB model to avoid possible convergence problems. This truncation is consistent in spirit with the fast polarization technique for explicit all-atom simulations of solutes and solvents developed in our lab.[7,8a] And the differences from the standard Poisson-Boltzmann technique have been outlined above.

The first-order version of the FB method, implemented by augmenting POSSIM software suite[8a] developed by Dr. Kaminski is applied to calculate pK_a values of substituted phenols and to find hydration energies of 40 other molecules simulated with the OPLS-AA force field.[14] The accuracy of 0.41 pH units and ca. 0.076 kcal/mol in the acidity constants and the hydration energies, respectively is achieved in these simulations. It should be noted that this accuracy is achieved with direct fitting to the pK_a and hydration energies data. Therefore, it is not claimed that the OPLS/FB methodology exceeds in accuracy to the best available quantum mechanical techniques, but only that it is physically correct enough to permit such a fitting and to assure a reasonable level of parameter transferability.

The rest of this chapter is divided into three sections. Section 2.2 discusses the fixed-charge Optimized Potential for Liquid Simulations (OPLS) force field formulation followed by

detailed methodology of continuum first-order Fuzzy Border solvation model and protocol for calculating absolute acidity constants of molecules. Section 2.3 discusses the results and discussions. Conclusions presented in section 2.4 followed by references.

2.2 Methods

2.2.1 *Force Field – Optimized Potential for Liquid Simulations-All Atom (OPLS-AA) force field* [14]

Professor William Jorgensen group has developed fixed-charge OPLS force field and is used for benchmarking in many molecular mechanics calculations. In OPLS force field the total energy (E_{total}) is evaluated as a sum of bond stretching ($E_{stretch}$), angle bending (E_{bend}), torsion energy ($E_{torsion}$) and electrostatic ($E_{electrostatic}$) and van der Waal (E_{vdW}) interactions evaluated for all the atoms in the system (**Equation 1**).

$$E_{total} = E_{stretch} + E_{bend} + E_{torsion} + E_{electrostatic} + E_{vdW} \quad \textbf{(1)}$$

The bond stretching, angle bending functions are represented by the harmonic terms (**Equation 2-3**):

$$E_{bond} = \sum_{bonds} K_r \left(r - r_{eq}\right)^2 \quad \textbf{(2)}$$

$$E_{angle} = \sum_{angles} K_\theta \left(\theta - \theta_{eq}\right)^2 \quad \textbf{(3)}$$

In **equation (2) and (3)**, r, r_{eq}, θ and θ_{eq} denote the actual and equilibrium values of bond lengths and angles. K_r and K_θ represent the force constants.

The torsional energy is given by the Fourier series (**Equation 4**)

$$E_{torsion} = \sum_i \frac{V_1^i}{2}[1 + \cos(\phi_i + f_i 1)] + \frac{V_2^i}{2}[1 + \cos(2\phi_i + f_i 2)] + \frac{V_3^i}{3}[1 + \cos(3\phi_i + f_i 3)] \quad \textbf{(4)}$$

In this intramolecular energy term presented in **equation (4)**, ϕ_i is the dihedral angle, V_1, V_2 and V_3 are the Fourier series coefficients and f1, f2 and f3 are phase angles.

The non-bonded Coulomb and Lennard-jones interactions between atoms a and b separated by three or more bonds is given in **equation 5**

$$E_{nb} = \sum_{i<j} \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) \right] f_{ij} \quad \textbf{(5)}$$

The non-bonded interactions are also evaluated for intermolecular pairs of atoms to obtain intermolecular energy.

$$E_{ab} = \sum_i^{on\ a} \sum_j^{on\ b} \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) \right] f_{ij} \quad \textbf{(6)}$$

The first term in **equation (5) and (6)** represents the electrostatic term calculated over all the atom pairs. The value of factor f_{ij} is 0.0 for 1,2 and 1,3 pairs of atoms belonging to the same

valence bond or angle and 0.5 for 1,4-interactions in atoms in the same dihedral. For all the other pairs of atoms, f_{ij} is set to 1.0.

The van der Waal interactions are calculated by Lennard Jones formalism given by second term in these equations. Geometric combining rules used for Lennard-Jones parameters are $\sigma_{ij} = (\sigma_i \sigma_j)^{1/2}$ and $\varepsilon_{ij} = (\varepsilon_i \varepsilon_j)^{1/2}$. In the AA (all-atom) model, for the charges *q* and the Lennard-Jones terms, an interaction site is placed on each atom.

2.2.2. *Fuzzy-Border (FB) continuum solvation model*

2.2.2.1 The Electrostatic Component of the Solvation Energy

The overall solvation energy ($\Delta G_{solvation}$) is calculated as a sum of the electrostatic ($\Delta G_{electrostatic}$) and non-polar terms ($\Delta G_{nonpolar}$):

$$\Delta G_{(solvation)} = \Delta G_{(electrostatic)} + \Delta G_{(nonpolar)} \qquad \textbf{(7)}$$

The electrostatic part of the energy is calculated by using an approximation to the Poisson-Boltzmann formalism.[4] Briefly, the solute-solvent surface can be considered as shown in **Figure 1**.

In **figure 1**, $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$ are the values of the electrostatic field in the solute and solvent regions respectively and **n** is the unit vector normal to the interface.

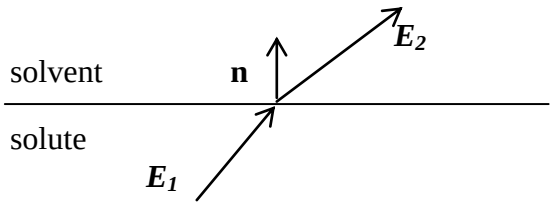


Figure 1: Electrostatic field at the solute-solvent interface.

The continuity of the normal component of the electric displacement across the surface requires that:

$$\varepsilon_1 \boldsymbol{E}_1.\boldsymbol{n} = \varepsilon_2 \boldsymbol{E}_2.\boldsymbol{n} \quad \textbf{(8)}$$

In **equation (8)** ε_1 and ε_2 stand for the dielectric constants inside and outside of the solute, respectively. Setting ε_1 =1, **Equation 8 can be rewritten** as (**Equation (9)**):

$$\boldsymbol{E}_1.\boldsymbol{n} = \varepsilon_2 \boldsymbol{E}_2.\boldsymbol{n} \quad \textbf{(9)}$$

According to the Gauss law,

$$4\pi\sigma = (\boldsymbol{E}_2 - \boldsymbol{E}_1).\boldsymbol{n} \quad \textbf{(10)}$$

with σ being the surface charge density at the interface. Combining **equations (9) and (10)**,

$$\sigma = -\frac{1}{4\pi}\left(1 - \frac{1}{\varepsilon}\right)\boldsymbol{E}_1.\boldsymbol{n} \quad \textbf{(11)}$$

Because the electrostatic field $\boldsymbol{E}_1$ itself depends on the surface charge density distribution, **Equation (11)** describes a self-consistent problem, just like the general electrostatic polarization case. The electrostatic part of the solvation energy can be then found as:

$$\Delta G_{(el)} = \frac{1}{2}\int_{S} \sigma \phi^{0} d^{2}r \qquad \textbf{(12)}$$

In **equation (12)** ϕ^{0} represents the electrostatic potential created by the charges and/or other electrostatic multipoles of the solute only (not by the polarized solvent) and the integration is carried out over the whole solute-solvent interface surface.

When the equation is solved numerically, the surface is represented by a discrete set of points *i*. In this case, **Equations (11) and (12)** are approximated by **Equations (13) and (14)**:

$$q_i = -\frac{1}{4\pi}\left(1 - \frac{1}{\varepsilon}\right) E_{1,i}.n_i \qquad \textbf{(13)}$$

$$\Delta G_{(el)} = \frac{1}{2}\sum_{i} q_i \phi_i^{0} \qquad \textbf{(14)}$$

The electrostatic field $E_{1,i}$ is calculated as (**Equation 15**):

$$E_{1,i} = \sum_{j} \frac{q_j R_{ij}}{R_{ij}^{3}} + \sum_{k \neq i} \frac{q_k R_{ik}}{R_{ik}^{3}} \qquad \textbf{(15)}$$

The first sum is taken over the solute charges (this expression can be easily extended to include higher-order multipoles), while the second one goes over the other solute-solvent interface points. R_{ij} stands for the vector from point *j* to point *i*. Specific details about our choice of the numeric grid points will be discussed in the next subsection.

The fast polarization approximation, which constitutes an integral part of the Fuzzy-Border (FB) model is introduced at this point. **Equations (13) and (15)** together form a self-consistent problem which can be solved iteratively. The first step consists of replacing $\boldsymbol{E}_{1,i}$ with $\boldsymbol{E}_{1,i}^{\mathbf{0}}$, the field created by the solute only and not by the polarized solute-solvent interface. If we stop at this stage, we obtain our first-order approximation. It still contains many-body interactions, since the electrostatic field is a vector quantity and it contains contributions from all the solute charges. But the problem is now analytical, not self-consistent, and the convergence is no longer an issue. If we now recompute the electrostatic field $\boldsymbol{E_{2,i}}$ taking into the account the field created by the interface charges using **Equation (15)**, obtain the new charges with the **Equation (13)**, but do not perform the next iteration, we obtain the second-order model, which is still safe from the electrostatic charges convergence problems since there are only two iterations and magnitudes of the charges cannot increase beyond the value they achieve at the first of the second iteration.

The rest of the formulation of the Fuzzy-Border model is given in the next subsection.

2.2.2.2. Choosing the Numerical Grid to Represent the Solute-Solvent Interface

A fixed cubic three-dimensional equally-spaced grid is used to minimize the noise resulting from grid rebuilding after moving a solute atom or a group of atoms. The interface between solute and solvent is assumed to consist of points with distances from R-Δ to R+Δ from the solute atom, as shown in **Figure 2**.

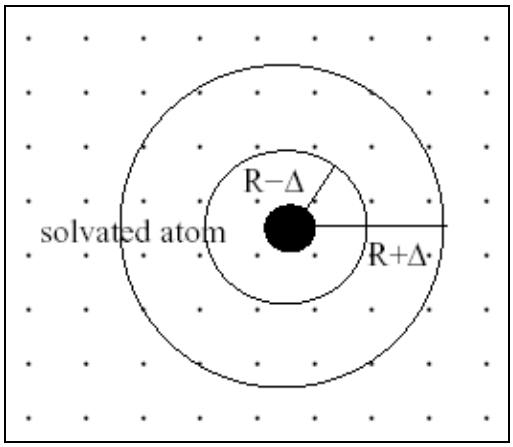


Figure 2: Schematic depiction of a solvated atom and the solvent grid.

The values of parameters R and Δ can be different for different solute atom types. Only those solvent points that were located in the solvent-accessible region were used. Therefore, the overall solvation surface is constructed with the knowledge of positions and solvation parameters of all the solute atoms. A point is defined as solvent-accessible if there is no solvent atom i which would be closer to the grid point than $R_i - \Delta_i$. Because the "real" radius of the solvation surface is R_i and not $R_i - \Delta_i$ or $R_i + \Delta_i$, the surface points j corresponding to the solute atom i, has weights w_j associated with them calculated as shown in **Equation (16)**:

$$w_j = a_{0,i}\left[\left(\frac{R_{ij}-R_i}{\Delta_i}\right)^4 - 2\left(\frac{R_{ij}-R_i}{\Delta_i}\right)^2 + 1\right] \qquad \textbf{(16)}$$

R_{ij} in **equation (16)** is the distance between the solvent grid point and the solute atom, $a_{0,i}$ is a parameter which depends on the solute atom type, and the whole weight is maximum at the nominal solvation radius R_i and decreases to zero at distances $R_i - \Delta_i$ and $R_i + \Delta_i$. Since one grid point can be within the solvation surface range of more than one solute charge, several weights w_{ji} may be needed for the same grid point j and different solute atoms i.

The weights are normalized in the following way (**Equation 17**):

$$w_j = \frac{\sum_i w_{ji}^2}{\sum_i w_{ji}} \qquad \textbf{(17)}$$

The unit normal vector n_{ji} for each solute atom i and the grid point j is assumed to be in the direction from the solute atom to the grid point. The overall unit normal vector for the grid point is then calculated as given in **Equation (18)**:

$$\boldsymbol{n_j = \sum_i n_{ji} w_{ji}} \qquad \textbf{(18)}$$

The overall FB continuum solvation formalism can now be written. Once the solvation surface grid points j are defined as described above, the zeroth-order electrostatic field at those points is found as shown in **Equation (19)**:

$$\boldsymbol{E^0_{2j} = \sum_i \frac{q_i R_{ij}}{R^3_{ij}}} \qquad \textbf{(19)}$$

The summation goes over all the solute points. The first-order FB charge on the grid point j is then found as shown in **Equation (20)**:

$$q_j^I = -A_{scale}\frac{1}{4}\pi(\varepsilon - 1)w_j E_{1,j}^0 . n_j \quad \textbf{(20)}$$

A_{scale} is a scaling factor and an adjustable parameter of the theory. The first-order electrostatic part of the solvation energy can then be calculated as shown in **Equation (21)**:

$$\Delta G_{(el)}^I = \frac{1}{2}\sum_j q_j^I \phi_j^0 \quad \textbf{(21)}$$

If the second-order approximation is to be produced, the first-order electrostatic field is found as described in **Equation (22)**:

$$E_{1,j}^I = E_{1,j}^0 + A_{self}\sum_{k \neq j}\frac{q_k R_{kj}}{R_{kj}^3} \quad \textbf{(22)}$$

The additional summation done over the solute-solvent interface point's *k*, and A_{self} being another adjustable parameter (with its value being the same for all the grid points). **Equation (20)** in then modified to include the first-order and not the zeroth-order field:

$$q_j^I = -A_{scale}\frac{1}{4}\pi(\varepsilon - 1)w_j E_{1,j}^I . n_j \quad \textbf{(23)}$$

The resulting second-order energy is found as shown in e**quation (24)** in accordance with e**quation (23)**:

$$\Delta G^{II}_{(el)} = \frac{1}{2}\sum_j q^{II}_j \phi^0_j \qquad \textbf{(24)}$$

It should also be noted that effect upon the energy of shifting the position of the fixed cubic grid by one half of the spacing between the grid points is assessed every time the hydration energy is calculated, and the energy is averaged with respect to this transformation.

Finally, the electrostatic part of the energy, regardless of whether it is calculated with the first- or second-order model, is multiplied by 332.0657418 in order to obtain the final result in kcal/mol.

This formulation of the electrostatic part of the Fuzzy-Border (FB) method lies between the Poisson-Boltzmann (PB) and Generalized Born (GB) models in the sense of being close to the physical correctness of PB one. The difference between the FB and PB techniques is clear – the former truncates some advanced iterations in the self-consistent procedure which is usually involved in solving the equations of the latter one. As far as the GB continuum solvation model is concerned, its surface formulation includes two terms:[4]

$$U = \sum_k U_{se}(q_k, r_k) + \sum_{i \neq j} U_{pr}(q_i, q_j, r_i, r_j) \qquad \textbf{(25)}$$

The second term in **equation (25)** represents screened pairwise electrostatic interactions:

$$U_{pr} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-D}}}, \alpha_{ij} = \sqrt{\alpha_i \alpha_j} \qquad D = \frac{r_{ij}^2}{(2\alpha_{ij})^2} \tag{26}$$

The first term in **equation (25)** stands for the single-charge solvation energy:

$$U_{se} = -\frac{q_k^2}{8\pi}\left(1 - \frac{1}{\varepsilon}\right)\int_S \frac{(R - r_k).n(R)}{|R - r_k|^4} d^2 R \tag{27}$$

It should be noted that this first term contains no reference to electrostatic interactions between the charge q_k and other charges of the solute (except in an indirect form, by shaping the solute-solvent interface S). And the value of the U_{se} term is what ultimately defines the Born parameter α in **equation (26)**. Therefore, the Generalized Born formalism contains no directly defined many-body electrostatic interactions, while the Fuzzy-Border one includes this part of the physical picture explicitly as the electrostatic field is calculated with all the solute charges, even in the first-order FB model. Thus, we place the FB technique between its PB and GB counterparts.

At the same time, it should be emphasized very strongly that such a relation among the physical bases of these three methods does not at all mean that the actual accuracy of computing hydration energies will be the best with the Poisson-Boltzmann formalism, somewhat worse with the Fuzzy-Border method and even worse with the Generalized Born one (as is witnessed by the data presented in the next Section of this paper). There are many factors which are affecting the accuracy. They include the size of the fitting set, the number of parameters used in the fitting, the fitting technique, etc. Generally speaking all physically reasonable models should be able to

produce reasonably accurate results for solvation energies of small molecular systems. And our aim is, first and foremost, to demonstrate that the Fuzzy-Border technique is in fact one of such reasonable and robust approaches. Moreover, the FB model is formulated without the true self-consistency of the true Poisson-Boltzmann method. Thus, in principle, all the expressions for the surface charges can be written analytically. This should make it possible to directly derive analytical gradients for this model in the future.

The grid generation for continuum solvation model is based on three steps to compute and visualize the solute molecular surface and its interaction with the solvent. Since a molecule do not have any definite boundaries, the surface and volume of a molecule can be described as van der Waals surface, solvent accessible surface area and molecular surface. The two dimensional representation of van der Waals surface, solvent accessible surface area and molecular surface is shown in the **Figure 3**.[15]

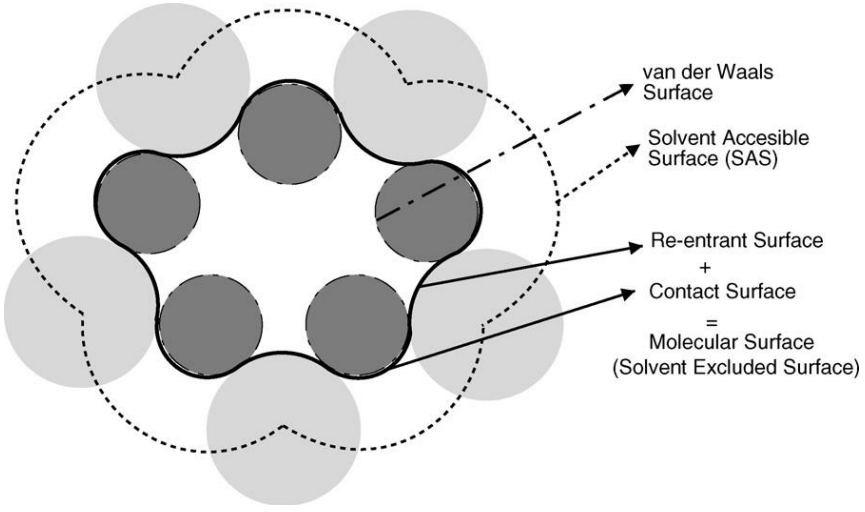


Figure 3: Two dimensional molecular surface representation[15]

Van der Waals surface of a molecule is the surface area of the atoms represented by spheres of van der Waals radius. Solvent accessible surface is the surface generated by rolling a probe of radius of a solvent molecule over the van der Waals surface. Molecular or the solvent excluded surface is the boundary of contact between the solvent surface and the van der Walls surface.

The solute as atomic coordinates is placed on three dimensional grid. The size of the grid in the three dimensions is held fixed in the simulations. The molecular surface of the solute is computed in three steps.

Firstly, outward marching front is used to compute the van der Waals and solvent accessible surface. All the grid cells are considered outside the two surfaces. The volume inside the solvent accessible surface area and the van der Waals is marked for each atom in the molecule by considering the centers of the cells whose centers fall inside the volume of these surfaces as inside as shown in **Figure 4**.

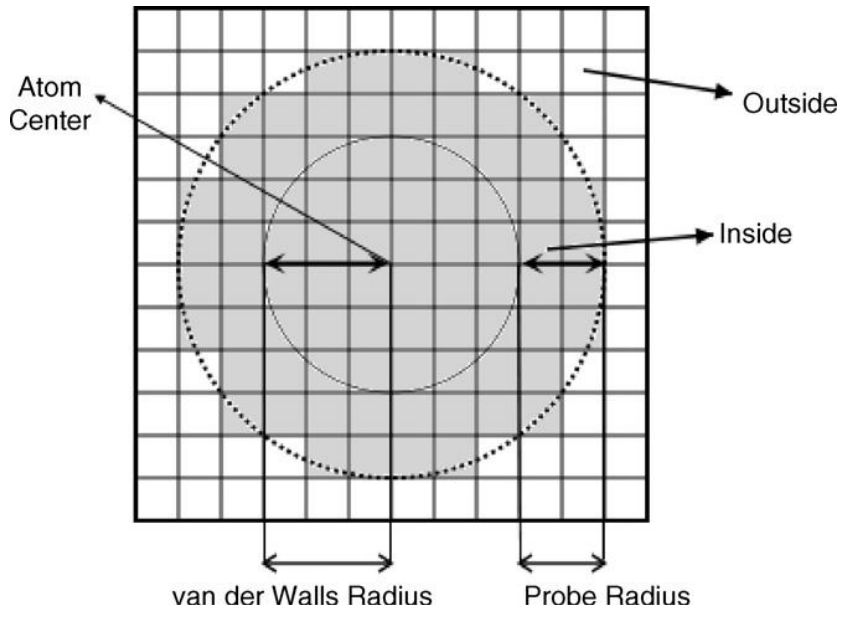


Figure 4: Location of van der Waals and solvent-accessible surface by marking the grid cells[15] whose centers fall inside the volume defined by the solvent-accessible surface as inside (shaded grid cells).

The second step is using inward marching front to mark out the parts inside the solvent molecule or the solvent excluded surface. The probe spheres centered in the solvent accessible area are found followed by identifying each grid centered in the probe sphere as outside or white as shown in the **Figure 5**. The grid cells that fall in the solvent marked inside in step one are now marked as outside

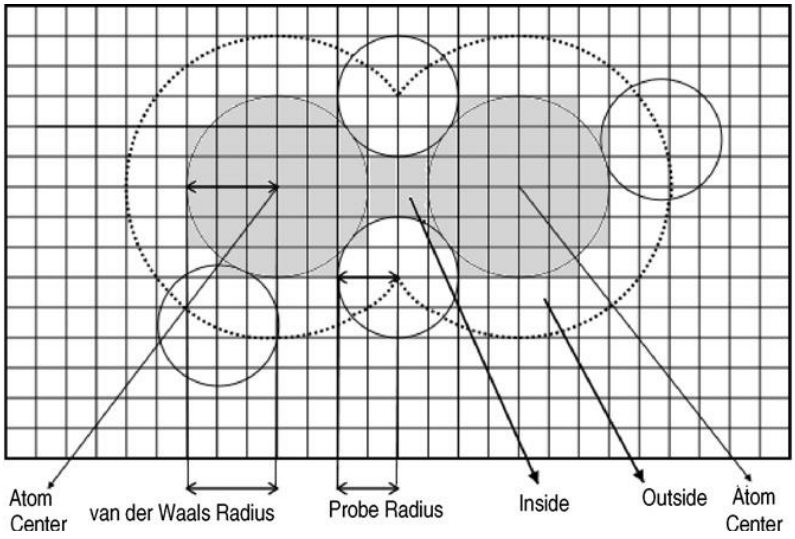


Figure 5: The grid cells whose centers fall inside the probe circles are marked as outside (white grid cells).[15]

The first two steps result in a grid illustrating the volume occupied by solvent excluded surface of the molecule. However, the volume occupied by inaccessible interior cavities inside the molecule need to be discarded while computing the molecular surface. The third step detect the interior cavities inside the molecule using Fast Marching method. This method first builds surface enclosing the molecule followed by shrinking the surface with a fixed signed speed. The grid cell marked as interior is used as the stopping criterion in the speed function. The evolving surface stops completely when in touch with the inside grid cells. In the shrinkage process, a grid cell is only visited once by the surface. This Fast Marching Method allows detection of outer

surface as well as the inaccessible interior cavities inside the molecule.

2.2.2.3. The Non-Polar Part of the Solvation Energy

The non-polar part of the solvation energy was calculated as a sum of two terms, one with a positive and one with a negative contribution (**Equation 28**):

$$\Delta G_{(np)} = \sum_j w_j A_{np} - \sum_i \sum_j w_j \frac{A_i^{LJ}}{R_{ij}^6} \qquad \textbf{(28)}$$

The first term contains a sum taken over all the grid points. This is essentially the overall solvent-accessible surface area (SASA) contribution which is commonly employed in continuum solvation models (see for example Reference 16) adapted to the Fuzzy-Border formalism. The second term is calculated with a double summation going over all the grid points *j* and all the solute atoms *i*. It approximates the attraction part of the Lennard-Jones energy for interactions between the solute and solvent atoms.

Once the electrostatic and non-polar terms of the solvation energy are found, the overall solvation energy can be calculated according to **equation (7)**.

2.2.3. The general protocol used in the pK_a calculations

The following closed thermodynamic cycle is used in order to calculate the acidity constants (**Figure 6**):

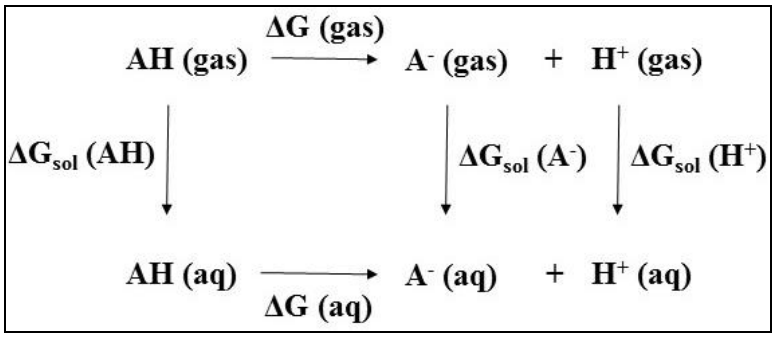


Figure 6: Thermodynamic cycle used to calculate pK_a

Thus,

$$\Delta G_{(gas)} + \Delta G_{sol}(A^-) + \Delta G_{sol}(H^+) - \Delta G_{(aq)} - \Delta G_{(sol)}(AH) = 0 \quad \textbf{(29)}$$

$$\Delta G_{(aq)} = \Delta G_{(gas)} + \Delta G_{sol}(H^+) + \Delta G_{sol}(A^-) - \Delta G_{(sol)}(AH) \quad \textbf{(30)}$$

$$\Delta G_{(aq)} = \Delta G_{(gas)} + \Delta G_{sol}(H^+) + \Delta G(aq, AH \rightarrow A^-) - \Delta G(gas, AH \rightarrow A^-) \quad \textbf{(31)}$$

Finally,

$$pK_a = \frac{\Delta G_{aq}}{(2.303RT)} \quad \textbf{(32)}$$

Because we concentrated on the accuracy of the solvation model in this work, the values of ΔG(gas) taken from literature are experimental values. The quantum mechanical values were adopted for the molecules whose experimental data was not available. Thus, we were calculating the ΔG(aq, AH→A$^-$) and ΔG(gas, AH→A$^-$) terms in **equation (31)**. They were found as differences of the corresponding energies, just as we did in the previous studies described in References 1-3 and 17. Moreover, since the proton free energy of solvation $\Delta G_{sol}(H^+)$ has the

same value for any molecular system, a carefully determined literature result[18] of -269.0 kcal/mol was used. This choice was the same as previous work in our group on calculating absolute pK_a values for substituted phenols with explicit solvation model and a polarizable force field.[17]

Values of the energies $G(A^-)$, G(A-H), $G(Acid^-)$, and G(Acid-H) were obtained via geometry optimizations with the Fuzzy-Border (FB) model. All the geometry optimizations were carried out with our POSSIM software suite[8a,19]. The pK_a values were obtained at 298.15 K.

2.3 Results and Discussions

The different combinations of the values of the solvation parameters were tried but the grid spacing was set to be 0.25Å in each dimension. All calculations were carried out for water as the solvent. The radius of the solvent molecule was set at 1.4Å, and the value of the dielectric constant was ε = 80.4.[20] All the calculations presented in this chapter were done with the first-order Fuzzy-Border model. The scaling factor A_{scale} = 0.07069. The non-polar factor A_{np} in the part of the non-polar solvation energy which corresponded to the positive contribution associated with the solvent-accessible surface area, had the same value of 3.747 for all the solute atoms and all the grid points at the solute-solvent interface. Finally, a value of the parameter Δ (the "fuzziness" of the border) of 0.25Å was picked for all the atom types considered.

OPLS-AA force field[14,15] was used for all the solute parameters employed in this study. The calculations were carried out with the POSSIM software[8a] modified to include the Fuzzy-Border (FB) continuum solvent model. The solvation energies were found as:

$$\Delta G_{(solv,calculated)} = E_{(solvated)} - E_{(gas)} \quad \textbf{(33)}$$

In **equation (33)** $E_{(solvated)}$ and $E_{(gas)}$ are the computed energies of the system in the aqueous solution and in gas-phase, respectively. Both these energies were obtained by energy minimizations, therefore, the resulting calculated hydration energy could differ from the nominal solvation energy obtained for the hydrated solute molecule in accordance with the **equation (33)**. All the geometry optimizations were completely unconstrained, and when applicable, the lowest-energy conformations were used.

The calculated pK_a values are listed in **Table 1**, the related hydration energies are given in **Table 2**, and the remaining energies of hydration are presented in **Table 3**. The values of the hydration parameters are shown in **Table 4**. In addition to calculating the hydration energies with the POSSIM/FB formalism, Poisson-Boltzmann (PBF) and Generalized Born (SGB) continuum solvent models as implemented in IMPACT software suite[8b,c] was also used to calculate hydration energies for comparison as shown in **Table 3**.

Table 1: Calculated and Experimental Values of Acidity Constants

System	pKa, OPLS/FB	pKa, experiment [a]
phenol	9.98	9.98
o-chlorophenol	8.57	8.56
m-chlorophenol	8.51	9.02
p-chlorophenol	10.07	9.38
m-cyanophenol	8.48	8.61
p-cyanophenol	7.42	7.95
m-nitrophenol	8.14	8.40
p-nitrophenol	7.87	7.15
o-methylphenol	10.02	10.29
m-methylphenol	9.11	10.08
p-methylphenol	10.50	10.14
Average Error	0.41	–

[a]Reference 17

Table 2: Values of Hydration Energy for Compounds Related to Phenol, in kcal/mol

System	Calculated Hydration Energies OPLS / Fuzzy-Border	Hydration Energy, Quantum Mechanics [a]
Phenol	-6.61	-7.21, -7.92
Phenoxide ion	-73.02	-73.26, -73.49
ortho-chlorophenol	-8.34	-4.20, -4.61
ortho-chlorophenoxide ion	-64.75	-67.32,-67.58
meta-chlorophenol	-8.49	-7.16, -7.73
meta-chlorophenoxide ion	-63.18	-65.83, -66.34
para-chlorophenol	-7.77	-7.50, -8.08
para-chlorophenoxide ion	-61.54	-67.13, -67.80
meta-cyanophenol	-17.08	-9.71, -10.48
meta-cyanophenoxide ion	-65.51	-63.99, -64.55
para-cyanophenol	-18.05	-10.42, -11.21
para-cyanophenoxide ion	-64.44	-61.35, -62.10
meta-nitrophenol	-9.20	-9.64, -10.54
meta-nitrophenoxide ion	-56.71	-63.20, -63.89
para-nitrophenol	-11.37	-10.65, -11.58
Para-nitrophenoxide ion	-56.44	-57.92, -58.89
ortho-methylphenol	-6.41	-6.60, -7.18

ortho-methylphenoxide ion	-73.70	-70.65, -70.91
meta-methylphenol	-6.46	-7.01, -7.70
meta-methylphenoxide ion	-75.79	-72.90, -73.26
para-methylphenol	-6.89	-7.04, -7.72
para-methylphenoxide ion	-75.71	-73.23, -74.06

[a]Reference 17

Table 3: Calculated and Experimental Values of Hydration Energy for Other Compounds, in kcal/mol

Compound	Hydration Energy, calculated			Hydration Energy, Experimental [a]
	Fuzzy-Border	Poisson-Boltzmann	Generalized Born	
CH_4	1.936	1.587	1.446	1.9 – 2.0[b]
C_2H_6	1.948	1.658	1.674	1.82
C_3H_8	2.005	1.760	1.813	1.96
C_4H_{10}	2.118	1.870	1.994	2.08
iso-C_4H_{10}	2.154	1.936	1.966	2.32
C_2H_4	1.056	1.332	1.361	1.27
1-propene	1.512	1.350	1.116	1.27
1-butene	1.429	1.463	1.228	1.38
butadiene	0.643	0.870	0.564	0.60
acetylene	–0.224	1.607	–0.076	–0.010
1-propyne	–0.190	0.963	–0.052	–0.310
1-butyne	–0.016	0.732	–0.042	–0.160
CH_3OH	–5.100	–5.410	–5.013	–5.11
C_2H_5OH	–4.989	–5.345	–4.037	–5.01
CH_3COCH_3	–3.900	–3.653	–3.709	–3.85
2-butanone	–3.675	–3.564	–3.521	–3.640
2-pentanone	–3.567	–3.251	–2.899	–3.530

3-pentanone	–3.537	–3.381	–3.636	–3.410
CH_3OCH_3	–2.092	–1.537	–1.470	–1.90
$C_2H_5OC_2H_5$	–1.728	–1.552	–1.765	–1.63
methyl amine	–4.566	–5.104	–2.803	–4.560
ethyl amine	–4.375	–4.309	–3.033	–4.500
n-propyl amine	–4.486	–4.331	–2.309	–4.390
n-butyl amine	–4.173	–5.750	–2.269	–4.290
dimethyl amine	–4.485	–4.533	–3.443	–4.290
diethylamine	–3.879	–3.901	–3.026	–4.070
ammonia	–4.301	–4.839	–0.660	–4.310
C_6H_6	–0.866	–1.466	–0.256	–0.870
toluene	–0.933	–1.917	–0.703	–0.89
C_6H_5OH	–6.612	–6.720	–5.624	–6.62
CH_3CONH_2	–9.797	–8.910	–8.349	–9.71
NMA	–10.069	–7.670	–10.637	–10.08
4-methyl-imidazole	–10.243	–11.356	–11.955	–10.25[c]
3-methyl-indole	–5.879	–9.104	–6.088	–5.88[c]
CH_3SH	–1.246	–1.423	–0.782	–1.24
C_2H_5SH	–1.299	–1.045	–0.803	–1.30
CH_3SCH_3	–1.528	–1.709	–1.297	–1.540

$C_2H_5SC_2H_5$	–1.451	–1.360	–1.133	–1.43
CH_3COOH	–6.702	–7.056	–6.051	–6.70
C_2H_5COOH	–6.448	–6.828	–6.440	–6.48
Average error	0.076	0.527	0.639	

[a]Experimental data are from Reference 19, except where noted. [b]Reference 23. [c]Reference 24.

Table 4: Fuzzy-Border Hydration Parameters

Atom	**OPLS-AA Atomtypes**[a]	R, **Å**	Δ, **Å**	a_0	A^{LJ}, **kcal/mol·Å^6**
Aliphatic C and C(=O) in NMA	135, 136, 137, 138, 148, 157, 181, 182, 206, 209, 210, 217, 235, 242, 505, 903, 906	1.900	0.25	0.02748	166.3
sp^2 C, alkenes	142, 143, 150	2.015	0.25	0.02748	196.5
Aliphatic H and H on sp^2 carbons	140, 144, 156, 185, 911	1.357	0.25	0.01034	1.327
sp C, alkynes	925, 927	2.050	0.25	0.01700	250.0
H, alkynes	926	1.500	0.25	0.0200	1.327
Aromatic C	145, 166, 500, 501, 502, 506, 507, 508, 514	2.050	0.25	0.02748	161.5
Aromatic H	146	1.320	0.25	0.01034	10.00
Polar H	155, 168, 204, 270, 240, 241, 504, 909	1.300	0.25	0.01034	1.327
N in amines	900, 901	1.650	0.25	0.00900	262.0
N in ammonia	127	1.357	0.25	0.00650	122.9
O, $(C_nH_{2n+1})OH$, ethers, O(H) in carboxylic acids, NMA	154, 180, 235, 268	1.735	0.25	0.02748	118,7

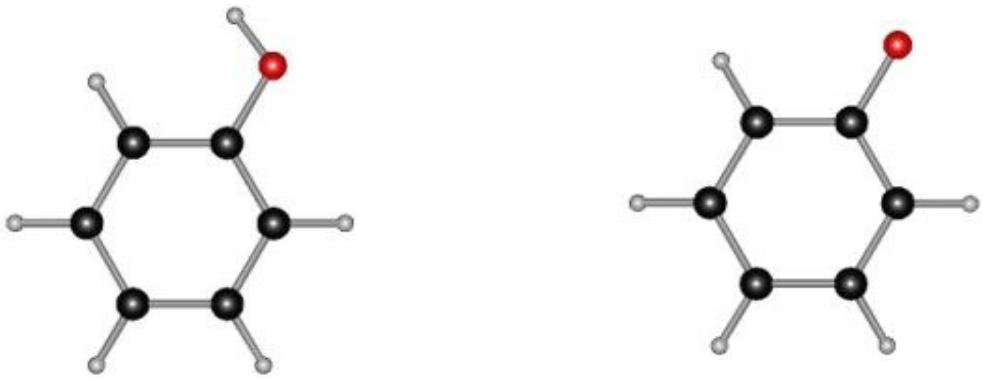
O, phenol	167	1.700	0.25	0.02748	143.3
O^- in phenoxide	420	2.690	0.25	0.01734	142.9
Cl, chlorophenols	264	1.905	0.25	0.00808	250.0
C (CN)	261	1.300	0.25	0.00600	100.0
N (CN)	262	1.825	0.25	0.02100	200.0
$O(NO_2)$	761	1.550	0.25	0.00821	110.0
$N(NO_2)$	767	1.370	0.25	0.00900	1.327
O, O(=C) in carboxylic acids and acetone	269, 281	1.750	0.25	0.02748	118.7
C(OOH) in carboxylic acids	267	1.950	0.25	0.02748	150.0
C(O), $–COO^-$, carboxylate ion	271	1.900	0.25	0.01000	150.0
O^-, $–COO^-$, carboxylate ion	272	1.700	0.25	0.01800	1.327
C(=O), acetone	280	2.050	0.25	0.02748	140.0
N, primary amines	900	1.700	0.25	0.02000	210.0

N, acetamide	237	1.870	0.25	0.01800	135.0
N, NMA	238	1.770	0.25	0.02000	190.0
NA (–N-H) in heterocycles	503	1.700	0.25	0.02000	167.0
NB in heterocycles	511	1.850	0.25	0.02000	167.0
S, thiols	200	2.050	0.25	0.07270	183.8
S, sulfides	202	2.150	0.25	0.04600	150.0

[a]Atomtypes used in this work, according to in implementation in BOSS, see Reference 15.

2.3.1 ***Hydration Energies of Benzene, Phenol and Phenoxide and the Phenol pKa Value***

Hydration energies of aromatic compounds and the calculation of pK_a of phenol required first developing parameters for the unsubstituted phenol (**Figure 7**). The natural course in developing aromatic parameters was undertaken by starting with benzene. The final FB hydration radii for the benzene carbon and hydrogen atomtypes were 2.050Å and 1.320Å, respectively. The values of the parameter a_0 for the aromatic carbon and hydrogen atoms were set at 0.02748 and 0.01034, and these were kept unchanged for all the further development of the carbon and hydrogen FB parameters to avoid over parameterization.



(a) (b)

Figure 7: Schematic description of a) phenol and b) phenoxide ion simulated with FB solvation model and OPLS-AA force field

The best values of the A^{LJ} parameters for the aromatic carbon and hydrogen atoms were found to be 161.5 and 10.00 kcal/mol· $Å^6$. The error in the calculated hydration energy of benzene (–0.866kcal/mol) as compared with the experimental data (–0.87kcal/mol)[19] was less than 0.01kcal/mol.

Then the FB hydration parameters for the –OH group in phenol were produced. The best performance was found with the phenol oxygen radius of 1.700Å. The parameter a_0 was equal to 0.02748 (the same as for the aromatic carbons), and the Lennard-Jones factor was 143.3 kcal/mol· $Å^6$. For the hydrogen, the values were 1.300Å, 0.01034 and 1.327. Moreover, this set of hydrogen parameters was found to be suitable for the other polar hydrogens as well, and the only difference in the aliphatic hydrogen was in a slight change in the radius. The error in the hydration energy of phenol was only 0.01kcal/mol, with the calculated value of –6.612 kcal/mol and the reference of –6.62 kcal/mol.[19]

The next step in our pK_a calculations required hydration parameters for the deprotonated phenol,

$C_6H_5O^-$, to be developed. We used the OPLS-AA atomtype 420 for the oxygen atom, with the Lennard-Jones parameters of σ = 3.15 Å and ε = 0.25 kcal/mol (the O^- in CH_3O^-). The hydration parameters for the aromatic ring remained unchanged, as only the hydration parameters of oxygen were fitted. The following strategy was adopted for the fitting of the parameters. The target hydration energy for the ion was chosen so as to lead to the experimental pK_a value of 9.98 pH units[17] of the unsubstituted phenol. This required the $C_6H_5O^-$ system to have a hydration energy of –73.02 kcal/mol with the reference ΔG(gas) of 349.0 kcal/mol.[21] The –73.02 kcal/mol solvation energy for the ion is consistent with the experimental range of –72 to –75 kcal/mol.[17] The target hydration energy with the O^- radius of 2.690 Å, a_0 = 0.01734 and the Lennard-Jones factor of 142.9 kcal/mol· $Å^6$ was obtained.

This lead to the phenol pK_a which was exactly same as the experimental value of 9.98 pH units.

2.3.2 *pK_a Values of the Substituted Phenols and Hydration Energies of Related Molecules*

The next molecules for which pK_a values were calculated contained phenol systems with methyl group at ortho- meta- and para positions (**Figure 8**). In order to demonstrate transferability of our hydration parameters and to avoid overparameterization, we decided to produce parameters for aliphatic carbon and hydrogen atoms first and then to use them in the methyl phenols as well in all the other compounds which contain alkyl groups without any further modifications.

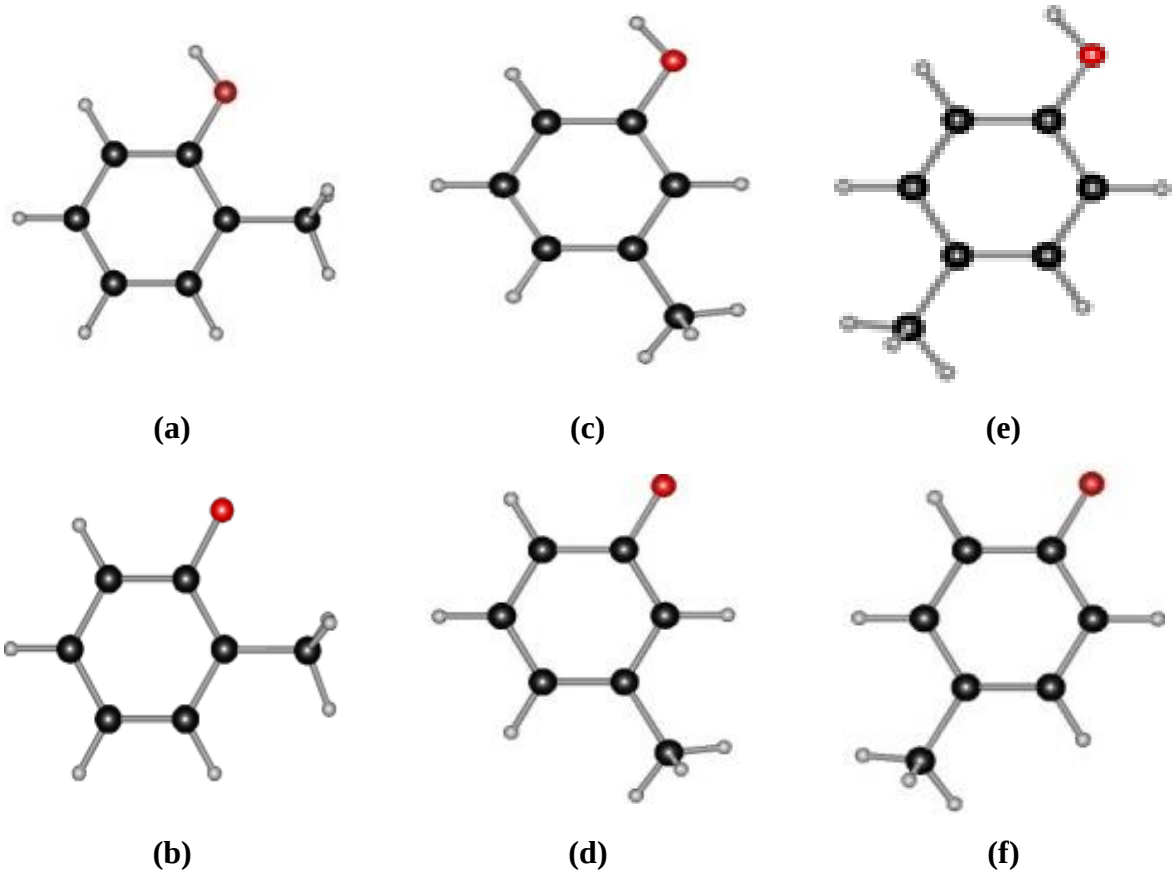


Figure 8: Schematic description of (a), (b) ortho-methylphenol and ortho-methylphenoxide ion, (c) (d) meta-methylphenol and meta-methylphenoxide (e) (f) para-methylphenol and para-methylphenoxide ion respectively simulated with FB solvation model and OPLS-AA force field

The series of species considered contained methane, ethane, propane, butane and iso-butane. The solvation parameters for all the aliphatic carbons were kept the same, and the same condition was observed for the aliphatic hydrogens. Moreover, we used the same aliphatic carbon and hydrogen hydration parameters in all cases when such a group was present, regardless of the chemical functionality of the remaining molecule (for example, in the $–CH_3$ groups in methyl phenol, NMA and methanol, in the $-C_2H_5$ group of ethanol and diethyl ether, the methyl group of the toluene molecule, etc.).

As has been mentioned earlier, the value of Δ was set to be the same for all the atomtypes

considered and equal to 0.25Å. The radius of the carbon atom was R = 1.9Å. In the OPLS formalism, the Lennard-Jones radius of an atom is $2^{1/6}$ times one half of the parameter which has a value of 3.5Å for the OPLS aliphatic carbon. Therefore, the Lennard-Jones radius for these atoms is 1.964Å. The solvation radius is fairly close to this number. While we did not specifically have a target of the Lennard-Jones radii as the solvation radii, it is worth noting that they were generally not too different. The value of a_0 for the aliphatic carbons in the FB model is equal to 0.02748 (the same as for the aromatic carbons and phenol oxygen), and the Lennard-Jones factor of 166.3 kcal/mol· $Å^6$. The corresponding hydrogen parameter values were 1.357 Å, 0.01034 and 1.327 kcal/mol· $Å^6$.

It can be seen from the data in **Table 3**, that the hydration energies for methane, ethane, propane, butane and iso-butane follow the general experimental trend. Moreover, we correctly reproduce the trend of hydration energy in iso-butane and n-butane, which is not always represented correctly by continuum solvation models.[19] The overall average error for the hydration energies of the aliphatic hydrocarbons was ca. 0.1 kcal/mol.

The pK_a values for the methyl phenols were in a very good agreement with the experimental data. The calculated values for the o-, m- and p-methyl phenols were found to be 10.02, 9.11 and 10.50 pH units, respectively, with their experimental counterparts being 10.29, 10.08 and 10.14 units. Therefore, the general trend of the m-methyl phenol being more acidic was followed (although our p-methyl phenol is somewhat more basic than it should be), and the overall error for these three acidity constants was 0.53 pH units, well within the range for which meaningful comparison with experiment. Moreover, it should be explicitly noted that no specific fitting for methyl phenols was carried out, with the hydration parameters taken directly from the

unsubstituted phenol calculations and the alkane hydration energy fitting. This attests to the robustness of our methodology and portability of the produced parameter values, which is especially important for the aliphatic groups which are present in a variety of compounds.

The next series of the compounds for which we calculated pK_a values were the o-, m-, and p-chlorophenols. In this case, the hydration parameters for the chlorine atom were fitted specifically to reproduce the experimental acidity constants of these substituted phenols (**Figure 9**).

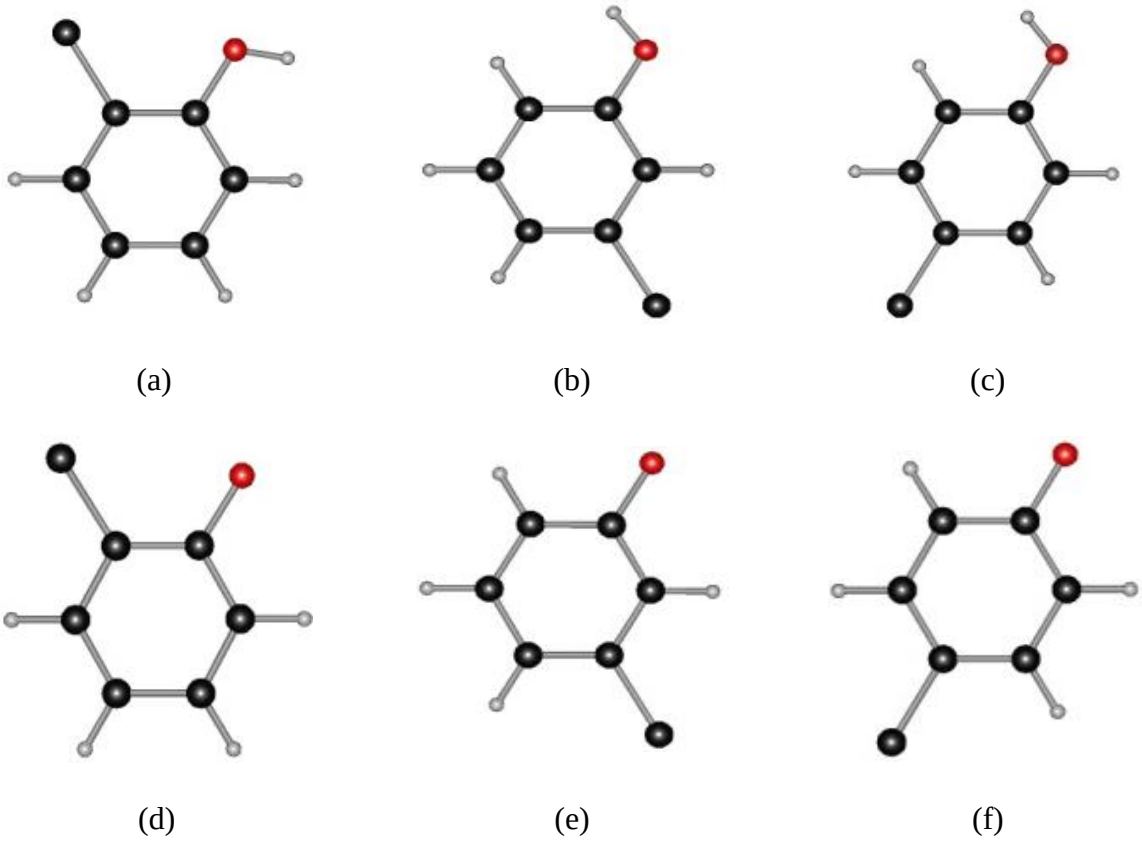


Figure 9: (a) ortho-cholorophenol (b) meta-chlorophenol (c) para-chlorophenol and their respective phenoxide ions (d)-(f) simulated with Fuzzy-Border continuum solvation model.

The parameter fitting has led to the following final values of the Cl^- hydration parameters: R =

1.905 Å, a_0 = 0.00808 and the A^{LJ} = 250.0 kcal/mol· $Å^6$. The final calculated pK_a values were 8.57, 8.51 and 10.07 pH units for the o-, m-, and p-chlorophenol, respectively. The experimental counterparts are 8.56, 9.02 and 9.38 units.[17] The average error is just under 0.40 pH units, and the calculated value of the acidity constant generally follows the experimental trend of growing from the ortho- to para-compound. In the case of m- and p-cyanophenols **(Figure 10)** we also did the substituent (-C≡N) hydration parameter fitting with the explicit goal of reproducing the experimental pKa values of these substituted phenols.

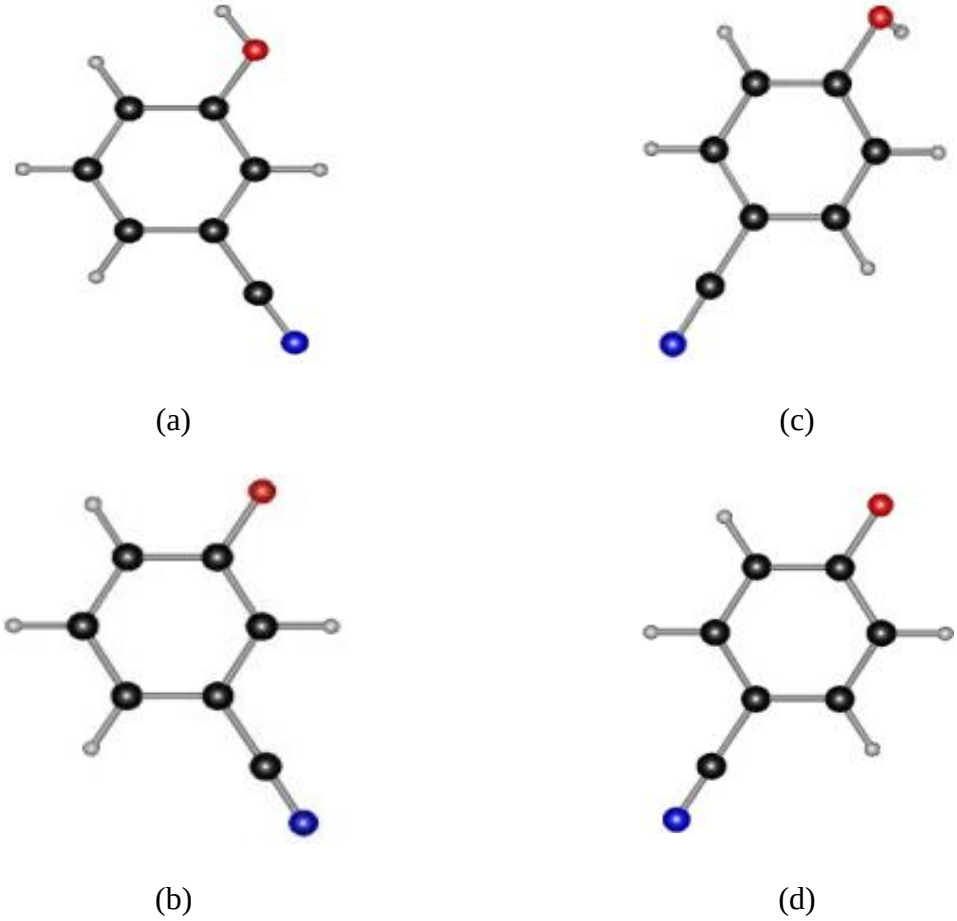


Figure 10: Schematic description of (a), (b) metacyanophenol, metacyanophenoxide ion and (c) (d) para-cyanophenol and para-cyanophenoxide ion respectively simulated with the FB continuum solvent model.

The final values of the parameters for the carbon and nitrogen atoms were R = 1.300 Å, a_0 =

0.00600 and the A^{LJ} = 100.0 kcal/mol· $Å^6$ and R = 1.825 Å, a_0 = 0.02100 and the A^{LJ} = 200.0 kcal/mol $Å^6$, respectively. The calculated values of the acidity constant for the m- and p-cyanophenols were 8.48 and 7.42 pH units (as can be seen from the data in **Table 1**). These values have the same relative order as the experimental numbers of 8.61 and 7.95 pH units[17] and the average error is 0.33 units.

The last set of the substituted phenols which we considered was the m- and p-nitrophenols (**Figure 11**).

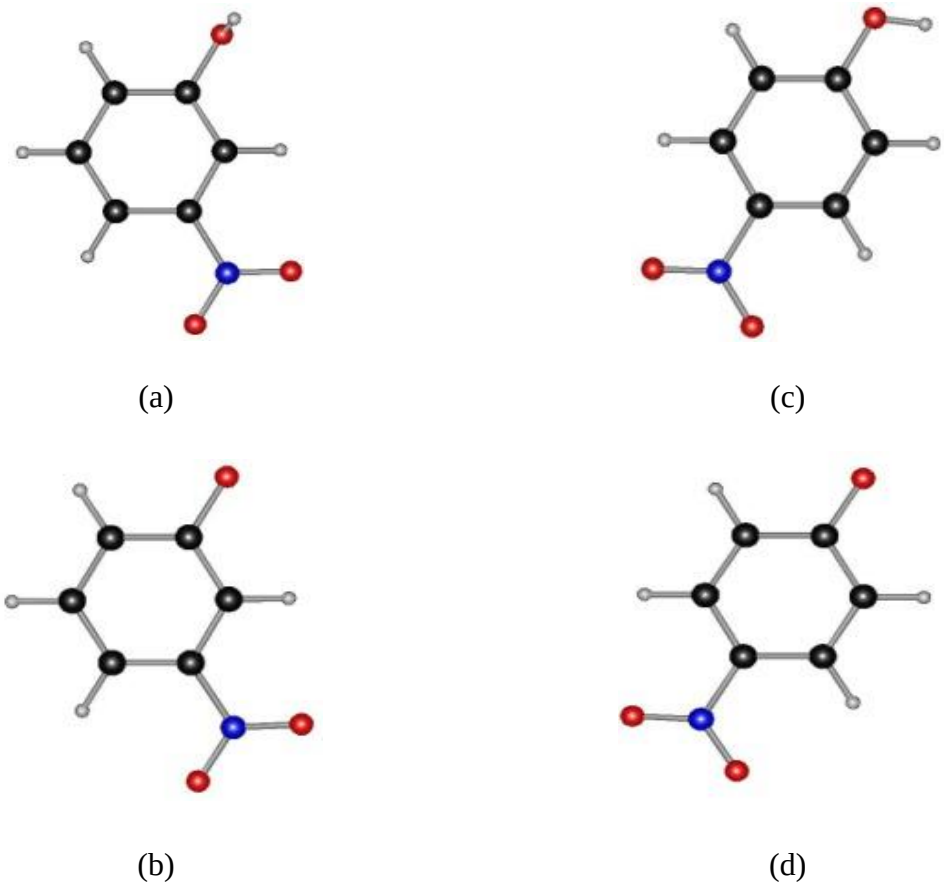


Figure 11: Schematic description of (a), (b) meta-nitrophenol and meta-nitrophenoxide (c) (d) para-nitrophenol and para-nitrophenoxide ion respectively simulated with parameterized FB model.

The experimentally measured pK_a values for m- and p-nitrophenol are 8.40 and 7.15 pH units

respectively.[17] The N and O hydration parameters for the nitro group was fitted to obtain the values of the acidity constants of 8.14 and 7.87 units. Therefore, the para- isomer is more acidic, in agreement with the experiment, and the average error is 0.49 pH units. The values of the solvation parameters for the nitrogen and oxygen atoms in the nitrogroup are R = 1.550 Å, a_0 = 0.00821 and the A^{LJ} = 110.0 kcal/mol· $Å^6$ and R = 1.370 Å, a_0 = 0.00900 and the A^{LJ} = 1.327 kcal/mol· $Å^6$, respectively.

The average error in the pK_a values of the substituted phenols, as calculated with the Fuzzy-Border model, is only 0.41 pH units. This is comparable with the best quantum error of 0.38 units for these molecules[17] and is better than some lower-level quantum mechanical data from the same reference. Of course, a direct comparison of these two sets of data would not be appropriate. A big advantage of the QM calculations is that they were performed without any specific refitting for the pK_a calculations, and many of our hydration parameters were fitted to reproduce the pK_a values (though even in these cases the relative order of pK_a values of phenols with substituents in different positions was still reasonably good, thus we can assume at least some transferability of the parameters in every single case). The values of the parameters R, a_0 and A^{LJ} were fitted for the following atomtypes: the O^- in phenoxyde, Cl atom, carbon and nitrogen of the –C≡N group, as well as nitrogen and oxygen atoms in $–NO_2$. The results have clearly demonstrated that our Fuzzy Border model can, with proper fitting of the parameters, reproduce magnitudes and relative orders of the absolute acidity constants rather well.

Moreover, listed in **Table 5** are the pK_a values for the unsubstituted phenol and para-chlorophenol, as calculated with the Fuzzy –Border continuum solvation model in this work and calculated with the explicit aqueous solvation.[16]

Table 5: pK_a Values for Phenol and *p*-chlorophenol

System	pK_a			
	Modified OPLS[a], explicit solvent	**OPLS/FB (this work)**	**QM[b]**	**Experiment[b]**
Phenol	4.50	9.98	9.88/10.23	9.98
p-chlorophenol	–1.06	10.07	9.84/9.77	9.38
Average Error	6.90	0.35	0.28/0.32	–

[a]Reference 16. [b]Reference 17.

Both sets of the results were obtained for the OPLS-AA solute model. It can be easily seen that the explicit solvent simulations produce a much greater average error of 6.90 pH units, while the FB model permits to reduce the deviation to the average of only 0.35. Although, once again, this is achieved by the specific pKa-targeted fitting of the solvation parameters, this is an additional reason to believe that the Fuzzy-Border model will be a useful tool in modeling of hydrated molecular systems described by the generally very successful OPLS-AA force field.

The gas-phase deprotonation energies for the phenol and substituted phenol molecules used in this work are given in **Table 6**.

Table 6: Gas-Phase Deprotonation Energies for the Phenol Systems and Propanoic and Butanoic Acids, in kcal/mol

Molecule	Deprotonation Energy	Source
Phenol	349.0	Reference 21
o-chlorophenol	337.1	Reference 17
m-chlorophenol	335.3	Reference 17
p-chlorophenol	336.5	Reference 17
m-cyanophenol	329.0	Reference 17
p-cyanophenol	325.5	Reference 17
m-nitrophenol	327.6	Reference 17
p-nitrophenol	324.8	References 17 and 21
o-methylphenol	349.95	Reference 25
m-methylphenol	350.75	Reference 25
p-methylphenol	352.13	References 21 and 25
propanoic acid	347.20	Reference 26
butanoic acid	347.26	Reference 27

2.3.3. ***Hydration Energies for the Non-Phenol Compounds***

Finally, the results of producing parameters for other compounds not related to the above pK_a calculations for the substituted phenols are reported here.

The four alkenes, ethylene, 1-propene, 1-butene and butadiene were simulated. The hydration

parameters for the hydrogen atoms were adopted directly from the alkanes and not refitted in any way. The sp^2 carbon parameters were slightly different from those for the alkane carbons, with R = 2.015 Å, a_0 = 0.02748 and the A^{LJ} = 196.5. The resulting hydration energies are listed in **Table 3**. It can be seen that the general agreement with the experimental data is good, the average error is only 0.14 kcal/mol, and the only mismatch in the order of these calculated solvation energies is observed for the ethylene, which is less hydrophobic than 1-propene (1.056 kcal/mol vs. 1.512 kcal/mol), while their experimental solvation energies are the same within 0.01 kcal/mol.[19]

The hydration parameters for the following alkynes: acetylene, 1-propyne and 1-butyne were also produced. In this case, solvation parameters for both the sp-hybridized carbons and for the hydrogens were refitted. The values for the carbon and hydrogen atoms are, respectively: R = 2.050 Å, a_0 = 0.01700 and the A^{LJ} = 250.0 and R = 1.500 Å, a_0 = 0.02000 and the A^{LJ} = 1.327. As can be seen from the data in Table 3, the average error in the solvation energy for these compounds is 0.16 kcal/mol.

Methanol and ethanol were simulated to represent aliphatic mono-alcohols. As shown in **Table 3**, the average error is only ca. 0.015 kcal/mol, with the trend of methanol being solvated better by about 0.1kcal/mol reproduced correctly. The solvation radius of the oxygen atom was 1.735Å. This correlates well with the OPLS-AA Lennard-Jones radius of 1.751Å for this atomtype.

As was mentioned above, all the polar hydrogen atoms have the same set of the first-order FB hydration parameters for OPLS solutes. Such atoms are assigned a van-der-Waals radius of zero in the OPLS-AA force field. The methanol and ethanol OH hydrogen was given a hydration radius of 1.3Å in our model.

The next class of compound that we considered was that of ketones. We simulated hydration for acetone, 2-butanone, 2-pentanone and 3-pentanone. As with all the other cases, the hydration parameters for analogous atoms in these systems were not dependent on the specific substance (thus, for example, the oxygen in acetone had exactly the same set of FB parameters as the oxygen atom in 3-pentanone). Moreover, the oxygen parameters were rather similar to those of the alcohols. The hydration radius of the ketone oxygen was slightly larger than that of the aliphatic alcohols (1.75 Å vs. 1.735 Å), otherwise the parameters were the same. The overall average error in the ketone hydration energies, as compared with their experimental counterparts,19 was just 0.065 kcal/mol. And the trend of the slight energy magnitude reduction from acetone (calculated solvation energy of –3.900 kcal/mol and experimental result of –3.85 kcal/mol) to 3-pentanone (calculated and experimental energies of –3.537 kcal/mol and –3.41 kcal/mol, respectively) was reproduced successfully. Moreover, the success with these compounds has demonstrated a good level of transferability of both our ketone and aliphatic hydration parameter values.

The dimethyl- and diethyl-ethers were considered next. As can be seen from the table, the errors were somewhat greater in this case, though still in the acceptable range, with the both hydration energies slightly overestimated, with an average error of ca. 0.15 kcal/mol. At the same time, the oxygen solvation parameters were taken directly from the alcohol oxygen, and the methyl and ethyl parameters were the same as produced in fitting hydration energies of the saturated hydrocarbons. This is an added proof of the robustness of the method. Moreover, the trend in reduction of the magnitude of the hydration energy with transition from the dimethyl- to diethyl-ether was reproduced correctly.

The next series of compounds in **Table 3** are amines – methyl amine, ethyl amine, n-propyl amine, n-butyl amine, dimethyl amine and diethylamine. The solvation parameters for the polar hydrogen atoms were the same as for the other polar hydrogens, as discussed above. Only the nitrogen solvation parameters were fitted. The resulting values are R = 1.650 Å, a_0 = 0.00900 and the A^{LJ} = 262.0 As shown in Table 3, the average error in the solvation energies is 0.12 kcal/mol, and the trend of the hydration energy magnitude decreasing in this series from methyl amine to diethyl amine is generally reproduced. Ammonia nitrogen parameters were fitted separately, with the final values of R = 1.357 Å, a_0 = 0.00650 and the A^{LJ} = 122.9. Deviation of the calculated hydration energy for ammonia (–4.301 kcal/mol) from the experimental value of –4.310 kcal/mol[19] is about 0.01 kcal/mol.

The benzene parameters described above were employed together with the parameters for alkanes to calculated the hydration energy of toluene, and found it within an error of 0.04kcal/mol from experiment. This result was obtained with no parameter refitting.

Acetamide and NMA are compounds which are important in their own rite and also have a great significance as building blocks of proteins. This is especially true for the NMA, which essentially represents the repeating unit in the protein and peptide backbones. Fitting parameters for amides often represents a challenge, and it is not unusual to have different parameter sets for the acetamide and NMA cases.[22] Our solvation parameters were the same for similar atoms in these two amides except for the nitrogens (which also have different atomtype designations in the OPLS-AA). The methyl groups had the standard FB solvation parameters for both the carbon and hydrogen atoms. The hydration parameters for the polar hydrogens were also preserved at

the same values as for the previously discussed molecule. We have discovered that the aliphatic version of the carbon hydration parameters and the aliphatic alcohol values for the oxygen work well for the amide C=O group. The only adjusted parameters were those of the amide nitrogen atoms. In both the cases, the radii were greater than those for the amines (1.770Å and 1.870Å for the NMA and acetamide nitrogens, respectively, compared to the 1.650Å for the R-NH_2). As shown in **Table 3**, the average error in the hydration energy of the amides was 0.05 kcal/mol.

In order to include examples of heterocyclic molecules and to have parameters for the histidine and tryptophan residues in the potential future development of a FB solvation parameter set for peptides and proteins, 4-methyl-imidazole and 3-methyl-indole were fitted. Once again, a good transferability of the solvation parameters was observed. All the aromatic carbon and hydrogen atomtypes retained the same values as those in benzene and toluene. The same is true with relation to the aliphatic carbons and hydrogens in these two molecules and all the other saturated hydrocarbon groups. The polar hydrogens had our standard FB polar hydrogen hydration parameter values. Only the nitrogen parameter values were refitted, and the parameters for the N(H) in both these molecules were the same. The average error in hydration energy for the two compounds was less than 0.01 kcal/mol.

Both the thiol and sulfide molecules, represented by the methane- and ethane-thiol and dimethyl- and diethyl-sulfides, respectively, used the standard hydration parameters for the methyl and ethyl groups, but the sulfur parameters were different for the two groups of compounds. The solvation radii of the both S atomtypes were greater than that of the aliphatic carbon. Values of these parameters are listed in Table 4. The average error in the hydration energy was only about 0.01 kcal/mol.

Finally, FB solvation parameters for two acids, acetic (CH_3COOH) and propanoic (C_2H_5COOH) were produced. The aliphatic tails had the standard FB parameter values for both carbon and hydrogen atoms. In the –COOH groups, the hydration parameters for the –OH part were transferred from the FB aliphatic alcohols without change, and the =O atom had parameters adopted directly from the acetone oxygen value. The only adjusted parameter set was that of the carbon, with the radius slightly greater than that of the aliphatic carbon and the Lennard-Jones parameter reduced by ca. 10%. As can be seen from the data in **table 3**, the resulting average error in the hydration energies for the acid molecules was only about 0.02 kcal/mol. Once again, this proves, albeit still anecdotal, of the generally good transferability of the FB parameters.

2.3.4 *Comparison of Fuzzy-Border Hydration Energies with Poisson-Boltzmann and Generalized Born Results*

The Fuzzy-Border (FB) results can be compared with those obtained with the Poisson-Boltzmann (PBF) and Generalized Born (SGB) formalisms. All the energies are listed in **Table 3**. It can be observed that the average error in the FB solvation energies (0.076 kcal/mol) is significantly smaller than those of the PBF and SGB models (0.527 kcal/mol and 0.639 kcal/mol, respectively).

The Poisson-Boltzmann continuum solvation gives a somewhat better result than the Generalized Born one, as can be expected from the more physically grounded formalism of the former. It should be emphasized very strongly that the smaller average error afforded by the Fuzzy-Border continuum solvent does not mean that the FB methodology is intrinsically better than the PBF and SGB ones. The latter models implemented in IMPACT have been parameterized for a larger set of molecules, and thus a somewhat greater error is natural to observe. What we do conclude

though is that out technique is robust enough, and, even if extending the currently available parameter sets to more classes of compounds does reduce the overall average accuracy, it is still likely to stay within the respectable range observed for the Poisson-Boltzmann and Generalized Born methods.

To further compare the FB, PBF and SGB results, the values of the electrostatic component of the hydration energy computed with these techniques are listed in **Table 7**.

Table 7: Calculated Values of Electrostatic Components of Hydration Energy for Other Compounds, in kcal/mol

Compound	Electrostatic Hydration Energy, calculated		
	Fuzzy-Border	**Poisson-Boltzmann**	**Generalized Born**
CH_4	–0.040	–0.043	–0.004
C_2H_6	–0.144	–0.098	–0.003
C_3H_8	–0.228	–0.123	0.017
C_4H_{10}	–0.288	–0.145	0.086
iso-C_4H_{10}	–0.299	–0.120	0.081
C_2H_4	–0.945	–0.642	–0.664
1-propene	–0.511	–0.607	–0.635
1-butene	–0.711	–0.612	–0.553
Butadiene	–1.145	–1.244	–1.199
Acetylene	–1.202	0.000	0.000
1-propyne	–0.853	–0.080	–0.084
1-butyne	–0.821	–0.249	–0.229
CH_3OH	–5.575	–6.473	–7.269
C_2H_5OH	–5.611	–7.193	–5.300
CH_3COCH_3	–4.996	–4.988	–5.224
2-butanone	–4.914	–4.942	–5.273
2-pentanone	–4.837	–4.750	–4.857
3-pentanone	–4.887	–4.776	–5.592
CH_3OCH_3	–2.776	–2.951	–3.106
$C_2H_5OC_2H_5$	–2.696	–2.874	–3.186
methyl amine	–2.349	–6.316	–3.309
ethyl amine	–2.439	–5.477	–3.231
n-propyl amine	–2.687	–5.285	–2.737

n-butyl amine	–2.560	–6.297	–2.721
dimethyl amine	–2.868	–5.082	–3.609
Diethylamine	–2.569	–4.536	–3.088
Ammonia	–3.078	–7.224	–2.158
C_6H_6	–2.108	–2.488	–2.508
Toluene	–2.551	–2.970	–2.989
C_6H_5OH	–5.52	–7.749	–7.521
CH_3CONH_2	–8.949	–10.175	–10.791
NMA	–7.650	–8.135	–9.161
4-methyl-imidazole	–7.695	–10.427	–10.989
3-methyl-indole	–4.180	–6.880	–7.529
CH_3SH	–4.902	–2.642	–2.303
C_2H_5SH	–4.667	–2.382	–2.069
CH_3SCH_3	–6.253	–3.005	–2.790
$C_2H_5SC_2H_5$	–5.985	–2.889	–2.549
CH_3COOH	–5.294	–8.346	–7.470
C_2H_5COOH	–5.106	–8.155	–7.975
Average signed difference from FB	–	–0.662	–0.065

From the comparison of electrostatic component of hydration energy calculated by FB, Poisson Boltzmann and Born methods in **table 7**, the following conclusions can be drawn. Generally speaking, the Poisson-Boltzmann electrostatic energy tends to be somewhat more negative than the Fuzzy-Border one, with the average signed difference of -0.662 kcal/mol. At the same time,

the Generalized Born electrostatic component is on average about the same as the Fuzzy-Border one (only 0.065 kcal/mol more negative). At the same time, deviations for particular components can be noticeably greater, and even the PBF model does not yield uniformly more negative results.

For example, the PBF electrostatic component for the compounds containing sulfur can be about two times smaller than that produced with the Fuzzy-Border model. At the same time, all the solutes were modeled with the same OPLS-AA force field. Therefore, we conclude that the differences in the electrostatic hydration energies given by the three continuum solvent models are not representing greater or smaller deviations from a physically correct set of results but rather are simply following from differences in fitting techniques employed to produce overall hydration energies, and the quality of these overall energies is adequately good for all the three methods (as can be seen from the results in **Table 3**).

2.3.5 Absolute Acidity Constants for Propanoic and Butanoic Acids

Finally, hydration parameters were fitted for reproducing absolute pK_a values of propanoic and butanoic acids. These compounds were chosen as relevant in calculating pK_a shifts of aspartic and glutamic acid residues of proteins (such as those calculated in Reference 1). Relevant results are presented in **Table 8**. The only two atom types for which solvation parameters were fitted are the -COO^- carbon and oxygen in the deprotonated forms of the acids.

The hydration energies of the protonated propanoic and butanoic acids as calculated with POSSIM/FB were equal to –6.443 kcal/mol and –6.197 kcal/mol, respectively (as can be seen from **Table 8**).

Table 8: Data and Results from pK_a Calculations for Propanoic and Butanoic Acids (Energies are in kcal/mol)

	Hydration Energy, Protonated Form		**Hydration Energy, Deprotonated Form**		**pK_a**	
System	**FB[a]**	**Reference[b]**	**FB[a]**	**Reference**	**FB[a]**	**Reference[c]**
propanoic acid	–6.443	–6.480	–77.964	–79.100[b]	4.90	4.87
butanoic acid	–6.197	–6.360	–78.112		4.66	4.83

[a]This work. [b]Reference 19. [c]Reference 1.

The hydration energies of propanoic and butanoic acid deviate by an average of only 0.1 kcal/mol from the experimental numbers. These data were obtained with the same parameters as the hydration energies for CH_3COOH and C_2H_5COOH, no refitting was done. The overall values of the acidity constants for these acid were 4.90 and 4.66 pH unites (compared to the reference 4.87 and 4.83 units). The agreement is good, but it was achieved with the direct fitting of the parameters for the deprotonated acids, therefore the ultimate quality of this result will have to be tested by using the same parameters with other systems in future work.

2.4. Conclusions

The first order Fuzzy-Border (FB) solvation model was applied with the OPLS-AA solute to calculate absolute pK_a values of several substituted phenols and hydration energies of a number of small molecules. The compounds were chosen to represent several classes which are important not only in themselves but also as building blocks in protein simulations. The FB model was implemented in a modified version of our POSSIM software suite for molecular simulations.

The overall average unsigned error in the calculated acidity constant values was equal to 0.41 pH units and the average error in the solvation energies was ca. 0.08 kcal/mol. While these results were achieved with fitting of the hydration parameters to the specific pK_a and hydration energy targets, the results still prove that the physical and numerical basis of the model is robust enough to permit such a good level of the performance, and the model can be expected to work well in further simulations of organic and biophysical systems. The parameter transferability also seems to be good.

The features of the FB model include utilizing a fixed three-dimensional grid for finding continuum solvation energy and an approximation to the Poisson-Boltzmann formalism which is designed to speed up the calculations and, more importantly, to remove the unfortunate potential problems which accompany the accuracy of the complete self-consistency of the standard Poisson-Boltzmann method and thus to avoid any issues related to convergence of the solvation energy. At the same time, the overall FB technique is still closer to the exact electrostatic model than the Generalized Born approximation.

References

1. MacDermaid, C. M.; Kaminski, G. A., Electrostatic polarization is crucial for reproducing pKa shifts of carboxylic residues in turkey ovomucoid third domain. The Journal of Physical Chemistry B 2007, 111 (30), 9036-9044.
2. Click, T. H.; Kaminski, G. A., Reproducing Basic p K a Values for Turkey Ovomucoid Third Domain Using a Polarizable Force Field. The Journal of Physical Chemistry B 2009, 113 (22), 7844-7850.
3. Click, T. H.; Ponomarev, S. Y.; Kaminski, G. A., Importance of electrostatic polarizability in calculating cysteine acidity constants and copper (I) binding energy of Bacillus subtilis CopZ. Journal of computational chemistry 2012, 33 (11), 1142-1151.
4. Ghosh, A.; Rapp, C. S.; Friesner, R. A., Generalized Born model based on a surface integral formulation. The Journal of Physical Chemistry B 1998, 102 (52), 10983-10990.
5. Bruccoleri, R. E.; Novotny, J.; Davis, M. E.; Sharp, K. A., Finite difference Poisson-Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. Journal of computational chemistry 1997, 18 (2), 268-276.
6. Grant, J. A.; Pickup, B. T.; Nicholls, A., A smooth permittivity function for Poisson–Boltzmann solvation methods. Journal of computational chemistry 2001, 22 (6), 608-640.
7. Kaminski, G. A.; Friesner, R. A.; Zhou, R., A computationally inexpensive modification of the point dipole electrostatic polarization model for molecular simulations. Journal of computational chemistry 2003, 24 (3), 267-276.
8. (a) Kaminski, G. A.; Ponomarev, S. Y.; Liu, A. B., Polarizable Simulations with Second-Order Interaction Model□ Force Field and Software for Fast Polarizable Calculations: Parameters for Small Model Systems and Free Energy Calculations. Journal of chemical theory and computation 2009, 5 (11), 2935-2943. (b) Impact version 3.0, Schrödinger, LLC, Portland, OR 2003 (c) Impact version 3.5, Schrödinger, LLC, Portland, OR 2004.
9. Barone, V.; Cossi, M.; Tomasi, J., A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. The Journal of chemical physics 1997, 107 (8), 3210-3221.
10. (a) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C., Refinement and parametrization of

COSMO-RS. The Journal of Physical Chemistry A 1998, 102 (26), 5074-5085. (b) Klamt, A., The COSMO and COSMO-RS solvation models. Wiley Interdisciplinary Reviews: Computational Molecular Science 2011, 1 (5), 699-709.

11. Cramer, C. J., and D. G. Truhlar. "In Trends and Perspectives in Modern Computational Science." Maroulis, G (**2006**): 112-140.
12. (a) Bruccoleri, R. E.; Novotny, J.; Davis, M. E.; Sharp, K. A., Finite difference Poisson-Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. Journal of computational chemistry 1997, 18 (2), 268-276. (b) Grant, J. A.; Pickup, B. T.; Nicholls, A., A smooth permittivity function for Poisson–Boltzmann solvation methods. Journal of computational chemistry 2001, 22 (6), 608-640.
13. Im, W.; Lee, M. S.; Brooks, C. L., Generalized born model with a simple smoothing function. Journal of computational chemistry 2003, 24 (14), 1691-1702.
14. (a) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. Journal of the American Chemical Society 1996, 118 (45), 11225-11236. (b) Jorgensen, W. L.; Tirado-Rives, J., Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. Proceedings of the National Academy of Sciences of the United States of America 2005, 102 (19), 6665-6670.
15. Can, T.; Chen, C.-I.; Wang, Y.-F., Efficient molecular surface generation using level-set methods. Journal of Molecular Graphics and Modelling 2006, 25 (4), 442-454.
16. Gallicchio, E.; Zhang, L. Y.; Levy, R. M., The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. Journal of computational chemistry 2002, 23 (5), 517-529.
17. Kaminski, G. A., Accurate prediction of absolute acidity constants in water with a polarizable force field: Substituted phenols, methanol, and imidazole. The Journal of Physical Chemistry B 2005, 109 (12), 5884-5890.
18. Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C., Absolute p K a determinations for substituted phenols. Journal of the American Chemical Society 2002, 124 (22), 6421-6427.
19. Ponomarev, S. Y.; Kaminski, G. A., Polarizable simulations with second-order interaction

model (POSSIM) force field: developing parameters for alanine peptides and protein backbone. Journal of chemical theory and computation 2011, 7 (5), 1415-1427.

20. Gallicchio, E.; Zhang, L. Y.; Levy, R. M., The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. Journal of computational chemistry 2002, 23 (5), 517-529.
21. Weast, R.; Astle, M., CRC Handbook of Chemistry and Physics (CRC, Boca Raton, FL), p. F216-F219 1979.
22. (a) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L., Free energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model. The Journal of Physical Chemistry 1994, 98 (49), 13077-13082. (b) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B., New model for calculation of solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. The Journal of Physical Chemistry 1996, 100 (28), 11775-11788.
23. Lopes, P. E.; Lamoureux, G.; Mackerell, A. D., Polarizable empirical force field for nitrogen-containing heteroaromatic compounds based on the classical Drude oscillator. Journal of computational chemistry 2009, 30 (12), 1821-1838.
24. Klicic, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C., Accurate prediction of acidity constants in aqueous solution via density functional theory and self-consistent reaction field methods. The Journal of Physical Chemistry A 2002, 106 (7), 1327-1335.
25. Fujio, M.; McIver Jr, R.; Taft, R. W., Effects of the acidities of phenols from specific substituent-solvent interactions. Inherent substituent parameters from gas-phase acidities. Journal of the American Chemical Society 1981, 103 (14), 4017-4029.
26. (a) S. M. Bachrach, Eds. Computational Organic Chemistry, Available at http://comporgchem.com/blog/?p-27#cyst (accessed May 10, **2012**); (b) Lemmon, E.; McLinden, M.; Friend, D.; Linstrom, P.; Mallard, W., NIST Chemistry WebBook. NIST Standard Reference Database 2005, 69.
27. Schmidt am Busch, M.; Knapp, E. W., Accurate pKa determination for a heterogeneous group of organic molecules. ChemPhysChem 2004, 5 (10), 1513-1522.
28. Maple, J. R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A., A polarizable force field and continuum solvation methodology for modeling of

protein-ligand interactions. Journal of Chemical Theory and Computation 2005, 1 (4), 694-715.

CHAPTER 3

Developing and parameterizing first-order Fuzzy-Border (FB) continuum solvation model with Polarizable Simulations Second-Order Interaction Model (POSSIM) force field and computing pK_a values of carboxylic and basic residues of OMTKY3 protein

Portion of this work was also presented at the following meeting:

"Computing acidity constants for turkey ovomucoid third domain protein using POSSIM (polarizable simulations second-order interaction model) force field and Fuzzy-Border (FB) continuum solvation model."

Ity Sharma and George A Kaminski, 248th American Chemical Society National Meeting & Exposition, San Francisco, California, 2014

3.1 Introduction

Structure, stability, solubility, and catalytic functioning of a protein are closely related to proton transfer reactions of its ionizable residues and, hence, to their pK_a values. These reactions are important since the locations of acid and basic sites on peptides affect the hydrogen bonding in the biomolecules and therefore impact their biological activity. The protonation states of biomolecules depend on the relative proton affinities of different sites and determining pK_a values is one way of gaining information on the protonation sites of the biomolecules.

Evaluation of pK_a values of protein residues are of particular interest in pharmaceutical applications and designing more robust industrial enzymes that are stable over a wide pH ranges. The pK_a values of ionizable functional groups such as –COOH, -SH, phenol, $-NH3^+$, imidazolium and guanidinium are significantly affected by neighboring residues in the protein. The experimental determination of pK_a for the above mentioned ionizable groups is not always straightforward. It is usually obtained by measuring spectroscopic properties of the residue as a function of pH. The acidity constants of amino acids with ionizable groups such as -SH and phenolic-OH (cysteine and tyrosine) can be measured using UV/vis spectroscopy.[1] For other amino acids, NMR is used to obtain individual pK_a.[2] However, assignment of NMR chemical shifts in proteins is not a trivial task. Therefore, accurate and rapid prediction of acidity constant shifts of individual protein residues is very important in computational protein research. However, the intrinsically many-body interactions in the protein residue makes it difficult for having one best standard technique to calculate these values. The protein conformation and the local environment of the residue in the protein affect the pK_a shifts of the residues.

There are numerous computational approaches to calculate pK_a values of proteins and protein residues within different proteins.[3-27] These can be broadly classified into three major methods. The first method comprises ab initio and DFT quantum mechanics (QM) to treat the residue or the part of protein containing the titrable group and the aqueous solvent as a dielectric continuum.[12, 16-18] This method is limited by the size of the protein that can be simulated quantum mechanically, the level of the theory used for quantum calculations and the parameterization of continuum solvation model to be employed with the quantum model used. These problems can be overcome using combined quantum mechanical/molecular mechanical (QM/MM) methods but this solution leads to the need to parameterize the solvation model, the correct choice of QM and MM regions and careful modeling of QM/MM coupling.

Second, pK_a values and shifts can be evaluated by employing a large database and fitting a linear free energy relationship.[22] This is useful only for cases in which the values are close to the fitting data set, and the issue of choosing the optimal way to fit the linear free energy arises. Finally, empirical force fields using explicit[11, 13, 25] or continuum solvent models[2-9, 17-21, 23, 25, 26] can be used to calculate pK_a shifts in proteins. The solvation energy and the energy of interaction of the residue with its environment can be calculated by a number of techniques. In some cases, the "effective dielectric constant" approach is used for simulating the protein. But these methods do not always work well, and the value of the dielectric constant for the interior of the protein varies significantly, sometimes reaching ca. 20, [5, 7, 8, 21, 23, 28] which raises the question of its physical meaning and undermines the ability of the method to be adequately applicable to a wide variety of cases. In addition, these calculations include electrostatic interactions of residues separated by large distances whereas experimental and theoretical evidence suggests immediate environment

of ionizable residue is most (and sometimes exclusively) important in calculating the values of acidity constants.[18, 29] The standard empirical fixed charge force fields like AMBER and CHARMM are widely used for calculating pK_a of protein residues. These force fields treat the polarization of the solute molecules in the mean-field manner and lack the explicit polarization term. Simonson *et al*[25] did molecular dynamics simulations to compute pK_a shifts of two aspartic acid residues (Asp20 and Asp26) in thioredoxin and Asp14 in ribonuclease A using AMBER and CHARMM force fields and explicit water and continuum solvation model. The average error using the explicit solvent model in these residues was 1.91 and 2.64 pH units respectively with calculated 3.2-4.5 unit maximum deviation. These results are not different from the pK_a of aspartic acid residues (Asp7 and Asp27) in OMTKY3 protein presented in reference 19 using the fixed-charge force field. The explicit polarization treatment absent in these calculations is necessary to account for all the protein electrostatic interactions and hence accurate determination of pK_a shifts of the residues.[25] The pK_a values for Asp residues in thioredoxin were also calculated with continuum solvation model with an average error of only 1.10-0.81pH units which cannot be considered as good as the results using the Polarizable Force Field (PFF) and Poisson-Boltzmann continuum solvation model (PBF) in reference 19.

Alternatively the polarization is accounted for in the force fields by incorporating conformation dependent charges for proteins for more accurate pK_a calculations. In one such pK_a calculation the authors use the force field with polarized protein-specific charges (PPC) for molecular dynamics simulations.[30] In PPC procedure, protein sub system is simulated using quantum mechanics to derive charges by fitting to the electrostatic potential. The error in the pK_a of Asp26 and Asp20 using the above methodology was 0.15 pH unit and 0.73 pH unit respectively. The

PPC charges producing accurate results use fixed charges specific for protein structure and environment, thus essentially providing tabulated polarization charge adjustment, while explicit treatment of polarization achieves the same goal automatically based on physical equations.

The average error in the pK_a of Asp26 and Asp20 calculated using PCC charges were also compared to the fixed-charge AMBER force field results with an error of 3.15 pH unit and 1.15 pH respectively. This show thermodynamic sampling in itself, as in reference 25, is not sufficient for reproducing pK_a of the protein residues but the adjustable electrostatics plays key role in such calculations.

Continuum solvation models have been effectively used to simulate the aqueous environment around the solute as they are relatively computationally inexpensive and do not require large number of water molecules to be present explicitly. The most popular and accurate implicit solvation model is the Poisson Boltzmann (PB) one based on the Poisson-Boltzmann equation. The electrostatic potential is usually computed by integration over the solute-solvent interface. Both the electrostatic field and the surface density affect each other resulting in self-consistent equations similar to the electrostatic polarization equation.[31] A further approximation of essentially the same model is known as the Generalized Born (GB) solvation technique.

The solution of the PB and GB model relies on solving the corresponding equations on numerical grid. This results in noise in the solvation energy due to grid rebuilding with the movements in the solute. Also, the self-consistent convergence in solving the PB equations is sometimes difficult to achieve on a numerical grid with arbitrary parameter values. The above

problems can be overcome by using a fixed equally spaced grid independent of the solute coordinates[32] and smoothing, antialiasing and proper choice of grid parameters.[32, 33]

Our Fuzzy-Border continuum solvation model[34] is capable of addressing these issues. It has two-fold distinguishing features. First, it uses fixed-position equally spaced 3-D grid in which coordinates of the solute are independent of the positions of nodes of the grid. Second, an approximation to the full scale Poisson-Boltzmann procedure was employed similar to our previously developed fast second-order polarization technique for solutes and explicit solutions and pure liquids.[35, 36] This methodology converts the self-consistent equations in the PB (or any full-scale polarization technique) model into an analytical one but is closer in reproducing many-body interactions to the PB than to the Generalized Born one.

In this chapter, the results of previously modified Fuzzy-Border (FB) continuous solvent model are introduced, its parameterization for the polarizable POSSIM force field and calculation of pK_a shifts of residues in OMTYK3. The turkey ovomucoid third domain (OMTKY3) protein was chosen since this has been extensively studied experimentally and computationally and its pK_a shifts have been measured and calculated.[37, 38] It is a 56-residue protein with five acidic (Asp7, Glu10, Glu19, Asp27 and Glu43) and six basic (Lys13, Arg21, Lys29, Lys34, His52 and Lys55) ionizable residues. The importance of including many body electrostatic polarizations explicitly in empirical force fields has already been demonstrated in calculations such as pK_a shifts, accurate determination of dimerization energies, modeling of sugar protein binding[39] and ion binding by proteins and small molecules.

The force field used for representing the solutes is POSSIM (Polarizable Simulations Second Order Interaction Model). It is a combination of inducible point dipole method and the fast second-order approach that has been successfully used to increase the speed of simulations by ca. an order of magnitude without any loss in accuracy.[36,40,41] The hydration energies, including those used in obtaining the pK_a values, are assessed with the modified Fuzzy-Border continuum solvation model. The FB approach has been used previously to compute hydration energy of small molecules and acidity constants of substituted phenols, but in that work the fixed charge OPLS-AA force field was employed for the solutes as discussed in reference34. The fixed-position grid points in the Fuzzy-Border limit the noise and the first order approximation for truncating the solvent polarization energy calculations is consistent with the second-order polarization scheme for the solutes and can also be used with any other formalism for the explicitly simulated molecules. Briefly, the modifications of the FB technique introduced is included, in addition to parameterizing it for the polarizable POSSIM force field to be used for the solutes, changing some general parameter values (for example, the spacing between the grid points) and adaptation of the multiple-marching level-set technique[42] modified for generation of the solvent-accessible grid points list suitable for FB calculations. The physical level of the true solvent energy approximation in our Fuzzy-Border methodology lies between Generalized Born and Poisson-Boltzmann solvation.

3.2 Methods

3.2.1 *Polarizable Simulations Second-order Interaction Model*

- **Formulation of POSSIM force field**

Polarizable Simulations Second-Order Interaction Model (POSSIM) force field was developed and successfully applied by Kaminski *et al* for studying biomolecular processes[36,40,41]. The polarizable POSSIM force field has an additional electrostatic polarization energy term in comparison to the fixed-charge OPLS force field. In general, the total energy of a molecule, E_{total} is described as sum of electrostatic, van der Waal energy, harmonic bond stretching, angle bending energy and torsional energy, **equation (1)**.

$$E_{total} = E_{electrostatic} + E_{vdW} + E_{stretch} + E_{bend} + E_{torsion} \quad \textbf{(1)}$$

The electrostatic energy added to the total energy in **POSSIM force field** includes the sum of **polarization energy, $\boldsymbol{E_{pol}}$,** and the contribution from interactions of permanent charges, $E_{additive}$, **equation (2)**. The total electrostatic energy is computed using point charges and induced point dipoles thus including all the interactions between the charge-charge, charge-dipole and dipole-dipole interactions.

$$E_{electrostatic} = \boldsymbol{E_{pol}} + E_{additive} \quad \textbf{(2)}$$

The **electrostatic polarization energy, $\boldsymbol{E_{pol}}$** is computed with inducible point dipole for the ith polarizable site, $\boldsymbol{\mu_i}$, shown in **equation (3).** E_i^0 is the electrostatic field due to the fixed atomic charges in the absence of the induced dipoles.

$$E_{pol} = -\frac{1}{2}\sum_{i} \mu_i E_i^0 \quad \textbf{(3)}$$

The induced dipole moment $\boldsymbol{\mu_i}$ is given as a product of scalar polarizabilites, α_i, and the total electric field, E_i^{total} as represented in **equation (4)**.

$$\mu_i = \alpha_i E_i^{tot} \quad \textbf{(4)}$$

The total electric field, E_i^{total} computed as **equation (5)** includes the field due to both the permanent charges and the induced dipoles.

$$E_i^{tot} = E_i^0 + \sum_{j \neq 1} T_{ij}\mu_j \quad (5)$$

T_{ij} in **equation (5)** is the dipole-dipole interaction tensor. It is calculated using **equation (6)** where I is the unit tensor and R_{ij} stands for the distance between atomic sites *i* and *j*.

$$T_{ij} = \frac{1}{R_{ij}^3}\left(\frac{3R_{ij}R_{ij}}{R_{ij}^2} - \mathrm{I}\right) \quad (6)$$

From **equations (4)** and **(5)**, the induced dipole moment on the ith polarizable site is calculated as **equation (7):**

$$\mu_i = \alpha_i E_i^0 + \alpha_i \sum_{j \neq 1} T_{ij} \mu_j \tag{7}$$

The **equation (7)** is self-consistent and is usually solved iteratively. The first three iterations lead to "first-order" **equation (8)**, "second-order" **equation (9)** and "third-order" **equation (10)** approximations for the induced dipoles:

$$\mu_i^0 = \alpha_i E_i^0 \tag{8}$$

$$\mu_i^I = \alpha_i E_i^0 + \alpha_i \sum_{j \neq 1} T_{ij} \mu_j^0 = \alpha_i E_i^0 + \alpha_i \sum_{j \neq 1} T_{ij} \alpha_j E_j^0 \tag{9}$$

$$\mu_i^{II} = \alpha_i E_i^0 + \alpha_i \sum_{j \neq 1} T_{ij} \mu_j^I = \alpha_i E_i^0 + \alpha_i \sum_{j \neq 1} T_{ij} \alpha_j E_j^0 + \alpha_i \sum_{j \neq 1} T_{ij} \alpha_i \sum_{k \neq j} T_{jk} \alpha_k E_k^0 \tag{10}$$

POSSIM employs the **second-order iteration** given in **equation (9)** and has shown to increase the computational speed by about an order of magnitude without any compensation in the accuracy. This **second order expression** also makes the equation analytical, thus eliminating the danger of polarization catastrophe.

The total electrostatic energy also includes the pairwise-additive Coulomb fixed charges interactions between the atoms *i* and *j*:

$$E_{additive} = \sum_{i \neq j} \frac{q_i q_j}{R_{ij}} f_{ij} \tag{11}$$

The factor f_{ij} is equal to zero for 1, 2 and 1, 3 pairs, (atoms which belong to the same valence bond or angle), 0.5 for 1,4 interactions (atoms in the same dihedral angle) and 1.0 for all other pairs.

Another distinguishing feature in POSSIM is the short-distance cutoff parameter, $\boldsymbol{R_{cut} = 0.8Å}$ to reduce unphysical growth of electrostatic interactions due to induced dipoles at distances close to each other and to the permanent electrostatic for small interatomic distances R_{ij}. If the overall distance between two atoms *i* and *j*, R_{ij}, is smaller than sum of these parameters $\boldsymbol{R_{min}^{ij} = R_{cut}^{i} + R_{cut}^{j}}$, smoothing function is used represented in **equation (12)**.

$$R_{ij}^{eff} = \left[1 - \left(\frac{R_{ij}}{R_{min}^{ij}}\right)^2 + \left(\frac{R_{ij}}{R_{min}^{ij}}\right)^3\right] \times R_{min}^{ij} \tag{12}$$

For the non-electrostatic part of POSSIM the standard Lennard-Jones formalism for van der Waal's energy is used:

$$E_{vdW} = \sum_{i \neq j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] f_{ij} \tag{13}$$

Following geometric combining rules are applied for Lennard-Jones coefficients: $\boldsymbol{\varepsilon_{ij} = (\varepsilon_i \,.\, \varepsilon_j)^{1/2}}$ and $\boldsymbol{\sigma_{ij} = (\sigma_i \,.\, \sigma_j)^{1/2}}$.

Standard harmonic formalism is used to calculate bond stretch and angle bend and the torsional term is represented in **equation (14)**.

$$E_{torsion} = \sum_i \frac{V_1^i}{2}[1 + \cos(\phi_i)] + \frac{V_2^i}{2}[1 - \cos(2\phi_i)] + \frac{V_3^i}{2}[1 + \cos(3\phi_i)] \quad (14)$$

The bond stretch and angle bend terms were adopted from the OPLS-AA force field as implemented in BOSS.[43]

➢ POSSIM Force Field Paramterization

Fitting the potential energy parameters for small molecules starts with producing atomic electrostatic polarizabilities. The three-body energies were used as targets as in the previous work to fit the polarizabilities. Briefly, the molecule is exposed to two dipolar electrostatic probes interacting with hydrogen-bonding sites. Each dipolar probe has charges of magnitude of +/- 0.78e that are positioned 0.58Å apart (for the dipole moment to be equal to that of nonpolarizable SPC/E water model[44] of 2.17D). An example of one such molecule with the probes is shown in the **figure 1**. The electrostatic part of the energy is calculated by using an approximation to the Poisson-Boltzmann formalism.

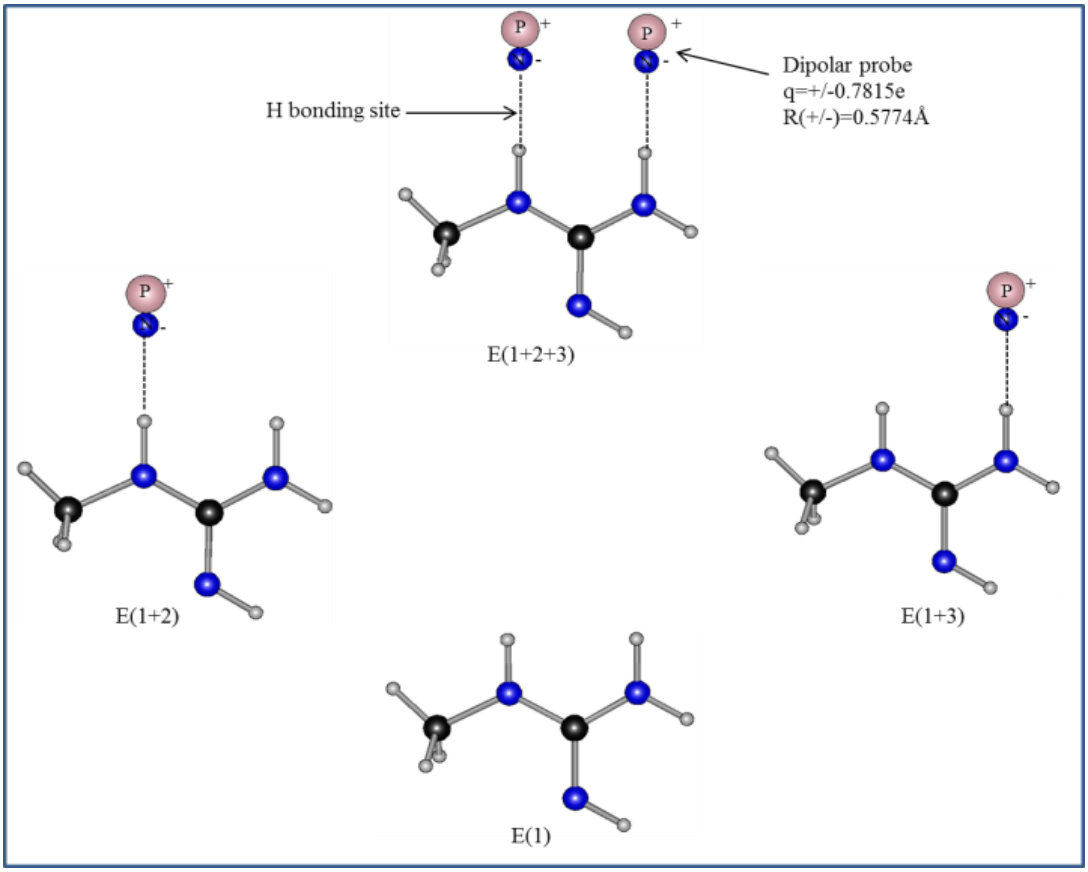


Figure 1: Calculating three-body energies of methylguanidine using dipolar probe

The target three-body energy is calculated according to **equation (15)**:

$$E_{3body} = \mathrm{E(1+2+3)} - \mathrm{E(1+2)} - \mathrm{E(1+3)} - \mathrm{E(2+3)} + \mathrm{E(1)} + \mathrm{E(2)} + \mathrm{E(3)} \quad (15)$$

Jaguar software suite[45] was used to compute the values of the QM energies using the density functional theory (DFT) with the B3LYP method and cc-pVTZ(-f) basis set. The resulting target three-body energies were the used to fit the isotropic atomic polarizabilities, α_i. The values of atomic polarizabilities were chosen so as to minimize the deviation of the energies calculated using POSSIM and quantum mechanical DFT three-body energies.

The next step in parameterization is producing permanent atomic charges and Lennard-Jones (LJ) parameters fitted to reproduce quantum mechanical dimerization energies and intermolecular distances. The atomic polarizabilities, α_i, produced in the previous step are kept

fixed. The dimerization energies in case of electrostatically neutral molecule were calculated for the homodimer and molecule binding to single water molecule at hydrogen bonding positions. The geometries of these dimers were optimized quantum mechanically by LMP2 optimizations using the cc-pVTZ(-f) and cc-pVQZ basis sets.

The torsional parameters in **equation (14)** were computed for dihedrals in methylguanidine molecule as well as for dihedrals in ethylamine and propylamine. These parameters were fitted to the constrained geometry optimizations at fixed dihedral angles calculated with LMP2/cc-pVTZ quantum mechanical level of theory.

3.2.2 *Fuzzy-Border Solvation Model*

The Fuzzy-Border continuum solvation model was employed together with the second-order polarization POSSIM force field in calculating protein pK_a shifts. This model limits the noise in the continuum solvation model using fixed position equally spaced grid points. It also uses first-order approximation for the solvent polarization developed previously to be consistent in spirit with in the second-order POSSIM.

The solvation energy is calculated as sum of electrostatic and non-polar parts of the solvation as given in **equation (16)**:

$$\Delta G_{solvation} = \Delta G_{electrostatic} + \Delta G_{nonpolar} \quad (16)$$

The electrostatic part of the total solvation energy is calculated as an approximation to the Poisson-Boltzmann formalism. It includes the integration over the whole solute-solvent interface of surface charge density, σ, and the electrostatic potential ϕ^0 created by the charges of the solute only, **equation (17)**.

$$\Delta G(el) = \frac{1}{2}\int_{S} \sigma \phi^0 d^2 r \tag{17}$$

The surface charge density $\boldsymbol{\sigma}$ at the interface is given by the self-consistent **equation (18)** as the electrostatic field E_1 in turn is dependent on surface charge density distribution.

$$\sigma = \frac{1}{4\pi}\left(1 - \frac{1}{\epsilon}\right)\boldsymbol{E}_1.\boldsymbol{n} \tag{18}$$

The numerical solution of these equations leads to the surface being represented by a discrete set of points *i*. This result in set of approximate self-consistent **equations (19)** and **(20)** solved iteratively.

$$\Delta G(el) = \frac{1}{2}\sum_{i} q_i \phi_i^0 \tag{19}$$

$$q_i = -\frac{1}{4\pi}\left(1 - \frac{1}{\epsilon}\right)\boldsymbol{E}_{1,i}.\boldsymbol{n}_i \tag{20}$$

The electrostatic field, $\boldsymbol{E_{1,i}}$ is computed as summation of fields produced by solute charges and solute-solvent interface points represented in **equation (21)**.

$$E_{1,i} = \sum_{j} \frac{q_j \boldsymbol{R_{ij}}}{R_{IJ}^3} + \sum_{k \neq 1} \frac{q_k \boldsymbol{R_{ik}}}{R_{ik}^3} \tag{21}$$

R_{ij} in this equation represents the vector pointing from *j* to *i*.

The first term in equation 21 correspond to the summation over all the solute charges whereas the second term represents summation for all the solute-solvent interface points.

The fast first order polarization approximation used in Fuzzy-Border model constitutes replacing $E_{1,i}$ with $E_{1,i}^{o}$ which is the field created by the solute alone. Although, this does not include the field due to the polarized solute-solvent interface, it still contains many-body interactions as electric field being a vector quantity includes interactions from all the solute charges. This approach converts the self-consistent **equations (20)** and (**21)** to analytical one, thus alleviating the convergence problem. The second-order FB model will include the field created by the first order interface charges (without iterations), thus replacing $E_{1,i}$ by $E_{2,i}$ using **equations (20)** and (2**1)**. This second-order model also has no convergence issue as only two iterations are involved and there is no increase in the magnitude of charges in the first of the second iteration.

The solute-solvent interface in FB model is represented by three dimensional cubic numerical grid. We aimed at reducing the noise arising from movement of solute atom or group of atoms by employing a fixed cubic three dimensional equally spaced grid. The interface between the solute and solvent is assumed to consist of points on the solvent-accessible solute surface with distances from R-Δ to R+Δ from the solute atoms (where R would be the distance for the exact location of the solute surface accessible by the solvent). Thus, the solvent-accessible surface is "fuzzy", with a non-zero thickness. The surface points j corresponding to the solute atom i has weights w_j associated with it as the real solvation surface radius is different from points in the R-Δ to R+Δ range that are not exactly at R, **equation (22)**:

$$w_j = a_{0,i}\left[\left(\frac{R_{ij} - R_i}{\Delta_i}\right)^4 - 2\left(\frac{R_{ij} - R_i}{\Delta_i}\right)^2 + 1\right] \quad (22)$$

In **equation (22),** R_{ij} represents the distance between the solvent grid point and the solute atom. The value of the parameter $a_{0,i}$ depends on the solute atom type. The weights are normalized and the unit vector between the grid point j and solute atom *i*, n_{ji} is calculated.

The zeroth-order electrostatic field at the surface grid points can now be calculated via summation over all the solute atoms, **equation (23)**.

$$E^0_{2,j} = \sum_i \frac{q_i R_{ij}}{R^3_{ij}} \quad (23)$$

The fuzzy border first order charge on the surface grid point *j* can be written as in **equation (24)**

$$q^I_j = -A_{scale}\frac{1}{4}\pi(\epsilon - 1)w_j E^0_{1,j}.n_j \quad (24)$$

A_{scale} is an adjustable scaling factor. The first-order electrostatic component of the solvation energy in fuzzy-border formalism can now be determined as

$$\Delta G(el)^I = \frac{1}{2}\sum_i q^I_j \phi^0_j \quad (25)$$

The non-polar part of the total solvation energy is calculated as sum of the two terms.

$$\Delta G(np) = \sum_j w_j A_{np} - \sum_i \sum_j w_j \frac{A_i^{LJ}}{R_{ij}^6} \quad (26)$$

The first term is the contribution of the solvent accessible surface area and is taken as summation over all the grid points. The second term in the above equation approximates the Lennard-Jones attraction between the solute and the solvent atoms and is taken in double summation over all the solute atoms *i* and the grid points *j*.

The solvent-accessible part of the grid is chosen using a Fuzzy-Border modification of the level-set method.[42] Briefly, the technique proceeds in three steps: first, an outward propagation from the interior of the solute creates the surface accessible to centers of solvent molecules. Second, an inward propagation is used to produce the contact surface (the truly solvent-excluded surface). The R referenced above refers to position of this very surface. Finally, another inward propagation is employed to identify interior cavities that can be distinguished from the outside solvent-accessible space.

3.2.3 ***General Scheme for pK_a calculations of protein residues***

The pK_a shifts[19] of protein residue can be calculated using pK_a values of the reference systems. The relative pK_a value is computed from the free energy difference for the residue A and its acidic or basic reference system (acid or side chain) in aqueous solution. Consider deprotonation reaction of acid as given in **equation (27)**:

$$acidH \rightarrow acid- + H+ \quad (27)$$

The change in the free energy for the above reaction can be written as:

$$\Delta G_1 = G(acid^-) + G(H^+) - G(acid - H) \quad (28)$$

The free energy change for the protein residue deprotonation reaction is, **equation (29):**

$$\Delta G_2 = G(A^-) + G(H^+) - G(A-H) \tag{29}$$

The values of the acidity constants for the acid, pK_a(acid) and the residue pK_a(A) for the above deprotonation process are:

$$pKa(acid) = \frac{\Delta G_1}{(2.303RT)} \tag{30}$$

$$pKa(A) = \frac{\Delta G_2}{(2.303RT)} \tag{31}$$

The difference between the acidity constants for the residue A and the reference system (acid) or the pK_a shift is calculated as:

$$\Delta pKa = pKa(A) - pKa(acid) \tag{32}$$

$$\Delta pKa = \frac{[G(A^-) - G(A-H) - G(acid^-) + G(acid-H)]}{2.303RT} \tag{33}$$

The pK_a of the residue A can therefore be written as **equation (34)**:

$$pKa(A) = \frac{pKa(acid) + [G(A^-) - G(A-H) - G(acid^-) + G(acid-H)]}{2.303RT} \tag{34}$$

The following thermodynamic cycle[13] can be used to find absolute pK_a values of reference acids AH with respect to which the pK_a shift can be determined:

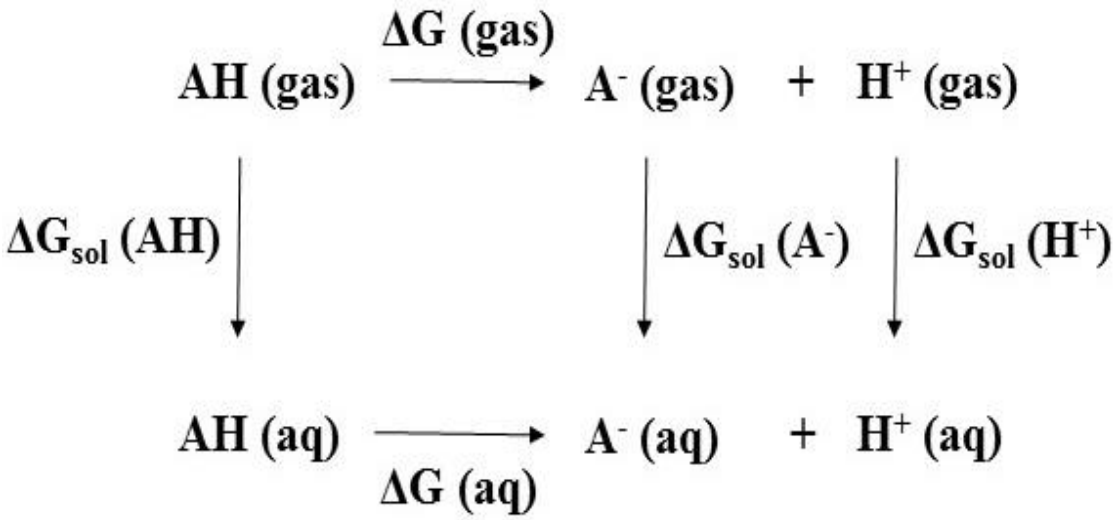


Figure 2: Thermodynamic cycle used to calculate pK_a

The energies of all the residues and the acid and basic reference systems, $G(A^-)$, G(A-H), $G(acid^-)$, G(A-H)are obtained via geometry optimizations with polarizable simulations second-order interaction model (POSSIM) force field and Fuzzy Border continuum solvation model using POSSIM software suite.[36,40,46] The experimental and calculated pK_a values listed in this study is at 298.15K. The above technique is similar to that used in the previous calculations of pK_a shifts of acid[19] and base [47] residues of the OMTKY3 protein.

3.3 Results and Discussions

3.3.1 ***POSSIM Force Field Parameterization***

Most of the POSSIM parameters for the pK_a calculations of acid and basic residues of OMTKY3 protein were adopted from the previous work.[36,40,41,48] New parameters were produced whenever they did not exist before. This included three-body energies, dimerization energies and torsional fitting for methylguanidine molecule needed for arginine pK_a calculations in OMTKY3 basic residues and torsional parameters for C-C-N-H, H-C-C-N dihedrals in ethylamine and C-C-C-N dihedral angle in propylamine for lysine pK_a calculations.

Methylguanidine molecule: ***Three-Body Energies, Dimerization Energies and Torsional Fitting***

The previously established procedure[36,40,41] was used to fit the electrostatic polarizabilities, permanent electrostatic charges, Lennard-Jones and the torsional parameters. All these calculations were carried out using our POSSIM software suite.

Three body energies were first fitted to the QM results to produce atomic polarizabilities. There were 6 hydrogen bonding positions of dipolar probes around methylguanidine and fifteen possible three-body energies. Four of these dipolar probes were acceptors and two had positive point charges at the hydrogen bonding distance from the nitrogen atoms as shown on **Figure 3**. The only fitted polarizability was that for the nitrogen atom with only two covalent bonds. The average absolute error in the three-body energies of methylguanidine was 0.151kcal/mol, which is consistent with our previously observed results.

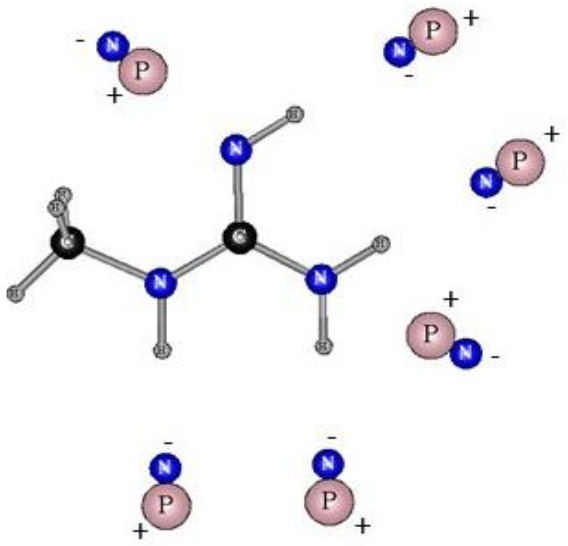
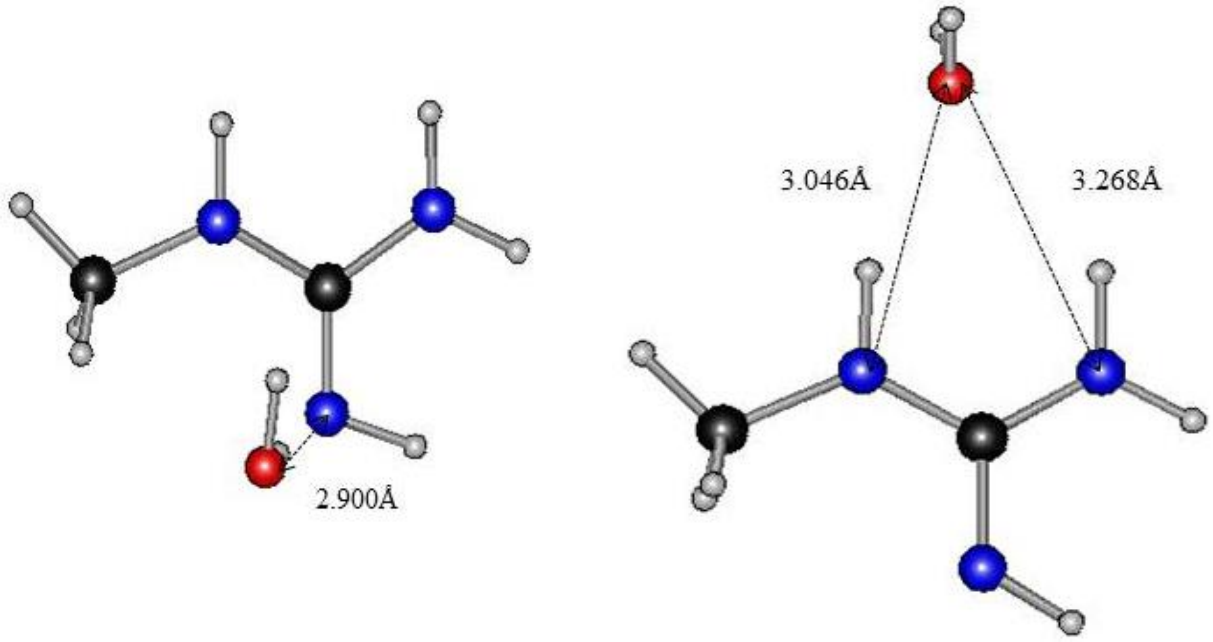


Figure 3: Dipolar probes used for calculating three-body energies for methylguanidine. Symbols P and N in the probe represent positive and negative charges, respectively.

The next step in the parameterization involved reproducing quantum mechanical methylguanidinium-water dimerization energies and distances. The Lennard-Jones parameters were fitted. The POSSIM water molecule was developed previously.[36] There were two methylguanidine-water dimers, and their geometries were optimized with the LMP2/cc-pVTZ(-f) level of theory using Jaguar program.[45] The geometries of these dimers simulated with POSSIM are shown in **Figure 4**.

The quantum mechanical distances are reproduced well within ca. 0.03 Å and the dimerization energies are fitted within the error range of 0.09 kcal/mol (**Table 1**).

Most of the torsional parameters for methylguanidine were adopted from the methylguanidinium ion.[41] Only the coefficients for key rotations of C-N-C-N(H) and H-N(H)-C-N torsions were obtained by fitting the POSSIM energies to LMP2/cc-pVTZ(-f) quantum mechanical values shown in **Table 2**. The nonbonded parameters and torsion parameters for methylguanidine are given in table 3 and 4 respectively.



(a) **(b)**

Figure 4: Gas-phase dimers of methylguanidine with water molecule

Table1: Computed dimerization energies (kcal/mol) and distances (Å) of methylguanidine-water dimers

Dimer	Energy		Distance	
	QM	POSSIM	QM	POSSIM
Methylguanidine-water(a)	-8.32	-8.21	2.91	2.90
Methylguanidine-water(b)	-6.43	-6.35	3.24 3.34	3.26 3.04

Table 2: Torsional energies of methylguanidine, kcal/mol

Molecule	Dihedral	Angle Values	Energy, QM	Energy, POSSIM
Methylguanidine	C-N-C-N(H)	0°	0.00	0.00
		15°	3.01	3.39
		30°	12.88	12.77
	H-N(H)-C-N	0°	0.00	0.00
		15°	6.41	6.25
		30°	11.55	8.99

In both the cases, the deviation in the torsional fitting was less than 0.1kcal/mol.

Table 3: Nonbonded parameters for methylguanidine in POSSIM force field

Atom	Atom Types	Charge, electrons	σ, Å	ε, kcal/mol	α^{-1}, $Å^{-3}$
C, (sp^3)	807	-0.116	3.500	0.066	0.5069
H(NH_2)	1001	0.338	0.0	0.0	9999.99
C, central	1002	0.310	3.290	0.170	2.20
N(H)	1003	-0.547	3.440	0.170	1.40
H(NH), on type 1003	1004	0.393	0.0	0.0	9999.99
N(H)	1006	-0.770	3.350	0.170	0.90
H(NH), on type 1006	1008	0.395	0.0	0.0	9999.99
N(H2)	1009	-0.693	3.170	0.170	1.40

Table 4: Torsional parameters (kcal/mol) in methylguanidine in POSSIM force field

Molecule	Dihedral angle	V_1	V_2	V_3
Methylguanidine	C-N-C-N(H)	0.000	7.500	-6.100
	H-N(H)-C-N	0.000	6.000	-1.000

Ethylamine and Propylamine molecule: *Torsional Fitting*

The torsion coefficients V_1 - V_3 for amines were produced for CCNH, HCCN in ethylamine and CCCN dihedrals in propylamine following the established procedure.[36] The gas phase LMP2/cc-pVTZ (-f) QM energies of the conformations were obtained with these dihedral angles restrained to their positions. These energies were then used to fit the torsional coefficients for POSSIM. The overall RMSD error in the torsional fitting of both the amines was less than 0.1kcal/mol as can be seen from data in **Table 5**. **Table 6** represents the torsional parameters in ethylamine and propylamine.

Table 5: Relative energies in kcal/mol, for ethylamine and propyl amine conformations

Molecule	dihedral	Angle values	Energy, QM	Energy, POSSIM
Ethylamine	C-C-N-H	60°	0.03	0.00
		120°	2.34	2.14
		180°	0.00	0.00
		240°	1.94	2.14
	H-C-C-N	0°	3.32	3.32
		60°	0.00	0.00
Propylamine	C-C-C-N	0°	5.08	4.72
		60°	0.01	0.02
		120°	3.61	3.96
		180°	0.00	0.00

Table 6: Torsional parameters (kcal/mol) in ethylamine and propylamine in POSSIM force field

Molecule	Dihedral angle	V_1	V_2	V_3
Ethylamine	C-C-N-H	0.252	0.100	0.362
	H-C-C-N	0.000	0.000	0.228
Propylamine	C-C-C-N	2.724	0.000	0.906

3.3.2 ***Fuzzy-Border Parameterization***

The numerical grid spacing in each dimension was chosen as 0.40Å in the reported FB model. All the simulations using FB model were carried out in water as the solvent. The radius of water molecule was taken as 1.4Å and the value of dielectric constant, ε, of water was set at 80.4 as in Reference 34. The scaling factor A_{scale} was set at 0.07069. The value of Δ or the "fuzziness" in the solvation radius was the same for all the atom types at 0.25Å. The nonpolar component of the total solvation energy, ΔG(np) has a factor, $A_{np,}$ that has the same value of 3.200 for all the solute atoms and all the grid-points at the solute-solvent interface.

All the simulations presented in this article are performed with revised first-order Fuzzy-Border model. These calculations are done with modified POSSIM software suite including revised FB continuum solvation model. POSSIM force field [36,40,41,46,48] was used for simulating all the solutes in the pK_a calculations of both the acid and the base residues of OMTKY3 protein. Values of the hydration parameters of the small molecules were produced by fitting to the experimental or quantum mechanical solvation energies of these systems. These solvation energies are calculated according to **equation (35)**:

$$\Delta G_{(solv,calculated)} = E_{(solvated)} - E_{(gas)} \quad (35)$$

In **equation (35)**, $E_{(solvated)}$ and $E_{(gas)}$ are the computed energies of the solute molecules in the aqueous and the gas phase respectively. Geometry optimizations were carried out, and the lowest energy conformer was chosen whenever applicable. The hydration energies of the small molecules parameterized with POSSIM force are listed in **Table 7**. The overall average error in the hydration energy of these molecules is 0.136kcal/mol. The fuzzy border hydration paramters with POSSIM force field for small molecules are presented in **Table 8**.

Table 7: Calculated and Experimental Values of Hydration Free Energies in kcal/mol

Compound	Hydration Energy, calculated	Hydration Energy, experimental[a]	Error
CH_4	1.904	1.9 – 2.0[b]	
C_2H_6	1.736	1.82	-0.084
C_3H_8	1.958	1.96	-0.002
C_4H_{10}	2.205	2.08	0.125
iso-C_4H_{10}	2.405	2.32	-0.085
CH_3OH	-5.22	-5.11	-0.110
C_2H_5OH	-4.84	-5.01	0.170
CH_3COOH	-6.854	-6.70	-0.150
C_2H_5COOH	-6.325	-6.48	0.150
CH_3CONH_2	-9.74	-9.71	-0.030
NMA	-10.09	-10.08	-0.010
C_6H_6	-0.872	-0.870	-0.002
Toluene	-0.936	-0.89	-0.046
C_6H_5OH	-6.62	-6.62	0.000
methylamine	-4.571	-4.560	-0.01
ethylamine	-4.430	-4.500	0.07
n-propylamine	-3.765	-4.390	0.63
n-butylamine	-3.327	-4.290	0.96
4-methylimidazole	-10.199	-10.250[b]	0.05
n-propylguanidine	-10.958	-10.920[c]	-0.04
Average error			0.136

[a]Reference 49, [b]Reference 34, [c]Reference 50

Table 8: Fuzzy-Border hydration parameters

Atom	OPLS-AA atom types	R, Å	Δ, Å	a_0	A^{LJ}, kcal/mol-Å^{-6}
Aliphatic C	135, 136, 137, 138, 148, 157, 223, 224, 242, 244, 235, 274, 292, 906	1.952	0.25	0.04000	240.0
Aliphatic H	140, 156, 911	1.415	0.25	0.04000	0.280
Aromatic C	145, 166	2.050	0.25	0.04000	128.7
Aromatic H	146	1.320	0.25	0.04000	12.00
Polar H	155, 168, 240, 241, 270, 909	1.200	0.25	0.04000	2.800
N in amine	900, 901				
O, C_nH_{2n+1}, carboxylic acids, NMA	154, 236, 268	1.725	0.25	0.04000	115.5
O, phenol	167	1.750	0.25	0.04000	75.40
N, acetamide	237	1.960	0.25	0.00600	0.200
N, NMA	238	1.700	0.25	0.04000	161.8
C(OOH) in carboxylic acids	267	1.942	0.25	0.04000	70.00
O, O(=C) in carboxylic acids	269	1.74	0.25	0.04000	57.00
C(O), -COO^-, carboxylate ion	271	1.900	0.25	0.02010	78.00
O^-, -COO^-, carboxylate ion	272	1.530	0.25	0.02010	96.00
N(RNH_3^+)	287	1.700	0.25	0.01900	13.00
H(RNH_3^+)	290	1.350	0.25	0.01900	22.00
N, primary amine	900	1.703	0.25	0.02010	237.7
NA, heterocycle	503	1.700	0.25	0.04000	228.0
NB, heterocycle	511	1.750	0.25	0.04000	200.1
NA, protonated heterocyle	512	1.700	0.25	0.02000	30.00
H, protonated	513	1.296	0.25	0.01423	1.000

heterocycle					
N(H_2), methylguanidinium	300	1.703	0.25	0.04000	100.0

3.3.3 *pK_a values of carboxylic residues of OMTKY3 protein and hydration energies of related molecules*

Coordinates of atoms in the five carboxylic residues (Asp7, Glu10, Glu19, Asp27 and Glu43) of the serine protease inhibitor turkey ovomucoid third domain (OMTKY3) were obtained from Protein Data Bank (PDB ID 1PPF). It has been demonstrated both experimentally and computationally that the local environment of a residue plays the most important role in defining the acidity constant shift of these residues.[18, 31] Therefore, only a part of the protein was explicitly simulated to compute the pK_a shifts. The hydrogen atoms were added with the IMPACT software suite while the carboxylic acid group protons that were added manually. The experimental pK_a values for the acid reference systems propanoic acid (Asp) and butanoic acid (Glu) is 4.87 and 4.83 units respectively.[51]

The initial conformations of the five systems (the parts of the protein containing Asp7, Glu10, Glu19, Asp27 and Glu43 residues) were obtained from the Reference 19. Propanoic acid (pK_a = 4.87) was used as the reference pK_a model for the aspartic acid residues and butanoic acid (pK_a = 4.83) was used for the glutamic acid residues of proteins respectively.

The deprotonated and the protonated acid forms for the aspartate and glutamate reference system are shown in **Figure 5**. The protonation of both the available sites of the acid was carried and the one with the lower total energy was used for the final pK_a calculations. In all the

structures shown, terminal carboxylic groups and the hydrogens in the molecular systems were kept flexible during the geometry optimizations.

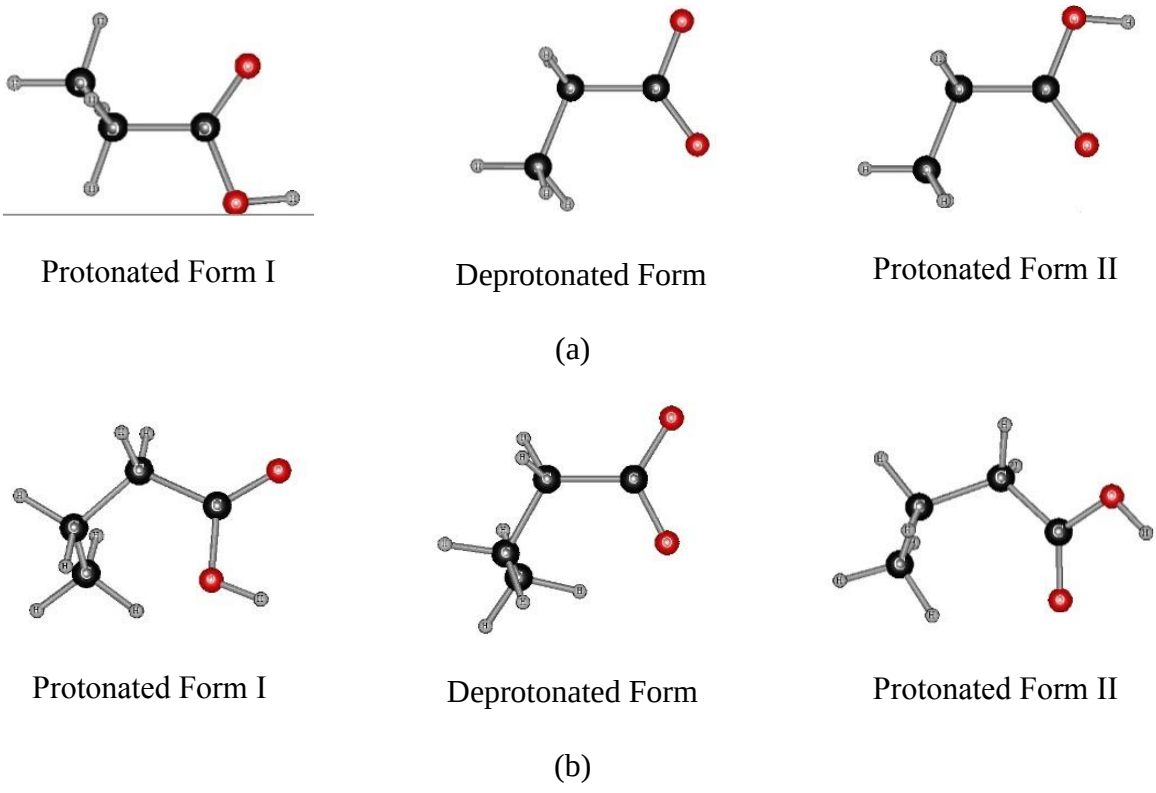


Figure 5: (a) Propanoic Acid used as the reference pK_a model for aspartate residues (pK_a = 4.87) (b) Butanoic Acid used as the reference pK_a model for glutamate residues (pK_a = 4.83)

The pK_a calculations of propanoic acid and the butanoic acid required calculating the protonation and deprotonation energies of the two acids in the gas and in solution.

First, the hydration parameters of aliphatic carbon and hydrogen atoms were produced by fitting to the experimental solvation energies of methane, ethane, propane, butane and isobutane. The solvation energies of the molecules were calculated according to the equation35. It can be seen from table 4 the calculated solvation energies of alkanes follow the general experimental trend

with the overall average error of 0.06kcal/mol. The solvation radius of carbon atom R is 1.952Å, which is in agreement with the Lennard-Jones radius in the OPLS formalism. The value of Lennard-Jones factor for the aliphatic carbon atoms is 240.0kcal/mol $Å^6$. The corresponding values for aliphatic hydrogen atoms are 1.415Å and 0.280kcal/mol $Å^6$. The value of a_0 for both aliphatic carbon and hydrogen atoms is 0.04000.

Next, the alcohols-methanol and ethanol were simulated to produce the hydration parameters of oxygen and the polar hydrogen atoms. The average error in the solvation energies of these molecules is 0.14kcal/mol. The solvation radius of oxygen atom is R is 1.725Å and the Lennard-Jones factor is 115.5 kcal/mol $Å^6$. The polar hydrogens in the two alcohols had radius, R = 1.200Å and the LJ factor is 2.800 kcal/mol $Å^6$. The value of a_0 for both oxygen and hydrogen atoms is 0.04000.

Fuzzy-border hydration parameters for C (COOH) and O, O (=C) in carboxylic acids were now produced by fitting to the solvation energies of ethanoic and propanoic acid. The overall average error in the solvation energies of the two acids – ethanoic and propanoic acid was 0.154 kcal/mol. The FB hydration radius for carboxylic carbon, C (-COOH), was set to be 1.942Å. The parameter a_0 = 0.04000 and the Lennard-Jones factor was 70.00 kcal/mol$Å^6$. The best values of the solvation energies of the acids were obtained with O (=C) radius of 1.740Å, a_0 = 0.04000 and the Lennard-Jones factor as 57.00 kcal/mol $Å^6$. These parameters were used to calculate the aqueous energy of protonated propanoic and the butanoic acids.

Fuzzy Border parameters for the carbon and oxygen of the carboxylate ion, ($-COO^-$) were required to calculate pK_a of carboxylic acids. We produced them by fitting to the experimental pK_a values of the two acids. The carbon radius of carboxylate ion is 1.900Å, and the Lennard-Jones factor is 78.00 kcal/mol $Å^6$. The oxygen atom of the carboxylate ion had the final radius of

1.530Å with the Lennard-Jones factor of 96.00 kcal/mol Å^6. The parameter a_0 was set to 0.02010 for the both the atoms. Summarized in **Table 9** are the energies of the protonated and deprotonated forms of the reference systems (propanoic acid and butanoic acid) and the calculated and experimental pK_a values.

The gas phase deprotonation energies ΔG(gas) for propanoic acid and butanoic acid were 347.20kcal/mol[50] and 347.26kcal/mol[51] respectively.

Table 9: Results from pK_a calculations of reference systems of acid residues of OMTKY3 protein using POSSIM force field and Fuzzy-Border (FB) continuum solvation Model

System		Energy,kcal/mol		pK_a	
		Protonated	Deprotonated	FB[a]	Expt[b]
Propanoic acid	gas	-23.68	-15.82	4.20	4.87
	aq	-29.86	-94.48		
Butanoic acid	gas	-24.49	-25.85	5.27	4.83
	aq	-29.97	-102.41		

[a] Calculations with POSSIM/FB (this work)
[b] Reference 19

The overall average error in the absolute pK_a of propanoic and butanoic acid was 0.55 pH units. These pK_a values of the two acids were obtained by direct fitting of the carbon and oxygen atom FB hydration parameters for the propanoate and butanoate ions.

For calculating the pK_a of protein residues, hydration parameters of other significant molecules required are NMA and acetamide. The aliphatic carbon and hydrogen atoms had the same solvation parameters as in the alkane molecules discussed above. Also, the hydration parameters of alcohol oxygen and polar hydrogens were transferable in these molecules. The

only atoms fitted in the two molecules were the amide nitrogen atoms. In acetamide molecule, N solvation parameters are 1.960Å, 0.00600 and 0.200kcal/mol where as in NMA they are 1.700Å, 0.04000 and 161.8kcal/mol. The average error in the solvation energies of the two amides is 0.02kcal/mol shown in table4.

The computed protonated and deprotonated energies of the carboxylic residues of turkey ovomucoid third domain are shown in **Table 10**. Fuzzy-Border hydration parameters developed for the propanoic and butanoic acids were used without any change.

Table 10: Results of carboxylic acid residues of OMTKY3 protein computed with Polarizable Simulations with Second-Interaction Model (POSSIM) and Fuzzy-Border (FB) continuum solvation Model using POSSIM software

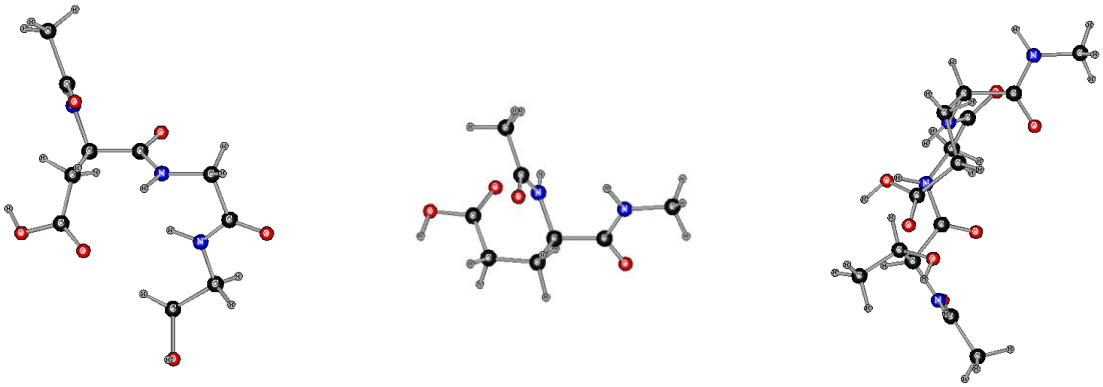
Carboxylic Residues OMTKY3	Energy, kcal/mol		pK_a	
	Protonated	Deprotonated	POSSIM/FB[a]	Experimental[b]
Asp7	-120.37	-188.45	2.33	2.67(0.06)
Glu10	-88.84	-162.57	3.89	4.11(0.09)
Glu19	-90.93	-165.38	3.36	3.21(0.02)
Asp27	-119.28	-188.26	1.66	2.28(0.07)
Glu43	-110.47	-183.64	4.29	4.81(0.10)

[a] Calculations with POSSIM/FB, this work
[b] Reference19

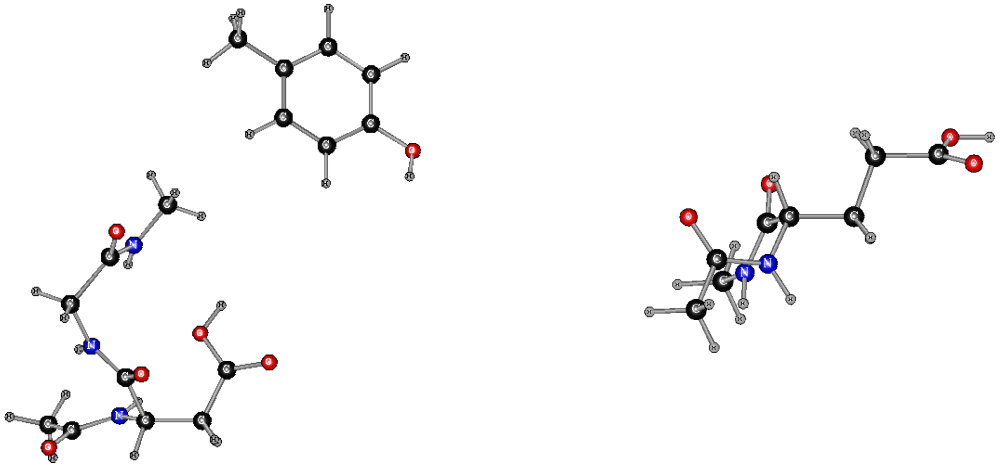
The error bars from the experimental results are given in the brackets in Table10. The experimental errors are small as compared to the POSSIM/FB results and although the deviations of the POSSIM/FB results from experiment can be partly due to the experimental errors, those experimental errors are not likely to be responsible for the most of it.

The effect of two hydrogen bonds with Ser 9 on results for the Asp7 residue can be seen in **Table 10**, with the overall energy change (and thus the pK_a) being the lowest in the series after asp27.

The pK_a of Asp7 is 2.33pH unit compared to experimental value of 2.67pH unit. The Glu10 residue has no hydrogen bonds with neighboring residues. The pK_a calculated for this residue is 3.89pH unit with an unsigned error of only 0.22pH unit from the experimental value. Glu19 residue shows a hydrogen bond from the hydrogen bonded to nitrogen of the backbone and also from the Thr17 residue in the crystal structure. Our model of Glu19 includes the Thr17 residue as well as the backbone part of Leu18.[19] The computed pK_a value for this residue is 3.36pH unit compared to the 3.21pH unit experimental value. The model of the Asp27 residue and its environment used in this study has three intramolecular hydrogen bonds as suggested by the PBD crystal structure of this residue. The predicted pK_a of Asp27 residue is 1.66pH unit compared to the experimental value of 2.28pH unit. The strong electrostatic interactions between the carboxyl group and the amide hydrogens are taken into account in the polarizable POSSIM and can be seen reproducing pK_a of this residue with error of 0.62pH units. The residue Glu43 has no hydrogen bonds for its carboxylic group. The pK_a value of this residue is 4.29pH unit compared to the experimental value of 4.81pH units. All the above results are obtained with fitting of the parameters to the pK_a values of the target atoms in the reference systems, and partial fitting to the protein residues.



Asp7 Glu10 Glu19



Asp27 Glu43

Figure 6: Models of the carboxylic acid residues and fragment of backbone of the OMTKY3 protein and their environment used in this work

Above results can also be compared with the pK_a values for the carboxylic residues obtained with the OPLS-SGB, OPLS-PBF and PFF-PBF.[19] These results are summarized in **Table 11**.

Table 11: Comparison between pK_a values for carboxylic acid residues of OMTKY3 protein calculated in this work with results of previous studies[19]

System	OPLS-SGB[a]		OPLS-PBF[a]		PFF-PBF[a]		POSSIM/FB[b]	
	pKa	Error	pKa	Error	pKa	Error	pKa	Error
Asp7	8.76	6.09	6.26	3.59	1.87	0.80	2.33	0.34
Glu10	1.59	2.52	4.95	0.84	4.41	0.30	3.89	0.22
Glu19	3.06	0.15	4.90	1.69	3.61	0.40	3.36	0.15
Asp27	14.16	11.88	11.60	9.32	1.66	0.62	1.66	0.62
Glu43	9.24	4.43	5.79	0.98	4.02	0.79	4.29	0.52
Average error		5.01		3.28		0.58		0.37

[a] Reference 19
[b] Calculations with POSSIM/FB, this work

The average error in the pK_a values of carboxylic acid residues is 0.37pH units obtained with POSSIM/FB technique. This result is by 0.21 pH units better than with the previous version of the polarizable force field for proteins which in itself was a significant improvement over the fixed-charges OPLS case. This demonstrates the robustness of the Fuzzy-Border continuum solvation model in such calculations. The results also confirm the importance of the explicit treatment of electrostatic polarization in reproducing or prediction protein pK_a shifts.

3.3.4 *pK_a values of basic residues of OMTKY3 protein and hydration energies of related molecules*

The pK_a of **six basic residues of OMTKY3 (Lys13, Arg21, Lys29, Lys34, His52, and Lys55)** were computed using POSSIM force field and FB continuum solvation model. The reference model systems for these residues are n-pentylamine (pK_a 10.6)[47] for lysine, by 4-ethylimidazole

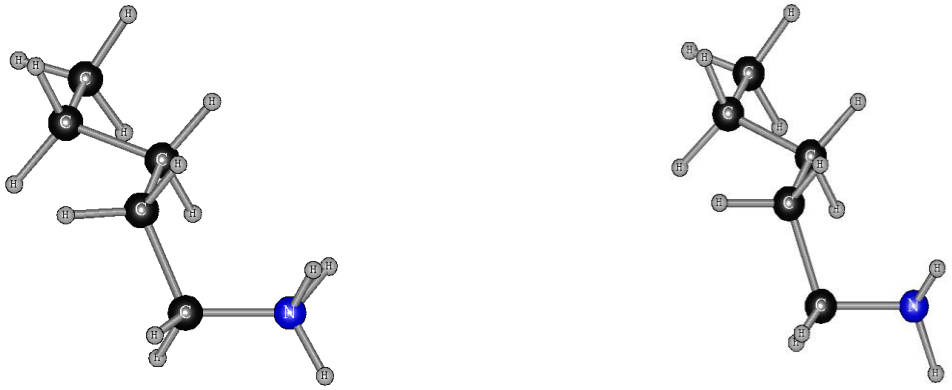
(pK_a 7.55)[47] for histidine and by n-butylguanidine (pK_a 12.48, from methylgyanidine)[47] for arginine.

The initial conformations for the six basic residues are obtained from the experimental OMTKY3 geometry (PDB ID 1omt[37]). These residues were capped with an acetyl group on the N-terminus and an N-methylamine group on the C-terminus except for Lys29 which will be discussed in the following section. The geometry optimizations of both the protonated and deprotonated forms were employed to determine their relative acidity constants.

Lys13, Lys29, Lys34 and Lys55 Residues: Fuzzy Border hydration parameters for amines were produced by fitting the nitrogen atom to the amines solvation energies. The primary amines chosen were methylamine, ethylamine, n-propylamine and n-butylamine. The solvation parameters for the polar amine hydrogen atoms were same as before. The overall average error in the hydration energies was 0.42kcal/mol and the trend of decreasing hydration energy with increasing carbon chain length was successfully reproduced as shown in table 4. The resulting hydration parameters for nitrogen atom are R= 1.703Å, a_0 = 0.2010 and A^{LJ} = 237.7kcal/mol.

The absolute acidity constant of lysine reference system (n-pentylamine) was computed by simulating its protonated and deprotonated forms, **Figure 7**, in gas and aqueous states.

Hydration parameters of atom N and H were produced by fitting to the pK_a value of n-pentylamine. The gas phase deprotonation energy of n-pentylamine used for the calculations is 212.5956 kcal/mol[54] The unsigned error in the calculated pK_a of n-pentylamine is 0.78 pH units as can be seen from **Table 12**.



(a) (b)

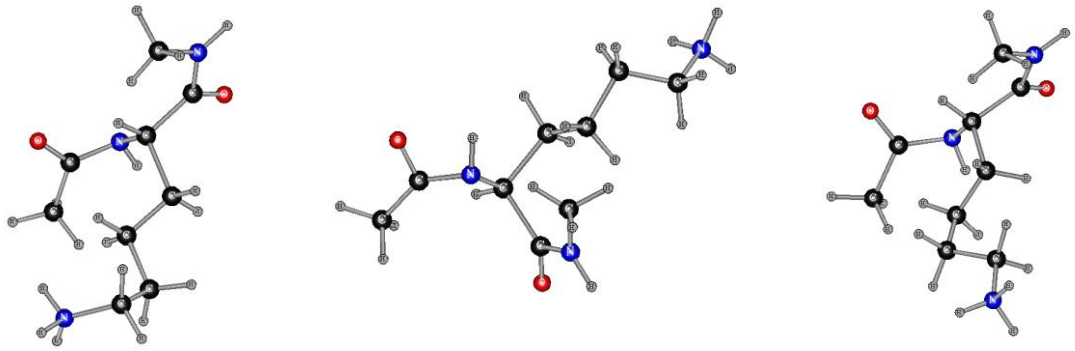
Figure 7: Optimized geometries of the protonated (a) and deprotonated (b) forms of pentylamine used as reference system for lysine residues

Table 12: Data and results from pK_a calculations of reference system n-pentylamine of lysine basic residues of OMTKY3 protein using POSSIM force field and Fuzzy-Border (FB) continuum solvation Model

System		Energy,kcal/mol		pK_a	
		Protonated	Deprotonated	FB[a]	Expt[b]
n-pentylamine	gas	37.61	22.35	11.41	10.63[c]
	aq	-36.83	19.87		

[c] Reference 47

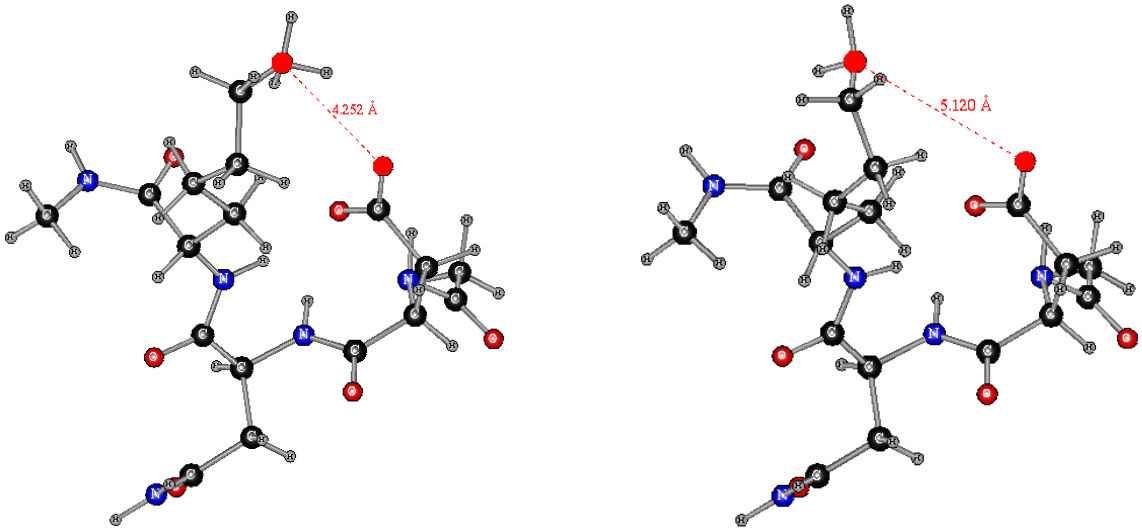
The previously[19,47] established procedure was followed and only a part of the protein in the immediate vicinity of the ionizable residues was used to calculate the pK_a shifts. Hydration parameters were employed to compute pK_a values of the four lysine residues in OMTKY3 protein without any additional parameterizations. The terminal functional group of the basic residue and the side chains were unconstrained for all the calculations. For Lys13, Lys34 and Lys55, there is no hydrogen bond with the other residues, **Figure 8**. The errors in the calculated pK_a for Lys13, Lys34 and Lys55 are 0.49, 0.22 and 1.07 pH units, respectively.



Lys13 Lys34 Lys55

Figure 8: Models of the Lys13, Lys34 and Lys55 residues and fragment of backbone of OMTKY3 protein used in the reported calculations

Because of the close proximity between the Lys29 and Asp27 residues in the crystal structure of OMTKY3, the model for Lys29 includes the sequence of residues Ac-Asp-Asn-Lys-NMe as shown on **Figure 9**.



(a) (b)

Figure 9: Lys29 chain geometries (a) protonated and (b) deprotonated Lys29 geometry without constraints on side chain after optimization.

Although the side chain in Lys29 has potential hydrogen bonding in the NMR structure, the free movement allowed in the side chain in neutral Lys29 structure permitted more accurate

calculation of its pK_a value. The distance between the O and N atoms in NMR structure and the optimized neutral Lys29 structures are 4.252Å and 5.120Å respectively. The overall error in the pK_a of Lys29 is an acceptable value of 0.06 pH units. The energies of the protonated, deprotonated lysine residues and the pK_a values are summarized in **Table 13**.

Table 13: Results of simulations of basic residues of OMTKY3 protein computed with Polarizable Simulations with Second-Interaction Model (POSSIM) and Fuzzy-Border (FB) continuum solvation Model using POSSIM software

System	Energy, kcal/mol		pK_a, Calculated[a]	pK_a, Experiment[b]	Error
	Protonated	Deprotonated			
Lys13	-93.30	-38.26	9.41	9.9	0.49
Arg21	-155.75	-88.75	12.39	12.80	0.41
Lys29	-321.07	-263.65	11.16	11.1	0.06
Lys34	-71.13	-15.45	9.88	10.1	0.22
His52	-103.40	-64.26 (N_ε) -22.69 (N_δ)	7.51	7.50	0.01
Lys55	-86.03	-30.14	10.03	11.1	1.07

[a] Calculations with POSSIM-FB, this work
[b] Reference 8

Arg21 Residue: The reference system for Arg21 is n-butylguanidine (pKa = 12.48[47]). The POSSIM potential energy parameters produced for methylguanidine were used to compute the pK_a shift of Arg21.

First the FB hydration parameters n-propylguanidine were produced. The standard hydration parameters of alkanes and polar hydrogens produced in this work were retained. Only the hydration parameters for the central C atom and the N (-NH, $-NH_2$) bonded to the central C atom were fitted to reproduce the solvation energy of n-propylguanidine. The average unsigned error in the solvation energy of n-propylguanidine was 0.04 kcal/mol given in **table 7**.

These solvation parameters were used to compute the pK_a of n-butylguanidine (**Figure 10**).

(a) (b)

Figure 10: Protonated (a) and deprotonated (b) forms of n-butylguanidine. These systems were used as reference system for the lysine residues

The computed pK_a of n-butylguanidine was 13.12pH units, compared to the experimental result of 12.48pH units[47] shown in **table14**.

Table 14: Results of simulations of pK_a calculations of reference system n-butylguanidine of arg21 basic residues of OMTKY3 protein using POSSIM force field and Fuzzy-Border (FB) continuum solvation Model

System		Energy,kcal/mol		pK_a	
		Protonated	Deprotonated	FB[a]	Expt[b]
n-butylguanidine	gas	-0.81	14.06	13.12	12.48[c]
	aq	-55.20	11.91		

[c] Reference 47

The same parameters were also used to compute pK_a of Arg21 (**Figure 11**).

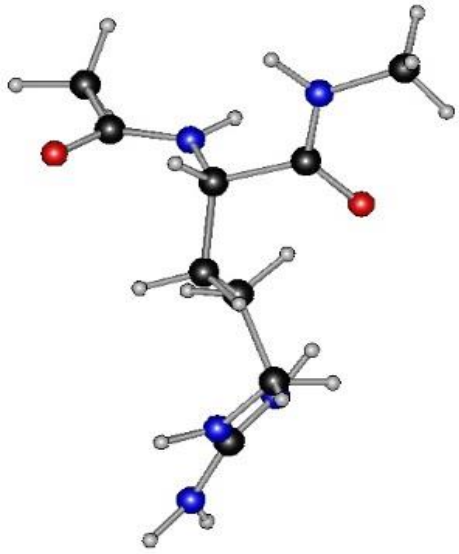


Figure 11: Arg21 residue and a fragment of the backbone of the OMTKY3 protein
The average quantum value of the n-methylguanidine [55] gas-phase deprotonation energy was used in calculating pK_a of Arg21. The error in the pK_a of Arg 21 was 0.41pH units.

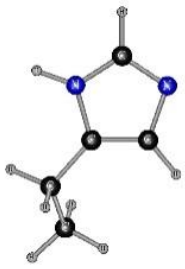
His52 Residue: 4-ethylimidazole was used as the reference model for calculating pK_a shift in His52 residue. The solvation parameters of the heterocyclic molecule are required to calculate its pK_a. First, the solvation parameters of benzene and phenol were produced. The C and H hydration radius in benzene are 2.050Å and 1.320Å respectively. The a_0 for both the atoms is 0.04000 and the Lennard-Jones factor is 128.7 kcal/mol and 12.00kcal/mol respectively. The average unsigned errror of 0.00kcal/mol was observed in the solvation energy of benzene as presented in table4. In case of phenol, the same parameters for aromatic carbon and hydrogen as well as polar hydrogen were employed. Only the oxygen atom was refitted. The solvation parameters of oxygen in phenol are R = 1.750Å, a_0 = 0.04000 and Lennard-Jones factor as 75.40kcal/mol. The average error in the solvation energy is less than 0.01kcal/mol, **table 7**.

The hydration energy of toluene was calculated using the hydration parameters of benzene without any further fitting. The overall average error in the toluene solvation energy was 0.05kcal/mol.

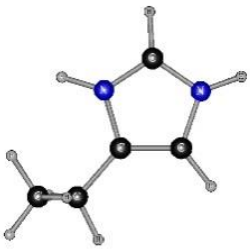
There was good transferability of hydration parameters of aromatic carbon and hydrogen for 4-methylimidazole. Only nitrogen atoms were fitted to the solvation energy. The calculated solvation energy of 4-methylimidazole is -10.20kcal/mol compared to the experimental value of -10.25kcal/mol. The solvation parameters of the nitrogen atoms are 1.700Å, 0.04000 and 121.7kcal/mol.

The same set of parameters was employed to calculate the energy of deprotonated 4-ethylimidazole molecule. Since there are two hydrogens covalently bonded to two different nitrogen atoms (N_δ and N_ε) in the imidazole ring, two deprotonated structures shown in **Figure 12** were simulated for both the His52 residue and the reference molecule 4-ethylimidazole.

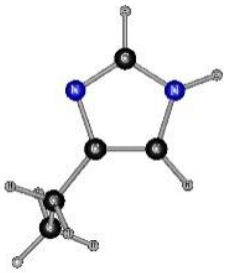
Only the hydration parameters for the nitrogen atoms in the 4-ethylimidazolium were fitted to reduce the pK_a error, with the final value of the latter being 0.79pH units **(Table 15)**. The gas-phase free energy of deprotonation was set at 216.6 kcal/mol.[56]



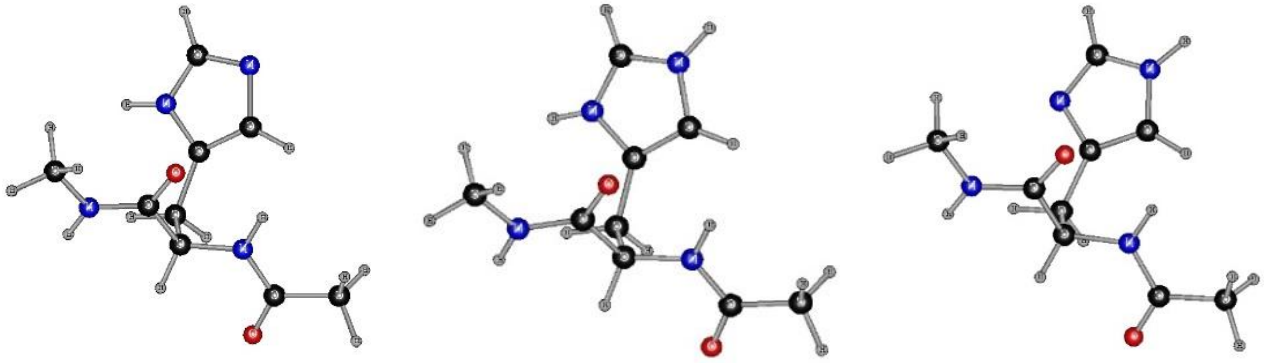
(a) *(N_δ)* 4-thylimidazole



(b) 4-ethylimidazolium



(c) *(N_ε)* 4-ethylimidazole



(d) *(N_δ)* histidine (e) histdine (f) *(N_ε)* histidine

(g) (h) *(i)*

Figure 12: 4-ethylimidazole, histidine reference system (deprotonated (N_δ) form (a), protonated form (b) and deprotonated (N_ε) form (c)) and histidine residue capped with acetyl group on N-terminus and N-methylamine on the C-terminus and fragment of backbone deprotonated (d), (f) and protonated (e)

Table 15: Results of pK_a calculations of reference system 4-ethylimidazole of histidine residue of OMTKY3 protein using POSSIM force field and Fuzzy-Border (FB) continuum solvation Model

System		Energy,kcal/mol			pK_a	
		Protonated	Deprotonated		FB[a]	Expt[b]
4-ethylimidazole	gas	18.15	-6.41 (N_ε)	4.31 (N_δ)	8.34	7.55[c]
	aq	-54.12	-14.92 (N_ε)	-1.72 (N_δ)		

[c] Reference 47

The same parameters were used to compute the acidity constant of His52 **(Table 13)**.

The deprotonated His52 structure with hydrogen covalently bonded to N_ε has lower energy value and was used to determine acidity constant of this residue. The calculated pK_a of His52 residue is 7.51pH units compared to experimental value of 7.50pH units.

The pK_a results of the basic residues of OMTKY3 produced with POSSIM force field and FB continuum solvation model can be compared with the results of previous pK_a calculations using SGB/OPLS, PBF/OPLS and PBF/PFF **(Table 16[47])**.

Table 16: Comparison Between pK_a Values for Residues of OMTKY3 Computed with Different Techniques

System	OPLS-SGB[b]		OPLS-PBF[b]		PFF-PBF[b]		POSSIM/FB[c]		Expt[a]	NMR[a]
	pK_a	Error Expt	pK_a	Error Expt	pK_a	Error Expt	pKa	Error Expt		
Lys13	16.7	6.9	9.0	0.8	10.4	0.5	9.4	0.5	9.9	11.2
Arg21	15.0	-	12.4	-	12.3	-	12.3	0.4[d]	-	12.8
Lys29	23.4	12.3	17.2	6.0	12.9	1.7	11.2	0.1	11.1	11.2
Lys34	16.3	6.2	9.9	0.3	9.9	0.3	9.8	0.2	10.1	11.7
His52	11.8	4.3	11.2	3.7	7.6	0.1	7.4	0.0	7.5	6.2
Lys55	15.7	4.6	9.5	1.6	10.1	1.0	10.0	1.1	11.1	11.3
Average		6.1		2.2		0.7		0.4		

[a] Reference 8, [b] Reference 47
[c] Calculations with POSSIM-FB, this work
[d] Arg21 error reported from NMR value

The overall average pK_a error in the basic residues using POSSIM/FB is 0.38pH units which is by 0.32 pH units more accurate than the previous pK_a calculations using PFF force field and PBF solvation model[47]. Although these results are produced by fitting to the experimental hydration energy and pK_a targets of the reference systems, it clearly demonstrates the Fuzzy-Border solvation model works well in predicting the protein acidity constants.

3.4. Conclusions

The Fuzzy-Border continuum solvation model was parameterized with POlarizable Simulations Second-Order Interaction Model (POSSIM) force field to reproduce hydration energies of small molecules of biological significance and compute pK_a values of carboxylic and the basic residues of OMTKY3 protein. The OMTKY3 protein was employed as a validation application. The Fuzzy-Border hydration parameters were obtained for small molecules and then transferred into the protein residues, and new POSSIM parameters for solutes were produced when required. All the simulations were carried out with a new version of POSSIM software including implementation of the FB model.

The distinguishing features of FB model for calculating solvation energies include a three-dimensional fixed position grid independent of solute coordinates and an approximation to the Poisson-Boltzmann (PB) formalism for assessing solvent polarization. This approximation is consistent with the second-order polarization formalism of POSSIM force field.

The pK_a of five carboxylic (Asp7, Glu10, Glu19, Asp27 and Glu43) and six basic (Lys13, Arg21, Lys29, Lys34, His52 and Lys55) residues of OMTKY3 were reproduced with an overall average error of 0.37 and 0.38pH units, respectively. This was an improvement from the previously reported values obtained with the PFF polarizable force field for proteins[19, 47] and this result validates both the Fuzzy-Border solvation model and the complete POSSIM protein force field[46] employed for simulation of the solutes. These errors are produced without explicitly fitting to the acidity constants of the residues. The overall average unsigned error in the hydration energy of small molecules was 0.136 kcal/mol.

The presented results demonstrate robustness of Fuzzy-Border continuum solvation model in calculating pK_a values of both the acidic and the basic residues of the protein. It

also offers another piece of evidence that explicit treatment of electrostatic polarization is crucial to permit accurate calculations of acidity constants.

References

1. Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A., Prediction and rationalization of protein p K a values using QM and QM/MM methods. The Journal of Physical Chemistry A 2005, 109 (30), 6634-6643.

2. Wuthrich, K., NMR of proteins and nucleic acids. Wiley: 1986.

3. Antosiewicz, J.; McCammon, J. A.; Gilson, M. K., Prediction of pH-dependent properties of proteins. Journal of molecular biology 1994, 238 (3), 415-436.

4. Antosiewicz, J.; McCammon, J. A.; Gilson, M. K., The determinants of p K as in proteins. Biochemistry 1996, 35 (24), 7819-7833.

5. Barth, P.; Alber, T.; Harbury, P., Accurate, conformation-dependent predictions of solvent effects on protein ionization constants. Proceedings of the National Academy of Sciences 2007, 104 (12), 4898-4903.

6. Dixit, S. B.; Bhasin, R.; Rajasekaran, E.; Jayaram, B., Solvation thermodynamics of amino acids assessment of the electrostatic contribution and force-field dependence. Journal of the Chemical Society, Faraday Transactions 1997, 93 (6), 1105-1113.

7. Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D., Empirical relationships between protein structure and carboxyl pKa values in proteins. Proteins: Structure, Function, and Bioinformatics 2002, 48 (2), 388-403.

8. Forsyth, W. R.; Gilson, M. K.; Antosiewicz, J.; Jaren, O. R.; Robertson, A. D., Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. Biochemistry 1998, 37 (24), 8643-8652.

9. Forsyth, W. R.; Robertson, A. D., Insensitivity of perturbed carboxyl p K a values in the ovomucoid third domain to charge replacement at a neighboring residue. Biochemistry 2000, 39 (27), 8067-8072.

10. Havranek, J. J.; Harbury, P. B., Tanford–Kirkwood electrostatics for protein modeling. Proceedings of the National Academy of Sciences 1999, 96 (20), 11145-11150.

11. Jorgensen, W. L.; Briggs, J. M., A priori pKa calculations and the hydration of organic anions. Journal of the American Chemical Society 1989, 111 (12), 4190-4197.

12. Kallies, B.; Mitzner, R., p K a values of amines in water from quantum mechanical calculations using a polarized dielectric continuum representation of the solvent. The

Journal of Physical Chemistry B 1997, 101 (15), 2959-2967.

13. Kaminski, G. A., Accurate prediction of absolute acidity constants in water with a polarizable force field: Substituted phenols, methanol, and imidazole. The Journal of Physical Chemistry B 2005, 109 (12), 5884-5890.
14. Kaminski, G. A.; Maple, J. R.; Murphy, R. B.; Braden, D. A.; Friesner, R. A., Pseudospectral local second-order Møller-Plesset methods for computation of hydrogen bonding energies of molecular pairs. Journal of Chemical Theory and Computation 2005, 1 (2), 248-254.
15. Khandogin, J.; York, D. M., Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. PROTEINS: Structure, Function, and Bioinformatics 2004, 56 (4), 724-737.
16. Klicic, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C., Accurate prediction of acidity constants in aqueous solution via density functional theory and self-consistent reaction field methods. The Journal of Physical Chemistry A 2002, 106 (7), 1327-1335.
17. Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H., The prediction of protein p K a's using QM/MM: the p K a of lysine 55 in turkey ovomucoid third domain. The Journal of Physical Chemistry B 2002, 106 (13), 3486-3494.
18. Li, H.; Robertson, A. D.; Jensen, J. H., The determinants of carboxyl pKa values in turkey ovomucoid third domain. Proteins: Structure, Function, and Bioinformatics 2004, 55 (3), 689-704.
19. MacDermaid, C. M.; Kaminski, G. A., Electrostatic polarization is crucial for reproducing pKa shifts of carboxylic residues in turkey ovomucoid third domain. The Journal of Physical Chemistry B 2007, 111 (30), 9036-9044.
20. Mehler, E. L.; Guarnieri, F., A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. Biophysical journal 1999, 77 (1), 3-22.
21. Nielsen, J. E.; Vriend, G., Optimizing the hydrogen-bond network in Poisson–Boltzmann equation-based pKa calculations. Proteins: Structure, Function, and Bioinformatics 2001, 43 (4), 403-412.
22. Ohno, K.; Sakurai, M., Linear-scaling molecular orbital calculations for the pKa values of ionizable residues in proteins. Journal of computational chemistry 2006, 27 (7), 906-916.

23. Sandberg, L.; Edholm, O., A fast and simple method to calculate protonation states in proteins. Proteins: Structure, Function, and Bioinformatics 1999, 36 (4), 474-483.
24. Schaller, W.; Robertson, A. D., pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. Biochemistry 1995, 34 (14), 4714-4723.
25. Simonson, T.; Carlsson, J.; Case, D. A., Proton binding to proteins: p K a calculations with explicit and implicit solvent models. Journal of the American Chemical Society 2004, 126 (13), 4167-4180.
26. Song, J.; Laskowski, M.; Qasim, M.; Markley, J. L., NMR determination of p K a values for Asp, Glu, His, and Lys mutants at each variable contiguous enzyme-inhibitor contact position of the turkey ovomucoid third domain. Biochemistry 2003, 42 (10), 2847-2856.
27. Zhang, M.; Vogel, H., Determination of the side chain pKa values of the lysine residues in calmodulin. Journal of Biological Chemistry 1993, 268 (30), 22420-22428.
28. Tanford, C.; Kirkwood, J. G., Theory of protein titration curves. I. General equations for impenetrable spheres. Journal of the American Chemical Society 1957, 79 (20), 5333-5339.
29. Harris, T. K.; Turner, G. J., Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. IUBMB life 2002, 53 (2), 85-98.
30. Ji, C.; Mei, Y.; Zhang, J. Z., Developing polarized protein-specific charges for protein dynamics: MD free energy calculation of pK a shifts for Asp 26/Asp 20 in Thioredoxin. Biophysical journal 2008, 95 (3), 1080-1088.
31. Ghosh, A.; Rapp, C. S.; Friesner, R. A., Generalized Born model based on a surface integral formulation. The Journal of Physical Chemistry B 1998, 102 (52), 10983-10990.
32. Bruccoleri, R. E.; Novotny, J.; Davis, M. E.; Sharp, K. A., Finite difference Poisson-Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. Journal of computational chemistry 1997, 18 (2), 268-276.
33. Grant, J. A.; Pickup, B. T.; Nicholls, A., A smooth permittivity function for Poisson–Boltzmann solvation methods. Journal of computational chemistry 2001, 22 (6), 608-640.
34. Sharma, I.; Kaminski, G. A., Calculating pKa values for substituted phenols and hydration energies for other compounds with the first-order fuzzy-border continuum solvation

model. Journal of computational chemistry 2012, 33 (30), 2388-2399.

35. Kaminski, G. A.; Friesner, R. A.; Zhou, R., A computationally inexpensive modification of the point dipole electrostatic polarization model for molecular simulations. Journal of computational chemistry 2003, 24 (3), 267-276.
36. Kaminski, G. A.; Ponomarev, S. Y.; Liu, A. B., Polarizable Simulations with Second-Order Interaction Model□ Force Field and Software for Fast Polarizable Calculations: Parameters for Small Model Systems and Free Energy Calculations. Journal of chemical theory and computation 2009, 5 (11), 2935-2943.
37. Hoogstraten, C. G.; Choe, S.; Westler, W. M.; Markley, J. L., Comparison of the accuracy of protein solution structures derived from conventional and network-edited NOESY data. Protein science: a publication of the Protein Society 1995, 4 (11), 2289.
38. Bode, W.; Wei, A.-Z.; Huber, R.; Meyer, E.; Travis, J.; Neumann, S., X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. The EMBO journal 1986, 5 (10), 2453.
39. Veluraja, K.; Margulis, C. J., Conformational dynamics of sialyl lewisx in aqueous solution and its interaction with selectine. a study by molecular dynamics. Journal of Biomolecular Structure and Dynamics 2005, 23 (1), 101-111.
40. Ponomarev, S. Y.; Kaminski, G. A., Polarizable simulations with second-order interaction model (POSSIM) force field: developing parameters for alanine peptides and protein backbone. Journal of chemical theory and computation 2011, 7 (5), 1415-1427.
41. Li, X.; Ponomarev, S. Y.; Sa, Q.; Sigalovsky, D. L.; Kaminski, G. A., Polarizable simulations with second order interaction model (POSSIM) force field: Developing parameters for protein side-chain analogues. Journal of computational chemistry 2013, 34 (14), 1241-1250.
42. Can, T.; Chen, C.-I.; Wang, Y.-F., Efficient molecular surface generation using level-set methods. Journal of Molecular Graphics and Modelling 2006, 25 (4), 442-454.
43. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. Journal of the American Chemical Society 1996, 118 (45), 11225-11236.
44. Berendsen, H.; Grigera, J.; Straatsma, T., The missing term in effective pair potentials. Journal of Physical Chemistry 1987, 91 (24), 6269-6271.

45. (a) Jaguar v3.5, Schrodinger, Inc.: Portland, OR, **1998**;
(b) Jaguar v4.2,Schrodinger, Inc.: Portland, OR, 2000;
(c) Jaguar, v7.6, Schrodinger, LLC: New York, NY, **2009**.
46. Li, X.; Ponomarev, S. Y.; Sigalovsky, D. L.; Cvitkovic, J. P.; Kaminski, G. A., POSSIM: Parameterizing complete second-order polarizable force field for proteins. Journal of chemical theory and computation 2014, 10 (11), 4896-4910.
47. Click, T. H.; Kaminski, G. A., Reproducing Basic p K a Values for Turkey Ovomucoid Third Domain Using a Polarizable Force Field. The Journal of Physical Chemistry B 2009, 113 (22), 7844-7850.
48. Ponomarev, S. Y.; Sa, Q.; Kaminski, G. A., Effects of Lysine Substitution on Stability of Polyalanine α Helix. Journal of chemical theory and computation 2012, 8 (11), 4691-4706.
49. Gallicchio, E.; Zhang, L. Y.; Levy, R. M., The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. Journal of computational chemistry 2002, 23 (5), 517-529.
50. Sitkoff, D.; Sharp, K. A.; Honig, B., Accurate calculation of hydration free energies using macroscopic solvent models. The Journal of Physical Chemistry 1994, 98 (7), 1978-1988.
51. Atkins, P.; de Paula, J. Physical Chemistry, 7th ed.; W. H. Freeman: New York, **2001**
52. NIST, NIST Chemistry WebBook, **2005**. Available at http://webbook.nist.gov.
53. Schmidt am Busch, M.; Knapp, E. W., Accurate pKa determination for a heterogeneous group of organic molecules. ChemPhysChem 2004, 5 (10), 1513-1522.
54. NIST, NIST Chemistry WebBook, **2005**. Available at http://webbook.nist.gov/cgi/cbook.cgi?ID=C110587&Units=SI&Mask=20#Ion-Energetics
55. Moser, A.; Range, K.; York, D. M., Accurate proton affinity and gas-phase basicity values for molecules important in biocatalysis. The Journal of Physical Chemistry B 2010, 114 (43), 13911-13921.
56. Lias, S. G.; Liebman, J. F.; Levin, R. D., Evaluated gas phase basicities and proton affinities of molecules; heats of formation of protonated molecules. Journal of physical and chemical reference data 1984, 13 (3), 695-808.

CHAPTER 4

Impact of pressure on conformational equlibria of N-acetyl-L-alanine-N'-methylamide in aqueous solution with Polarizable Simulations Second-order Interaction Model (POSSIM) and OPLS-AA force fields

4.1 Introduction

Interaction of low power ultrasound intensities with biological tissue is known to have effects that has several therapeutic advantages. It can be used in numerous biomedical applications such as treatment of cancer[1], physiotherapy[2], transdermal drug delivery[3] and thrombolysis[4]. It is also known to facilitate drug delivery by increasing cell membrane permeability.[5] These applications rely on damaging effects of mechanical stress caused by shock waves on biological materials.[6] The shock waves can be induced by cavitation followed by collapse of cavitation bubbles in ultrasound treatments.[7,8] The cavitation or formation and growth of fluid bubbles are caused by decrease in local static pressure in the fluid flow and these bubbles collapse due to rapid change in this pressure thus generating highly energetic large amplitude shock waves.

The thorough understanding of the shock waves and their interaction with the biological soft matter is the key to develop new applications such as treatment of complex diseases like cancer and many new emerging techniques such as gene therapy and drug delivery.

There has been significant progress in development of biomolecular modeling and computational methods[9] to have an additional insight into processes such as interactions of shock waves with cell membranes[10] at the molecular level. The large size of the biological molecules and its immediate environment such as cofactor, membrane or protein limits the accurate quantitative predictions of binding or free energies or dynamical features of subcellular structures. But molecular modelling can still provide qualitative picture of changes in the biological systems at the all-atom level in response to the external environment.

Monte Carlo (MC) and molecular dynamics (MD) simulations are the two popularly used methods to study computational biology. The quantum mechanical simulations of biological molecules offers an accurate means of studying these molecules but is challenging owing to their

large size and the need to include its solvent environment or a cofactor bound to protein or membrane. The requirement of more computer resources and best level of quantum theory for specific application limits the applicability of quantum calculations. The molecular dynamics or Monte Carlo methods usually employ the molecular mechanics parameters or empirical force fields. Empirical force fields including the explicit treatment of many body interactions in particular the electrostatic polarization have become an important tool to simulate the structural, dynamic and equilibrium thermodynamic properties of biomolecules. Also, empirical force fields with explicit electrostatic polarization energy term offer accurate assessment of energy for larger biomolecules.

Simple model system alanine dipeptide, Ace-Ala-Nme, has been the subject for many experimental and theoretical studies of the backbone conformational equilibria in complex proteins and peptides. The molecular conformational equilibrium is critical in both the chemical and biological phenomenon in the aqueous phase and is widely studied as it plays key role in the structural chemistry of chain molecules. The molecular conformation equilibrium in water is particularly important in case of biological systems. The external factors such as temperature, pressure or solvent affect the equilibrium conformations and this molecular flexibility is important in the chemical reactions as well as the biological phenomenon in liquid phase.

Alanine dipeptide is used as a model molecule to parameterize and validate molecular mechanics empirical force fields.[11] The similarities in the structural features of alanine dipeptide and polypeptide backbone are methyl group bonded to carbon atom at α position, two peptide groups having NH and CO that can make hydrogen bonds with each other or the polar solvent molecules and highly flexible ϕ and ψ backbone dihedral angles representing the rotations around C-N-C^{α}-C and N-C^{α}-C-N bonds, figure 1. The peptide unit has another dihedral angle ω representing the

rotation around -CO-NH- is generally linear (~180°) due to the partial electron delocalization around the C-N bond. The conformations of alanine dipeptide can be approximated by the specific backbone ϕ and ψ dihedral angles around the central carbon atom.

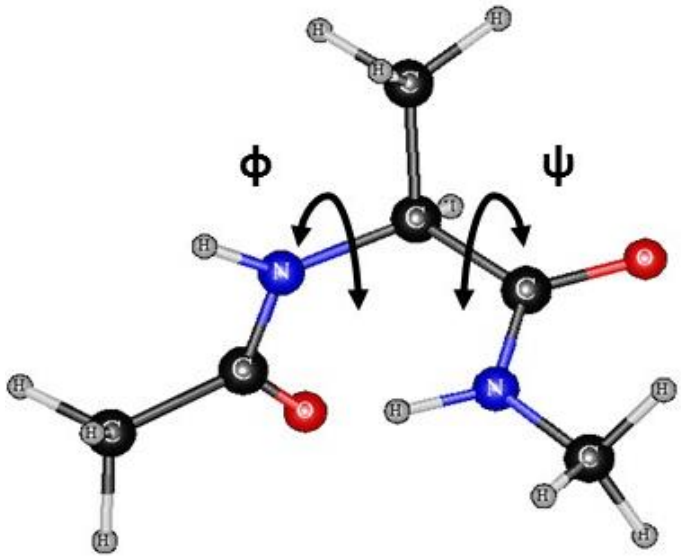


Figure 1: Alanine dipeptide molecule showing backbone ϕ and ψ dihedral angles.

The conformations of alanine dipeptide have been studied in both vacuum and various solvents by theoretical methods such as MO[12], MD[13], RISM[14] and experimental NMR[15] and Raman[16] spectroscopy. These studies suggest the presence of four conformers, intramolecular hydrogen bonded equatorial type 7 membered ring $C7_{eq}$, extended structure in C5, right handed alpha helix α_R, and polyglycine type PII conformer in solution. The molecular dynamics simulations of alanine dipeptide conformers using CHARM22 force field and TIP3P water has shown that PII and α_R are predominant in the solution [13(d)] where as $C7_{eq}$ and C5 are in majority in the vacuum.[12-16] $C7_{eq}$ was also found to be in higher concentration in chloroform than in aqueous solution. Overall, $C7_{eq}$ conformer has the minimum energy in vacuum and is probably dominant in nonpolar solvents.

We have evaluated the effects of pressure on the conformations of alanine dipeptide in water using both the POlarizable Simulations Second-order Interaction Model (POSSIM) and fixed charge OPLS-AA force field. These simulations were carried out using the POSSIM software suite.

4.2 Methods

Monte Carlo (MC) simulations of alanine dipeptide conformers in solvent were carried out with fixed charge OPLS and polarizable POSSIM force fields. The details of OPLS and POSSIM force field are discussed in chap2 and chap3 respectively. The simulations of eight alanine dipeptide conformers (α_L, α_R, α', β_2, C5, $C7_{eq}$, $C7_{ax}$ and PII) in aqueous phase are run with POSSIM software suite using the fixed-charges OPLS and polarizable POSSIM force fields. These simulations are run at a pressure of 0.1MPa and 250MPa. The initial structures of alanine dipeptide conformers were set at the quantum mechanical values for (ϕ, ψ) at minimum energy points and the simulations proceeded with a) all the degrees of freedom unconstrained and b) all degrees of freedom unconstrained except backbone dihedral angles (ϕ, ψ).

The periodic box of 512 water molecules was generated for alanine dipeptide simulations in water. The box size of water is 25.2 × 25.2 × 25.2Å.

The dipole cutoff of 7 Å was used with quadratic smoothing of the interaction energy to zero over the last 0.5 Å before the cutoff distance. The cutoff distance of 8.5 Å was used for solute-solute, solvent-solvent and solute-solvent charge interactions and 100 Å for the intramolecular interactions. In order to ensure convergence, atleast 20×10^6 configurations of Monte Carlo simulations were run. Volume moves were attempted every 260 configurations.

All Monte Carlo simulations were carried out at temperature of 298K.

4.3 Results and Discussions

Monte Carlo simulations of the left and right handed helical α_L and α_R, α', sheet type β_2, C5, seven membered ring $C7_{eq}$, $C7_{ax}$ and polyglycine type PII alanine dipeptide conformers were carried out using POSSIM force field at 298K and pressure of 0.1MPa and 250MPa. POSSIM force field has been parameterized previously against the alanine dipeptide molecule using the established procedure.[17] These parameters were used to study the conformational equilibrium of alanine dipeptide in water and also to validate the applicability of the parameters in water in reproducing the conformers. Simulations with OPLS force field were also done for comparison. Secondly, the effect of high pressure on the general stability of the conformers in aqueous state along with their two backbone torsion angles ϕ and ψ were also examined with the POSSIM and OPLS force field.

The starting conformations of all the eight conformers of alanine dipeptide (α_L, α_R, α', β_2, C5, C7eq, C7ax and PII) with the backbone dihedral angles (ϕ, ψ) at the quantum mechanical values were taken from the reference 17.

4.3.1 ***Fixed dihedral angles (φ, ψ) at quantum mechanical values with POSSIM force field***

Monte Carlo simulations were carried on all the conformers of alanine dipeptide (α_L, α_R, α', β_2, C5, $C7_{eq}$, $C7_{ax}$ and PII) in water at pressure of 0.1MPa and 250MPa at 298K. The backbone ϕ, ψ dihedral angles were constrained at their quantum mechanical values in these simulations. The other degrees of freedom of the conformers were completely unconstrained. Figure 2 depicts the Monte Carlo simulation of α_R conformer in water in NPT ensemble (isobaric, isothermal) at 298K and 0.1MPa atmospheric pressure, with fixed (ϕ, ψ) dihedral angles.

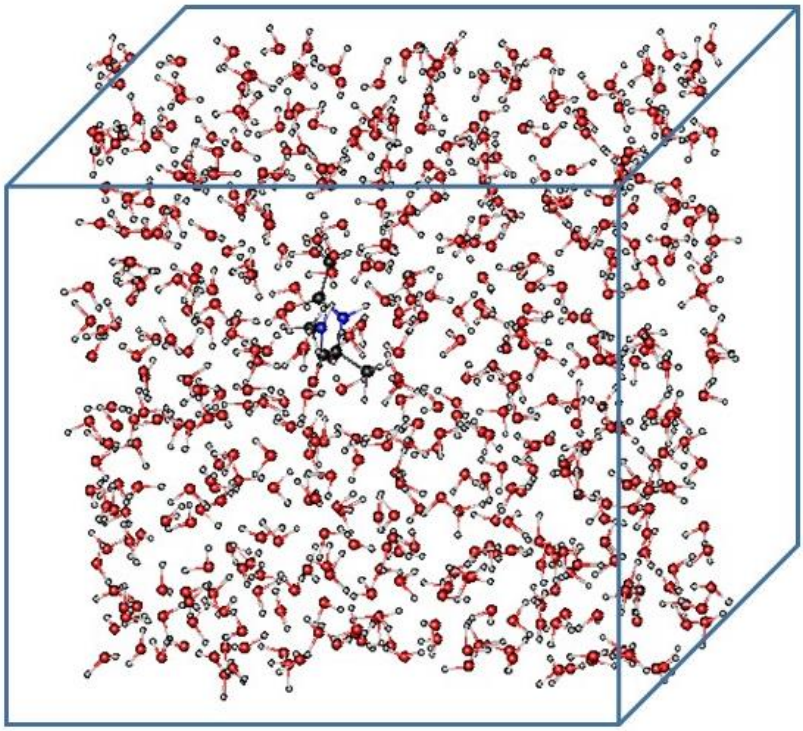


Figure 2: A box used in Monte Carlo simulation of α_R conformer in water at 0.1MPa and 298K with POSSIM force field using POSSIM software suite

The relative total average energy of the conformers at the two pressures 0.1MPa and 250MPa is plotted against the Monte Carlo simulation length as shown in figure 3. The numerical value of total free energies of the conformers in figure 3 is not significant for two reasons. Firstly the chemical structure of the conformers are different and second the force fields contain terms which may lead to different baseline energy values.[18] Qualitatively, at the pressure of 0.1MPa, the order of lowest energy conformers due to high dielectric solvent effects with polarizable POSSIM force field is $C7_{ax} < C5 < C7_{eq} < \alpha_R < \beta_2, < \alpha' < PII < \alpha_L$. As mentioned before, theoretical and experimental studies of conformations of alanine dipeptide in different solvents suggest the presence of four conformers namely $C7_{eq}$, C5, α_R, PII in aqueous solution at 1atm and 298K. The two conformers predominant in aqueous solution out of the above mentioned four

conformers are PII and α_R where as $C7_{eq}$ and C5 are in majority in the vacuum.[19] It is also well known that molecular mechanical methods do not always reproduce correct geometry of the $\alpha_{L,}$ $\alpha_{R,}$ β_2, and PII alanine dipeptide conformers very well.[17]

At high pressure of 250MPa, the stability order of the alanine dipeptide conformers is PII > β_2 > $\alpha^{'}$ > α_R > $C7_{ax}$ > $C7_{eq}$ > α_L > C5 as seen in figure 3(a). POSSIM force field indicate the increased stability of PII conformer compared to α_R and other conformers at higher pressure. Figure 3 and 4 are the plots of total free energy and volume of the eight conformers at the pressures 0.1MPa and 250MPa respectively.

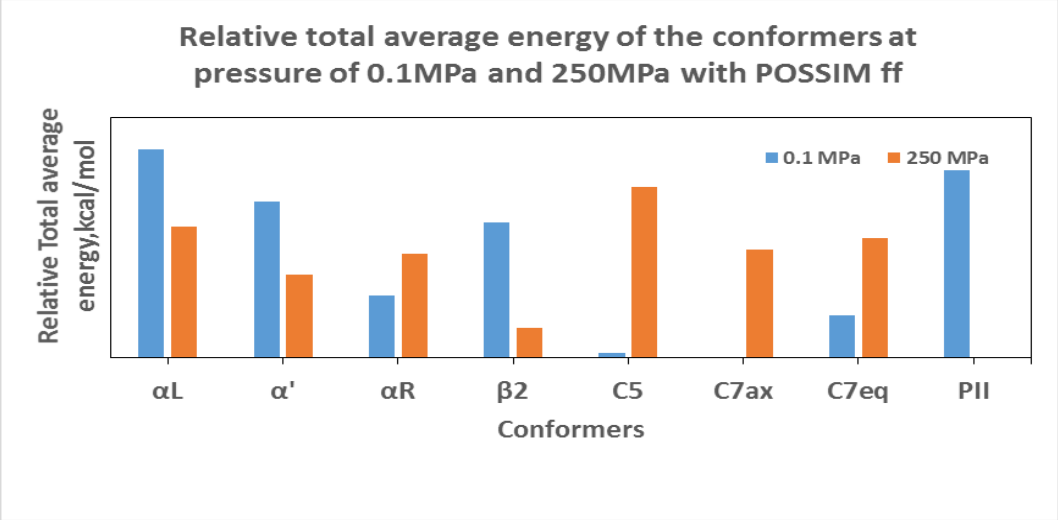


Figure 3: Plot of relative total energy of conformers ($\alpha_{L,}$ $\alpha_{R,}$ $\alpha^{'}$, β_2, C5, $C7_{eq}$, $C7_{ax}$ and PII) at 0.1MPa and 250MPa with POSSIM force field

The relative total average volume of the alanine dipeptide conformers at pressures of 0.1MPa and 250MPa with POSSIM force field is shown in the figure 4. The average volume of the conformers at 0.1MPa pressure is $C7_{ax}$ < $C7_{eq}$ < PII < α_R < β_2 < C5 < α_L < $\alpha^{'}$.

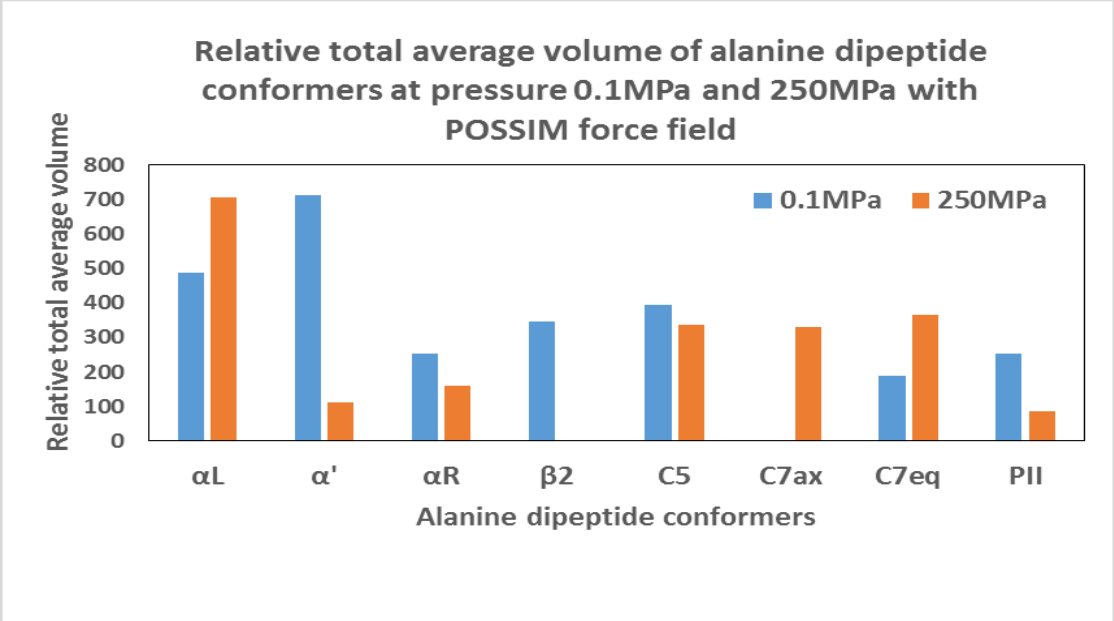
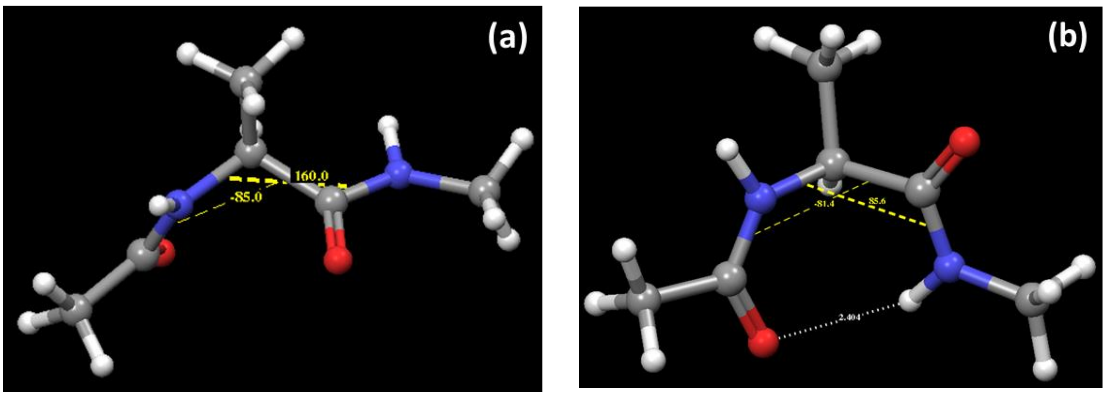


Figure 4: Plot of relative total average volume of conformers (α_L, α_R, α', β_2, C5, $C7_{eq}$, $C7_{ax}$ and PII) at 0.1MPa and 250MPa with POSSIM force field

The structures of alanine dipeptide conformers with the backbone dihedral angles (ϕ, ψ) fixed at the quantum values at 0.1MPa pressure with POSSIM force field are shown in the figure 5.



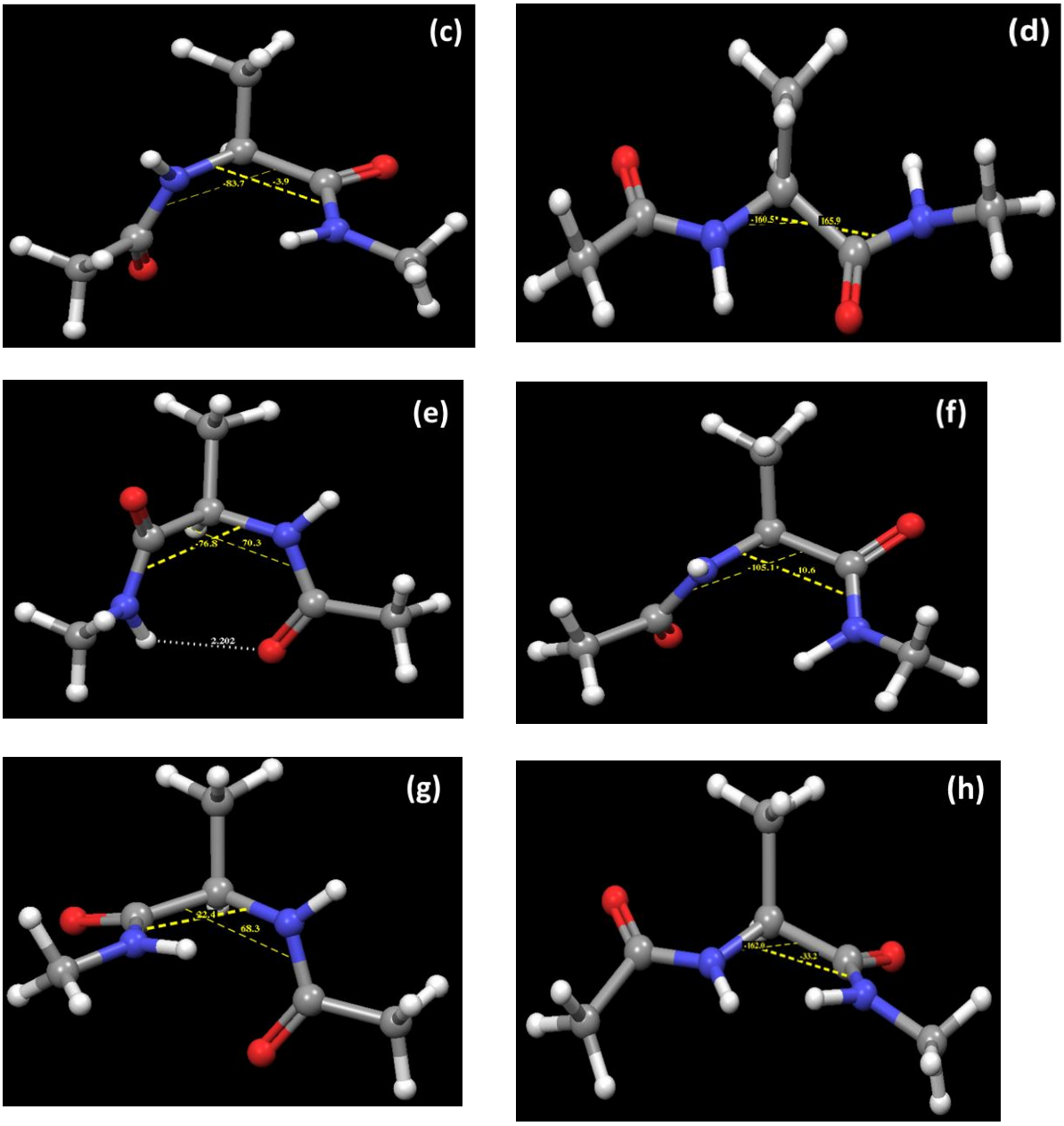


Figure 5: Alanine dipeptide conformers (a) PII (b) $C7_{eq}$ (c) α_R (d) C5 (e) $C7_{ax}$ (f) β_2 (g) α_L (h) α' constrained at the (ϕ, ψ) torsion angles at the quantum mechanical values in aqueous solution at 0.1MPa and 298K. TIP3P waters are omitted for clarity.

The smallest total average volume is observed for the seven membered ring $C7_{ax}$ conformer followed by $C7_{eq}$ conformer in aqueous solution. This is attributed to the intramolecular hydrogen bonding (2.404Å in C7eq in figure 5(b)) in the seven membered conformers between

the acetyl group carbonyl oxygen atom and the methyl amide group NH hydrogen atom. The other conformers lack this intramolecular hydrogen bonding, although the acetyl carbonyl and methyl amide NH groups are exposed to the water molecules. The smallest volume of $C7_{ax}$ and $C7_{eq}$ conformers in water is considered to be mainly due to the intramolecular hydrogen bonding than the solute-solvent electrostatic interactions whereas the compact volume of the most extended C5 conformer is due to the preferable electrostatic interactions with the water molecules. [20]

4.3.2 ***Unconstrained dihedral angles (φ, ψ) at quantum mechanical values with POSSIM force field***

The additional insight into the difference between the conformational equilibrium of alanine dipeptide conformers at the two pressures of 0.1MPa and 250MPa is obtained by running Monte Carlo simulations of the conformers with unconstrained (ϕ, ψ) dihedral angles. The plot of average (ϕ, ψ) dihedral angles of the conformers as a function of number of Monte Carlo simulation length at the pressures 0.1MPa and 250MPa respectively are shown in figure 6 and 7 respectively. . Each value of the backbone angles ϕ and ψ is averaged over the last 5×10^6 Monte Carlo configurations. The straight lines in the figures 6 and figure 7 represent the quantum value of the dihedral angles ϕ and ψ respectively. Table 1 lists the ϕ, ψ torsion values obtained with the polarizable POSSIM force field for alanine dipeptide conformers at the 0.1MPa and 250MPa values of pressure.

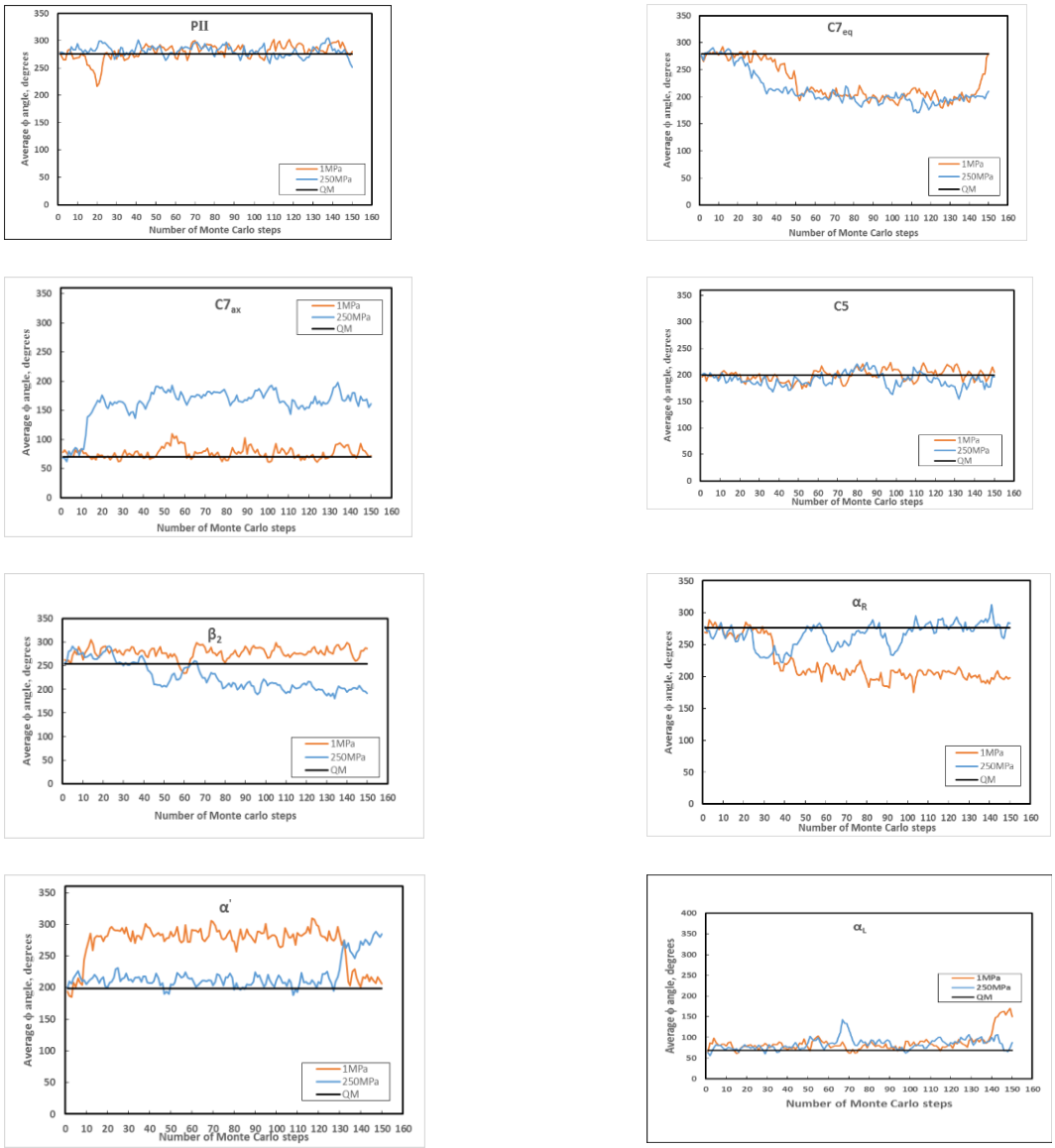


Figure 6: Averaged ϕ dihedral angle (last $5x10^6$ MC configurations) in alanine dipeptide conformers as a function of number of Monte Carlo steps with POSSIM force field**.**

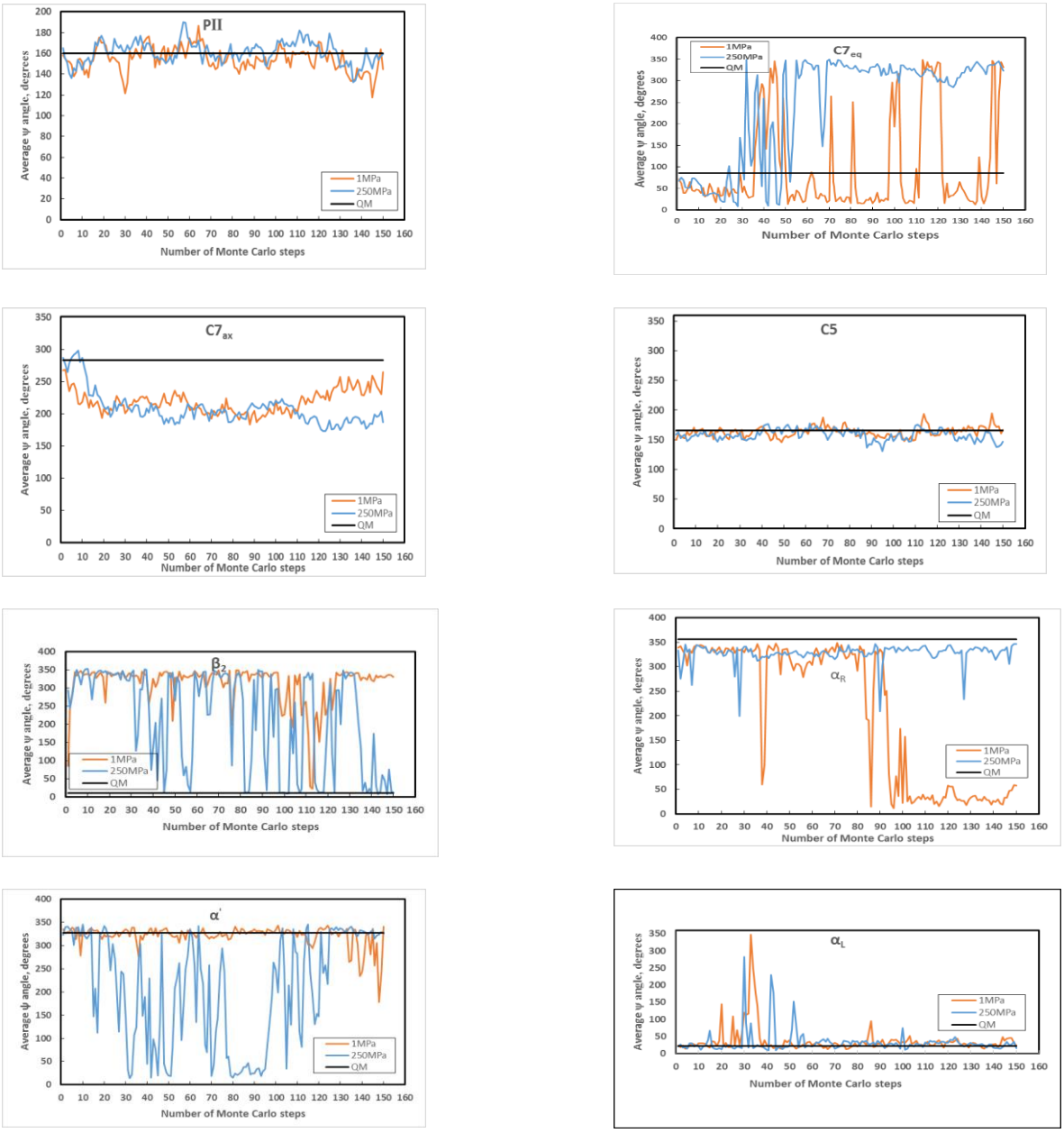


Figure 7: Averaged ψ dihedral angle (last $5x10^6$ MC configurations) in alanine dipeptide conformers as a function of number of Monte Carlo steps with POSSIM force field

Table 1: Dihedral angles *(ϕ,ψ)* for the alanine dipeptide in TIP3P water at 0.1MPa and 250 MPa pressure with POSSIM force field

Conformer	ϕ			ψ		
	QM	0.1MPa	250MPa	QM	0.1MPa	250MPa
PII	-85.0	-79.7	-108.7	160.0	145.0	159.7
$C7_{eq}$	-81.4	-79.6	-149.6	85.6	-29.2	-37.5
$C7_{ax}$	70.3	71.8	161.2	-76.8	-95.8	-173.0
C5	-160.5	-154.9	-163.3	165.9	166.8	147.2
β_2	-105.1	-73.6	-168.8	10.6	-30.9	12.0
α_R	-83.7	-162.0	-76.2	-3.9	57.2	-13.5
α'	-162.0	-154.0	-75.3	-33.2	-18.9	-34.8
α_L	68.3	149.8	88.3	22.4	23.9	15.6

At 0.1MPa, both (*ϕ, ψ*) torsion angles are fairly stable for PII and C5 conformers with average deviation of (5.3, 15.0) and (5.6, 0.9) from the quantum mechanical (QM) values. For $C7_{eq}$ conformer, torsion angle *ϕ* is more stable than *ψ* compared to the QM value with the change of (1.8, 56.4) degrees. The deviation in the (*ϕ, ψ*) dihedral angles of α_R conformer from the quantum mechanical values is (78.3, 53.3). The overall average deviation of the dihedral angle *ϕ* and *ψ* conformers at 0.1MPa and 298K from the quantum mechanical values is 26.7 and 22.6 respectively for POSSIM force field, table 1.

At higher pressure of 250MPa, the change in the (*ϕ, ψ*) angles of $C7_{eq}$, PII and C5 conformer (70, 8.3), (29.0, 14.7) and (8.4, 19.6) respectively, table 1.

4.3.3 *Fixed dihedral angles (φ, ψ) at quantum mechanical values with OPLS force field*

The alanine dipeptide conformers in water were also simulated with OPLS force field with isobaric isothermal ensemble at pressure of 0.1MPa and 250MPa at 298K. The relative total average energy of the conformers in water at the pressure of 0.1MPa and 250MPa at 298K is

shown in figure 8. All the ϕ, ψ dihedral angles are fixed at the quantum values of the conformers. The plot indicate the left handed helix α_L conformer as the most stable conformer in aqueous solution followed by $C7_{eq}$, α_R and PII at pressure of 0.1MPa. The effect of high pressure on the stability trend of the alanine dipeptide conformers is not very significant in comparison to the POSSIM force field. The order of stability at 250 MPa is $\alpha_L > C7_{eq} > C7_{ax} >$ PII $> C5 > \beta_2 > \alpha_R > \alpha'$.

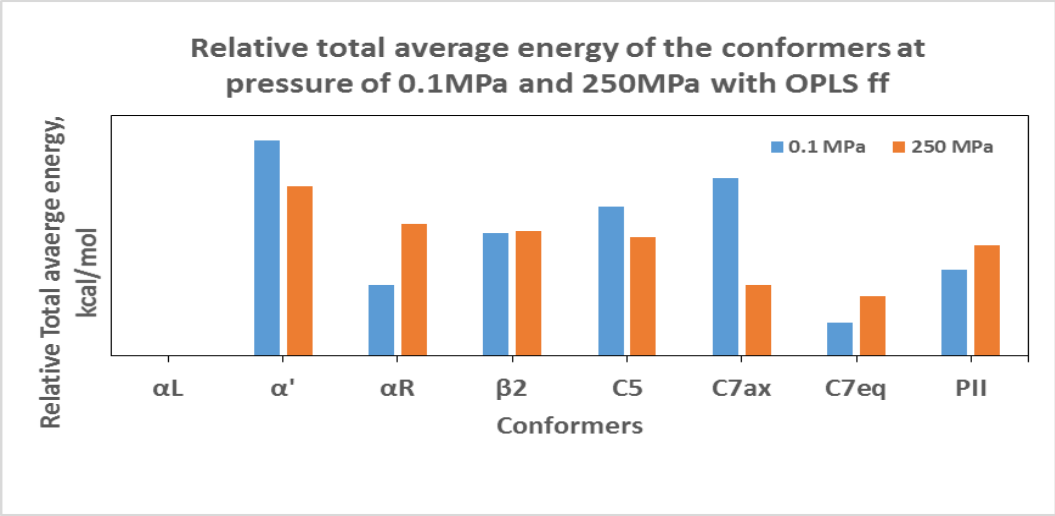


Figure 8: Plot of relative total energy of conformers (α_L, α_R, α', β_2, C5, $C7_{eq}$, $C7_{ax}$ and PII) at 0.1MPa pressure with OPLS force field

The total average volume of the conformers at the two pressures 0.1MPa and 250MPa is shown in figure 9.

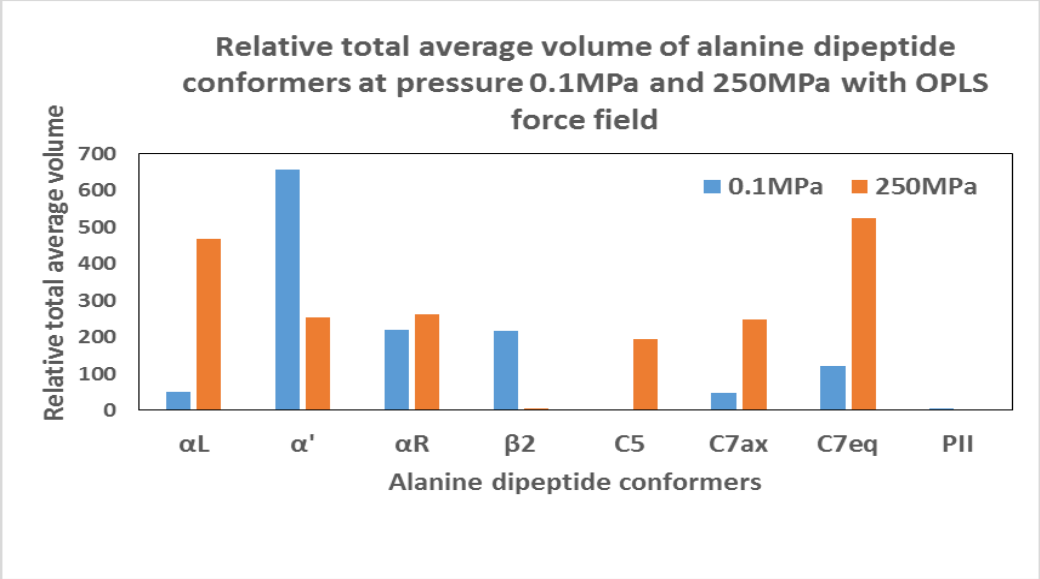


Figure 9: Plot of relative total average volume of conformers (α_L, α_R, α', β_2, C5, $C7_{eq}$, $C7_{ax}$ and PII) at 0.1MPa and 250MPa with OPLS force field

At 0.1MPa, the average volume of the conformers follow the trend C5 < PII < $C7_{ax}$ < α_L < $C7_{eq}$ < β_2 < α_R < α' where as at high pressure of 250MPa, the average volume of the conformers is PII < β_2 < C5 < $C7_{ax}$ < α' < α_R < α_L < $C7_{eq}$.

4.3.4 ***Unconstrained dihedral angles (φ, ψ) at quantum mechanical values with OPLS force field***

The alanine dipeptide conformations with unconstrained ϕ and ψ dihedral angles were also run at 0.1MPa and 250MPa pressure respectively with nonpolarizable OPLS force field. Figure 10 and figure 11 represent the plots of average ϕ and ψ angles of the conformers as a function of simulation length with OPLS force field. Table 2 summarizes the backbone (ϕ,ψ) dihedral angles at pressure 0.1MPa and 250 MPa.

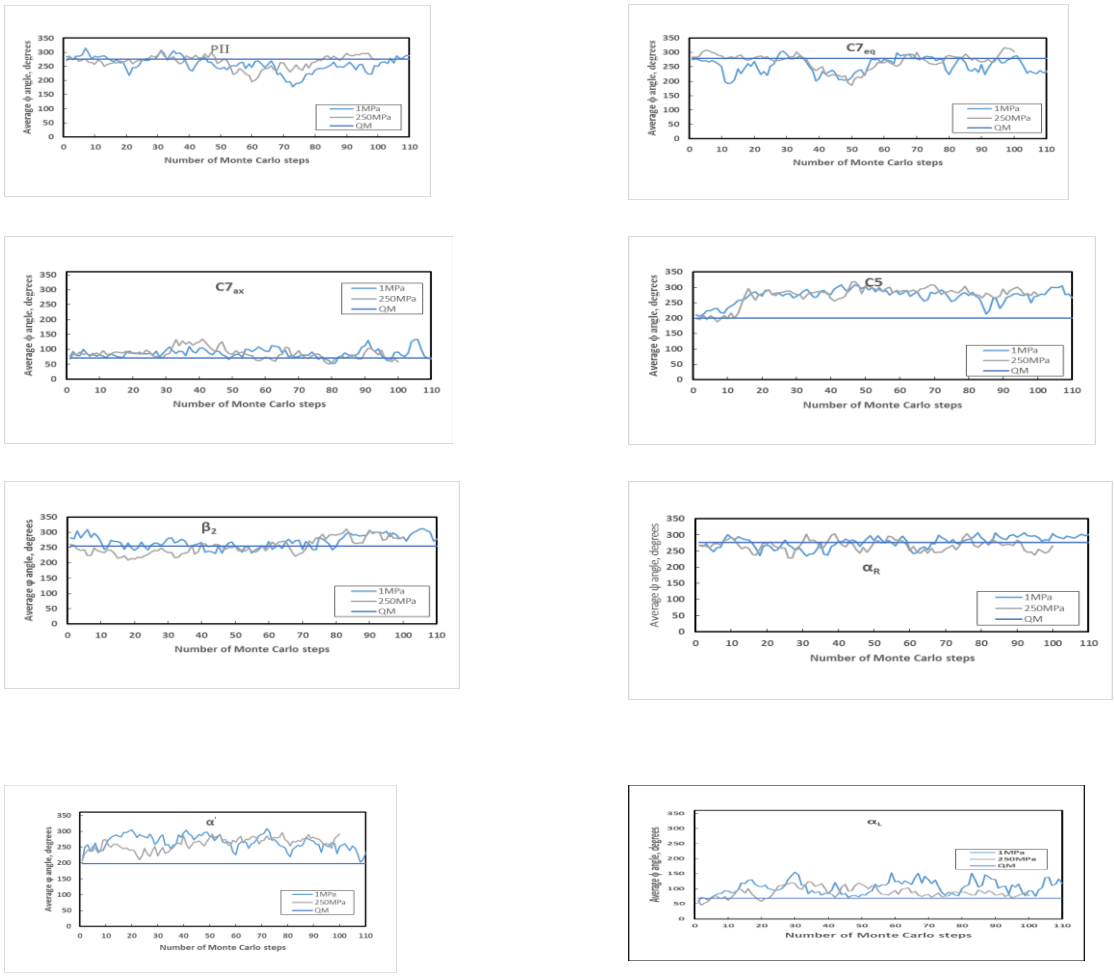


Figure 10: Averaged ϕ dihedral angle (last $5x10^6$ MC configurations) in alanine dipeptide conformers as a function of number of Monte Carlo simulation length with OPLS force field

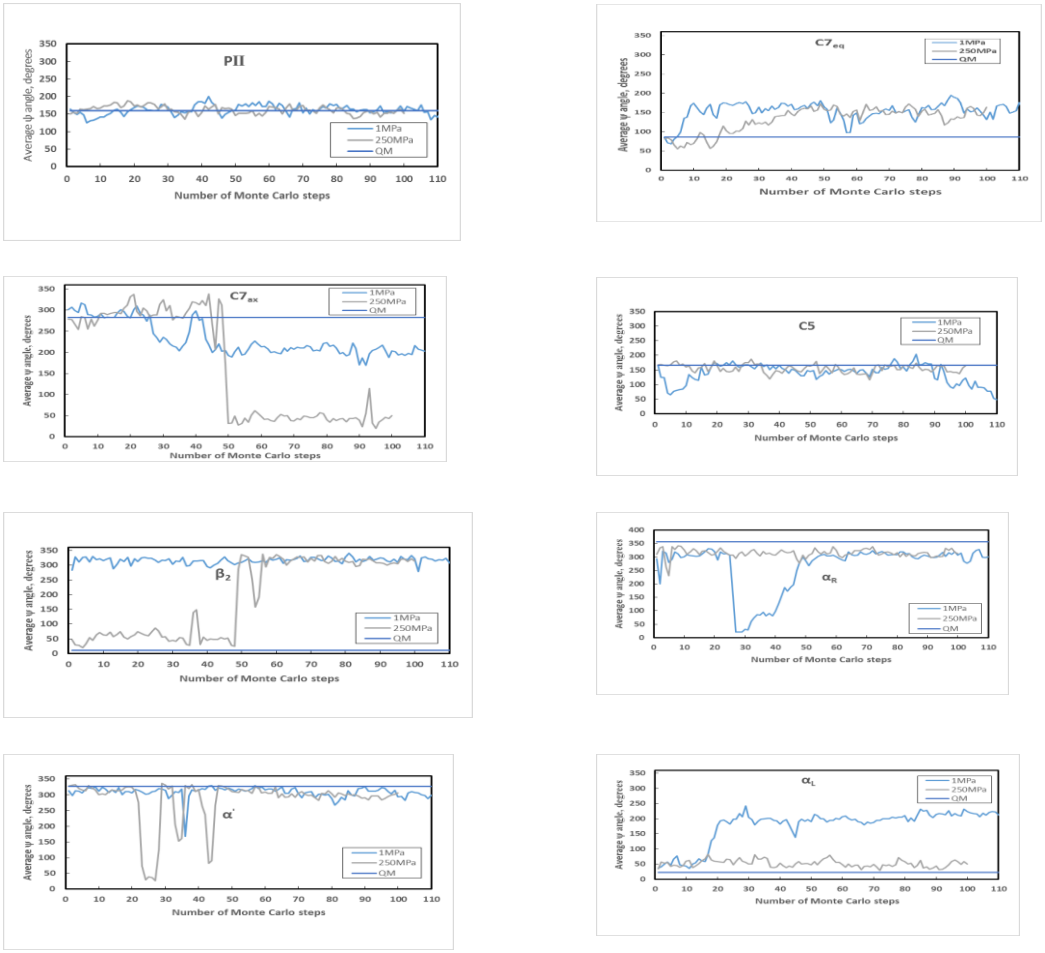


Figure 11: Averaged ψ dihedral angle (last $5x10^6$ MC configurations) in alanine dipeptide conformers as a function of number of Monte Carlo simulation length with OPLS force field

Table 2: Dihedral angles (ϕ,ψ) for the alanine dipeptide in TIP3P water at 0.1MPa and 250 MPa pressure with OPLS force field

Conformer	ϕ			ψ		
	QM	0.1MPa	250MPa	QM	0.1MPa	250MPa
PII	-85.0	-75.2	-86.5	160.0	118.8	151.7
$C7_{eq}$	-81.4	-127.9	-57.4	85.6	175.0	164.2
$C7_{ax}$	70.3	119.2	58.2	-76.8	169.7	49.5
C5	-160.5	-76.8	-84.1	165.9	-53.6	165.1
β_2	-105.1	-110.3	-76.3	10.6	24.8	-46.1
α_R	-83.7	-94.8	-93.8	-3.9	-62.8	-46.8
α'	-162.0	-59.0	-68.2	-33.2	- 49.7	-54.3
α_L	68.3	82.8	88.3	22.4	-163.3	50.5

The computed average deviation of the dihedral angles ϕ and ψ of alanine dipeptide conformations from the quantum values is 40.3 and 70.8 degrees respectively at 0.1MPa pressure (Table 2) which is higher compared to the 26.7 and 22.6 degrees shown by POSSIM force field. Both the torsion angles ϕ and ψ were unstable for the conformers $C7_{ax}$, $C7_{eq}$, C5 and PII conformations which are experimentally found to be stable in aqueous solution.

The effect of high pressure on the conformers with OPLS force field is minimal except for the conformers $C7_{ax}$ and β_2. The change in the torsion angles (ϕ, ψ) for $C7_{ax}$ and β_2 is (61, 120.2) and (34, 21.3) respectively. For $C7_{eq}$ and C5, the change in dihedral angles (70.5, 10.8) and (7.3, 111.5) are observed.

4.4. Conclusions

Polarizable POSSIM force field was used to study alanine dipeptide conformations (α_L, α_R, α', β_2, C5, $C7_{eq}$, $C7_{ax}$ and PII) in TIP3P water at pressure of 0.1MPa and 250MPa at 298K respectively. The same study was also conducted with fixed charge OPLS¬AA force field for comparison. Monte Carlo simulations were performed on alanine dipeptide conformers with their backbone φ, ψ dihedral angles both fixed first at the quantum mechanical values. All other degrees of freedom were unconstrained.

Simulations of alanine dipeptide conformers in water with constrained φ, ψ dihedral angles with POSSIM force field predicted qualitative results fairly in agreement with the literature results. At 0.1MPa pressure and 298K, the conformers $C7_{ax}$, $C7_{eq}$ and α_R have the lowest energies where as at high pressure of 250MPa PII is the stable conformation. Experimentally, Raman spectra of alanine dipeptide in aqueous solution suggest the presence of PII, α_R conformers along with C7eq. At high pressure of 250MPa, PII is the most stable conformation as predicted by the POSSIM simualtion results.

The total average volume of the conformers with POSSIM force field at constrained dihedral angles at 0.1MPa is lowest for the seven membered ring conformers, $C7_{ax}$ and $C7_{eq}$ followed by PII conformer. The lowest volume of C7eq has been shown to be the result of intramolecular H bonding causing it smaller in the solvent exclusion volume unlike the extended C5 conformer which preferably interacts with the water molecules. The small volume of PII conformer in aqueous solution has been linked to the preferable hydrogen bonding in aqueous solvent.

POSSIM force field with unconstrained dihedral angle predicts stable dihedral angles of PII and C5 conformations of alanine dipeptide as compared to the other conformers at both the pressure of 0.1Mpa and 250MPa.

In comparison, Monte Carlo simulations of fixed dihedral angles alanine dipeptide in water at 0.1MPa with OPLS force field predicted α_L as the most stable conformer followed by $C7_{eq}$, α_R and PII at both the pressure of 0.1MPa. At high pressure of 250MPa, OPLS predicts the relatively high total average energy of α_R conformer in comparison to the other four conformers mentioned.

The average deviations in the torsion angles (φ, ψ) in alanine dipeptide conformations at 0.1MPa with POSSIM and OPLS force field are force field are (26.7, 22.6) and (40.3, 70.8) degrees respectively.

These results indicate that polarizable POSSIM and OPLS force field reproduces the qualitative results of alanine dipeptide in water fairly well. Further parameterization of alanine dipeptide torsion angles in water is required with POSSIM force field.

References

1. Hill, C.; Ter Haar, G., High intensity focused ultrasound—potential for cancer treatment. The British journal of radiology 1995, 68 (816), 1296-1303.
2. Hill, C. R., Physical principles of medical ultrasonics. Ellis Horwood: 1986.
3. Mitragotri, S.; Blankschtein, D.; Langer, R., Ultrasound-mediated transdermal protein delivery. Science 1995, 269 (5225), 850-853.
4. Everbach, E. C.; Francis, C. W., Cavitational mechanisms in ultrasound-accelerated thrombolysis at 1 MHz. Ultrasound in medicine & biology 2000, 26 (7), 1153-1160.
5. (a) Kodama, T.; Doukas, A. G.; Hamblin, M. R., Shock wave-mediated molecular delivery into cells. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 2002, 1542 (1), 186-194. (b) Kodama, T.; Doukas, A. G.; Hamblin, M. R., Delivery of ribosome-inactivating protein toxin into cancer cells with shock waves. Cancer letters 2003, 189 (1), 69-75. (c) Frampton, J. P.; Fan, Z.; Simon, A.; Chen, D.; Deng, C. X.; Takayama, S., Microbubbles: Aqueous Two-Phase System Patterning of Microbubbles: Localized Induction of Apoptosis in Sonoporated Cells (Adv. Funct. Mater. 27/2013). Advanced Functional Materials 2013, 23 (27), 3366-3366.
6. Suhr, F.; Delhasse, Y.; Bungartz, G.; Schmidt, A.; Pfannkuche, K.; Bloch, W., Cell biological effects of mechanical stimulations generated by focused extracorporeal shock wave applications on cultured human bone marrow stromal cells. Stem cell research 2013, 11 (2), 951-964.
7. (a) Koşar, A.; Oral, O.; Itah, Z.; Gozuacik, D., Bubbly cavitating flow generation and investigation of its erosional nature for biomedical applications. Biomedical Engineering, IEEE Transactions on 2011, 58 (5), 1337-1346. (b) Lukes, P.; Sunka, P.; Hoffer, P.; Stelmashuk, V.; Pouckova, P.; Zadinova, M.; Zeman, J.; Dibdiak, L.; Kolarova, H.; Tomankova, K., Focused tandem shock waves in water and their potential application in cancer treatment. Shock Waves 2014, 24 (1), 51-57.
8. (a) Miller, M. W.; Miller, D. L.; Brayman, A. A., A review of in vitro bioeffects of inertial ultrasonic cavitation from a mechanistic perspective. Ultrasound in medicine & biology 1996, 22 (9), 1131-1154. (b) Kimmel, E., Cavitation bioeffects. Critical Reviews™ in Biomedical Engineering 2006, 34 (2).

9. (a) Collepardo-Guevara, R., Biomolecular modeling and simulation: The productive trajectory of a field. 2011. (b) van der Kamp, M. W.; Shaw, K. E.; Woods, C. J.; Mulholland, A. J., Biomolecular simulation and modelling: status, progress and prospects. Journal of The Royal Society Interface 2008, 5 (Suppl 3), 173-190. (c) Lim, C.; Zhou, E.; Quek, S., Mechanical models for living cells—a review. Journal of biomechanics 2006, 39 (2), 195-216. (d) Siu, S. W.; Vácha, R.; Jungwirth, P.; Böckmann, R. A., Biomolecular simulations of membranes: physical properties from different force fields. The Journal of chemical physics 2008, 128 (12), 125103.
10. Koshiyama, K.; Kodama, T.; Yano, T.; Fujikawa, S., Molecular dynamics simulation of structural changes of lipid bilayers induced by shock waves: Effects of incident angles. Biochimica et Biophysica Acta (BBA)-Biomembranes 2008, 1778 (6), 1423-1428.
11. Brooks, C.; Case, D. A., Simulations of peptide conformational dynamics and thermodynamics. Chemical Reviews 1993, 93 (7), 2487-2502.
12. (a) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks III, C. L.; Pople, J. A., Theoretical study of blocked glycine and alanine peptide analogs. Journal of the American Chemical Society 1991, 113 (16), 5989-5997. (b) Boehm, H. J.; Brode, S., Ab initio SCF calculations on low-energy conformers of N-acetyl-N'-methylalaninamide and N-acetyl-N'-methylglycinamide. Journal of the American Chemical Society 1991, 113 (19), 7129-7135. (c) Gould, I. R.; Cornell, W. D.; Hillier, I. H., A quantum mechanical investigation of the conformational energetics of the alanine and glycine dipeptides in the gas phase and in aqueous solution. Journal of the American Chemical Society 1994, 116 (20), 9250-9256. (d) Deng, Z.; Polavarapu, P.; Ford, S.; Hecht, L.; Barron, L.; Ewig, C.; Jalkanen, K., Solution-phase conformations of N-acetyl-N'-methyl-L-alaninamide from vibrational Raman optical activity. The Journal of Physical Chemistry 1996, 100 (6), 2025-2034. (e) Han, W.-G.; Jalkanen, K.; Elstner, M.; Suhai, S., Theoretical study of aqueous N-acetyl-L-alanine N'-methylamide: Structures and Raman, VCD, and ROA spectra. The Journal of Physical Chemistry B 1998, 102 (14), 2587-2602.
13. (a) Tobias, D. J.; Brooks III, C. L., Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution: A comparison of theoretical results. The Journal of Physical Chemistry 1992, 96 (9), 3864-3870. (b) Chipot, C.; Pohorille, A., Conformational equilibria of terminally blocked single amino acids at the water-hexane

interface. A molecular dynamics study. The Journal of Physical Chemistry B 1998, 102 (1), 281-290. (c) Apostolakis, J.; Ferrara, P.; Caflisch, A., Calculation of conformational transitions and barriers in solvated systems: application to the alanine dipeptide in water. The Journal of chemical physics 1999, 110 (4), 2099-2108. (d) Smith, P. E., The alanine dipeptide free energy surface in solution. The Journal of chemical physics 1999, 111 (12), 5568-5579.

14. Pettitt, B. M.; Karplus, M., Conformational free energy of hydration for the alanine dipeptide: thermodynamic analysis. The Journal of Physical Chemistry 1988, 92 (13), 3994-3997.
15. (a) Madison, V.; Kopple, K. D., Solvent-dependent conformational distributions of some dipeptides. Journal of the American Chemical Society 1980, 102 (15), 4855-4863. (b) Poon, C.-D.; Samulski, E. T.; Weise, C. F.; Weisshaar, J. C., Do bridging water molecules dictate the structure of a model dipeptide in aqueous solution? Journal of the American Chemical Society 2000, 122 (23), 5642-5643.
16. Avignon, M.; Garrigou-Lagrange, C.; Bothorel, P., Conformational analysis of dipeptides in aqueous solution. II. Molecular structure of glycine and alanine dipeptides by depolarized Rayleigh scattering and laser Raman spectroscopy. Biopolymers 1973, 12 (7), 1651-1669.
17. Ponomarev, S. Y.; Kaminski, G. A., Polarizable simulations with second-order interaction model (POSSIM) force field: developing parameters for alanine peptides and protein backbone. Journal of chemical theory and computation 2011, 7 (5), 1415-1427.
18. Ponomarev, S. Y.; Sa, Q.; Kaminski, G. A., Effects of Lysine Substitution on Stability of Polyalanine α Helix. Journal of chemical theory and computation 2012, 8 (11), 4691-4706.
19. Takekiyo, T.; Imai, T.; Kato, M.; Taniguchi, Y., Temperature and pressure effects on conformational equilibria of alanine dipeptide in aqueous solution. Biopolymers 2004, 73 (2), 283-290.

Chapter 5

Future Directions

5.1 Introduction

The long term objective of the work presented in this dissertation is development and application of chemically accurate and computationally efficient continuum solvation model for polarizable POSSIM and fixed-charge OPLS force fields for proteins. The previous chapters show development of hydration parameters of small organic molecules of biological significance by reproducing the solvation energies of the molecules. The overall average error in the hydration energies of the small molecules using OPLS force field is 0.08kcal/mol. The parameters were also fitted to the target atoms to calculate the absolute acidity constant of the substituted phenols with the overall unsigned average error of 0.41pH units. Following the successful implementation of first-order FB model of solvation with OPLS force field, it was expanded to work with polarizable POSSIM force field. The empirical hydration parameters were fitted for the solvation energies of the small molecules with an average error of 0.136kcal/mol. In addition, the FB model was also validated by reproducing the pKa shifts of five carboxylic and six basic residues of OMTKY3 protein within the acceptable error range by fitting the required atoms. There was a good transferability of the parameters of other atoms. In future, the FB model can be applied to reproduce acidity constants of other proteins. The acidity constants of ribonuclease sa protein with POSSIM force field can be computed using this solvation model. Secondly, FB model of solvation can be used to calculate the binding free energy of protein-ligand complexes. Thirdly the refined Fuzzy Border model including the second order approximation can be used to calculate the protein ligand binding affinity of HIV inhibitors.

5.2 Calculation of pK_a values of proteins using polarizable POSSIM force field

The continuum Fuzzy-Border solvation model can be used to accurately predict pKa values of ribonuclease Sa (Rnase Sa) protein using polarizable POSSIM force field[1]. The pKa's for ribonuclease Sa protein and its fragment have been extensively studied experimentally and computationally[2, 3, 4], which will provide a very solid benchmark for assessing validity of the FB continuum model of solvation. The Ribonuclease Sa (Rnase Sa) is a small 96 amino acid globular protein isolated from the bacterial strain Streptomyces aureofaciens. It is an acidic protein with 7 Asp, 5 Glu, 2 His, 5 Arg and no lysine residues.

It been explicitly demonstrated that the electrostatic interactions play dominant role in pK_a shifts in Rnase Sa[5(a), 5(b)]. Also, the Asp residues in Rnase Sa have wide range of pKa values from 2.4 to 7.4[5(a)]. Therefore, application of the fast polarizable technique of POSSIM force field combined with implicit solvation model is expected to be advantageous both as an application and as a test for the model.

The pKa calculations of the Rnase Sa will help to establish robustness of the Fuzzy-Border solvation technique. It will once again prove that FB solvation model combined with POSSIM force field is adequate for calculations of pKa of proteins which will enable in long term to create an automated acidity constant predictor for proteins and ligands.

5.3 Calculation of binding free energy of protein-ligand complexes

The FB continuum model of solvation with polarizable POSSIM force field can be applied for computing binding affinities for protein–protein complexes or for protein–ligand interactions. This will provide an opportunity to not only validate the performance of continuum model of

solvation in calculations of protein-ligand binding affinity but also to predict currently unknown protein pka values. In particular, the continuum model of solvation/POSSIM force field can be used to evaluate the binding affinity of farnesyl transferase inhibitors (FTIs). Farnesyl transferase inhibitors are class of biologically active anti-cancer drugs that inhibit farnesylation of target proteins including Ras[6, 7, 8, 9]. The synthetic route of producing some of the potent natural inhibitors is challenging due to the structure of the inhibitors. The analysis of binding affinities of FTIs with accurate POSSIM force field and continuum solvent will help to propose more potent synthetic farnesyl transferase inhibitors.

The following thermodynamic cycle, **figure 1**, can be used to obtain relative binding energy of the protein-ligand complex[10].

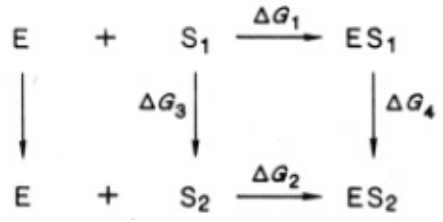


Figure 1 Thermodynamic cycle to calculate relative binding free energy[10]

From the figure 1, the binding affinity of protein ligand complex can be calculated according to the equation (1)[10].

$$\Delta\Delta G_b = \Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4 \quad (1)$$

Protein farnesyl transferase is an enzyme coupling a 15-carbon isoprenyl group to Ras proteins, (**Figure 2**) catalyzing farnesylation of the protein p21, which is a product of Ras onco-gene. The p21 undergoes a one-amino acid mutation resulting in permanent activation followed by uncontrolled cell growth and division. Thus, inhibiting the farnesyl transferase is critical step in cessation of cell growth and thus development of successful anti-cancer drugs. This has resulted in synthesis of large number of farnesyl transferase inhibitors.

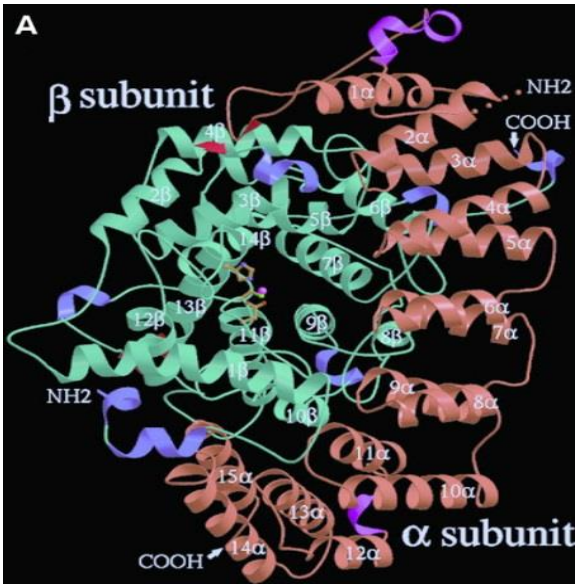


Figure 2: Structure of the FTase heterodimer

Based on their mechanism of action, FTIs can be divided into different groups. The direction here is study of CAAX peptidomimetics and analogues of naturally discovered CP-type inhibitors.

This computational study of the binding affinity of FTIs will help to understand the underlying mechanisms of interactions and thus help in developing potent synthetic farnesyl transferase inhibitors.

5.4 Applying second order FB model to compute binding affinity of HIV inhibitors

The FB continuum formulation can be refined by incorporating the **second order approximation**[11] for more accurate calculation of protein ligand binding affinity interactions. The details of the second-order fuzzy border formulation are discussed in chapter 2. This second order FB model of solvation with POSSIM force field can be applied to predict binding affinity of HIV-1 reverse transcriptase (RT) inhibitors.

Acquired immunodeficiency syndrome (AIDS) is caused by human immunodeficiency virus (HIV). According to the World Health Organization (WHO), there were approximately 35 million people worldwide living with HIV/AIDS in 2013.[12] The complexity of the disease makes it difficult to induce breakthrough in the treatment of AIDS. The HIV virus is a retrovirus, the viral RNA strand produces single stranded RNA and double stranded DNA through reverse transcriptase enzyme (RT). Thus RT (Figure 3) being a multifunctional enzyme is the most important target to block replication and stop evolution of virus inside host cell.

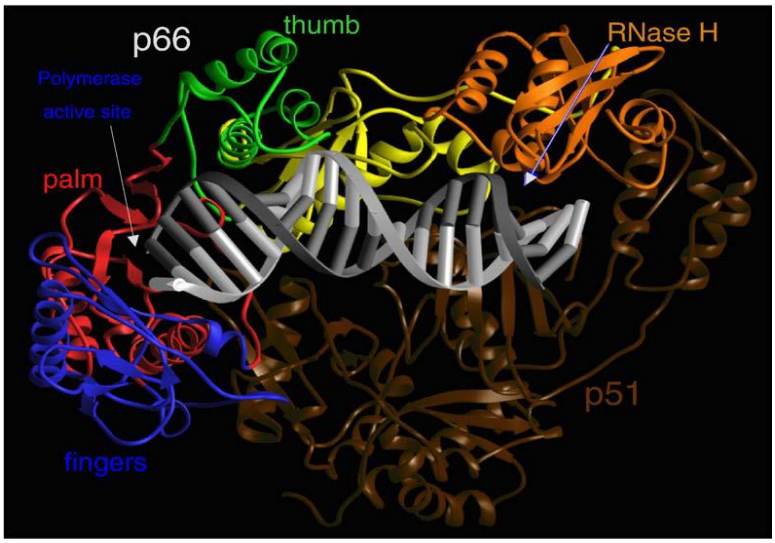
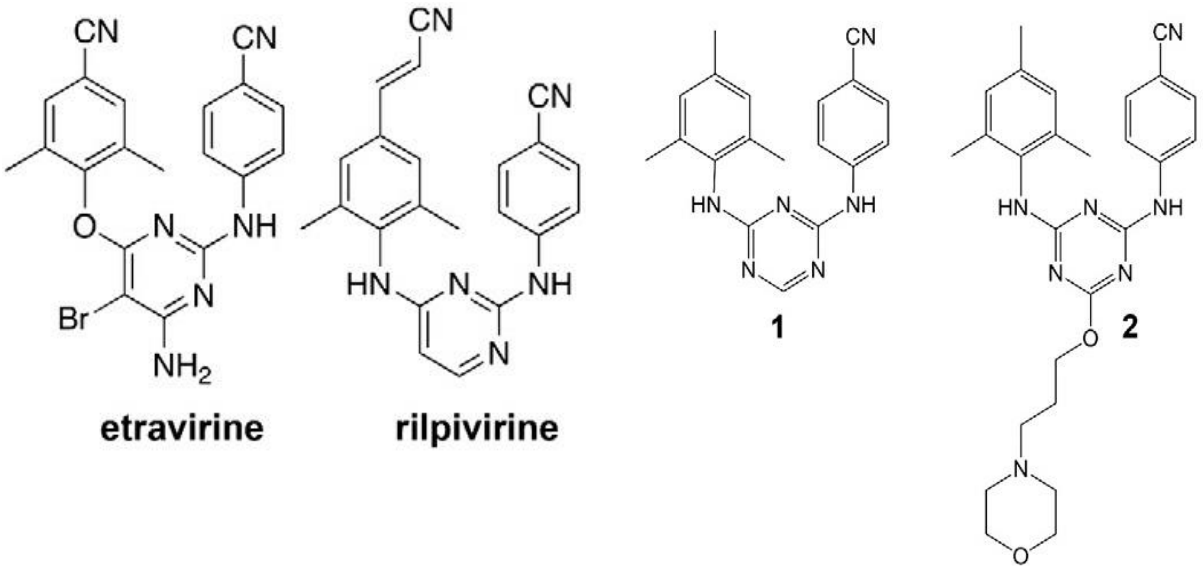


Figure 3: Ribbon representation of HIV-1 RT in a complex with nucleic acid. The fingers, palm, thumb, connection, and RNase H subdomains of the p66 subunit are shown in blue, red, green, yellow, and orange, respectively. The p51 subunit is shown in dark brown. The template and primer DNA strands are shown in light gray and dark gray, respectively. [13]

There are two main types of anti-AIDS/HIV drugs differing in structure and mechanism of the action targeting RT enzyme – nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). NNRTIs, an important class of antiretroviral therapy, bind to hydrophobic RT cavity 10Å away from polymerase active site causing conformational change and thus loss of catalytic function in the RT enzyme. There is a

need for new NNRTIs that are less toxic and in particular inhibit HIV-1 strains resistant to NNRTIs currently present.[13]

The FB methodology combined with polarizable POSSIM formulation can be used to study the binding interactions of diaryltriazine (DATAs) inhibitors.[14] The DATAs inhibitors are structurally related to FDA approved second generation diarylpyrimidine (DAPY) inhibitors with very low solubility such as etravirine and rilpivirine (Figure 4). The study of binding interactions between the structurally modified DATA inhibitors and HIV-RT enzyme can help in development of novel inhibitors with improved pharmacological properties.



(a) (b)

Figure 4: (a) Structures of DAPY analogues of NNRTIs (b) Compound 1 and compound 2 are DATA derivatives. [14]

References

1. Kaminski, G. A.; Ponomarev, S. Y.; Liu, A. B., Polarizable Simulations with Second-Order Interaction Model Force Field and Software for Fast Polarizable Calculations: Parameters for Small Model Systems and Free Energy Calculations. Journal of chemical theory and computation 2009, 5 (11), 2935-2943.

2. Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D., Empirical relationships between protein structure and carboxyl pKa values in proteins. Proteins: Structure, Function, and Bioinformatics 2002, 48 (2), 388-403.

3. Li, H.; Robertson, A. D.; Jensen, J. H., The determinants of carboxyl pKa values in turkey ovomucoid third domain. Proteins: Structure, Function, and Bioinformatics 2004, 55 (3), 689-704.

4. (a) Forsyth, W. R.; Gilson, M. K.; Antosiewicz, J.; Jaren, O. R.; Robertson, A. D., Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. Biochemistry 1998, 37 (24), 8643-8652. (b) Fisher, B. M.; Schultz, L. W.; Raines, R. T., Coulombic effects of remote subsites on the active site of ribonuclease A. Biochemistry 1998, 37 (50), 17386-17401. (c) Quirk, D. J.; Raines, R. T., His... Asp catalytic dyad of ribonuclease A: histidine pK a values in the wild-type, D121N, and D121A enzymes. Biophysical journal 1999, 76 (3), 1571-1579. (d) Edgcomb, S. P.; Murphy, K. P., Variability in the pKa of histidine side-chains correlates with burial within proteins. Proteins: Structure, Function, and Bioinformatics 2002, 49 (1), 1-6.

5. (a) Laurents, D. V.; Huyghues-Despointes, B. M.; Bruix, M.; Thurlkill, R. L.; Schell, D.; Newsom, S.; Grimsley, G. R.; Shaw, K. L.; Treviño, S.; Rico, M., Charge–charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI= 3.5) and a basic variant (pI= 10.2). Journal of molecular biology 2003, 325 (5), 1077-1092. (b) Huyghues-Despointes, B. M.; Thurlkill, R. L.; Daily, M. D.; Schell, D.; Briggs, J. M.; Antosiewicz, J. M.; Pace, C. N.; Scholtz, J. M., pK values of histidine residues in ribonuclease Sa: effect of salt and net charge. Journal of molecular biology 2003, 325 (5), 1093-1105.

6. Stinson, S. “New Natural Products Have Unusual Structures”, *Chem. Eng. News*, *73*, 29-29, 1995.

7. (a) Prendergast, G. C., Farnesyltransferase inhibitors: antineoplastic mechanism and clinical prospects. Current opinion in cell biology 2000, 12 (2), 166-173. (b) Ohkanda, J.; Knowles, D. B.; Blaskovich, M. A.; Sebti, S. M.; Hamilton, A. D., Inhibitors of protein farnesyltransferase as novel anticancer agents. Current topics in medicinal chemistry 2002, 2 (3), 303-323. (c) Adjei, A. A.; Rowinsky, E. K., Novel anticancer agents in clinical development. Cancer biology & therapy 2003, 2 (sup1), 4-14. (d) Andresen, B. M.; Couturier, M.; Cronin, B.; D'Occhio, M.; Ewing, M. D.; Guinn, M.; Hawkins, J. M.; Jasys, V. J.; LaGreca, S. D.; Lyssikatos, J. P., Streamlined processes for the synthesis of a farnesyl transferase inhibitor drug candidate. Organic process research & development 2004, 8 (4), 643-650.

8. (a) Dabrah, T. T.; Kaneko, T.; Massefski, W.; Whipple, E. B., CP-225,917 and CP-263,114: Novel ras farnesylation inhibitors from an unidentified fungus. 2. Structure elucidation. Journal of the American Chemical Society 1997, 119 (7), 1594-1598. (b) Carlacci, L., Conformational analysis of a farnesyltransferase peptide inhibitor, CVIM. Journal of computer-aided molecular design 2000, 14 (4), 369-382.

9. (a) Park, H.-W.; Boduluri, S. R.; Moomaw, J. F.; Casey, P. J.; Beese, L. S., Crystal structure of protein farnesyltransferase at 2.25 angstrom resolution. Science 1997, 275 (5307), 1800-1805. (b) Long, S. B.; Casey, P. J.; Beese, L. S., Cocrystal structure of protein farnesyltransferase complexed with a farnesyl diphosphate substrate. Biochemistry 1998, 37 (27), 9612-9618. (c) Dunten, P.; Kammlott, U.; Crowther, R.; Weber, D.; Palermo, R.; Birktoft, J., Protein farnesyltransferase: structure and implications for substrate binding. Biochemistry 1998, 37 (22), 7907-7912. (d) Zahn, T. J.; Eilers, M.; Guo, Z.; Ksebati, M. B.; Simon, M.; Scholten, J. D.; Smith, S. O.; Gibbs, R. A., Evaluation of isoprenoid conformation in solution and in the active site of protein-farnesyl transferase using carbon-13 labeling in conjunction with solution-and solid-state NMR. Journal of the American Chemical Society 2000, 122 (30), 7153-7164. (e) Long, S. B.; Casey, P. J.; Beese, L. S., The basis for K-Ras4B binding specificity to protein farnesyl-transferase revealed by 2 Å resolution ternary complex structures. Structure

2000, 8 (2), 209-222. (f) Reid, T. S.; Long, S. B.; Beese, L. S., Crystallographic analysis reveals that anticancer clinical candidate L-778,123 inhibits protein farnesyltransferase and geranylgeranyltransferase-I by different binding modes. Biochemistry 2004, 43 (28), 9000-9008.

10. Kaminski, G. A.; Jorgensen, W. L., Host–guest chemistry of rotaxanes and catenanes: application of a polarizable all-atom force field to cyclobis (paraquat-p-phenylene) complexes with disubstituted benzenes and biphenyls†. Journal of the Chemical Society, Perkin Transactions 2 1999, (11), 2365-2375.

11. Sharma, I.; Kaminski, G. A., Calculating pKa values for substituted phenols and hydration energies for other compounds with the first-order fuzzy-border continuum solvation model. Journal of computational chemistry 2012, 33 (30), 2388-2399.

12. https://www.aids.gov/hiv-aids-basics/hiv-aids-101/global-statistics/

13. Sarafianos, S. G.; Marchand B.; Das K.; Himmel D. M.; Parniak M. A.; Hughes S. H.; Arnold E., Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition. Journal of Molecular Biology 2009, 385 (3) 693-713

14. Mislak, A. C.; Frey, K. M.; Bollini, M.; Jorgensen, W. L.; Anderson, K. S., A mechanistic and structural investigation of modified derivatives of the diaryltriazine class of NNRTIs targeting HIV-1 reverse transcriptase. Biochimica et Biophysica Acta (BBA)-General Subjects 2014, 1840 (7), 2203-2211.