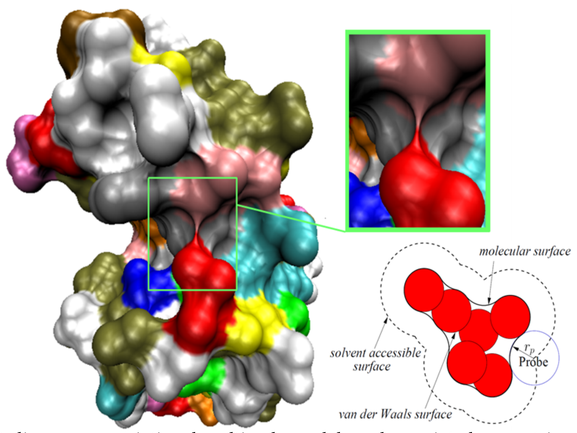
Variational Multiscale Modeling of Biomolecular Complexes

Kelin Xia[1], Xin Feng[2], Yiying Tong[2], and Guo-wei Wei[1,3]

[1]*Department of Mathematics, Michigan State University, MI 48824, USA*
[2]*Department of Computer Science and Engineering, Michigan State University, MI 48824, USA*
[3]*Department of Biochemistry and Molecular Biology, Michigan State University, MI 48824, USA*

Introduction

Multiscale modeling is of paramount importance to the understanding of biomolecular structure, function, dynamics and transport. Geometric modeling provides structural representations of molecular data from the Protein Data Bank (PDB) and the Electron Microscopy Data Bank (EMDB). Commonly used geometric models, such as molecular surface (MS), van der Waals surface, and solvent accessible surface are *ad hoc* devision of solvent and solute regions and lead to troublesome geometric singularities, as demonstrated in the figure below. At fundamental level, solvent and solute electron densities overlap each other and there is no sharp solvent-solute interface.



Surface definitions and defects

Molecular surface of protein (ID:1PPL) and its geometric singularity.

We discuss our variational multiscale models and associated geometric modeling of biomolecular complexes, based on differential geometry of surfaces and geometric measure theory. Our models give rise to singularity-free surface representation, curvature characterization, electrostatic mapping, solvation energy and binding affinity analysis of biomolecules.

Variational Multiscale Models

Almost all the biological processes in a living cell occur in aqueous surroundings, because up to 65%-90% of human cell mass is water. In our multiscale solvation model, a total free energy functional is constructed to include polar and nonpolar free energies.

- **The free energy functional** of the solvation process is,

$$G_{\text{total}}[S,\Phi] = \int \left\{ \gamma|\nabla S| + pS + S\left[-\frac{\epsilon_m}{2}|\nabla\Phi|^2 + \Phi\ \rho_m\right] + (1-S)\left[-\frac{\epsilon_s}{2}|\nabla\Phi|^2 - k_BT\sum_\alpha \rho_{\alpha 0}\left(e^{-\frac{q_\alpha\Phi+U_\alpha-\mu_{\alpha 0}}{k_BT}}-1\right)\right]\right\} d\mathbf{r},$$

where γ is the surface tension, S is the hypersurface and can be viewed as a characteristic function of the solute domain, p is the hydrodynamic pressure, and U_α denotes the solvent-solute non-electrostatic interactions, such as the van der Waals interaction. Here Φ is the electrostatic potential, ϵ_s and ϵ_m are the dielectric constants of the solvent and solute, respectively, ρ_m represents the fixed charge density of the solute and k_BT is the thermal energy. By applying the variational principle to minimize the total solvation free energy with respect to Φ and S, two equations are generated.

- **The Generalized Poisson-Boltzmann Equation** describes the electrostatic potential,

$$-\nabla\cdot(\epsilon(S)\nabla\Phi) = S\rho_m + (1-S)\sum_\alpha q_\alpha\rho_{\alpha 0}e^{-\frac{q_\alpha\Phi+U_\alpha-\mu_{\alpha 0}}{k_BT}},$$

where $\rho_{\alpha 0}$ denotes the bulk concentration, and $\mu_{\alpha 0}$ is a relative reference chemical potential.

- **The generalized Laplace-Beltrami equation** governs the surface formation under potential driven geometric flows,

$$\frac{\partial S}{\partial t} = |\nabla S|\left[\nabla\cdot\left(\gamma\frac{\nabla S}{|\nabla S|}\right) + V_1\right],$$

where the potential driven term V_1 is given by

$$V_1 = -p + \frac{\epsilon_m}{2}|\nabla\Phi|^2 - \Phi\ \rho_m - \frac{\epsilon_s}{2}|\nabla\Phi|^2 - k_BT\sum_\alpha \rho_{\alpha 0}\left(e^{-\frac{q_\alpha\Phi+U_\alpha-\mu_{\alpha 0}}{k_BT}}-1\right).$$

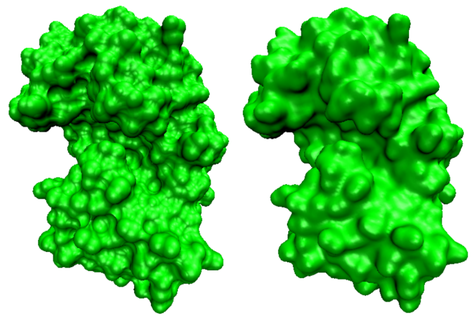
The external potential term can be adjusted to take into consideration of other effects. For instance, if we assume that the biomolecular system is far from equilibrium and account for the chemical potential related energy in the free energy functional, we result in a potential driven term of the form

$$V_2 = -p + U + \frac{\epsilon_m}{2}|\nabla\Phi|^2 - \Phi\ \rho_m - \frac{\epsilon_s}{2}|\nabla\Phi|^2 + \Phi\sum_\alpha \rho_\alpha q_\alpha + \sum_\alpha\left[k_BT\left(\rho_\alpha\ln\frac{\rho_\alpha}{\rho_{\alpha 0}} - \rho_\alpha + \rho_{\alpha 0}\right) - \mu_{\alpha 0}\rho_\alpha\right].$$

Our model can be easily modified to account for other physical interactions.

Surface and Electrostatic Analysis

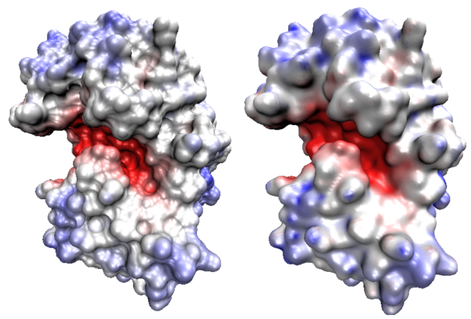
The solution of the generalized Laplace-Beltrami equation gives rise to biomolecular surfaces of controllable resolutions. These surfaces are free from geometric singularities. Their resolutions can be tuned to avoid the local atomic fluctuation in the curvature analysis.



Variational multiresolution surface

A comparison of the molecular surface (Left Chart) and our variational surface (Right Chart) of a protein (ID:1PPL).

The Generalized Poisson-Boltzmann equation is solved to obtain electrostatic potential, which offers an indication of possible binding sites.



Electrostatic Potential

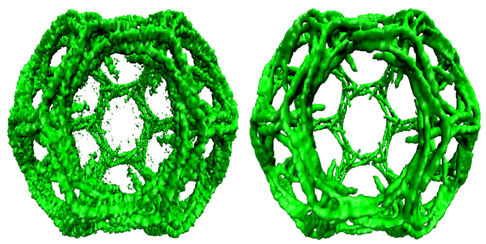
A comparison of the electrostatic potentials on the molecular surface (Left Chart) and our variational surface (Right Chart) of a protein (ID:1PPL).

EMD Data Preprocessing

The EMDB collects protein structural information from electron microscopy. The associated data are in a volumetric format and usually suffer from low signal-to-noise rate (SNR). Therefore, the noise reduction of EMDB data is mandatory. High order geometric flows, which can more efficiently suppress the high-frequency components, are employed for EMDB data analysis. A specific form of arbitrarily high order geometric PDEs is given by,

$$\frac{\partial S}{\partial t} = (-1)^q\sqrt{g(|\nabla\nabla^{2q}S|)}\nabla\cdot\left(\frac{\nabla(\nabla^{2q}S)}{\sqrt{g(|\nabla\nabla^{2q}S|)}}\right) + P(S,|\nabla S|),$$

where $g(|\nabla\nabla^{2q}S|) = 1 + |\nabla\nabla^{2q}S|^2$ is the generalized Gram determinant. When $q=0$, we arrive at a generalized mean curvature flow.



EMDB data noise reduction

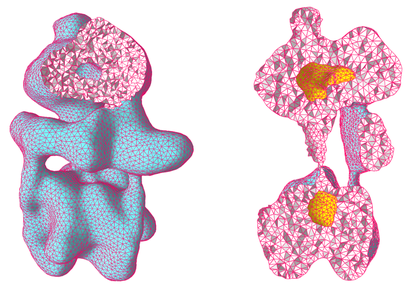
Noise reduction of the EMDB data (ID emd5119).

References

[1] Z. Chen, N. Baker, and G. W. Wei, *Differential geometry based solvation model I: Eulerian formulation*, Journal of Computational Physics, 229: 8231-8258 (2010).
[2] G. W. Wei, Q. Zheng, Z. Chen, and K. L. Xia, *Variational multiscale models for charge transport*, SIAM Review. 54, 699-754 (2012).
[3] X. Feng, K. L. Xia, Y. Tong, and G. W. Wei, *Geometric modeling of organelles, subcellular structures and multiprotein complexes*, International Journal for Numerical Methods in Biomedical Engineering, 28, 1198-1223 (2012)
[4] K. L. Xia, X. Feng, Z. Chen, Y. Tong, and G. W. Wei, *Multiscale geometric modeling of macromolecules I: Cartesian representation*,Journal of Computational Physics, 257, 912-936 (2014).

Geometric Modeling — Meshing

The Lagrangian representation of protein surfaces derived from our variational multiscale models can be used for volumetric meshing. The Delaunay triangulation algorithm is implemented.

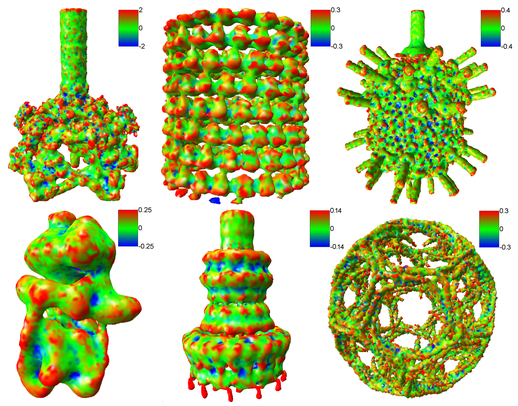


Mesh detail

The meshing of protein complex (emd1590). High quality mesh is constructed with the Delaunay triangulation in TetGen.

Geometric Modeling — Curvature

The surfaces are characterized by using Gaussian, mean, minimal principal and maximal principal curvatures, which indicate potential binding sites.

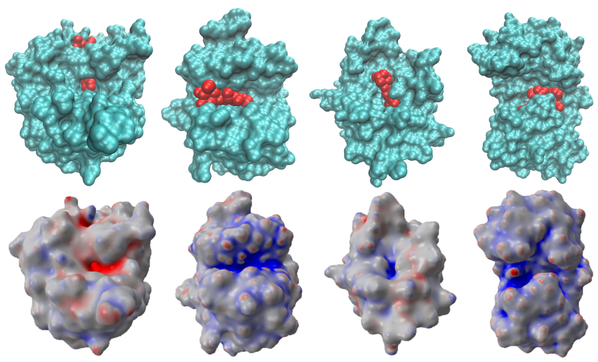


Mean curvature

Mean curvature maps of six multiprotein complexes.

Binding-Site Prediction

The product of minimal curvature and electrostatic potential indicates binding sites.



Binding site prediction

The upper row shows experimental binding sizes, and the lower row illustrates our predictions.

Conclusion

Our variational multiscale modeling demonstrates a great promising for the geometric, physical, and biological analysis of biomolecular complexes.

Acknowledgment

This work was partially supported by NSF grants CCF-0936830, IIS-0953096 and DMS-1160352 (FRG), and NIH grants R01GM-090208.

http://www.math.msu.edu/~wei/ xiakelin@msu.edu fengxin@msu.edu ytong@msu.edu wei@math.msu.edu