With the urgent need for new medical approaches due to increased bacterial resistance to antibiotics, antimicrobial peptides (AMPs) have been considered as potential treatments for infections. Experiments indicate that combinations of several types of AMPs might be more effective at inhibiting bacterial growth with reduced toxicity and a lower likelihood of inducing bacteria resistance. The molecular mechanisms of AMP-AMP synergistic antimicrobial activity, however, remain not well understood. Here, we present a theoretical approach that allows us to relate the physicochemical properties of AMPs and their antimicrobial cooperativity. A concept of physicochemical similarity is introduced, and it is found that less similar AMPs with respect to certain physicochemical properties lead to greater synergy because of their complementary antibacterial actions. The analysis of correlations between the similarity and the antimicrobial properties allows us to effectively separate synergistic from non-synergistic AMPs pairs. Our theoretical approach can be used for the rational design of more effective AMPs combinations for specific bacterial targets, for clarifying the mechanisms of bacterial elimination, and for a better understanding of cooperativity phenomena in biological systems.

197-Plat
De novo design of a DNA-binding protein capable of multi-site recognition
Lee A. Solomon, Jacklin Sedhom.
Department of Chemistry and Biochemistry, George Mason University, Manassas, VA, USA.
DNA binding is critically important to nearly all genetic functions. Proteins must bind their target sequence with high affinity and selectivity, or the cell or organism could die. Certain proteins recognize a set of unique but related sequences (e.g. lac repressor, and lambda-phage CI). Our group studies these DNA-protein binding phenomena using *de novo* design, in which we recreate a desired function in an artificial protein through stepwise mutations. Through this bottom-up approach, *de novo* design better highlights the required protein-engineering and biophysical principles for function by removing the complexity inherent in natural proteins. We are interested in uncovering the minimal requirements needed for a single protein to recognize a set of unique-but-related DNA sequences. We first investigated the information presented by the DNA itself. We developed an algorithm that converts input DNA sequences into an array of non-covalent interactions accessible in the DNA major groove. It then aligns those arrays to provide a map of the contacts shared amongst all of the input DNA sequences. We first used this to expose the common information amongst the six sites the CI protein binds and predict alternative binding sites for CI in other phage genomes. This provided us with consensus patterns of information for us to design our protein against. We then looked at various structural parameters of the protein itself including the surface charge, recognition helix placement, and number of contacts to the DNA establishing a series of affinities based on these combinations. Finally, we showed sequence selective binding for our de novo protein to phage DNA based on the recognition helix. We hope to use this technology to combat food-borne viral pathologies, and anticipate other groups using it to combat other genetic diseases.

198-Plat
Directed evolution of dynamic, multi-state, and computational proteins
Sahand Rahi, Vojislav Gligorovski.
EPFL, Lausanne, Switzerland.
Directed evolution is a powerful method in biological engineering. Current approaches were devised for evolving steady-state properties such as enzymatic activity or fluorescence. A fundamental problem remains how to evolve dynamic, multi-state, or computational functionalities, e.g., folding times, on/off kinetics, state-specific activities, or switching and logic gating. This would require screening or selecting for multiple states of a protein of interest (POI) and their transition kinetics. We created a directed evolution paradigm which is germane for these goals ('optovolution'). The needed selection pressures are generated in vivo by an artificial signaling cascade where light controls the POI through optogenetics, switching between off (0) and on (1) states. In turn, the POI controls a Cdk1 cyclin, which has been engineered to be required to oscillate (1-0-1-0···) to drive the cell cycle in budding yeast. In this way, evolution acts on different input-output relations, transient states, and dynamics of the POI in every cellular replication cycle. Optovolution is continuous, based on self-selection, and was robust and efficient in our tests. We applied optovolution to evolve dozens of new variants of the popular LOV transcription factor El222, needed for diverse applications. These mutants are stronger, less

leaky, or have green- or UV-shifted excitation spectra, which was thought to be impossible and which is a requirement for lowering phototoxicity and multiplexing.

199-Plat
Defining the physicochemical basis for proton selectivity through the *de novo* design of functional proton channels
Vincent Silverman, Gisselle Prida, Arora Rohrbach, **Huong T. Kratochvil**.
Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.
Using *de novo* protein design, we engineer proteins from first principles to guide our understanding of natural protein structure and function. Specifically, our work utilizes design principles to define the physicochemical basis for proton selectivity and conductivity. Here, we designed functional proton channels to explore several mechanisms of proton selective transport across membranes. One such mechanism involves the formation of short-lived hydrogen-bonded water wires in starkly hydrophobic segments of the channel. Our previous work revealed that these transient water wires define proton channel selectivity and conductivity: we successfully engineered functional proton channels by incorporating polar Gln residues to keystone positions in an apolar pentameric pore. Multiscale reactive molecular dynamics simulations of our X-ray structures indicated that the introduction of these polar Gln residues lowered the energy barrier for the formation of the key proton-mediated water wires that enable the movement of protons across the otherwise apolar channel lumen. Our combined experimental and computational biophysics study demonstrated that these short-lived networks are necessary for the precise translocation of the proton. We now extend this work to explore how the length of pore hydration affects observed proton conduction rates. Molecular dynamics simulations of our new proton channel designs reveal unique hydration patterns in the channel lumen with the introduction of small polar Ser in key positions. Further, we introduce ionizable sidechains (i.e. His, Glu, Asp) into novel designed channels to define the roles of these amino acids in the selective transport of protons. Our unique design approach enables us to refine our understanding of the molecular mechanisms of proton channel selectivity and conductivity.

200-Plat
Tuning protein function: Rheostats, toggles, and neutrals, oh my!
Liskin Swint-Kruse.
Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, Kansas City, KS, USA.
Protein evolution occurs one amino acid change at a time. In some proteins, multiple amino acid changes must occur before protein function is significantly altered; other protein functions are quite sensitive to single amino acid changes. Indeed, in some proteins, a wide "tunable" range—from wild-type (or better), to partially-impaired, to dead—can be accessed using a set of single substitutions. In some cases, this range can even be accessed by site-saturating substitutions at single positions ("rheostat" positions). Understanding which protein functions are sensitive/resistant to single substitutions—and which proteins contain rheostat positions—is needed to advance bioengineering and personalized medicine. However, predicting the locations of rheostat positions and identifying which protein functions are sensitive to single substitutions is challenging. Many rheostat positions are far from active/binding sites, and their substitutions lead to a variety of non-canonical outcomes. Functional sensitivity not readily identified from protein conformation or the type of function; for example, the substrate binding affinities for two pyruvate kinase homologs, which have very similar tertiary structures and catalyze identical chemical reactions, show tunable ranges that differ by more than an order of magnitude. Intriguingly, dynamic coupling calculations show promising correlations between measured functional changes and the locations of long-range rheostat positions. Combined, results suggest that emergent properties of coupled amino acid networks could produce the complex outcomes observed for rheostat substitutions that enable functional tuning.

201-Plat
Dynamics of surfactant protein B at the alveolar air-liquid interface: Insights from molecular modeling and simulations
Mona Minkara, Tyler Locke.
Department of Bioengineering, Northeastern University, Boston, MA, USA.
In this study, we address the knowledge gap surrounding the structural characteristics of surfactant protein B (SP-B), a pivotal component of pulmonary

surfactant, vital for the lipid restructuring necessary for maintaining respiratory function. Utilizing comparative modeling with homologous Saposin-family proteins, predicted structures for SP-B in both open and closed conformations were generated. To represent the diverse conditions within the alveoli, molecular dynamics simulations were performed for these conformations in varying solvent environments, including water and chloroform. Across a minimum of 500 ns production simulation time, we gathered data on overall root-mean-square deviation (RMSD), per-residue root mean square fluctuation (RMSF), specific geometric parameters, and solvent distribution. The results highlight the relative stability of the closed conformation in water and reveal a considerable conformational change in the open conformation due to hydrophobic forces in water. Our findings elucidate the affinity of different protein regions to hydrophobic and hydrophilic environments, enhancing our understanding of SP-B's structural-functional attributes and potentially guiding the development of improved therapies for pulmonary diseases linked to surfactant dysfunction.

202-FlashTalk
HLA3DB: Comprehensive annotation of peptide/HLA complexes enables blind structure prediction of T cell epitopes

Sagar Gupta[1,2], Santrupti Nerli[1], Sreeja Kutti Kandy[1], Glenn L. Mersky[1], Nikolaos Sgourakis[1,3].
[1]Center for Computational and Genomic Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA, [2]College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA, [3]Department of Biochemistry and Biophysics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.

The class I proteins of the major histocompatibility complex (MHC-I) display epitopic peptides derived from endogenous proteins on the cell surface for immune surveillance. Accurate modeling of peptide/HLA (pHLA, the human MHC) structures has been mired by conformational diversity of the central peptide residues, which are critical for recognition by T cell receptors. Here, analysis of X-ray crystal structures within a curated database (HLA3DB) shows that pHLA complexes encompassing multiple HLA allotypes present a discrete set of peptide backbone conformations. Leveraging these representative backbones, we employ a regression model trained on terms of a physically relevant energy function to develop a comparative modeling approach for nonamer peptide/HLA structures named RepPred. Our method outperforms the top pHLA modeling approach by up to 19% in terms of structural accuracy, and consistently predicts blind targets not included in our training set. Insights from our work provide a framework for linking conformational diversity with antigen immunogenicity and receptor cross-reactivity.

203-FlashTalk
Design of protein fold-switches using alchemical mutation free energy calculations

Lisa Schmidt, Bert L. de Groot.
Department of Theoretical and Computational Biophysics, Max Planck Institute for Multidisciplinary Sciences, Goettingen, Germany.

Protein fold-switches are important in many biological processes and exhibit great structural diversity. This makes them perfect for protein design studies with the aim to make programmable proteins with new functions. Even though protein design has made remarkable progress in the last years, designing fold switching proteins *de novo* is still a challenging task. We apply alchemical free energy simulations with PMX to calculate the mutation free energy differences between different protein conformations. Combining the relative free energy differences of different mutations, their inter residue contact profiles and NMR data we are able to find residues and interactions that are important for the stabilization of different protein folds and map out fold switching behavior. Compared to long equilibrium simulations often used to investigate interaction networks, this method is faster and can be applied for large scale screening of interaction networks thereby helping to understand fold switching behavior and design new fold switches.

204-FlashTalk
Direct prediction of intrinsically disordered protein conformational properties from sequence

Jeffrey M. Lotthammer, Garrett M. Ginell, Daniel Griffith, Ryan Emenecker, Alex S. Holehouse.
Department of Biochemistry and Molecular Biophysics, Washington University in St. Louis, St. Louis, MO, USA.

Intrinsically disordered regions (IDRs) are ubiquitous across all domains of life and play a range of functional roles. While folded domains are generally well-described by a stable 3D structure, IDRs exist in a collection of interconverting states known as an ensemble. This structural heterogeneity means IDRs are largely absent from the PDB, contributing to a lack of computational approaches to predict ensemble conformational properties from sequence. Here we combine rational sequence design, large-scale molecular simulations, and deep learning to develop ALBATROSS, a deep learning model for predicting ensemble dimensions of IDRs—including the radius of gyration, end-to-end distance, polymer scaling exponent, and ensemble asphericity—directly from sequence. ALBATROSS enables the instantaneous prediction of ensemble average properties at proteome-wide scale. ALBATROSS is lightweight, easy-to-use, and accessible as both a locally installable software package and a point-and-click style interface in the cloud. We first demonstrate the applicability of our predictors by examining the generalizability of sequence-ensemble relationships in IDRs. Then, we leverage the high-throughput nature of ALBATROSS to characterize sequence-specific biophysical behavior of IDRs within and between proteomes.

Platform: Computational Methods and Machine Learning, Artificial Intelligence, and Bioinformatics I

206-Plat
Expectation maximised molecular dynamics: An unsupervised machine learning approach toward rapid estimation of biomolecular transition barriers

Neelanjana Sengupta[1], Pallab Dutta[1], Abhay Kshirsagar[1], Parth Bibekar[2].
[1]Department of Biological Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur, India, [2]Indian Institute of Science Education and Research Kolkata, Mohanpur, India.

Modern-day efforts towards uncovering emergent protein structure-function relationships leverage combined data from experiments and computations. A persistent challenge remains in the capture of the location and the magnitude of high transition barriers in complex energy landscapes. We have recently developed an algorithm named 'Eexpectation maximized molecular dynamics' (or EMMD) that incorporates a statistical inference-based approach in estimating free energy barriers related to rate-limiting transitions. The method bypasses conventional thermodynamic sampling by connecting metastable basins using Bayesian likelihood maximization. A tunable self-feedback protocol in incorporated to prevent unnecessary sampling that does not effectively contribute to the underlying distributions. The algorithm demonstrates significant efficiency in predicting experimentally known free energy barriers in disease relevant proteins such as a kinase and an intrinsically disordered protein. EMMD demonstrates the efficacy of an unsupervised machine learning approach that can be leveraged to significantly improve collective variable search and accuracy within routine enhanced sampling protocols.

207-Plat
Deep-SPT, a deep learning toolbox for single particle tracking in 3D, reveals how biological motion encodes function

Jacob Kæstel-Hansen[1,2], Tom Kirchhausen[3], **Nikos S. Hatzakis**[1,2].
[1]Department of Chemistry, University of Copenhagen, Copenhagen, Denmark, [2]Department of Chemistry, Novo Nordisk Foundation Center for Optimised Oligo Escape, Copenhagen, Denmark, [3]Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, MA, USA.

Biological motion is highly heterogeneous and displays a variety of diffusion types that may vary drastically across both systems and time and are dependent on regulatory cues, spatial localization and nanoscale interactions. Our current understanding of heterogeneous diffusion primarily relies on ensemble techniques reporting the behavior of a large ensemble of biomolecules, which masks heterogeneous behavior and thus the actual underlying biology. Single particle methods enable direct observation of single-particle tracking (SPT) has enabled the quantitative analysis of dynamic biological processes with nanometer spatial and millisecond temporal resolution, revealing dynamic behaviors such as cell entry pathways and trafficking of biologicals previously masked in ensemble averaging.Herein, we