1853-Wkshp
A Physicist's Approach to Statistical Analyses of Biological Data
Patrice Koehl.
Computer Science and Genome Center, University of California, Davis, Davis, CA, USA.
The ongoing transformation of biology to a quantitative discipline has drastically increased our opportunities to unravel the mechanisms that relate the dynamics of biological systems to their functions as it allows for the investigation of such systems at spatial and temporal scales never observed before. The biggest challenge today is to assimilate the wealth of information generated in this process into a conceptual framework. We face issues with the volume of data generated (a Big Data challenge) as well as with the complexity of the systems they represent. In this talk I will show examples for which a combination of mathematics, physics, and biology provides solutions to these challenges. I will focus specifically on the concept of networks in biology, their morphologies and dynamic behaviors.

1854-Wkshp
Glycan Biosynthesis: Structure, Information, and Heterogeneity
Anjali Jaiman, **Mukund Thattai**.
Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Bangalore, India.
The surfaces of all living cells are decorated with branched sugar polymers known as glycans. These information-rich structures confer cells with a recognizable molecular identity, and underlie many specific cell-cell interactions. Analytic methods - including NMR, mass-spectrometry, and glycan arrays - now permit the routine profiling of glycans associated with various cells or proteins. This has stimulated efforts to build comprehensive and searchable glycan databases. However, from an informatics perspective glycans present multiple challenges. First, whereas nucleotide and amino acid chains are efficiently represented as strings, sugars can polymerize into complex tree-like objects. The potential combinatorial space of glycans is therefore much larger than that of proteins. Second, many specific molecular interactions appear to be mediated by groups of closely-related glycan variants rather than by a single well-defined structure. This phenomenon of ''micro-heterogeneity'' makes it difficult to rigorously characterize the glycan repertoire of a cell. In this workshop, I will use ideas from algorithmic self-assembly to show that glycan structure and diversity are best understood through the lens of glycan biosynthesis. I will demonstrate that a specific glycan structure is the outcome of glycosyltransferase enzymes acting according to simple rules in a specific order, like workers on a factory floor. Errors in this process produce a well-defined spectrum of glycan by-products, precisely matching the observed micro-heterogeneity in real glycan profiles. This predictive theoretical framework allows us to use glycans as sensitive cell-biological probes. It provides a unifying perspective within which the rich and growing datasets of glycan structures can be organized and fully utilized.

1855-Wkshp
Large-Scale Machine Learning Approaches for Molecular Biophysics
Arvind Ramanathan[1], Chakra S. Chennubhotla[2], Pratul K. Agarwal[3], Christopher B. Stanley[4].
[1]Computer Science, Oak Ridge National Lab, Oak Ridge, TN, USA, [2]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA, [3]Computer Science and Mathematics Division, Oak Ridge National Lab, Oak Ridge, TN, USA, [4]Neutron sciences Division, Oak Ridge National Lab, Oak Ridge, TN, USA.
Extracting knowledge from large, heterogeneous, unstructured and high-dimensional data is one of the major challenges for large-scale machine learning algorithms. In this talk, I will present our recent results developing unsupervised machine learning approaches to explore such data sets. A large number of these datasets follow heavy-tailed distributions, characterized by long-range dependencies. We quantify the tails of these distributions using higher order statistics and use tensor-based representations to build data mining algorithms for: (1) online detection of events that signify anomalies in spatio-temporal patterns; (2) building low- dimensional latent variable models to capture the intrinsic multiscale structure; and (3) hierarchical clustering and visual organization of the data to gain relevant insights. We will illustrate these approaches on a variety of applications including the integration of sparse experimental observations with atomistic-scale information for understanding the function of cellular systems. We will also discuss how these approaches can be widely applied to other domains.

Workshop: Advances in Computing Large Systems

1856-Wkshp
Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations
Angel E. Garcia[1], Jacob Miner[2], Alan A. Chen[3].
[1]Dept Phys/Appl Phys/Aston, Rensselaer Polytechnic Inst, Troy, NY, USA, [2]Dept of Biology, Rensselaer Polytechnic Inst, Troy, NY, USA, [3]Dept of Chemistry, SUNY Albany, Albany, NY, USA.
Structured RNAs exhibit a distinct preference for loops of precisely 4 nucleotides. Approximately 70% of these ''tetraloops'' are comprised of just three specific loop sequences: UUCG, GCAA, or CUUG. The abundance of these sequences is thermodynamic in origin, as each motif forms a unique network of non-canonical interactions within their loops that stabilize the folded state. Modification to the Amber force field enables the de-novo folding of three hyperstable RNA tetraloops to 1-3 Å RMSD from their experimentally determined structures using molecular dynamics simulations initialized in the unfolded state. To study the thermodynamics and kinetics of folding on an RNA tetraloop we simulated the (rGCAA) tetraloop with stem lengths of two (octamer), and four (dodecamer) C-G base pairs. The thermodynamics is obtained from replica exchange molecular dynamics (REMD) simulations with overall sampling exceeding 300 microseconds. The kinetics of the octamer was studied in a 100 microseconds molecular dynamics simulation using the Anton supercomputer. The thermodynamics reveal that the octamer folds and unfolds reversibly. However, the dodecamer behaves glassy and adopts multiple metastable loop configurations that do not unfold in extensive REMD simulations. The ability to recapitulate the signature non-canonical interactions of the three most abundant hyperstable stem-loop motifs represents a significant step towards the accurate description of nucleic acid tertiary structures, dynamics and stability using unbiased all-atom molecular dynamics simulations.

1857-Wkshp
Bacterial Outer Membranes and Interactions with Membrane Proteins
Wonpil Im.
Department of Molecular Biosciences and Center for Bioinformatics, The University of Kansas, Lawrence, KS, USA.
Because the bacteria's outer membrane (OM) acts as an effective barrier against the permeation of both hydrophobic and hydrophilic compounds, gram-negative cell permeation is one of the biggest challenges to the discovery of novel antibiotics for bacterial infections and antibiotic resistance worldwide. The bacterial OM is a unique and highly asymmetric lipid bilayer composed of phospholipids in the inner leaflet and lipopolysaccharide (LPS) in the outer leaflet. An LPS molecule is a complex amphiphatic compound consisting of lipid A, a core oligosaccharide, and an O-antigen polysaccharide. Despite the direct relationship of gram-negative bacteria to the public health and also the fact that there are over 190 identified O-serotypes for Escherichia coli, our molecular-level understanding of how the bacterial OMs behave and work for various types of bacteria, how membrane proteins behave in the OM, and how known drug molecules and potential drugs can enter through the OMs is rudimentary at best. This talk presents our ongoing efforts on all-atom modeling and simulations of these complex bacterial OMs with and without various outer membrane proteins using the CHARMM36 (protein, lipid, carbohydrate) force fields. In addition, various technical aspects and perspectives are also discussed together with future developments of CHARMM-GUI LPS Modeler and OM Builder.

1858-Wkshp
Protein Folding and Recognition in the Cell – an in Silico Approach
Margaret S. Cheung[1,2].
[1]University of Houston, Houston, TX, USA, [2]Rice University, Center for Theoretical Biological Physics, Houston, TX, USA.
I will review the approach of coarse-grained molecular simulations for the investigation of protein folding and protein-protein interactions in a cellular environment, particularly the one for the research from my group. We used a low-resolution model for the representation of proteins and macromolecules that mimic a jam-packed space inside a cell. We made these low-resolution models act like ''a real thing'' by keeping the physics in its dynamics and the principle of chemical interactions between macromolecules in the simulations. With this approach, we were able to characterize the mechanism of protein folding and protein-protein interactions that involve structurally large rearrangement in the presence of dominant forces inside a cell, such as the volume exclusion from the macromolecular crowding effect and the ionic strength. Based on simple ideas for modeling a large system, I will report several new discoveries and testable predictions from our computational studies.