Research Statement

Bing Liu
liubing@pitt.edu

Understanding how biological systems function is a grand challenge facing science. The rewards of success will range from better medical therapies to new generation of biofuels. Conventional biology research culminates in the emerging of fields with the suffix *-omics*, such as genomics, proteomics and metabolomics. These fields seek to describe complete sets of knowledge about the constituent elements of biological systems. However, it is still far from an in-depth understanding of biological processes, because the constituent elements do not function alone but exist in complex assemblies and are interacting as highly regulated networks. Hence a basic aim of *systems biology* is to understand how specific biological functions emerge from the dynamics of biological networks. It is commonly recognized that computational modeling and analysis methods will play a crucial role in this endeavor.

Given this outlook, the central theme of my research is on *the development of computational modeling and analysis techniques to study biological systems at the systems level*. My work builds mathematical models to describe biological systems and employs artificial intelligence and formal verification techniques to analyze their dynamical behaviors. I use probabilistic frameworks to address the stochasticity in biological systems, and develop algorithms to construct model structure, estimate unknown parameters, discover new biology, as well as design precision medicine. I also leverage the power of high-performance computing techniques to enable the modeling and analysis of large-scale multicellular systems. As an integral part of my research, I collaborate closely with biologists and clinicians to study various systems and tackle real-world biological problems that are crucial to medicine and healthcare. I believe that my research will help move the state-of-the-art of systems biology forward and will have a substantial impact on our healthcare, food supplies and many other issues that are essential to our survival.

I now discuss the two aspects of my current systems biology research. These projects are a testament to my overall vision of using the power of computer science to solve biological problems. I then outline future directions that I am interested in pursuing in the next state of my research career.

Current Research

Modeling, Simulation and Analysis Techniques

A probabilistic approximation framework. Cellular processes are governed by a multitude of signaling pathways. A signaling pathway can be viewed as a complex network of biochemical reactions. A standard approach of modeling the dynamics of a pathway is through a system of ordinary differential equations (ODEs) [1]. However, such ODE systems often require a large number of numerical simulations to carry out model calibration and analysis tasks. To get around this, I developed a probabilistic approximation method by which an ODE system is reduced to a dynamics Bayesian network (DBN) [2]. It helps to bridge the gap between the accuracy of the results obtained by ODE simulation and the limited precision of experimental data used for model construction and validation. Consequently, tasks such as parameter estimation, sensitivity analysis, etc. can be performed efficiently using the advanced Bayesian inference techniques [3] we developed, instead of resorting to a large number of ODE simulations.

Case studies on two signaling pathways including the nerve growth factor (NGF) and epidermal growth factor (EGF) signaling pathways and the segmentation clock network have demonstrated that our method is promising in terms of efficiency and accuracy [4]. I then applied this method in a combined computational

and experimental study of an innate immune system [5]. 71 unknown rate constants were estimated by well-fitting hundreds of time serials data points and the experimentally verified model predictions helped to elucidate regulatory mechanisms and discover new biology [6].

To handle even larger models, we implemented our approach as a GPU-based tool. A sophisticated compilation strategy was used to carefully balance parallelism with memory accesses to obtain good performance [7]. Benchmark results show that it runs several orders of magnitude faster than a 10-PC-cluster implementation [8]. In a case study of the myosin light chain (MLC) pathway model consisting of 105 equations and 164 unknown parameters, it took 38 h whereas a 10-PC cluster implementation turned out to be infeasible [8].

Apart from parameter estimation and sensitivity analysis, the DBN approximation also enables probabilistic verification of dynamical properties. In [8], we also equipped the DBN approximation scheme with a model checking procedure enables probabilistic verification of dynamical properties.

A key component for various analyses of the DBN approximation is an approximate Bayesian inferencing algorithm called factored frontier (FF) algorithm. FF is employed since exact inference is computationally infeasible for large DBNs. The accuracy of this inferencing procedure can be improved –at additional computational cost– by using a parametrized variant of the FF algorithm called hybrid factored frontier (HFF) algorithm we presented in [3]. At each time slice, HFF explicitly maintains the probabilities of a number of global states called spikes. The results show that by increasing the number of spikes one can reduce errors while the additional computational effort required is only quadratic in the number of spikes.

Model checking based analysis techniques. Model checking is a *Turing Award* winning technique developed by Dr. Edmund Clarke for verifying properties of the behavior of discrete systems automatically. It have been fruitfully used to detect subtle bugs in a variety of hardware and software applications, ranging from microprocessor designs to satellite-control software [9]. I have been working with Dr. Clarke on developing model checking based techniques for analyzing the behavior of biological systems [10].

I developed a statistical model checking (SMC) based method for the calibration and analysis of biopathway models that explicitly address cell-to-cell variability [11]. I use SMC to verify dynamical properties of an ODE system accompanied by a prior distribution. Our specification logic can encode both qualitative behaviors and experimental data, which enables novel SMC based parameter estimation and sensitivity analysis procedures. In the case study of a very large network that contains 100 unknown parameters, our method yielded good parameter estimates by fitting both quantitative data and qualitative knowledge. The new property based sensitivity analysis powered by SMC also led to important biological insights that would be difficult to obtain using conventional approaches.

I also developed a computational framework for analyzing biological systems that possess multiple operational modes, which are often modeled using *hybrid automata*. A hybrid automaton has multiple modes, each of which a systems of ODEs is associated with. The analysis of hybrid systems is challenging; especially when the mode dynamics are governed by nonlinear ODEs. In [12], I developed a δ-decision based framework to tackle the parameter synthesis problem of biological hybrid automata models. The applicability of this method has been demonstrated through models of the cardiac cell action potential. The results show that our parameter synthesis framework is convenient and efficient for model selection as well as identifying crucial parameter ranges related to cardiac disorders. In a recent study [13], I extended this method to enable the identification of personalized therapeutic strategies for prostate cancer.

Furthermore, we also developed a probabilistic analysis method by approximating the mode transitions as stochastic events [14]. Our theoretical results form the basis for a SMC procedure which enables various analysis tasks to be carried out. The case studies on cardiac cell dynamics and the circadian rhythm indicate that our scheme can be applied in a number of realistic settings

Integrative Study of Biological Systems

An integral part of my research is to collaborate with biologists to build mathematical models and use the computational techniques I developed to solve real-world biological problems. In the past a few years, I modeled a number of crucial biological systems to obtain a quantitative understanding of the dynamics of their underlying mechanisms, towards the development of (poly)pharmacological strategies for a variety of human diseases.

Innate immune system. Complement system is the frontline of human immune system, which quickly detects invading microbes and alerts the host to eliminate the hostile substances. Inadequate or excessive complement activities may lead to immunerelated diseases. I led a team consists of computer scientists and biologists and developed a detailed computational model of the human complement system [5]. Using our DBN approximation techniques, we found that C4BP induces differential inhibition on the classical and lectin complement pathways and acts mainly by facilitating the decay of the C3 convertase. Our results also highlighted the importance of infection-mediated microenvironmental perturbations, which alter the pH and calcium levels. All these predictions were validated empirically [5].

In a recent work [15], I led the same team and conducted a systems modeling study of the Toll-like receptors (TLR) pathways, which recognize pathogen-associated molecular patterns (PAMPs) on viruses and trigger innate immune response by producing cytokines. The innate immune response is highly dependent on the timing of encountering PAMPs, suggestive of a short-term 'memory'. I developed the first calibrated mathematical model for the kinetics of TLR3 and TLR7 pathway crosstalk [15]. Our model analysis demonstrates that the JAK-STAT pathway mediates cytokine synergy thus boosting the immune response, while maintaining homeostasis to avoid excessive inflammatory response. This 'cytokine rheostat' mechanism enables macrophages to fine tune their response to multiple, temporally-separated infection events involving the TLR3-TLR7 pathways.

The insights we gained through the above works help to elucidate the regulatory mechanisms of the innate immune system and potentially contribute to the development of immunomodulation therapies.

Cell death & disease. Cellular stresses or intrinsic/extrinsic signals can induce different forms of cell death such as apoptosis, necroptosis, and ferroptosis, which are governed by multiple signaling pathways and their crosstalks. The modulation of cell death has been identified as an important therapeutic target for diverse diseases, including radiation diseases, neurodegenerative diseases, liver diseases, cancers, etc.

For instance, developing pharmacological strategies for controlling ionizing radiation (IR)-induced cell death is important for both mitigating radiation damage and alleviating the side effects of anti-cancer radiotherapy manifested in surrounding tissue morbidity. Exposure cells to ionizing radiation (IR) often triggers the onset of p53-dependent apoptotic pathways. In collaboration with radiation oncologists at University of Pittsburgh Medical Center (UPMC), I built a stochastic model of p53 induced apoptosis comprised of coupled modules of nuclear p53 activation, mitochondrial cytochrome c release and cytosolic caspase activation [16]. Our model analysis shows that immediate administration of PUMA inhibitors following IR exposure effectively suppresses excessive cell death, provided that there is a strong caspase/Bid feedback loop; however, the efficacy of the treatment diminishes with increasing delay in treatment implementation. In contrast, the combined inhibition of Bid and Bax elicits an anti-apoptotic response that is effective over a range of time delays.

IR exposure also causes necroptosis, a newly discovered non-apoptotic cell death. The cell fate decision between apoptosis and necroptosis is governed by a complex and intertwined signaling network. In a follow-up work [17], we collaborated with clinicians at UPMC and developed the first calibrated ODE model for apoptosis and necroptosis pathways and their crosstalk mediated by damaged associated molecular patterns

(DAMPs). Our results highlight the role of FLIP in regulating cell fate and suggested that inhibiting caspase-8 and cytochrome c could effectively suppress excessive cell death. These results provide novel insights into the development of drug combinations for mitigating the severe radiation damage.

It is known that autophagy can protect cells by maintaining cellular homeostasis and relieving various cytotoxic stresses. In order to selectively prevent the death of normal cells and induce the death of cancer cells, we developed a unified model of autophagy-apoptosis signaling network that involves mTOR signaling, inositol signaling, G-protein signaling, PI3K-AKT signaling, calcium signaling, intrinsic apoptosis pathways and the crosstalks among them [18]. We found that cytoplasmic Ca^{2+} fine tunes autophagy and apoptosis responses and its role is conferred by CaMKKβ. Our results reveal a time-dependent dual role of p53 in regulating the cell-fate determination. We also predicted drug combinations for improving the efficacy of cancer therapies.

Autophagy specific to the elimination of damaged mitochondria is called *mitophagy*. Our collaborators at the Department of Environmental and Occupational Health identified that cardiolipin externalization to the outer mitochondrial membrane acts as an elimination signal for mitophagy in neuronal cells. The collapse of asymmetric distribution of CLs may also lead to apoptosis depending on the stress level. To understand the role of CL in regulating mitochondrial homeostasis in response to cellular stresses, I built a validated rule-based model of the dynamics of cardiolipin pathways. Our model reveals the complexed role of H_2O_2 in cell fate determination. This work might result in a platform for the drug development for the early stages of radiation diseases.

Alpha1-antitrapsin deficiency (ATD). Alpha1-antitrypsin (AT) is a major serine protease inhibitor in the circulation and extracellular fluids, which protects tissues from collateral damage. Alpha1-antitrypsin deficiency (ATD), which leads to decreased circulating AT levels, is the most common genetic cause of liver disease in children and can also cause lung disease in adults. Currently, the only available treatment for ATD induced liver disease is transplantation. Therefore, there is a great need for the development of pharmacological strategies for reducing the accumulation of hepatotoxic ATZ. In collaboration with Pittsburgh Children's Hospital, I developed a kinetic model that captures the dynamics of ATZ degradation via ER-associated degradation (ERAD) pathway and autophagy pathway [19]. Our model reproduces the dose-response effects of carbamazepine and fluspirilene and enabled us to identify crucial drug targets and treatment strategies for enhancing ATZ clearance in liver.

Asthma. Asthma is a chronic inflammatory disease of the lungs which affects 300 million people worldwide kills 300,000 people per year. The current treatment for asthma such as inhaled corticosteroids and b2-agonists are effective in general but still failed in many, including the subtype of *Type 2 Hi asthma*. We are collaborating with clinicians in the Asthma institute and performing quantitative modeling and analysis of the 15 LO1 pathway in order to identify potential intervention strategies. We modeled the interactions between 15 LO1, 15 HETE-PE and PEBP1, the interactions of PEBP1 with Raf-1 and GRK2, as well as those between Raf-1 and MEK/ERK, and between GRK2 and beta-2AR. Our results suggest that the binding of 15LO1 with unphosphorylated monomeric PEBP1 is crucial and could be potential drug target for inhibiting the mucus production and desensitization of B2AR.

Prostate cancer Prostate cancer is the second leading cause of cancer-related deaths among men in United States. Hormone therapy in the form of androgen deprivation has been a cornerstone of the management of advanced prostate cancer for several decades. Recent clinical studies suggest that the efficacy of hormone therapy for prostate cancer depends on the characteristics of individual patients. In a recent work [13], I modeled the population dynamics of heterogeneous prostate cancer cells in response to androgen suppression as a nonlinear hybrid automaton. We estimated personalized kinetic parameters to characterize patients

and employ δ-reachability analysis to predict patient-specific therapeutic strategies for postponing the potential cancer relapse. The results show that our model is in good agreement with the published clinical data in literature, and may lead to a prognostic tool for prostate cancer therapy.

Synaptic signaling A human brain contains 100 billion neurons which make trillions of connections though synapses. Neural networks are formed dynamically, which facilitate the information processing and storage. Long-lasting synaptic plasticity is an essential mechanism for brain plasticity, which serves learning and memory. We are constructing a computational model to capture the dynamics of synaptic signaling network, in the collaboration with neurologists at CalTech. Our model takes into account the dynamics of AMPA and NMPA receptors, the activation of CaMKII, MAPK, AKT pathways, as well as the downstream protein synthesis. Our stochastic simulation results reproduce existing data and lead to interesting insights into the long-term potentiation and depression [20].

Pancreatic cancer. pancreatic cancer is the fourth cause of cancer death in the United States. Recently, the research of pancreatic cancer has been shifted to the investigation of the microenvironment of the pancreatic cancer cells. We have developed a multicellular model consisting of intracellular signaling networks of pancreatic cancer cells and stellate cells respectively as well as intercellular interactions among them. Through logical modeling and stochastic simulation, we analyzed the roles of important oncoproteins and tumor suppressors. We also evaluated potential drug intervention schemes for pancreatic cancer [21].

Type I diabetes (T1D). T1D affects millions of people in the world. Osteoporosis is a serious complication of T1D. I am currently collaborating with biologists from Michigan State University on identifying mechanism of T1D induced bone loss. We hypothesize that signals move between the gut and bone (and possibly liver) to modify bone density under healthy versus diseased conditions. I developed a computational model by integrating the current knowledge about calcium homeostasis, vitamin D metabolism, bone modeling and the pathophysiology of T1D. Our goal is to identify novel intestinal-bone regulatory pathways and therapeutic targets.

Future Research

A dozen years ago, a major contributor of Human Genome Project stated that the ultimate test of understanding biology would be to create a computer model of a cell. Recently, researchers at Stanford reported the first whole-cell computational model in the world to describes the life cycle of the bacterium *M. genitalium* who has the smallest genome. Although this preliminary attempt is stirring, many challenges remain towards whole-cell/multicellular models of higher organisms. My long-term goal is to develop core technologies which will enable us to model more complicated organisms, from higher microorganisms to eventually human tissues. Over the short-term horizon, my future research will continue to develop modeling and simulation frameworks, model calibration and analysis methods, and high-performance computing applications for systems biology. I will also continue to work closely with biologists and clinicians to answer big questions in biology and medicine. Below are two example projects I expect to pursue in the coming years.

An Integrative Modeling Platform for Polypharmacological Strategy Development

Most of my research thus far has focused on modeling the dynamics of disease-related cellular systems and identify therapeutic targets and strategies. I plan to integrate the computational techniques I have been developing, towards a unified platform for researchers to systematically construct models, perform analyses, as well as identify polypharmacological strategies.

In many settings, the cell can possess distinct functional modes with specific biopathways active in each mode. Further, in multicellular systems, cells can communicate with each other through inter-cellular pathways. Using a monolithic model of a single biochemical network in such settings will result in an unstructured, messy and opaque model. Consequently, calibration and analysis of such large models will be computationally intractable.

I plan to establish a new framework for the study of multi-mode multicellular systems. The framework will combine the power of rule-based stochastic simulation engines and scalable probabilistic approximation techniques. I will extend the DBN and develop a novel formalism named multi-mode dynamic Bayesian network (MDBN). The dynamics of a single cell will be probabilistically modeled and approximated as a MDBN. At multicellular level, I will encapsulate the MDBN approximation for each single cell together with its spatial location as an *agent*, and develop a rule-based stochastic modeling framework to simulate cell migration, proliferation, apoptosis, differentiation, as well as various cell-cell communications.

The challenging part of this problem is coming up with efficient inference techniques on MDBN to enable parameter estimation, sensitivity analysis, and probabilistic verification procedures. More research is also needed to devise better sampling methods for constructing the MDBN approximation. To test this framework for usability and practicality, I will work with my current collaborators to study pancreatic cancer microenvironment and the chemotaxis of bacterial population.

The modeling component of the proposed platform will enable the identification of drug targets through sensitivity analysis techniques. To predict and optimize potential drugs for selected targets, I plan to incorporate machine learning [22] and chemoinformatics approaches including druggability simulations and probabilistic-matrix factorization methodology into the platform. These functionalities will enable the user to perform quantitative molecular and systems pharmacology analysis and predict repurposed drugs.

The sequence, timing and combination in which drugs are administrated have a substantial influence on overall therapeutic results. A unique feature of the proposed platform is to allow the user to designing polypharmacological strategies or combination therapies and identifying the optimal timings for successive administering of one or more drugs.

I plan to extend the analysis techniques we developed for hybrid automata [12, 13, 14] to handle the proposed MDBN formalism. To identify the optimal drug treatment schedule that ensures both efficiency and safety, I will formulate a constraint optimization problem. An objective function will be devised to evaluate the efficiency. Constraints will be constructed using crucial indicators for toxicity or adverse effects. The search space will be high-dimensional, where each point in the search space will correspond to a vector of drug-time pairs. I will adopt global optimization methods such as evolutionary strategy for optimizing the drug delivery schedules. The platform also allows the user to incororpate experimental datasets to monitor the dynamics of biomarkers after drug treatment, and adaptively optimize therapeutic strategy.

A Virtual Immune System for Personalized Medicine

Despite huge recent advances in medical sciences, including many drugs that target the immune system, scientist still do not fully understand this complex system. It not only orchestrates the processes by which our bodies fight invading pathogens, but also cause autoimmune disease such as diabetes and rheumatoid arthritis. I have been collaborating with immunologist for years to study innate immune systems such as the complement system [5] and Toll-like receptors pathways [15]. At present, we are ideally positioned to extend our models and build a novel comprehensive virtual immune system that could help the development of personalized immunotherapies, having extensive experience in using computational methods to gain biological insights. The model will consist of the signaling networks of T-cells, B-cells, macrophages, etc. I believe the proposed computational framework based on MDBNs can play a very helpful role in terms of modeling the whole immune network and performing *in silico* experiments to answer "what if" questions, generating hypothesis and sharpening the choices for experimental design. I will continue my collaboration

with UPMC and NUS to obtain patients data. The resulting model will be trained for individual patient to enable the further design of personalized treatment strategies [13] using the proposed integrative platform. I believe the proposed virtual immune system will have a crucial role to play in helping the pharmaceutical industry to avoid late-stage failures and thus to lower the cost of drug development.

References

[1] Liu B, Thiagarajan PS, Modeling and analysis of biopathways dynamics, *Journal of Bioinformatics and Computational Biology* **10**:1231001, 2012.

[2] Liu B, Thiagarajan PS, Hsu D, Probabilistic approximations of signaling pathway dynamics, in *CMSB'09, Bologna, Italy*, pp. 251–265, 2009.

[3] Palaniappan SK, Akshay S, Liu B, Genest B, Thiagarajan PS, A hybrid factored frontier algorithm for dynamic Bayesian networks with a biopathways application, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **9**:1352-1365, 2012.

[4] Liu B, Hsu D, Thiagarajan PS, Probabilistic approximations of ODEs based bio-pathway dynamics, *Theoretical Computer Science* **412**:2188–2206, 2011.

[5] Liu B, Zhang J, Tan PY, Hsu D, Blom AM, Leong B, Sethi S, Ho B, Ding JL, Thiagarajan PS, A computational and experimental study of the regulatory mechanisms of the complement system, *PLoS Computational Biology* **7**:e1001059, 2011.

[6] Liu B, *Computational Modeling of Biological Pathways: Probabilistic Approximation and Analysis Techniques*, LAP LAMBERT Academic Publishing, 2012.

[7] Hagiescu A, Liu B, Palaniappan SK, Chattopadhyay B, Cui Z, Wong WF, Thiagarajan PS, GPU code generation for ODE-based applications with phased shared-data access patterns, *ACM Transactions on Architecture and Code Optimization* **10**:551-5519, 2013.

[8] Liu B, Hagiescu A, Palaniappan SK, Chattopadhyay B, Cui Z, Wong WF, Thiagarajan PS, Approximate probabilistic analysis of biopathway dynamics, *Bioinformatics* **28**:1508–1516, 2012.

[9] Clarke EM, *Model Checking*, MIT Press, 1999.

[10] Liu B, Kong S, Gao S, Clarke EM. Parameter identification using delta-decisions for hybrid systems in biology. *CMU SCS Technical Report, CMU-CS-13-136*.

[11] Palaniappan SK, Gyori B, Liu B, Hsu D, Thiagarajan PS, Statistical model checking based calibration and analysis of biopathway models, in *CMSB'13, Klosterneuburg, Austria*, pp. 1–15, 2013.

[12] Liu B, Kong S, Gao S, Zuliani P, Clark EM, Parameter synthesis for cardiac cell hybrid models using delta-decisions, in *CMSB'14, Manchester, UK*, pp. 99-113, 2014.

[13] Liu B, Kong S, Gao S, Zuliani P, Clarke EM, Towards prostate personalized cancer therapy using delta-reachability analysis, in *HSCC'15, Seattle, USA*, pp. 227-232, 2015.

[14] Gyori B, Liu B, Liang Y, Paul S, Ramanathan R, Thiagarajan PS, Probabilistic approximation and analysis of hybrid systems, in *HSB'09, Madrid, Spain*, pp. 1–15, 2015.

[15] Liu B, Liu Q, Palaniappan S, Bahar I, Thiagarajan PS, Ding JL. Innate immune memory and homeostasis may be conferred through TLR3-TLR7 pathway crosstalk. *Science Signaling, in revision*.

[16] Liu B, Bhatt D, Oltvai ZN, Greenberger JS, Bahar I, Significance of p53 dynamics in regulating apoptosis in response to ionizing radiation, and polypharmacological strategies, *Nature Scientific Reports* **4**:6245, 2014.

[17] Liu B, Che Z, Bahar I. Mathematical modeling of cell fate decision between necroptosis and apoptosis. *In preparation*.

[18] Liu B, Oltvai ZN, Bayir H, Kagan VE, Greenberger JS, Bahar I. Computational modeling of cell fate decision by coordination of autophagy and apoptosis. *In preparation*.

[19] Liu B, Oltvai ZN, Pak SC, Silverman GA, Perlmutter DH, Bahar I. Computational modeling reveals polypharmacological strategies for regulating alpha1-antitrypsin z mutant protein disposal. *In preparation*.

[20] Liu B, Kaya C, Kennedy MB, Bahar I. Stochastic modeling reveals the significance of camkii dynamics in synaptic plasticity. *In preparation*.

[21] Wang Q, Liu B, Miskov-Zivanov N, Faeder J, Lotze M, Clarke E. A multicellular model of pancreatic cancer microenvironment. *In preparation*.

[22] Zhou H, Gao S, Nguyen NN, Fan M, Jin J, Liu B, Zhao L, Xiong G, Tan M, Li S, Wong L, Stringent homology-based prediction of H. sapiens-M. tuberculosis H37Rv protein-protein interactions, *Biology Direct* **9**:1-52, 2014.