SESSION 1D

FRONTIERS IN PEPTIDE AND PROTEIN SCIENCE

Session Moderator: Rachel Klevit, Biochemistry
MGH 228
12:30 PM to 2:15 PM

* Note: Titles in order of presentation.

Learning the Ideal Similarity Matrix for Peptide Sequences with Given Functionality
Dylan Hylander, Sophomore, Engineering Undeclared
Mentor: Siddharth Rath, Materials science and engineering, Genetically Engineered Materials Science and Engineering Center
Mentor: Mehmet Sarikaya, Materials Science & Engineering

The Genetically Engineered Materials Science and Engineering Center (GEMSEC) labs revolve around designing and synthesizing genetically engineered peptides for inorganic materials (GEPIs). Experimentally characterizing GEPIs can be slow, and therefore a computational method that can predict functionalities would greatly accelerate the development of bio/inorganic interface design and implementations. The Pairwise Similarity Score is a proven predictor of relative binding affinity and has been used to predict GEPIs specific for quartz, gold, hydroxyapatite, and MoS2. In previous work, a similarity matrix was updated based on whether a peptide (Strong or Weak binding) had higher similarity to strong peptides and less similarity with weak peptides. Our method instead obtains the most ideal similarity matrix via stochastic gradient descent to best predict the relative binding affinities. The values in an amino-acid similarity matrix are randomly initialized and subsequently updated until convergence by minimizing the errors in binding affinity prediction. 5-fold cross-validation is used as a metric to evaluate performance on test data. We expect to observe higher predictability with this learned similarity matrix than using a literature matrix. This would compound work done by the high throughput screening, confirming count numbers observed during phage display are correlated with their actual binding affinity, while using a novel large dataset to test known successful predictive models. All in all, the work carried out in this project accelerates the development pace of bio-nano-devices of the future.

The research is supported by NSF/DMR-DMREF program under the Materials Genome Initiative.

SESSION 1D

FRONTIERS IN PEPTIDE AND PROTEIN SCIENCE

Session Moderator: Rachel Klevit, Biochemistry
MGH 228
12:30 PM to 2:15 PM

* Note: Titles in order of presentation.

VSEPR Encoding of Peptide Structures for Predicting Binding-Affinity
Jonathan Taylor (Jonathan) Francis Landau, Junior, Mathematics
Ximing Lu, Junior, Computer Science (Data Science), Statistics
Undergraduate Research Conference Travel Awardee
Mentor: Mehmet Sarikaya, Materials Science & Engineering
Mentor: Siddharth Rath, Materials science and engineering, Genetically Engineered Materials Science and Engineering Center

The goal of this project is to encode peptides, i.e., short amino acid sequences, in terms of smaller molecular components such as their VSEPR (Valence Shell Electron Pair Repulsion) features for training interpretable models with reasonable predictability of functionality. This enables us to go beyond the limitations imposed by treating peptides as sequences of letters, thereby enabling a generalized encoding that works for lipids and other biomolecules that are of interest in a comparable scenario. Biological processes are rarely disjoint and often complicated which lends justification to our approach. Current methods for binding affinity prediction, such as one-hot encoding, where letter-based sequences are converted to a binary representation, do not take into account molecular level features. Combined with a neural network, such a simple encoding is better at predicting affinities of short peptides, e.g., 5-9 Amino acids long, but with an increase in length from 9 to 10, the predictability suffers an exponential drop. Several alternatives have been employed in literature, but they also suffer from the negative impact of distal effects. In the VSEPR approach, encoding peptides in terms

of their component functional-group geometries enables us to encode the actual physical length, rather than the number of amino acids. This leads to an overlap between peptides of different length, thereby reducing the fall in predictability. In this encoding, we create 5 channeled matrices with each channel corresponding to 'central-atom connectivity', 'bond-types', 'bond-lengths', 'bond-angles' and 'lone-pairs' that is then fed through a Deep Residual-Neural-Network. The metrics used to evaluate the models are Pearson-Correlation, Spearman-Rank-Correlation-Coefficient, and Area-under-Receiver-Operating-Curve. With this technique, we were able to consistently predict binding affinities of peptides without an appreciable loss between 9 or 10 length peptides. This method would allow one to create length invariant encodings, not limited to just peptides, significantly improving the practicality of using such a model. The research is supported by NSF/DMR-DMREF program under Materials Genome Initiative.

SESSION 1D

FRONTIERS IN PEPTIDE AND PROTEIN SCIENCE

Session Moderator: Rachel Klevit, Biochemistry
MGH 228
12:30 PM to 2:15 PM
* Note: Titles in order of presentation.

Molecular Dynamic Folding Propensities of Genetically-Designed Dodecapeptides on Single Layer Atomic 2D Solids
Tatum Grace Hennig, Senior, Atmospheric Sciences: Chemistry
Mentor: Mehmet Sarikaya, Materials Science & Engineering

Our laboratory, GEMSEC, which operates at the intersections of biology-materials-informatics fields, is developing materials and methods to seamlessly bridge biology with solid-state devices towards establishing the foundations of future hybrid devices, e.g., bioelectronics, bionanosensors, and biomolecular fuel cells. Towards this goal, we use the smallest functional biomolecule, peptide, combined with the smallest functional solid in materials science, i.e., single atomic layer materials. Herein, we study the interactions of genetically designed peptides with surfaces of graphene, a semimetal. A phage display library-selected peptide, GrBP5, is a graphene-binding dodecapeptide that has a wide range of applications. Since peptides have short amino acid sequences, they are known to display intrinsically disordered structures in solution. Here we study the conformational propensities of the WT peptide and its rationally designed mutants under a variety of experimental conditions (pH, concentration, temperature, time, etc.) to understand their behavior on solid surfaces that includes surface phenomena from binding, surface diffusion, intermolecular interaction and self-organization. Molecular dynamics (MD) simulations of WT-GrBP5 and its mutants have been completed in water and on graphene for 200ns, 20,000 timeframes under different temperatures and pH values that range from 5 to 55 oC and 3.5 to 10.0, respectively. The analyses, including the RMSD maps and Ramachandran plots, show explicit folding propensities, stable and unstable structures, for a given sequence under a given set of experimental conditions. The computational modeling, backed up by experimental validations carried out under similar conditions, are leading to the design of novel peptide sequences with predictable behavior under desired environmental conditions. The fundamental understanding of the differences in conformational behavior of GrBP5 mutants are now extended to other solid-binding peptides that are specific to semiconductor and insulator single layer materials providing the much essential information for the design of hybrid devices of the future. The research supported by NSF/DMR-DMREF program.

POSTER SESSION 3

MGH 241, Easel 135
2:30 PM to 4:00 PM

The Foundation of Biomineralization via Genetically-Designed Peptide-Ion Interactions
Andrea Ming Hwei Dao, Senior, Chemical Engineering
Levinson Emerging Scholar, NASA Space Grant Scholar
Aniruddh Saxena, Junior, Bioengineering
Mary Gates Scholar, UW Honors Program
Yousef Mohammed Baioumy, Senior, Chemical Engineering
Mentor: Mehmet Sarikaya, Materials Science & Engineering
Mentor: Deniz Tanil Yucesoy, Materials Science and Engineering

Biological mineralization is the formation of minerals in hard tissues guided by proteins. Unique aspects of these minerals include the molecular control of hierarchical structure, intricate architectures, and multifunctional properties for inspiration in bionanotechnology and nanomedicine applications. Numerous biomineralization strategies have been developed in hard tissue regeneration therapies. However, there is currently no in-depth understanding of how proteins regulate the synthesis of these inorganics or the physiological formation of the minerals. The ability to control mineral formation for biomedical applications, therefore, is still limited to the use of a few mineral-directing proteins extracted from tissues. Biomineralization can also be controlled using short peptide domains derived from natural proteins known to have a regulatory role in mineralization. Our laboratory has designed peptides derived from amelogenin (ADPs), the key protein

in tooth formation, using combinatorial selection and computational design, whose utility in rebuilding hydroxyapatite (HAp) mineral on tooth has been demonstrated in numerous case studies. The goal here is to understand the fundamental mechanisms of biomineralization guided by ADP5 and develop a methodology to form HAp with exclusive control of its growth kinetics and mineral crystallography. We designed mutants of ADP5 to investigate changes in mineralization kinetics, nucleation, and morphology. In the current study, we are establishing the conditions for ion-peptide interactions on the onset pH for mineral nucleation using calcium/phosphate and mutant ADPs. The goal is to gain insights into the correlation between sequence domains and biomineralization outcomes eventually facilitating greater control over the reaction and further optimize remineralization approach. The developed method has a high potential to develop non-invasive oral health care materials and methods by restoring mineral loss, the root cause of dental ailments and, eventually, help bring clinical and over-the-counter dental products into the market with preventive, restorative, therapeutic, and cosmetic characteristics. Sponsored by SoD Spencer Funds.

POSTER SESSION 3

Balcony, Easel 102

2:30 PM to 4:00 PM

Fabrication of a High-Throughput Whispering Gallery Mode Dip Sensor for Peptide Characterization

John Taylor (John) Hamann, Senior, Mechanical Engineering
Willem L Weertman, Graduate,
Mentor: Mehmet Sarikaya, Materials Science & Engineering
Mentor: Richard Lee, Materials Science & Engineering

Whispering Gallery Mode (WGM) sensors have unprecedented sensitivity in the optical detection of label-free biomolecules. These sensors can detect surface adsorption and have been used to detect single molecule adsorption and interaction processes. By observing resonance shifts during molecular interactions, WGM sensors can characterize a molecule's surface adsorption. The goal of this project is to develop a robust WGM dip sensor array controlled by a three-axis stage in order to perform high-throughput characterization of peptide binding and adsorption within a 96-well plate format. The peak of spectral absorbance is the WGM resonance, and as this changes with surface adsorption we measured a spectral shift. Using this spectral shift in combination with the known concentration of our peptide species, we determined binding kinetics. The WGM sensor was used to characterize different peptide sequences to further understand the effects of peptide mutations on binding kinetics. A single microsphere resonator was used as proof of principle and will eventually be adapted to an array of eight WGM microsphere resonators to generate large amounts of data. This high throughput approach will provide the much needed large amount of quality data that is necessary for the development and adaptation of machine learning and applied statistical analysis algorithms toward the eventual development of artificial intelligence platforms in material science. The project is supported by NSF-DMREF through the Materials Genome Initiative.

POSTER SESSION 3

MGH 241, Easel 136

2:30 PM to 4:00 PM

A Neural Network for Predicting Peptide Binding Affinity

Francesca Caroline Green, Senior, Materials Science & Engineering
Louis Stokes Alliance for Minority Participation, NASA Space Grant Scholar
Mentor: Mehmet Sarikaya, Materials Science & Engineering
Mentor: Siddharth Rath, Materials science and engineering, Genetically Engineered Materials Science and Engineering Center

Our Lab, GEMSEC, uses molecular biology, bioinformatics, genome sciences, and engineering for de novo design of short amino acid sequences for various applications such as tooth remineralization strategies in dentistry, biosensing in cancer diagnostics, and bioelectronics in single-molecule detection. Designing and constructing peptides for a desired function begins with selecting the appropriate sequence of amino acids with the predictive conformation that affects the function. In this project we use latent-space representation (matrix factorization) in conjunction with a simple neural network to create a model that is able to predict peptide binding affinity to several alleles of MHC-I protein. Python was used to encode amino acids by creating data frames defining the functional groups within them, differing by n-terminus, intermediate, and c-terminus of each amino acid and their placement along the backbone of each structure. A tensor was created using the data frames describing each amino acid to encode the 9- and 10-length sequences of thousands of unique peptides from the Immune Epitope Database. Each chemical structure and peptide sequence can be described by k attributes, or latent features. Matrix factorization was used to discover the latent features and send this feature encoding to a neural network (NN) to determine binding affinity. The goal is to minimize the mean-squared-error by stochastic gradient descent in a supervised learning protocol. The two modules of matrix factorization and NN provide an optimum between interpretability and predictability simultaneously.The successful prediction of peptide binding affinity towards nanoscale targets provides novel opportunities for drug design towards

targeted public health initiatives and in technology applications such as bio/nano hybrid devices. This research is supported by NSF/DMR-DMREF program under the Materials Genome Initiative.

POSTER SESSION 3

MGH 241, Easel 133

2:30 PM to 4:00 PM

Matrix Factorization on Total Similarity Scores for Peptide Binding

Chris Laing Pecunies, Senior, Mat Sci & Engr: Nanosci & Moleculr Engr
Mentor: Mehmet Sarikaya, Materials Science & Engineering

The study of peptide self-assembly on solid surfaces has the potential to catalyze numerous nanotechnological advances such as biosensors and nanoelectronics. A comprehensive understanding of the factors that influence peptide binding to solids would allow for expansive integration of biomolecules and solid-state devices, and organic-inorganic interface bridging will permit greater information flow between biological systems and technological devices. The Genetically Engineered Materials Science (GEMSEC) lab has engineered peptides that are capable of binding to substrates such as graphene, monolayer molybdenum disulfide, and boron nitride nanosheets. Utilizing experimentally determined binding affinities of these binding peptides alongside a database of biochemical and physicochemical properties of amino acids, we have developed a method to computationally predict short amino acid sequences that preferentially bind to atomically flat surfaces. Matrix factorization and linear regression is used to train a model capable of predicting an experimentally observed peptide count number (observed during sequencing of eluate of phage display biopanning) from 8 Total Similarity Scores (TSS) that are calculated from 8 novel similarity matrices. This model is then used to predict the ranking of actual binding affinities of genetically engineered peptides to monolayer molybdenum disulfide from fluorescence microscopy experiments. The ability to predict solid binding by peptides will facilitate further research into peptide structure and functional properties upon adsorption. Ultimately, these methods will be implemented in a cohesive software platform using machine learning and signal processing tools to allow determination of sequence-property linkages and pattern recognition in a larger bioinformatics context, and allow for nanomedical and nanotechnological advances at the intersection of materials science, biology, and genetics. Supported by the NSF-DMREF program through the Materials Research Initiative.