Conference Program

The Fourth TSIMF Conference on Computational and Mathematical Bioinformatics and Biophysics

Tsinghua Sanya International Mathematics Forum

December 13 – 16, 2021, (Beijing Time)
December 12 – 15, 2021, (US Central Time)

Zoom Meeting ID: 848 1435 3966

Password: TSIMF

Zoom Link:
https://ua-edu.zoom.us/j/84814353966?pwd=ZTBQSDVrQ3EvbmE2ZTNBM0dRUy83Zz09

Organizing Committee:
Stephen Shing-Toung Yau, Tsinghua University, China
Guowei Wei, Michigan State University, USA
Changchuan Yin, University of Illinois at Chicago, USA
Shan Zhao, University of Alabama, USA

DAY 1

December 13, 2021, (Beijing Time)

December 12, 2021, (US Central Time)

Beijing Time	**US Central Time**	**Session Chair: John Z.H. Zhang**, Shenzhen Institute of Advanced Technology, China
8:30 – 8:50 AM	6:30 – 6:50 PM	**Zoom Registration**
8:50 – 8:55 AM	6:50 – 6:55 PM	**Welcoming Remark** Stephen Shing-Toung Yau, Tsinghua University, China
8:55 – 9:25 AM	6:55 – 7:25 PM	**Tamar Schlick**, New York University, USA The complex conformational landscape of the SARS-CoV-2 Frameshifting RNA element
9:30 – 9:55 AM	7:30 – 7:55 PM	**Weihua Geng**, Southern Methodist University, USA A Cartesian FMM-accelerated Galerkin boundary integral Poisson-Boltzmann solver
10:00 – 10:25 AM	8:00 – 8:25 PM	**Xiaoqin Zou**, University of Missouri - Columbia, USA New strategies to predict protein-peptide interactions
10:30 – 10:55 AM	8:30 – 8:55 PM	**Huan-Xiang Zhou**, University of Illinois at Chicago, USA Local interactions and transient secondary structures govern backbone dynamics of intrinsically disordered proteins
11:00 – 11:25 AM	9:00 – 9:25 PM	**Lin Li**, University of Texas at El Paso, USA Developing and applying computational approach to investigate the viral structures
11:30 – 11:55 AM	9:30 – 9:55 PM	**Wenrui Hao**, Penn State University, USA Computational models of cardiovascular disease

Zoom Link: https://ua-edu.zoom.us/j/84814353966?pwd=ZTBQSDVrQ3EvbmE2ZTNBM0dRUy83Zz09

DAY 2

December 14, 2021, (Beijing Time)

December 13, 2021, (US Central Time)

Beijing Time	US Central Time	**Session Chair: Jie Liang**, University of Illinois at Chicago, USA
8:50 – 9:00 AM	6:50 – 7:00 PM	**Zoom Registration**
9:00 – 9:25 AM	7:00 – 7:25 PM	**Jie Liang**, University of Illinois at Chicago, USA Probability landscapes and global flow maps of discrete flux of stochastic reaction networks: exact construction, computation, and topological analysis
9:30 – 9:55 AM	7:30 – 7:55 PM	**Qi Wu**, Institute of Microbiology, CAS, China Using Fourier transform to connect response of the trait and the fitness to natural selection
10:00 – 10:25 AM	8:00 – 8:25 PM	**Dongqing Wei**, Shanghai Jiaotong University, China AIDD and drug candidates by super-computing
10:30 – 10:55 AM	8:30 – 8:55 PM	**Zixuan Cang**, North Carolina State University, USA Mapping cell-cell communications in spatial transcriptomics data
11:00 – 11:25 AM	9:00 – 9:25 PM	**Xi Chen**, Xi'an University of Finance and Economics, China Gaussian and Non-Gaussian Colored Noise Induced Escape in a Tumor-Immune Model
11:30 – 11:55 AM	9:30 – 9:55 PM	**Rui Dong**, Tsinghua University & BIMSA, China Full chromosomal relationships between populations and the origin of humans

Zoom Link: https://ua-edu.zoom.us/j/84814353966?pwd=ZTBQSDVrQ3EvbmE2ZTNBM0dRUy83Zz09

DAY 3

December 15, 2021, (Beijing Time)

December 14, 2021, (US Central Time)

Beijing Time	US Central Time	**Session Chair: Alexey Onufriev**, Virginia Tech, USA
8:50 – 9:00 AM	6:50 – 7:00 PM	**Zoom Registration**
9:00 – 9:25 AM	7:00 – 7:25 PM	**John Z.H. Zhang**, Shenzhen Institute of Advanced Technology, China Structure and interaction of protein-ligand complex
9:30 – 9:55 AM	7:30 – 7:55 PM	**Alexey Onufriev**, Virginia Tech, USA The Mathematics of Polymer Deformation: Energy Convex Hull Theory
10:00 – 10:25 AM	8:00 – 8:25 PM	**Jian Jiang**, Wuhan Textile University, China Geometric Graph Learning for Toxicity Prediction
10:30 – 10:55 AM	8:30 – 8:55 PM	**Kelin Xia**, Nanyang Technological University, Singapore Neighborhood complex based machine learning (NCML) models for drug design
11:00 – 11:25 AM	9:00 – 9:25 PM	**Duc Nguyen**, University of Kentucky, USA Mathematical-based graph neural network for drug design
11:30 – 11:55 AM	9:30 – 9:55 PM	**Yongcheng Zhou**, Colorado State University, USA Membrane Pore formation and dual phase field modeling

Zoom Link: https://ua-edu.zoom.us/j/84814353966?pwd=ZTBQSDVrQ3EvbmE2ZTNBM0dRUy83Zz09

DAY 4

December 16, 2021, (Beijing Time)

December 15, 2021, (US Central Time)

Beijing Time	**US Central Time**	**Session Chair: Shi-Jie Chen**, University of Missouri, USA
8:20 – 8:30 AM	6:20 – 6:30 PM	**Zoom Registration**
8:30 – 8:55 AM	6:30 – 6:55 PM	**Dmytro Kozakov**, Stony Brook University, USA TBA
9:00 – 9:25 AM	7:00 – 7:25 PM	**Guowei Wei**, Michigan State University, USA Forecasting vaccine-breakthrough SARS-CoV-2 variants
9:30 – 9:55 AM	7:30 – 7:55 PM	**Shi-Jie Chen**, University of Missouri, USA MgNet: Decoding RNA-metal ion interactions using a deep learning graphical convolutional neural network
10:00 – 10:25 AM	8:00 – 8:25 PM	**Ya Jia**, Central China Normal University, China A mathematical model for multiple phenotypic states of breast cancer cell
10:30 – 10:55 AM	8:30 – 8:55 PM	**Zhiliang Xu**, University of Notre Dame, USA Mathematical modeling of blood clotting and cell motion
11:00 – 11:25 AM	9:00 – 9:25 PM	**Jinzhi Lei**, Tiangong University, China Individual cell-based modeling of tumor cell plasticity-induced immune escape after CAR-T therapy
11:30 – 11:55 AM	9:30 – 9:55 PM	**Jiali Gao**, University of Minnesota, USA Multistate density functional theory as a matrix functional of densities for excited states

Zoom Link: https://ua-edu.zoom.us/j/84814353966?pwd=ZTBQSDVrQ3EvbmE2ZTNBM0dRUy83Zz09

Abstracts of Invited Talks

Mapping cell-cell communications in spatial transcriptomics data

Zixuan Cang, North Carolina State University, USA

Abstract: The recent development of high-resolution omics level technologies has reshaped modern biological research. These high-dimensional and noisy datasets are accumulating with a fast pace. Efficient and biologically meaningful algorithms are needed to extract biological insights from these raw datasets. In this talk, I will discuss using optimal transport, a powerful geometric data analysis method, to integrate multimodal omics datasets and to infer cell-cell communications, a crucial process that drives the correct developments and functions of biological systems. I will also talk about a new formulation of optimal transport called supervised optimal transport inspired by these biological applications.

MgNet: Decoding RNA-metal ion interactions using a deep learning graphical convolutional neural network

Shi-Jie Chen, University of Missouri, USA

Abstract: Magnesium ions (Mg2+) play a vital biological role as cofactors, interacting with RNA molecules to facilitate RNA folding and function. Previous attempts to accurately model RNA-Mg2+ binding have been plagued by difficulties arising from the challenge of accurately locating Mg2+ binding sites. Using experimental RNA structural data, we developed and applied MgNet, a machine-learning model, to predict Mg2+ binding sites in RNA molecules. This approach exploits local binding information associated with each nucleotide. In particular, electrostatic and 3D-shape (RNA volume) features are used to capture the key interaction patterns for the network to predict the density distribution of Mg2+ around the RNA molecules. Five-fold cross-validation on a dataset of 177 selected Mg2+-containing structures and comparisons with three different types of methods validate the approach. Results show that this new approach predicts Mg2+ binding sites with higher accuracy and efficiency. We use saliency analysis for eight different Mg2+ binding motifs to reveal the coordinating atoms of Mg2+ ions. Furthermore, learning the relevant physical mechanism through in-depth training on the known ion-RNA complexes, MgNet uncovers new Mg2+ binding motifs.

Gaussian and Non-Gaussian Colored Noise Induced Escape in a Tumor-Immune Model

Xi Chen, Xi'an University of Finance and Economics, China

Abstract: We investigate the mean first passage time of a tumor-immune model with Gaussian colored noise by the two analytic approximation methods of singular perturbation analysis and small correlation time approximation. For the first time, it is shown that the singular perturbation analysis is accurate in the sense of retaining linear term of the small correlation time parameter, while the small correlation time approximation keeps all the even higher-order terms of the same small parameter, but it neglects the linear leading order term. This contrast suggests that the

singular perturbation method has a better accuracy than the small correlation approximation method when the correlation time parameter is small. As a further application of the singular perturbation method, the mean first passage time in the case of non-Gaussian noise is also deduced and discussed. It is shown that as the strength of immunization or the non-Gaussian deviation parameter increases, the mean first passage time decreases, and thus both enhancing immunization and applying heavy-tailed random perturbation can accelerate the extinction of tumor cells.

Full chromosomal relationships between populations and the origin of humans

Rui Dong, Tsinghua University & BIMSA, China

Abstract: A comprehensive description of human genomes is essential for understanding human evolution and relationships between modern populations. However, most published literature focuses on local alignment comparison of several genes rather than the complete evolutionary record of individual genomes. Combining with data from the 1000 Genomes Project, we successfully reconstruct 2504 individual genomes and propose Divided Natural Vector method to analyzing the distribution of nucleotides in the genomes. Comparisons based on autosomes, sex chromosomes and mitochondrial genomes reveal the genetic relationships between populations. We confirm the "out-of-Africa" hypothesis and assert that humans most likely originated in Eritrea. The reconstructed genomes are stored on our server and can be further used for any genome-scale analysis of humans. This project provides the complete genomes of thousands of individuals and lays the groundwork for genome-level analyses of the genetic relationships between populations and the origin of humans.

Multistate density functional theory as a matrix functional of densities for excited states.

Jiali Gao, University of Minnesota, USA

Abstract: The Hohenberg and Kohn theorems established the relationship of the ground state energy as a functional of its density, whereas properties for the excited states are only extracted from linear-response time-dependent approaches in DFT. It's a long-standing goal to develop a time-independent theory within the framework of DFT for electronically excited states. In this paper, I will present two theorems that establish the Hamiltonian as a matrix functional of state and transition densities and a variational principle that minimization of the ensemble energy of the subspace spanned by the N-lowest eigenstate leads to the exact solutions of the excitation energies and state vectors. We illustrate the application of the method to a photoreceptor protein to understand the mechanism of excited state energy transfer.

A Cartesian FMM-accelerated Galerkin boundary integral Poisson-Boltzmann solver

Weihua Geng, Southern Methodist University, USA

Abstract: The Poisson-Boltzmann model is an effective and popular approach for modeling solvated biomolecules in continuum solvent with dissolved electrolytes. In this paper, we report

our recent work in developing a Galerkin boundary integral method for solving the Poisson-Boltzmann (PB) equation. The solver has combined advantages in accuracy, efficiency, and memory usage as it applies a well-posed boundary integral formulation to circumvent many numerical difficulties associated with the PB equation and uses an $O(N)$ Cartesian Fast Multipole Method (FMM) to accelerate the GMRES iteration. In addition, special numerical treatments such as adaptive FMM order, block diagonal preconditioners, Galerkin discretization, and Duffy's transformation are combined to improve the performance of the solver, which is validated on benchmark Kirkwood's sphere and a series of testing proteins.

Computational models of cardiovascular disease

Wenrui Hao, Penn State University, USA

Abstract: In this talk, I will introduce several computational models of cardiovascular disease including both atherosclerosis and aortic aneurysm growth to quantitatively predict the long-term cardiovascular risk. These models integrate both the multi-layered structure of the arterial wall and the aneurysm pathophysiology together. The heterogeneous multiscale method is employed to tackle different time scales while the finite element method is adopted to the deformation of the hyperelastic arterial wall all the time. A three-dimensional realistic cardiovascular FSI problem with an aortic aneurysm growth based upon the patients' CT scan data is simulated to validate a medically reasonable long-term prediction.

A mathematical model for multiple phenotypic states of breast cancer cell

Ya Jia, Central China Normal University, China

Abstract: Quantitative modeling of microscopic genes regulatory mechanisms in an individual cell is a crucial step towards understanding various macroscopic physiological phenomena of cell populations. Based on the regulatory mechanisms of genes zeb1 and cdh1 in single breast cancer cell, we propose a mathematical model. By constructing the effective landscape of stationary probability for the genes expressions, it is found that (i) each breast cancer cell has three phenotypic states (i.e., the stem-like, basal, and luminal states) which correspond to three attractions of the probability landscape. (ii) The interconversion between phenotypic states can be induced by the noise intensity and the property of phenotypic switching is quantified by the mean first passage time. (iii) Under certain conditions, the probabilities of each cancer cell appearing in the three states are consistent with the macroscopic phenotypic equilibrium proportions in the breast cancer SUM159 cell line. (iv) Our kinetic model involving the TGF-β signal can also qualitatively explain several macroscopic physiological phenomena of breast cancer cells, such as the TGF-β paradox in tumor therapy, the five clinical subtypes of breast cancer cells, and the effects of transient TGF-β on breast cancer metastasis.

Geometric Graph Learning for Toxicity Prediction

Jian Jiang, Wuhan Textile University, China

Abstract: Toxicity analysis is a major challenge in drug design and discovery. Recently significant progress has been made through machine learning due to its accuracy, efficiency, and

lower cost. US Toxicology in the 21st Century (Tox21) screened a large library of compounds, including approximately 12 000 environmental chemicals and drugs, for different mechanisms responsible for eliciting toxic effects. The Tox21 Data Challenge offered a platform to evaluate different computational methods for toxicity predictions. Inspired by the success of multiscale weighted colored graph (MWCG) theory in protein−ligand binding affinity predictions, we consider MWCG theory for toxicity analysis. In the present work, we develop a geometric graph learning toxicity (GGL-Tox) model by integrating MWCG features and the gradient boosting decision tree (GBDT) algorithm. The benchmark tests of the Tox21 Data Challenge are employed to demonstrate the utility and usefulness of the proposed GGL-Tox model. An extensive comparison with other state-of-the-art models indicates that GGL-Tox is an accurate and efficient model for toxicity analysis and prediction.

TBA

Dmytro Kozakov, Stony Brook University, USA

Abstract: TBA

Individual cell-based modeling of tumor cell plasticity-induced immune escape after CAR-T therapy

Jinzhi Lei, Tiangong University, China

Abstract: This talk will introduce an individual cell-based computational model based on major assumptions of the tumor cells heterogeneity and plasticity as well as the heterogeneous responses to CAR-T treatment. Model simulation and animal experiments suggest a mechanics of tumor cell plasticity-induced immune escape after CAR-T therapy through a CART cell-induced transition of tumor cells to hematopoietic stem-like cells and myeloid-like cells. The proposed computational model framework was successfully developed to recapitulate the individual evolutionary dynamics and potentially allows to predict the outcomes of CAR-T treatment through model simulation based on early-stage observations of tumor burden and tumor cells analysis.

Developing and applying computational approach to investigate the viral structures

Lin Li, University of Texas at El Paso, USA

Abstract: How the giant DNA viruses assemble thousands of proteins accurately to build their protein shells, the capsids, remains largely unknown. In the last three decades, many giant DNA viruses have been discovered, which present a unique and essential research frontier for the study of self-assembly and regulation of supramolecular assemblies. Revealing the mechanisms of giant virus assembly will help to discover the mysteries of many self-assembly biology problems. Paramecium bursaria Chlorella virus 1 (PBCV-1) is one of the most intensively studied giant viruses. Based on the high-resolution structure of PBCV-1, we implemented computational biophysics approaches to investigate the physics interactions among capsomers, which are giant virus capsid building units. This talk will show some fundamental physics mechanisms in the viral capsid assembly. Among all the binding modes, the weakest binding

mode is located at the boundary of trisymmetron. This explains why the seam between two neighboring trisymmetrons becomes the breaking line when a giant virus capsid dissociates. Formulas generated for the total number of each binding mode within one capsid show the mode within trisymmetron are dominating the stabilization of the capsid which is consistent to previous observation. Besides the viral capsid assembly, methods introduced in this study can be applied to study more complicated assembly process for other biomolecular structures.

Probability landscapes and global flow maps of discrete flux of stochastic reaction networks: exact construction, computation, and topological analysis

Jie Liang, University of Illinois at Chicago, USA

Abstract: Distinct cellular phenotypes are often encoded in the epigenetic wiring of the genetic regulatory networks, and stochasticity in reactions of these networks in an important contributor to the diverse cellular behavior of isogenic cells in different tissues. While the stochastic reaction kinetics model provides a fundamental framework to investigate stochasticity in reaction networks, solving the underlying discrete Chemical Master Equation (dCME) is challenging. It is difficult to accurately estimate probabilities of rare events using trajectory based method such as Gillespie stochastic simulations. Here we discuss how time-evolving and steady state probability landscape can be exactly computed using the n-simplex optimal state enumeration algorithm and the ACME (accurate chemical master equation) method (www.bioacme.org, DOI. 10.1137/15M1034180), without Monte Carlo simulation or Fokker-Planck/Langevin approximation. We also discuss how to guarantee small errors and how to estimate the best achievable accuracy with a given laptop/supercomputer, through an a priori calculated error bound when truncation of the state space inevitably occurs for complex networks. In addition, we describe a new formulation of discrete probability flux and velocity under proper boundary conditions and how their global flow maps can be exactly computed. We further discuss how the topology of the high-dimensional probability landscape can be accurately characterized using newly developed algorithm of persistent homology, so peaks, basins, and cycles of different dimensions are accounted for. We give examples on analysis of phenomenological characterization of cellular decision networks, including bi-stability, epigenetic states, and the robustness of wild type versus mutants networks. In addition, we discuss the origin of multimodality in feedforward networks that lacks feedback loop nor cooperativities. Furthermore, we discuss how ACME computation can reveal the origin of the reservoir of HIV latently infected cells and uncover detailed mechanism of probabilistic intra-cellular control of latency and transactivation. Moreover, we discuss how new approaches for health intervention can been formulated based on large scale computation exploring changes in the probability landscapes of networks under different perturbations, in the context of clinical treatment of the “shock and kill” and the “block and lock” strategies.

Mathematical-based graph neural network for drug design

Duc Nguyen, University of Kentucky, USA

Abstract: Graph neural network (GNN) has become the centerpiece of deep learning architectures for graph-specific tasks touching a variety of real world applications, such as social media, financial fraud detection, neuroscience, and many more. As the graph is a natural

language to represent molecules, one can use the molecular graph directly in GNN. Consequently, GNN has earned more attention and emerged as the most popular deep learning model in biological sciences. However, most variants of the graph neural network have been limited themself to the prediction of small molecular properties, namely quantum mechanics, hydrophobicity, and toxicity as their graph nodes only manage to encode the simple fingerprints. In the past few years, our groups have developed powerful mathematical representations for the diverse biological datasets in the low-dimensional spaces, namely differential geometry, geometric and algebraic graphs. By integrating these mathematical features with graph neural networks, we arrived at novel models not only perform well on virtual screening targeting important drug properties but also have the ability to design new drugs at an unprecedented speed. The proposed models have further enhanced the accuracies of our winner scoring functions in D3R Grand Challenges, a worldwide annual competition series in computer-aided drug design.

The Mathematics of Polymer Deformation: Energy Convex Hull Theory

Alexey Onufriev, Virginia Tech, USA

Abstract: Most biologically relevant DNA is strongly deformed. While weak deformation of polymers is well described by Hooke's Law and Euler's Elastica Theory, the strong deformation regime is not yet well understood: some polymers, notably the DNA, show unexpected behavior. In particular, experimental evidence on the looping of DNA fragments shorter than 100 base-pairs points to the fact that strongly bent DNA -- most relevant from a biological perspective -- is considerably more flexible than expected. Likewise, when stretched beyond a certain point, the tension in the double-helical DNA stops increasing, even though the polymer keeps on stretching. We propose a general quantitative framework of polymer deformation that includes both the weak and strong bending regimes on the same footing, based on a single general principle. As the deformation increases beyond a certain (polymer-specific) point, the change in the convexity properties of the effective deformation energy of the polymer makes the harmonic deformation energetically unfavorable: in this strong deformation regime the energy of the polymer varies linearly with the average deformation, as the system follows the convex hull of the deformation energy function. Predictions of the theory are discussed in the context of recent experiments on DNA cyclization and stretching. Completely counter-intuitively, probability of forming very short DNA loops is predicted to increase with decreasing loop length. Next, we go beyond Euler's Elastica Theory by considering non-convex deformation energy in general, which brings new and unexpected twists to the 300 year old theory of elastic deflections of structures.

The complex conformational landscape of the SARS-CoV-2 Frameshifting RNA element

Tamar Schlick, New York University, USA

Abstract: A combination of graph-based modeling for RNAs with pseudoknots, chemical reactivity experiments, and microsecond molecular dynamics simulations will be described to untangle the complex conformational landscape of the frameshifting RNA element of SARS-CoV-2 and suggest new avenues for anti-viral therapy.

Reference:

T Schlick et al, JACS (2021): https://pubs.acs.org/doi/abs/10.1021/jacs.1c03003
T Schlick et al, BJ (2020): https://www.cell.com/biophysj/pdf/S0006-3495(20)30814-6.pdf
with New and Notable commentary:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7857021/

AIDD and drug candidates by super-computing

Dongqing Wei, Shanghai Jiaotong University, China

Abstract: A new agonist of a membrane protein, α7nAChR was discovered, i.e., wgx50, which is extracted from the Sichuan pepper, a result from extensive so called network pharmacology based in AIDD. Extensive experimental studies show it could combine with α7nAChR on nerve cells, induce depolymerization of Aβ, inhibit Aβ-induced neurocyte apoptosis, and suppress the release of TNF-αand IL-1β from microglia. In vivo experiments showed that it could improve the cognition ability in APP-Transgenic Mice. These results suggest that wgx50 is a promising drug candidate for AD treatment and other applications in the area of anti-aging.

Forecasting vaccine-breakthrough SARS-CoV-2 variants

Guowei Wei, Michigan State University, USA

Abstract: Understanding the mechanisms of SARS-CoV-2 transmission and evolution is one of the greatest challenges of our time. We accurately forecasted two vital spike protein receptor-binding domain (RBD) residues (452 and 501) responsible for all the prevailing SARS-CoV-2 variants in May 2020 (Journal of Molecular Biology (2020) 432, 5212). In the same article, we predicted 1149 most likely RBD mutations our of 3686 possible ones. All 737 known RBD mutations fell into this category. We also unveiled that natural selection through infectivity is the mechanism governing SARS-CoV-2 evolution out of a variety of competing theories, which laid a foundation for us to accurately forecast emerging SARS-CoV-2 variants. We show that vaccine-resistance is a new transmission pathway due to vaccinations. We forecast that two complementary transmission pathways, i.e., infectivity and vaccine-resistance, will prolong our battle with COVID-19 for years. Continuously confirmed by new experiments, our predictions of emerging variants and their disruption of antibodies are based on all available viral genomes isolated from millions of patients, tens of thousands of mutational data, all available antibodies (currently over 150), algebraic topology, and deep learning.

Using Fourier transform to connect response of the trait and the fitness to natural selection

Qi Wu, Institute of Microbiology, CAS, China

Abstract: Since Darwin's masterpiece "the origin of species" in 1859, natural selection has always been the central subject of evolutionary biology. However, the quantitative scenario of how traits respond to the natural selection is far from perfect, especially in complicated cases such as multigenic determination with non-additive epistatic effect, non normal distribution in population, and long-term response. In this paper, we introduced the Fourier transform to connect quantitative trait with its fitness in the polygenic evolution of one population, and used the uncertainty principle of Fourier transform to describe the integrated response of trait and fitness to natural selection in

one inequality. The model is suitable for the long-term evolutionary dynamics of the polygenic adaptation, allowing the interaction between genes (epistasis) without assuming the normal distribution of traits in populations, and no complex mathematical tools necessary, such as higher-order moments or cumulants, but the variance of phenotype and fitness only. With the simulation results, the validity of the inequality relation has been verified in the case of single population under various factors of population dynamics. This formula specifies the existence of a certain minimum value for the integrated responses of both fitness and trait to selection. We call it "the principle of least response" (PLR) to natural selection. In deduction of the PLR inequality, we have introduced a constant L to express the minimum unit of variable of the selective response and a variable m0 to denote the degree of the trait/fitness maintaining a certain linear change of the response to selection, which are corresponded with the Planck constant and particle mass in the inequality relation of Heisenberg's uncertainty principle in quantum mechanics in both implication and mathematical form. It strongly suggests an extraordinary similarity between the evolutionary dynamics and the quantum mechanics. We hope that it will bring some inspiration to the theoretical analysis of polygenic evolution dynamics in future.

Neighborhood complex based machine learning (NCML) models for drug design

Kelin Xia, Nanyang Technological University, Singapore

Abstract: The importance of drug design cannot be overemphasized. Recently, artificial intelligence (AI) based drug design has begun to gain momentum due to the great advancement in experimental data, computational power and learning models. However, a major issue remains for all AI-based learning models is efficient molecular representations. Here we propose Neighborhood complex (NC) based molecular featurization (or feature engineering), for the first time. In particular, we reveal deep connections between NC and Dowker complex (DC) for molecular interaction based bipartite graphs, for the first time. Further, NC-based persistent spectral models are developed and the associated persistent attributes are used as molecular descriptors or fingerprints. To test our models, we consider protein-ligand binding affinity prediction. Our NC based machine learning (NCML) models, in particular, NC-based gradient boosting tree (NC-GBT), are tested on three most-commonly used datasets, i.e., including PDBbind-v2007, PDBbind-v2013 and PDBbindv2016, and extensively compared with other existing state-of-the-art models. It has been found that our NCML models can achieve state of-the-art results.

Mathematical modeling of blood clotting and cell motion

Zhiliang Xu, University of Notre Dame, USA

Abstract: In this talk, I will first discuss a macroscale phase-field model for studying mechanical stability of developing clots. In particular, simulations show structures of the blood clot are remodeled due to flow shear. Then I will present a microscale phase-field model for simulating motion and shape transformation of vesicles under flow conditions. Lastly, I will discuss a purely data-driven deep neural network algorithm for PDEs.

Structure and interaction of protein-ligand complex

John Z.H. Zhang, Shenzhen Institute of Advanced Technology, China

Abstract: In this talk, we present some recent work in protein-ligand and protein-protein interactions. The reported work involved development of machine learning methods to accurately predict complex structures in protein-ligand and protein-peptide systems as well as methods to predict binding energies in protein-protein interaction.

Local interactions and transient secondary structures govern backbone dynamics of intrinsically disordered proteins

Huan-Xiang Zhou, University of Illinois at Chicago, USA

Abstract: The lack of well-defined structures in intrinsically disordered proteins (IDPs) calls for a fundamental reassessment of how their amino-acid sequences code for functions. Some attention has been paid to nascent structures, but a missing link is sequence-dependent backbone dynamics, which we previously showed to arise from the formation of correlated segments stabilized by polyproline II (PPII) helices and salt bridges in the ChiZ disordered N-terminal region.1 To define general rules governing sequence-dependent backbone dynamics, we performed molecular dynamics simulations on eight IDPs. Nearly all above-average transverse relaxation rates and heteronuclear Overhauser enhancements along the sequence were attributable to interactions between side chains and formation of secondary structures. PPII stretches are the most common form of transient secondary structures, and are found in most IDPs and are often stabilized by local interactions. However, in some IDPs, stable α-helices are also present. These locally rigidified elements may code for nascent structures, whereas segments with fast dynamics may readily adapt to binding partners.2

1. Hicks, A. et al. Biomolecules 10, 946 (2020).
2. Hicks, A. et al. JACS Au 1, 66-78 (2021).

Membrane Pore formation and dual phase field modeling

Yongcheng Zhou, Colorado State University, USA

Abstract: Membrane pore formation is critical step in programmed cell death (PCD), during which the enclosed cellular environment is disrupted, allowing the passive flows molecules. The structure and dynamic evolution of the apoptotic membrane pores remain open questions despite the intense research in the past decades. In this talk we will highlight the current challenges in remodeling of membrane pore formation, describe a geometrical parameterization of energetic modeling of protein-membrane interactions, and its computational implementation of the model using dual phase field: one for relatively rigid protein and the other for fluidic bilayer membrane.

New strategies to predict protein-peptide interactions

Xiaoqin Zou, University of Missouri – Columbia, USA

Abstract: Peptides are short polymer chains consisting of amino acids. They are flexible and often change conformations when they bind to proteins. Protein-peptide interactions play an important role in many cellular processes. In silico prediction of protein-peptide complex

structure is highly desirable for mechanistic investigation of these processes and for therapeutic design. However, it is challenging to predict all-atom structures of protein-peptide complexes without any knowledge about the binding site and the bound peptide conformation, because of the large degrees of freedom involved in the system. In this talk, I will present our recent development of new strategies for predicting protein-peptide complex structures, based on the integration of information-driven modeling and physics/chemistry-based computational modeling of the interaction modes. The peptide is treated as a flexible structure during modeling, and the search space includes the whole surface of the protein. The methods have been systematically and extensively tested, and the results will be presented. The methods are computationally efficient. They can be used either as a standing-alone tool for large-scale protein-peptide docking or as an initial-stage sampling tool for protein-peptide structure refinement programs.

Abstracts of the Special Issue

A special issue entitled "Computational and Mathematical Bioinformatics and Biophysics" is dedicated to this conference, and will be published in Communications in Information & Systems. Four manuscripts have been accepted/submitted to this special issue. For your information, the abstracts of these manuscripts are given below.

Physics-guided multiple regression analysis for calculating electrostatic free energies of proteins in different reference states

Tania Hazra, Department of Mathematics, Misericordia University, Dallas PA 18612, USA
Shan Zhao, Department of Mathematics, University of Alabama, Tuscaloosa AL 35406, USA

Abstract: An implicit solvent modeling problem is studied in this work, i.e., by calculating the electrostatic free energy between water and a new reference state, how to recover the original solvation free energy between water and vacuum states. Such a recovery is considered for the super-Gaussian Poisson-Boltzmann (PB) model [T. Hazra, S. Ahmed-Ullah, S. Wang, E. Alexov, and S. Zhao, Journal of Mathematical Biology, (2019) 79:631-672], which is a heterogeneous dielectric model to mimic the conformational changes of a macromolecule. Nevertheless, while the dielectric function should physically decrease in the vacuum state as it leaves the macromolecular region, the super-Gaussian dielectric function has an inflation over the narrow band of the solute-solvent boundary. To avoid such a non-monotonicity issue, a new reference state with a large enough dielectric value is employed in the super-Gaussian PB model. Based on the electrostatic free energy calculated using this new reference state, a multiple regression model is developed in this paper to estimate the original free energy. The proposed regression model is built physically by accounting for the contribution of each individual atom explicitly, which is modeled via the analytical result of the Kirkwood sphere. Moreover, a regression analysis is conducted for four simple physical descriptors that are related to electrostatic interactions between solute and solvent, i.e., the total number of atoms, the total charge, and the area and volume of the solvent excluded surface (SES). By using a data set of 74 proteins, the dependence of these four descriptors is analyzed. Numerical results indicate that the multiple regression model performs well in estimating the electrostatic free energies.

Computing electrostatic binding energy with the TABI Poisson-Boltzmann solver

Leighton Wilson [1], Jingzhen Hu [2], Jiahui Chen [3], Robert Krasny [1], and Weihua Geng [4]

[1] Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA.
[2] Department of Mathematics, Duke University, Durham, NC 27710, USA.
[3] Department of Mathematics, Michigan State University, MI 48824, USA.
[4] Department of Mathematics, Southern Methodist University, Dallas, TX 75275, USA.

Abstract: We present computations of electrostatic binding energy DDGelec of solvated biomolecular complexes using the treecode-accelerated boundary integral (TABI) Poisson-Boltzmann solver. TABI computes the electrostatic potential on the triangulated molecular surface of a complex and its monomers, and further processing yields the solvation free energy needed to compute the binding energy. Among two codes examined for surface triangulation, MSMS and NanoShaper, the latter is more accurate, efficient, and robust in TABI calculations. The accuracy of the computed DDGelec is sensitive to the accuracy of the PB solver due to cancellation of digits, and the error can be efficiently reduced by extrapolating low triangulation density results to the high density limit.

Neural-PDE: A RNN based neural network for solving time dependent PDEs

Yihao Hu [1], Tong Zhao [2], Shixin Xu [3], Lizhen Lin [1], and Zhiliang Xu [1]

[1] Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46545, USA.
[2] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46545, USA.
[3] Duke Kunshan University, Kunshan, Jiangsu 215316, P.R. China.

Abstract: Partial differential equations (PDEs) play a crucial role in studying a vast number of problems in science and engineering. Numerically solving nonlinear and/or high-dimensional PDEs is frequently a challenging task. Inspired by the traditional finite difference and finite elements methods and emerging advancements in machine learning, we propose a sequence-to-sequence learning (Seq2Seq) framework called Neural-PDE, which allows one to automatically learn governing rules of any time-dependent PDE system from existing data by using a bidirectional LSTM encoder, and predict the solutions in next n time steps. One critical feature of our proposed framework is that the Neural-PDE is able to simultaneously learn and simulate all variables of interest in a PDE system. We test the Neural-PDE by a range of examples, from one-dimensional PDEs to a multi-dimensional and nonlinear complex fluids model. The results show that the Neural-PDE is capable of learning the initial conditions, boundary conditions and differential operators defining the initial-boundary-value problem of a PDE system without the knowledge of the specific form of the PDE system. In our experiments, the Neural-PDE can efficiently extract the dynamics within 20 epochs training and produce accurate predictions. Furthermore, unlike the traditional machine learning approaches for learning PDEs, such as CNN and MLP, which require great quantity of parameters for model precision, the Neural-PDE shares parameters among all time steps, and thus considerably reduces computational complexity and leads to a fast learning algorithm.

Gaussian and Non-Gaussian Colored Noise Induced Escape in a Tumor-Immune Model

Xi Chen [1] and Yanmei Kang [2]

[1] School of Statistics, Xi'an University of Finance and Economics, Xi'an 710100, P.R. China.

[2] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, P.R. China.

Abstract: We investigate the mean first passage time of a tumor-immune model with Gaussian colored noise by the two analytic approximation methods of singular perturbation analysis and small correlation time approximation. For the first time, it is shown that the singular perturbation analysis is accurate in the sense of retaining linear term of the small correlation time parameter, while the small correlation time approximation keeps all the even higher order terms of the same small parameter, but it neglects the linear leading order term. This contrast suggests that the singular perturbation method has a better accuracy than the small correlation approximation method when the correlation time parameter is small. As a further application of the singular perturbation method, the mean first passage time in the case of non-Gaussian noise is also deduced and discussed. It is shown that as the strength of immunization or the non-Gaussian deviation parameter increases, the mean first passage time decreases, and thus both enhancing immunization and applying heavy-tailed random perturbation can accelerate the extinction of tumor cells.