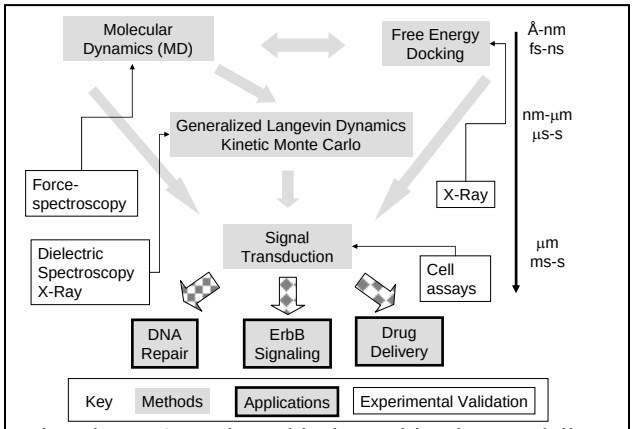
Molecular Systems Biology via Multiscale Modeling and High-Performance Computing

Synopsis of Research Program: In the realm of fundamental and applied sciences, the interface between nanotechnology and biotechnology is evolving rapidly. At this interface, a quantitative description of the underlying fundamental processes is inherently a multiscale problem. In my laboratory, we strive to achieve a multiscale description of equilibrium and dynamic processes associated single biomolecules, biomolecular assemblies, as well as cooperative interactions of multiple biological entities (proteins, nucleic acids, and membranes), of particular relevance to cell signaling pathways implicated in cancers. In the expansive scope of computational biology, our multiscale modeling studies complement genomics and bioinformatics based methods; together, these tool sets provide the foundations for quantitative biological and biomedical research.

Studies in my laboratory are motivated by biological phenomena such as DNA repair, receptor-mediated cellular signal transduction, and targeted drug delivery. The origin of such phenomena can be traced to the interactions prevalent in proteins, nucleic acids and lipids at the atomic or nanoscale, however the actual manifestation of the biophysical and biochemical response is at the mesoscale (~μm, ~ms). Therefore, there is an imminent need for developing a computational technology for addressing such problems at multiple scales. We thus adopt a hierarchical multiscale approach (see flowchart 1) including molecular dynamics (MD), Monte Carlo (MC), mixed quantum mechanics molecular mechanics (QMMM), and free energy/path-based methods (umbrella sampling, transition path sampling) on fully atomistic explicitly solvated biomolecular systems. We also adopt a spatially resolved hybrid deterministic/ stochastic method by combining a field theoretic (continuum) description for cell membrane dynamics and a discrete (lattice) description for the protein dynamics to study dynamical processes in cell membranes mediated by signal transduction. Our multiscale approach is generalizable to a variety of processes in nanobiotechnology and systems biology applications.



Flowchart 1: Hierarchical multiscale modeling; validation is through collaborative experiments.

At Penn, my laboratory has initiated several collaborations with experimentalists to synergize our computational findings. In SOM, we work together with Professor Mark Lemmon of Biochemistry and Biophysics, David Eckmann of Anesthesia, and recently with Xiaowei Xu of Pathology. In SAS, we recently have begun collaborations with Tobias Baumgart and Eric Meggers both in Chemistry. These collaborations and others have earned me secondary appointments in Chemical and Biomolecular Engineering, and in Biochemistry and Biophysics, and memberships in the Genomics and Computational Biology Graduate Group, and in the Institute of Targeted Medicine and Therapeutics (ITMAT). Outside of Penn, I collaborate with Dr. Samuel Wilson, deputy director of NIH/NIEHS, and am initiating a collaboration with Dr. Lorena Beese of Duke University (who is a seminar speaker in BE for spring 2007). These associations compliment our contributions, and together they form the foundations on which we are continuing to expand our research program.

Brief Description of Ongoing Research Areas:

DNA replication/repair: From a single-molecule perspective, we seek to understand how protein architecture leads to catalytic properties such as efficiency and specificity. In the context of polymerases

these traits inherently govern the fidelity — the faithful replication or repair of DNA by polymerases — in cognate (native) and non-cognate (when DNA lesions or carcinogens are present) settings. Structural and energetic aspects of these interactions are well characterized, but the dynamical attributes are not very well appreciated. We strive to delineate how the dynamical characteristics of the polymerase system affects the reaction pathway of correct nucleotide incorporation and how differences in the dynamic behavior impact non-cognate (incorrect or translesion) incorporation. The central hypothesis we would like to test using our multiscale approach is that the context specific preorganization of the catalytic site of the polymerase in the active conformation (i.e., correct versus incorrect nucleotides in native versus damaged DNA) is influenced by the slow dynamical modes of the system and that the coupling of the slow modes with the fast reactive degrees of freedom can help lower the activation barrier during catalysis and possibly even introduce a source of discrimination, thereby influencing error control. Our goals are expected to reveal the causative mechanisms of mutational hot-spots, and to quantify signal transduction in DNA repair.

Receptor-mediated signaling: The ErbB family receptor — which includes the epidermal growth factor receptor (EGFR or ErbB1), ErbB2, ErbB3, and ErbB4 — over-expression generally correlates with occurrence of clinical cancers. Therefore, small molecule receptor tyrosine kinase (RTK) inhibitors for the Erb family RTKs are of significant interest as cancer therapeutics. However, approved drugs such as the inhibitor gefitinib (Iressa) targeting ErbB1 show promise only in a small (demographic) sub-population and clinical responses varied among population samples. Epidemiological studies correlate enhanced sensitivity of the drug to somatic mutations in the ErbB1. Yet, the underlying biochemical basis involving the drug sensitivity to the mutations is still not clear. The ERb family of receptors activates a multi-layered network mediating crucial pathways leading to cell proliferation, differentiation, migration, and metabolism, in response to many related activating growth factors. Thus, a quantitative description of the receptor activation from a molecular perspective, as well as the signal transduction resulting form protein-protein interactions in networks, are both essential to pin-down the origin of inhibitor sensitivity and drug resistance.

The process of ligand-induced dimerization and activation of the extracellular domain is now well understood. Precisely how this ligand-induced dimerization event is coupled to, and leads to, activation of the intracellular tyrosine kinase domain is only now emerging from recent studies. We propose to carry out multiscale simulations to help identify the role, cause, and significance of drug sensitizing mutations of the Erb family RTKs. We seek to computationally study the modes of activation of Erb RTKs, and how the mutation landscape alters the activation mechanism. Our results will thus serve to identify constitutively active mutants, and will help formulate a paradigm to understand the Erb family RTK dependence (and sensitivity) to small molecule inhibitors. Our objective is to optimize small molecule therapeutic inhibition strategies.

Design of nanocarriers for targeted drug delivery: The strategic approach to increase the dose of the drug reaching diseased tissue while simultaneously decreasing the dose reaching normal tissue relies on the technology of injecting the nanocarriers into the blood stream close to the disease tissue. Yet the optimization of a therapeutic intervention based on such targeting is plagued by the complexities induced by the interplay between nanocarriers and cell surface leading to binding and internalization. These are not easily discerned from experiments alone due to critical events occurring at multiple length and time scales. We seek to develop a multiscale model to quantify the key interactions in transvascular drug delivery, namely, nanocarriers binding to endothelial cells and subsequent events leading to nanocarrier internalization. Our modeling integrates the effects of mechanical forces, receptor-ligand interactions, and physicochemical processes such as membrane dynamics as well as intra-membrane (lateral) diffusion to determine the timescale of nanocarrier arrest on the target cell. We employ our simulations as a means to determine the optimal properties for the design of the nanocarriers, namely, nanocarrier size, receptor and ligand surface density, contact surface area between the vehicle and target

cells. The modeling will incorporate and bridge multiple scales (e.g., continuum mechanics, nanoscopic stochastic methods). These results will guide the optimal engineering design, as well as the translation of drug delivery systems for targeted disease treatment.

Teaching Program: Complementing my interdisciplinary research program in quantitative biology, the educational program I am building leverages the existing channels and systems at Penn to build a rigorous and visionary program encompassing theoretical biology, computational biology, and experimental biology. I have an established track record of developing one new course and revamping one existing course in the bioengineering curriculum at Penn: BE559, a multidisciplinary graduate course with a mandatory computational laboratory on multiscale modeling; and BE324, an undergraduate course on physical principles in bioengineering. BE559 typically has student enrolments from diverse disciplines: 4-5 engineering disciplines within SEAS, 2 within SAS, and 1 within SOM. The evaluation scores for this course as been as high as 3.6 for instructor quality, and 3.5 for course quality. It is also notable that the research problems explored in my lab are ported as homework in the computational laboratory for BE559. BE324 provides a rigorous foundation to BE undergraduates and has been revamped from ground-up to be in-line with a two-part quantitative series on Physical Chemistry (BE324: taught by me) and Transport (BE350: taught by John Schotland). BE324 obtained a score of 3.2 for instructor quality and 2.8 for course quality. In addition to these courses, I have served as a client for BE senior design, obliged independent-study (BE099 and BE599) requests from students on several occasions for customized courses or to meet specific needs. I have also served as a guest-lecturer for BE100 and CBE562.

Capitalizing on this track record, I propose to establish a new *BE 4XX/5XX* course on *Signal Transduction in Biochemical Systems.* Leveraging the expertise of my lab in studying single molecule signaling and signal transduction, I propose to launch this completely new course primarily targeting bioengineering students, but also accessible to other engineering and science disciplines. The course will provide comprehensive and rigorous theoretical and hands-on modeling experience on the treatment of biochemical signal transduction pathways in the cellular context. The novelty of BE400 will be further exemplified by a *mandatory computational laboratory*, which I will design solely for this course, whose purpose is to provide the students of BE 4XX/5XX with hands-on experience with advanced research tools in the quantitative biology literature. Here, students will run real simulations and embark on problem solving using high-performance codes developed by me over the years, and others available through academic licensing. This course will also serve to launch the research tools I have developed in my laboratory into the classroom.

In addition, I propose to establish a "Molecular CAVE": a 3-dimensional scientific visualization laboratory for the visual exploration for molecular and biomolecular systems by utilizing the CAVE technology. This technology will enable the display of time-resolved molecular motions as 3-dimensional animations in the CAVE. Moreover, via an interface to the motion tracker, the 3-dimensional environment will be interactive. The Molecular CAVE laboratory will house a set of visualization and animation modules to introduce the subject of molecular structure, protein, DNA, RNA, conformations, biomolecular dynamics, and how their form and motion determine their function. A repertoire of animations assembled from our research studies is utilized in developing an *Interactive Computational Molecular Biology workshop* to be taught in the Molecular Cave laboratory. *Funding for this* project will be saught from the Greater Philadelphia Bioinformatics Alliance, the Penn Nanobio Interface Center, and Bioadvance Consortium.

Outreach: To maximize the impact of my research and educational activities, I actively seek to disseminate our research findings as well as research tools through our laboratory's website at Penn, through the Greater Philadelphia Bioinformatics Alliance, and through internationally recognized channels; I have taught in a multiscale modeling and simulations workshop in CECAM (Lyon, France). To popularize my research ideology, I plan on conducting and teaching workshops on signal transduction, multiscale modeling, and computational drug design at regional, national, and international venues.