UNIVERSITY of WASHINGTON

Undergraduate Research Symposium May 17, 2019 Mary Gates Hall

Online Proceedings

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.

Genetically Encoded Photocleavable Linkers for Protein Release from Biomaterials

Alder Colleen Strange, Senior, Biochemistry, Individualized Studies, Psychology
UW Honors Program
Mentor: Cole DeForest, Chemical Engineering & Bioengineering
Mentor: Jared Shadish, Chemical Engineering

Precise spatiotemporal control over biochemical cue presentation is necessary to mimic the complex, heterogenous environments found in biological systems. Achieving this level of control within engineered microenvironments would allow for the manipulation of cell growth and differentiation, which could be utilized in tissue engineering and drug delivery. To this end, we developed a method that utilizes fusion proteins made from a novel PhotoCleavable protein linker (PhoCl) and a protein of interest (POI). This method allows for spatiotemporal control of POI release from hydrogels in response to cytocompatible violet light (λ = 405). This system is flexible, as PhoCl can be conjugated to many different POIs, including fluorescent proteins, enzymes, and growth factors, and was found to not affect protein function. Additionally, PhoCl undergoes a green-to-red transition after photocleavage, allowing for real-time tracking and quantification of POI release. As PhoCl cleaves in response to visible light, which is less damaging to cell function and has a greater tissue penetration depth than the traditionally used UV light, PhoCl fusion proteins hold promise for use *in vivo*. To demonstrate the feasibility of this system, PhoCl fusion proteins were formed with several fluorescent proteins (e.g., mRuby, sfGFP, mCerulean). Conjugating these fusion proteins into gels and exposing them to patterned light produced spatiotemporal localized release of proteins with micron scale resolution, which was demonstrated through fluorescent imaging of the photopatterned gels. To support the potential *in vivo* applications of this system, PhoCl was also used in mammalian cell studies with epidermal growth factor (EGF). These studies showed the expected increased cell

growth in response to photomediated EGF release. This illustrates the potential versatility of the PhoCl system in biological applications, thus supporting the relevance of this novel system to tissue engineering and drug delivery methods.

Inhibition of a D*e Novo* Synthetic Amyloidogenic Peptide's Aggregation by Naturally Occuring Amyloids

Charles Haoyi Lin, Senior, Biology (Molecular, Cellular & Developmental), Biochemistry
UW Honors Program
Mentor: Valerie Daggett, Bioengineering

Amyloid diseases are characterized by the aggregation and buildup of proteins in vital tissues and organs. Insoluble β-sheet amyloid fibrils were previously thought to be the major underlying cause of tissue degeneration and cell death. However, recent experimental evidence suggests that soluble oligomers, which form during protein aggregation and before polymerization into fibrils, are the principal cause of toxicity in mammalian cells. These toxic oligomeric protein assemblies are believed to share a common sequence-independent secondary protein backbone structure known as α-sheet. This project proposes the investigation of a synthetic peptide known as AP3 that is capable of forming toxic oligomers and β-sheet amyloid fibrils. This peptide was *de novo* designed with a completely randomized sequence which preserves the underlying chirality that produces α-sheet character leading to its exhibition of amyloidogenic properties under acidic conditions. Furthermore, AP3 aggregation was shown to be inhibited by three naturally occuring amyloid proteins implicated in their respective dieseases: Amyloid Beta (Alzheimer's), IAPP (Type II Diabetes), and Transthyretin (Cardiac Amyloidosis). Analysis using dot-blot assays, soluble oligomer binding assays (SOBA), and BLITz assays will provide additional insight into the behavioral, binding, and kinetic properties of AP3. Upon further evaluation, we aim to demonstrate the ability of AP3 to serve as a synthetic model for naturally occuring amyloids and provide a better understanding of amyloidogenesis as well as the interactions between amyloidogenic species. This research will prove useful in the creation of more effective amyloid inhibitors and treatments for amyloid diseases.

Reprogramming *S. cerevisiae* Mating and Sporulation for High-Throughput Enrichment of Peptide Binders

Riley Maeliann (Riley) Stockard, Senior, Bioengineering
Levinson Emerging Scholar, Mary Gates Scholar, UW Honors Program, Washington Research Foundation Fellow
Mentor: Eric Klavins, Electrical Engineering
Mentor: David Younger, Electrical Engineering

Since 1982, with the introduction of insulin as the first recombinant protein therapeutic, peptide and protein drugs have grown to encompass 10% of the pharmaceutical market and are the fastest expanding class of drugs. Advantages of using peptides over small-molecule drugs include high potency, selectivity, and capability to be engineered for a diverse range of targets, most commonly to disrupt or facilitate key protein-protein interactions (PPIs) in the human body for a therapeutic effect. To search for strong binders for a therapeutic target, combinatorial peptide libraries of up to billions of different sequences are synthesized and screened against promising targets. Due to the enormous library size, screening for high affinity binders often requires multiple rounds of enrichment in order to isolate the most potent molecules, a laborious and potentially bottlenecking step in developing protein drugs. Current approaches that have strategies for enrichment, such as phage display and yeast surface display, are limited to screening a library of binders against one target instead of multiple targets (library-on-library). This proposal describes the development of a peptide binder screen utilizing a simple workflow of repeated mating and sporulation of genetically engineered S. cerevisiae, or baker's yeast. This technology improves the throughput of established screening methods through a library-on-library format that efficiently isolates high-affinity peptide binding interactions.

Wrangler: Toward Improved Peptide Design

Jennifer Ann (Jenny) Ferina, Senior, Bioengineering
UW Honors Program
Mentor: Valerie Daggett, Bioengineering
Mentor: Matthew Childers, Bioengineering

While there is an abundance of computational tools for protein design, the vast majority of them focus on static structures. Data for the most probable rotamers of an amino acid are often obtained from mining its appearances across crystal structures, which do not take into account the native dynamics or conditions of a protein. Additionally, conformational secondary structure propensities are often derived from the frequency that an amino acid appears within a certain range of dihedral backbone angles in static structures. However, these data do not emulate realistic conditions, because proteins are dynamic and surrounded by solvent; thus, incorporation of dynamic data should improve the design process. Wrangler, our in-house software for peptide design, integrates dynamic data for both rotamers and secondary structure propensities. The data are obtained from molecular dynamics (MD) simulations in explicit water, taking into account interactions with surrounding solvent to better sample the most probable conformations. Several amyloid peptide aggregation inhibitors were designed using Wrangler's scoring metrics to validate the software. These inhibitors were shown via MD simulations to better retain secondary structure than our current standards. Future work involves synthesizing the peptides in the wet lab to determine whether they experimentally perform better than our current inhibitors.

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.

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.