

# COLLABORATION PLAN AND RESEARCH STATEMENT

NATHAN A. BAKER

## 1. INTRODUCTION

I am a Laboratory Fellow in the Applied Statistics and Computational Modeling Group at Pacific Northwest National Laboratory (PNNL). Prior to moving to PNNL, I was a tenured Associate Professor of Biochemistry and Molecular Biophysics in the Center for Computational Biology at Washington University School of Medicine in St. Louis. My research focuses on the development of new algorithms and mathematical methods in biophysics, nanotechnology, and informatics. Current research projects include the development of computational methods for modeling solvation in biomolecular systems, mathematical methods for mesoscale materials modeling, and new methods for machine learning in signature discovery [SDI, BBB<sup>+</sup>13]. This research summary and collaboration plan will focus on the first two topics. My research is currently funded by the National Institutes of Health and the Department of Energy.

I am very interested in the opportunity to engage more closely with Brown Applied Mathematics faculty beyond my ongoing interactions with George Karniadakis and Martin Maxey. The most immediate area for potential collaboration is with Anastasios Matzavinos. While at Washington University, I had begun to develop systems biology models for nitric oxide species in blood and tissue for the purpose of understanding and predicting the role of oxidative stress in shock. In particular, I was developing very stiff high-dimensional reaction-diffusion equations to describe diffusion and chemical interconversion of oxygen and nitric oxide derivatives across red blood cell and blood vessel boundaries. This work ended early because of student issues and my move to PNNL; however, I am anxious to restart it. Anastasios' interests in modeling oxygen diffusion through tissue are complementary to the nitric oxide modeling problem: nitric oxide plays an important role in vasodilation and the resulting oxygenation of blood and the surrounding tissue. I am looking forward to talking with Anastasios about possible collaborations in this area. Additionally, I have common research interests with several other Brown DAM faculty. For example, Basilis Gidas' interests in biomolecular structural informatics fits well with my work on using physical (PDE-based) quantities to gain additional insight into protein structure and function [LYZ<sup>+</sup>14, ZBK<sup>+</sup>06]. Chip Lawrence's research in bioinformatics fits with my past as well as ongoing work; I currently collaborate with his former student LeeAnn McCue at PNNL. Finally, Matt Harrison's interests in pattern theory fits well with the work that I've led for the PNNL Signature Discovery Initiative [SDI].

There are several mechanisms through which I hope to interact with Brown Applied Mathematics faculty. First, joint proposals and projects are an important vehicle for supporting collaborative research. I am interested in pursuing NSF and NIH proposals with Brown faculty in the mathematical biology areas highlighted above. In particular, I have successfully competed for funding through NSF DMS 13-570 in the past and have found it to be an excellent mechanism for supporting mathematical biology research. Additionally, I am interested in exploring new funding opportunities from DARPA and IARPA with Brown DAM faculty. Second, I would be very interested in co-mentoring Brown Applied Mathematics students as part of the collaborative research described

above. Finally, I am interested in future opportunities to give seminars or teach short courses on mathematical biology topics.

## 2. ONGOING RESEARCH

**2.1. Models and methods for biomolecular solvation and electrostatics.** I have an ongoing interest in the development of improved theories and computational methods for biomolecular solvation and electrostatics. At the molecular level, nearly all biological interactions occur in the presence of water and ions. Therefore, accurate models for interactions between proteins, nucleic acids, and small molecules (e.g., drugs or metabolites) requires accurate treatments of the influence of water and ions on these systems. In principle, the most accurate solvation models would treat solvent in its full atomic or molecular detail. However, such *explicit solvent* methods are often infeasible for use in biomolecular systems because of the number of degrees of freedom involved and the computational expense of generating thermodynamically meaningful averages over all solvent and ion positions in the system. *Implicit solvent* models replace the solvent and ionic degrees of freedom with continua and thus reduce the computational expense of the problem by significantly reducing the dimensionality of the system. My research focuses on the theoretical formulation of accurate and physically realistic implicit solvent models and the development of robust numerical approaches to solve the resulting equations.

For the past 15 years, I have worked on numerical methods for the Poisson-Boltzmann equation, one of the most popular implicit solvent models. The Poisson-Boltzmann equation is a second-order elliptic nonlinear partial differential equation (PDE) with significant complexity introduced by the complicated geometry of biomolecules and the stiff nature of the second-order (dielectric) coefficient which changes by several orders of magnitude across the biomolecular surface [RCT<sup>+</sup>12]. Our APBS software package [APBS] implements a variety of numerical methods for the Poisson-Boltzmann equation and related models of solvation. This software, developed in collaboration with several research groups in biology and mathematics, includes finite element [HBW00, BHW00, BSHM01], finite difference multigrid [BSJ<sup>+</sup>01], and boundary element approaches for solving the Poisson-Boltzmann and related equations. The goal of APBS is to make sophisticated numerical methods available to biological researchers unfamiliar with the underlying equations. These software packages have been applied to a wide range of biological research applications and continue to be widely used by a community of approximately 25,000 users worldwide. One specific application my group has emphasized is the accurate prediction of protein titration states [CFH<sup>+</sup>11, GHPB14], a challenging  $2^N$ -variable discrete optimization problem for which we are currently exploring graph theoretic solution approaches that reduce computational effort to  $\mathcal{O}(N^4)$ .

In addition to continuing to support the integration of new mathematical methods into APBS and PDB2PQR, much of my current research focuses on the improvement of the underlying solvation models. Most recently, I have enjoyed a very productive collaboration with Prof. Guowei Wei (Michigan State Univ.) to improve implicit solvent models by creating a self-consistent model which couples the polar solvation modeled by the Poisson-Boltzmann equation with important nonpolar terms such as dispersion and cavity formation energies [WB06]. The resulting model has the form of coupled nonlinear Laplace-Beltrami equations for the solvent distribution and the traditional Poisson-Boltzmann equation for the electrostatic potential [CBW10, CBW11]. These equations are solved using a geometric flow approach and have demonstrated significantly improved performance in predicting solvation energies for a range of small molecules [DCHL<sup>+</sup>13, TCC<sup>+</sup>13]. My research continues to focus on the expansion of the underlying theory (e.g., development of nonlocal integro-differential models for ion correlation) as well the development of more efficient numerical methods for solving the resulting equations to enable its application to macromolecules. Recently, we have begun a very productive collaboration with Profs. Lois Pollack (Cornell Univ.)

and Alexey Onufriev (Virginia Tech Univ.) to test these improved theories on experimental data collected by the Pollack lab on very challenging nucleic acid molecular complexes with high charge densities [TPK<sup>+</sup>14]. These tests have exposed important deficiencies in several implicit solvent models and provide an essential dataset for our ongoing effort to improve solvation models.

**2.2. Mathematical methods for mesoscale materials modeling.** I am the PNNL program lead for the Collaboratory on Mathematics for Mesoscopic Modeling of Materials (CM4), led by Prof. George Karniadakis (PI) [CM4]. The goal of CM4 is the development of general mathematical and numerical methods to model complex materials systems with emergent phenomena that are not currently understood. Such systems have hierarchical interactions within materials that result in unique structural and response properties. The CM4 integrated mathematical approach will connect molecular details to macroscopic understanding of the collective response of materials, their transport, assembly, stability, structure, and phase. The presence of broken symmetries and inhomogeneities (in both time and space) requires detailed modeling of interfaces. In turn, these interfaces require understanding the coupling between scales, as well as the connectivity and correlation in hierarchical structures. CM4 consists of six mathematics research tasks and a task focused on the integration of the resulting mathematical methods in the form of scalable software applied to important materials science research challenges. My role in CM4 is to support this integration by identifying challenge problems which can incorporate the different mathematical methods while addressing interesting research questions in materials science. In addition to guiding all challenge problem tasks, I am actively and directly involved in the following specific research areas.

The first area focuses on the development of uncertainty quantification methods for high-dimensional biomolecular systems. Although biomolecular structures are subject to a wide range of uncertainties, this information is rarely incorporated into computational biology calculations because of the high dimensionality of the structures. We have developed an approach based on generalized polynomial chaos and “active subspace” random variables to incorporate conformational uncertainty into simple geometry-dependent solvation energy calculations [LYZ<sup>+</sup>14]. We are currently expanding this work to assess more complicated PDE-based solvation properties based on the models described in the previous section.

The second area emphasizes development of scalable numerical methods for diffusion and reaction in colloidal and biomolecular systems. Our initial work on this area has focused on use of smoothed particle hydrodynamics, incorporating a new treatment for Robin boundary conditions to simulation ligand binding and reaction in enzyme systems [PDB15]. We are currently expanding these approaches to model large systems of enzymes, as found in bioreactors and related catalytic systems.

Additionally, although not directly funded by CM4, I am also working with CM4 researchers to explore the application CM4-developed methods to the modeling of uncertainty and charge noise in silicon materials used in quantum computing qubit devices.

## REFERENCES

- [APBS] *APBS: Adaptive Poisson-Boltzmann Solver*, <http://www.poissonboltzmann.org>.
- [BBB<sup>+</sup>13] N.A. Baker, J.L. Barr, G. T. Bonheyo, C. A. Joslyn, Krishnaswami K., M. E. Oxley, R. Quadrel, L. H. Sego, M. F. Tardiff, and A. S. Wynne, *Research towards a systematic signature discovery process*, Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on (2013), 301–308.
- [BHW00] N. A. Baker, M. J. Holst, and F. Wang, *Adaptive multilevel finite element solution of the Poisson-Boltzmann equation II. Refinement at solvent-accessible surfaces in biomolecular systems.*, Journal of Computational Chemistry **21** (2000), 1343–1352.

- [BSHM01] N. A. Baker, D. Sept, M. J. Holst, and J. A. McCammon, *The adaptive multilevel finite element solution of the Poisson-Boltzmann equation on massively parallel computers*, IBM Journal of Research and Development **45** (2001), 427–438.
- [BSJ<sup>+</sup>01] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, *Electrostatics of nanosystems: application to microtubules and the ribosome*, Proceedings of the National Academy of Sciences of the United States of America **98** (2001), 10037–10041.
- [CBW10] Z. Chen, N. A. Baker, and G. W. Wei, *Differential geometry based solvation model I: Eulerian formulation*, Journal of Computational Physics **229** (2010), 8231–8258.
- [CBW11] ———, *Differential geometry based solvation model II: Lagrangian formulation*, Journal of Mathematical Biology **63** (2011), 1139–1200.
- [CFH<sup>+</sup>11] T. Carstensen, D. Farrell, Y. Huang, N. A. Baker, and J. E. Nielsen, *On the development of protein  $pK_a$  calculation algorithms*, Proteins **79** (2011), 3287–3298.
- [CM4] *CM4: Collaboratory on Mathematics for Mesoscopic Modeling of Materials*, <http://www.pnnl.gov/computing/cm4>.
- [DCHL<sup>+</sup>13] M. D. Daily, J. Chun, A. Heredia-Langner, G.-W. Wei, and N. A. Baker, *Origin of parameter degeneracy and molecular shape relationships in geometric-flow calculations of solvation free energies*, The Journal of Chemical Physics **139** (2013), 204108.
- [GHPB14] L. J. Gosink, E. A. Hogan, T. C. Pulsipher, and N. A. Baker, *Bayesian model aggregation for ensemble-based estimates of protein  $pK_a$  values*, Proteins **82** (2014), 354–363.
- [HBW00] M. J. Holst, N. A. Baker, and F. Wang, *Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I. Algorithms and examples*, Journal of Computational Chemistry **21** (2000), 1319–1342.
- [LYZ<sup>+</sup>14] H. Lei, X. Yang, B. Zheng, G. Lin, and N. A. Baker, *Quantifying the influence of conformational uncertainty in biomolecular solvation* (2014), Submitted.
- [PDB15] W. Pan, M. Daily, and N. A. Baker, *Numerical calculation of protein-ligand binding rates through solution of the Smoluchowski equation using smooth particle hydrodynamics*, BMC Biophysics (2015), in press.
- [RCT<sup>+</sup>12] P. Ren, J. Chun, D. G. Thomas, M. J. Schnieders, M. Marucho, J. Zhang, and N. A. Baker, *Biomolecular electrostatics and solvation: a computational perspective*, Quarterly Reviews of Biophysics **45** (2012), 427–491.
- [SDI] *The PNNL Signature Discovery Initiative*, <http://signatures.pnnl.gov>.
- [TCC<sup>+</sup>13] D. G. Thomas, J. Chun, Z. Chen, G.-W. Wei, and N. A. Baker, *Parameterization of a geometric flow implicit solvation model*, Journal of Computational Chemistry **34** (2013), 687–695.
- [TPK<sup>+</sup>14] I. S. Tolokh, S. A. Pabit, A. M. Katz, Y. Chen, A. Drozdetski, N. A. Baker, L. Pollack, and A. V. Onufriev, *Why double-stranded RNA resists condensation*, Nucleic Acids Research **42** (2014), 10823–10831.
- [WB06] J. A. Wagoner and N. A. Baker, *Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms*, Proceedings of the National Academy of Sciences of the United States of America **103** (2006), 8331–8336.
- [ZBK<sup>+</sup>06] X. Zhang, C. L. Bajaj, B. Kwon, T. J. Dolinsky, J. E. Nielsen, and N. A. Baker, *Application of new multi-resolution methods for the comparison of biomolecular electrostatic properties in the absence of global structural similarity*, SIAM Multiscale Modeling & Simulation **5** (2006), 1196–1213.

COMPUTATIONAL AND STATISTICAL ANALYTICS DIVISION, PACIFIC NORTHWEST NATIONAL LABORATORY, PO Box 999, MSID K7-20, RICHLAND, WA 99352. PHONE: +1-509-375-3997.

*E-mail address:* [nathan.baker@pnnl.gov](mailto:nathan.baker@pnnl.gov)

*URL:* <http://goo.gl/U9xu4t>