John von Neumann Institute for Computing

Simulating Biomolecules in Cellular Environments

M. Feig, S. Tanizaki, J. Chocholoušová, M. Sayadi,
J. W. Clifford, B. D. Connelly, S. Mukherjee, S. M. Law

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 23-30, 2008.

http://www.fz-juelich.de/nic-series/volume40

Simulating Biomolecules in Cellular Environments

Michael Feig[1,2], **Seiichiro Tanizaki**[1,5], **Jana Chocholoušová**[1,6], **Maryam Sayadi**[2], **Jacob W. Clifford**[3], **Brian D. Connelly**[4], **Shayantani Mukherjee**[1], **and Sean M. Law**[1]

[1] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA
E-mail: {feig, shayan, slaw}@msu.edu

[2] Department of Chemistry, Michigan State University
E-mail: sayadima@msu.edu

[3] Department of Physics, Michigan State University
E-mail: jacobc@msu.edu

[4] Department of Computer Science and Engineering, Michigan State University
E-mail: connel42@cse.msu.edu

[5] *current address*: Department of Chemistry and Biochemistry, University of Texas, Arlington, Arlington, TX 76019, USA
E-mail: tanizaki@uta.edu

[6] *current address*: Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, 166 10 Prague 6, Czech Republic
E-mail: chocholousova@uochb.cas.cz

Simulations of biomolecules with realistic representations of cellular environments remain challenging. Implicit solvent methods can reduce the system degrees of freedom and accelerate conformational sampling. Implicit solvent methods can be developed based on a decomposition of the solvation free energy into electrostatic and non-polar contribution. The electrostatic contribution based on continuum electrostatics theory can be conveniently calculated according to the generalized Born (GB) formalism. While the GB formalism is well established in aqueous solvent, applications to dense cellular environments and heterogeneous biological membrane environments are discussed.

1 Introduction

Biological function is often not fully understood until a picture of biomolecular structure and dynamics at the atomic level is developed. Structures of proteins and nucleic acids have become widely available from X-ray crystallography and NMR spectroscopy while atomic-level insight into single molecule dynamics has been gained primarily from computational molecular dynamics studies. Molecular dynamics simulations are well established for the study of single biomolecules over nanosecond time scales, but it remains challenging to study larger biomolecular complexes and longer, biologically more relevant time scales[1].

A significant part of the computational cost of molecular dynamics simulations stems from solvent-solvent and solute-solvent interactions. In the canonical approach the solvent environments is represented in an explicit fashion often resulting in systems with many more solvent atoms than solute atoms. One approach for accelerating simulations of biomolecules involves the application of mean-field descriptions of solute-solvent interactions instead of explicit solvent[2]. A common strategy is the decomposition of the solvation

free energy into polar and non-polar components according to equation 1:

$$\Delta G_{\text{solvation}} = \Delta G_{\text{solvation,polar}} + \Delta G_{\text{solvation,non-polar}} \quad (1)$$

The polar comntribution of the solvation free energy is due to electrostatic solute-solvent interactions and can be obtained by invoking continuum electrostatic theory[3]. In this formalism, a solvated biomolecular system may be described as a set of explicit solute charges embedded in a low-dielectric cavity that is surrounded by a continuum high-dielectric environment. Such a system is described rigorously by the Poisson equation 2, which relates the electrostatic potential ϕ to a distribution of charges, ρ, and dielectric constants ϵ.

$$\nabla \cdot [\epsilon(r) \nabla \phi(r)] = -4\pi\rho(r) \quad (2)$$

It turns out that direct solution of the Poisson equation with finite difference methods is also computationally expensive. Instead, the empirical generalized Born (GB) formalism is more commonly employed to approximate the electrostatic solvation free energy from Poisson theory at a fraction of the computational cost[4].

Many flavors of the GB formalism have been proposed over the last two decades. All are based on the GB equation 3 proposed orignally by Still et al.[5]:

$$\Delta G_{\text{solvation,GB}} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j e^{-r_{ij}^2 / F \alpha_i \alpha_j}}} \quad (3)$$

Different GB implementations vary in the calculation of the GB radii α_i. Acoording to the Coulomb field approximation, the α_i are essentially obtained from an integral of $1/r^4$ over the solute cavity[6]. The most accurate GB methods estimate the integral directly and include additional correction terms to account for deficiencies of the Coulomb field approximation in larger molecules[7].

The non-polar contribution to the solvation free energy consists of the cost of solute cavity formation and contributions from solute-solvent van der Waals interactions. These two terms may be considered separately but are often combined into a single simple term based on the solvent-accessible surface area (SASA) according to Eq. 4[8] with typical values of γ ranging from $5 - 30$ cal/mol/Å^2.[8,9]

$$\Delta G_{\text{solvation, non-polar}} = \gamma \cdot SASA \quad (4)$$

2 Simulations of Biomolecules in Implicit Aqueous Solvent

Simulations of biomolecules in implicit solvent become possible by simply adding the solvation free energy to a vacuum molecular mechanics potential with the partial atomic charges used in the calculation of the electrostatic solvation free energy taken from the force field according to Eq. 5. The resulting forces may then be integrated according to Langevin dynamics[10], Eq. 6, to obtain trajectories that are coupled to solvent through stochastic collisions R and frictional forces according to the friction coefficient γ.

$$U_{implicit} = U_{MM} + \Delta G_{\text{solvation,GB}} + \gamma \cdot \text{SASA} \quad (5)$$

$$F = -\nabla U_{implicit} + R - \gamma \dot{r}(t) = m\ddot{r}(t) \quad (6)$$

Implicit solvent simulations represent a compromise between computational efficiency and the level of realism that can be achieved with a mean-field solvent representation. Model

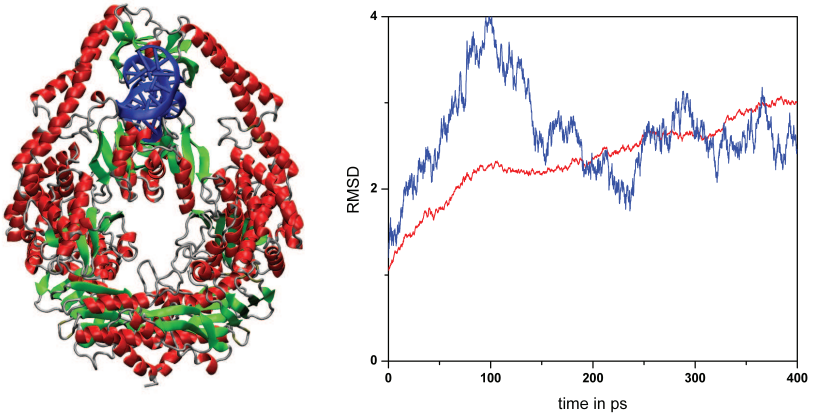

Figure 1. Left: Crystal structure of MutS in complex with DNA (PDB ID: 1E3M). Right: RMSD of C_α atoms (red) and DNA phosphorous atoms (blue) from the experimental structure in molecular dynamics simulations with implicit solvent using the GBMV method in CHARMM[7].

deficiencies may materialize due to the continuum nature of the implicit solvent model and simplicity of the non-polar term or due to limitiations in approximating the continuum electrostatic model with a particular GB implementation.

The quality of implicit solvent simulations may be evaluated by comparing long molecular dynamics simulations with implicit and explicit solvent with experimental data[11]. Very similar conformational sampling is observed in long simulations of protein G and ubiquitin over tens of nanoseconds. Average root mean square deviations (RMSD) from the experimental structures from X-ray crystallography of less than 1 Å with both implicit and explicit solvent[11].

Stable implicit solvent simulations of larger complexes and of nucleic acids[13] are also possible. As an example, simulations of the MutS dimer in complex with mismatched DNA is shown in Figure 1. The simulations are relatively short but reach only 3 Å after 400 ps which is considered good for such a large complex with flexible domains. Explicit solvent simulations of MutS reached a similar deviation from the X-ray structure within the first few hundred picoseconds (data not shown). Successful implicit solvent simulations of nucleic acid systems are remarkable because of strong solute-solvent interactions with the poly-ionic nucleic acids.

It turns out that the GBMV implicit solvent method used here does not offer any computational advantage on a time per integration step basis for the large MutS system because the ratio of solvent to solute atoms decreases with increasing solute size for a single, approximately spherical solute molecule. Previous timing tests have found that implicit solvent simulations with the relatively expensive but accurate GBMV method are only faster for single proteins with up to 200-300 residues[14]. However, implicit solvent offers additional advantages, in particular the ability to traverse conformational space more rapidly when using Langevin dynamics with low friction coefficients[11]. Furthermore, implicit solvent is the only practical solution for simulations of multiple freely diffusing biomolecules within a given solvent environment.

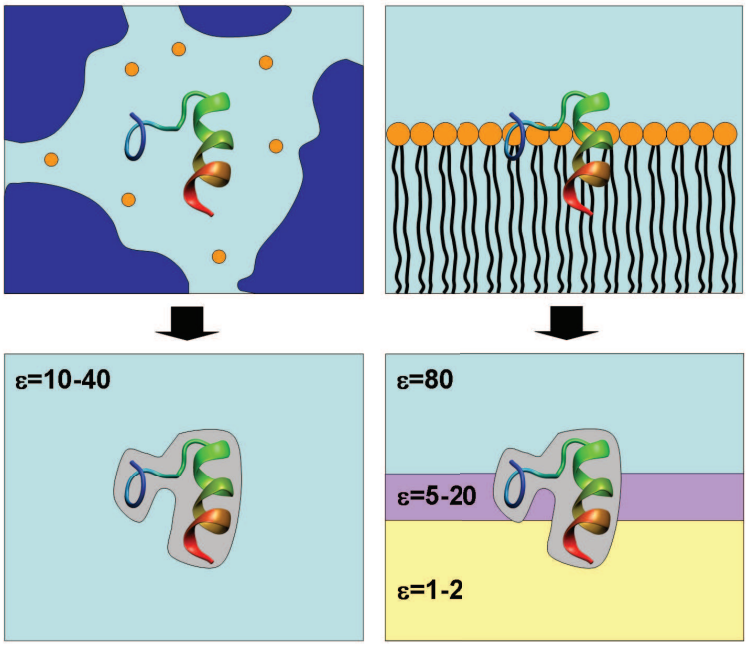

Figure 2. Schematic illustration of implicit modeling of dense cellular environments (left) and heterogeneous membrane bilayers (right) with continuum electrostatics.

3 Simulations of Biomolecules in Implicit Cellular Environments

Dilute aqueous solvent is generally not a good model of crowded cellular environments[15]. The dense concentration of biomolecules and co-solvents presents a complex solvent environment with reduced polarizability compared to water and steric hindrance due to crowding. As a first approximation, dense cellular environments can be modeled in an implicit fashion by assuming a reduced dielectric constant[17] and an increased cost of cavity formation to reflect crowding effects (see Fig. 2). This raises the question how the conformational sampling of peptides and proteins varies in environments with reduced dielectric response. Based on typical dielectric constants of proteins, the effects of co-solvents, and a dielectric modulation of water[16] in crowded environments, one may estimate that the effective dielectric constant of dense cellular environments lies in the range of $\epsilon = 10 - 40$.

Continuum electrostatics methods can readily accommodate a reduced dielectric constant of the environment. Based on physical insight it is expected that reduced dielectric screening enhances charge-charge interactions and in particular the formation of secondary structure elements through hydrogen bonding while the formation of hydrophobic cores becomes less favorable in low dielectric environments. Implicit solvent simulations with dielectric constants between 5 and 80 using a slightly modified GB formalism have confirmed these assumptions for poly-alanine and the amphipathic peptide melittin. However, the simulation results also indicate that even relatively minor changes in the dielectric constant can affect the conformational sampling of melittin in more subtle ways[17].

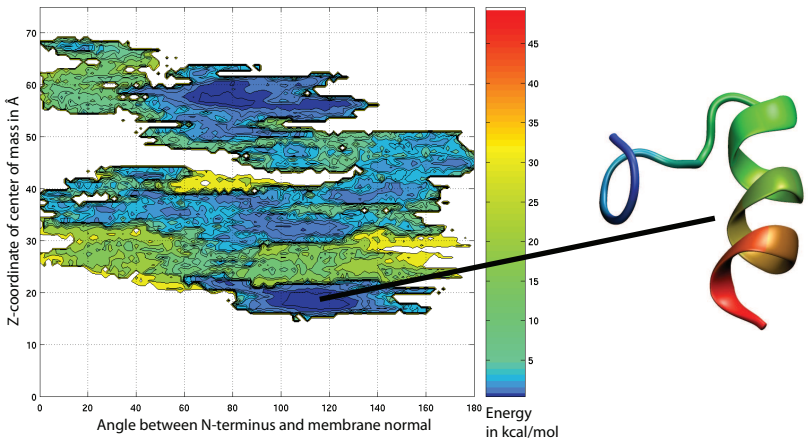

Figure 3. Conformational sampling of influenza fusion peptide with neutral N- and C-termini from temperature replica exchange simulations with the HDGB implicit membrane model. Each of the eight replicas, spaced from 300 to 500K, was run for 15 ns. The potential of mean force in kcal/mol is shown as a function of the angle of the N-terminal part of the peptide relative to the membrane normal and the z-coordinate of the center of mass in Å with z=0 corresponding to the membrane center. The dominant conformation is shown on the right.

4 Simulations of Biomolecules in Implicit Membrane Environments

Biological membrane environments involve heterogeneous environments with a hydrophobic core and a polar surrounding aqueous solvent environment. Implicit membrane models therefore require a spatially varying electrostatic and non-polar contribution to the solvation free energy. The electrostatic solvation free energy can be calculated based on a layered dielectric system[18] with a low dielectric core with ϵ near 1, an intermediate dielectric region near the membrane-water interface, and a high delectric environment elsewhere. Such a model is readily implemented with Poisson theory but presents challenges for the standard GB formalism. The application of heterogeneous dielectric environments becomes possible after introduction of an effective dielectric profile $\epsilon(z)$ and application of the modified GB equation 7:

$$\Delta G_{\text{solvation,HDGB}} = -\frac{1}{2}\sum_{i,j}\left(1-\frac{1}{(\epsilon_i+\epsilon_j)/2}\right)\frac{q_i q_j}{\sqrt{r_{ij}^2+\alpha_i\alpha_j e^{-r_{ij}^2/F\alpha_i\alpha_j}}} \qquad (7)$$

The non-polar contribution to the solvation free energy also varies between an essentially zero cost in the membrane interior to a significant cost of cavity formation in water.

Using this scheme, it has been possible to perform implicit solvent simulations of membrane-bound peptides and proteins[18,19] and accurately reproduce explicit solvent membrane insertion profiles of amino acid side chain analogs[18]. As an example of first applications, Fig. 3 shows extensive conformational sampling of influenza fusion peptide near a membrane interface.

5 Summary and Outlook

Implicit solvent formalisms can be used successfully to simulate biomolecules in cellular environments ranging from simple aqueous solvent to dense cellular environments and heterogeneous dielectric environments. These methods open the door for the simulation of sub-cellular processes in atomic detail and over biologically relevant time scales.

Acknowledgments

Financial support from NSF CAREER grant 0447799 and the Alfred P. Sloan Foundation and access to computational resources through the High Performance Computing Center at Michigan State University are acknowledged.

References

1. M. Karplus, *Molecular dynamics simulations of biomolecules*, Acc. Chem. Res. **35**, 321–323, 2002.
2. B. Roux, T. Simonson, *Implicit Solvent Models* Biophys. Chem. **78**, 1–20, 1999.
3. B. Honig, A. Nicholls, *Classical Electrostatics in Biology and Chemistry*, Science **268**, 1144–1149, 1995.
4. M. Feig, C. L. Brooks, III, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*, Curr. Op. Struct. Biol. **14**, 217–224, 2004.
5. W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics*, J. Amer. Chem. Soc. **112**, 6127–6129, 1990.
6. D. Bashford, D. A. Case, *Generalized Born models of macromolecular solvation effects*, Ann. Rev. Phys. Chem. **51**, 129–152, 2000.
7. M. S. Lee, F. R. Salsbury, Jr., C. L. Brooks, III, *Novel generalized Born methods* J. Chem. Phys. **116**, 10606–10614, 2002.
8. D. Sitkoff, K. A. Sharp, B. Honig, *Accurate calculation of hydration free-energies using macroscopic solvent models*, J. Phys. Chem. **98**, 1978–1988, 1994.
9. W. Im, M. S. Lee, C. L. Brooks, III, *Generalized Born model with a simple smoothing function*, J. Comp. Chem. **24**, 1691–1702, 2003.
10. C. L. Brooks, III, M. Berkowitz, S. A. Adelman, *Generalized Langevin theory for many-body problems in chemical-dynamics - gas, surface collisions, vibrational-energy relaxation in solides, and recombination reactions in liquids*, J. Chem. Phys. **73**, 4353–4364, 1980.
11. M. Feig, *Kinetics from implicit solvent simulations of biomolecules as a function of viscosity*, J. Chem. Theory Comput. **3**, 1734–1748, 2007.
12. T. A. Darden, D. York, L. G. Pedersen, *Particle mesh Ewald: An N log(N) method for Ewald sums in large systems*, J. Chem. Phys. **98**, 10089–10092, 1993.
13. J. Chocholousova, M. Feig, *Implicit solvent simulations of DNA and DNA-protein complexes: Agreement with explicit solvent vs. experiment*, J. Phys. Chem. B **110**, 17240–17251, 2006.

14. M. Feig, J. Chocholousova, S. Tanizaki, *Extending the horizon towards the efficient modeling of large biomolecular complexes in atomic detail*, Theor. Chem. Acc. **116**, 194–205, 2006.
15. A. B. Fulton, *How crowded is the cytoplasm?*, Cell **30**, 345–347, 1982.
16. F. Despa, A. Fernandez, R. S. Berry, *Dielectric modulation of biological water*, Phys. Rev. Lett., **93**, 2002.
17. S. Tanizaki, J. W. Clifford, B. D. Connelly, M. Feig, *Conformational sampling of peptides in cellular environments*, Biophys. J. **94**, 747–759, 2008.
18. S. Tanizaki, M. Feig, *A generalized Born formalism for heterogeneous dielectric environments: Application to the implicit modeling of biological membranes*, J. Chem. Phys. **122**, 124706, 2005.
19. S. Tanizaki, M. Feig, *Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model*, J. Phys. Chem. B **110**, 548–556, 2006.