

Summer RosettaCon 2026 REU Poster Abstracts

Colter Bufford

Southern Nazarene University

REU: Gray Lab, Johns Hopkins University

Benchmarking Deep Learning and Physics-Based Models for Orthogonal Protein Specificity Prediction

Orthogonal protein pairs have two protein pairs where one pair was engineered from the other such that each pair only binds to their partner and not to the other pair. These systems have therapeutic potential, with applications in cell-based therapies and synthetic signaling networks. Engineering such orthogonal pairs requires identifying key residues that can be mutated to inhibit binding to the native partner while introducing binding to its orthogonal counterpart, without allosterically perturbing the rest of the protein. Previous work has built orthogonal systems using physics-based models like Rosetta to assess Gibbs free energy changes from promising mutations, but predictive accuracy is limited by the approximate nature of these score functions and their difficulty capturing entropic and long-range allosteric effects.

Steven Engelbretson

University of Texas Austin

REU: IPD, University of Washington

Small-Molecule Binder Design: Paclitaxel as a Case Study

The current era of *de novo* small molecule protein binder design involves many successful discrete and substitutable steps. However, research groups currently face high compute costs to curate strategies for particular molecules and low success rates. Here we introduce a new design strategy that aims to establish a generalizable zero-shot small molecule binder design pipeline that is capable of significantly reducing turnaround time for individual targets. Designed pipelines must be tested against many ligand targets to assess the accuracy of the pipeline—through experimental validation of results—and find optimal workflows, design parameters, and output filters. The small molecule drug Paclitaxel possesses many features that make it a desirable target to test pipelines, such as having a small and rigid structure, while possessing diverse functional groups with an overall low polarity. The presence of high stereocenters also presents an important challenge, as the current generation of structure prediction models often fail to precisely represent the chirality of non-protein molecules. Final designs of the pipeline will then be experimentally validated through Surface Plasmon Resonance to estimate target binding affinity. Additionally, successful high-affinity binders can be leveraged to curate a database for the development of future protein-ligand interaction models.

Allie Gibson

Georgetown University

REU: Whitehead Lab, University of Colorado Boulder

Investigating Antibody Molecular Recognition Patterns for Public Viral Epitopes

Antibodies are crucial to establishing population-level resistance to disease. Developing computational models that could predict antibody structures would be incredibly useful for vaccinology; however, the variability of antibody structure makes prediction difficult. Analyzing antibody binding to public epitopes would allow for the identification of peptide patterns that elicit public responses, thus serving as validation for an experimental dataset that could be used to train models to predict antibody structure. In this project, truncations of peptides from influenza, Epstein-Barr, and SARS-CoV-2 viruses were expressed to be screened against a library of antibody antigen binding fragments.

Lia Headley

Howard University

REU: Schoeder Lab, Leipzig University

De Novo Design and Glycan-Aware Optimization of EGFRvIII Binders

EGFRvIII is a constitutively active oncogenic receptor found in glioblastoma, arising from deletion of exons 2–7 that removes the extracellular ligand-binding domain and drives ligand-independent, tumor-promoting signaling. Computational protein binder design has traditionally treated target surfaces as static, often overlooking glycan architecture that can shape molecular recognition and binding accessibility.

The purpose of this study is to determine whether explicitly modeling glycosylation during de novo protein design improves the predicted binding performance of protein binders targeting EGFRvIII. We hypothesize that glycan-aware design strategies will generate binders with stronger predicted binding energetics and higher structural confidence than glycan-blind approaches.

This computational structural biology study used RFdiffusion to generate 100 mini-protein binder backbones targeting hotspots A563, A566, A578, and A585 on the EGFRvIII structure (PDB 8UKW), followed by sequence optimization with ProteinMPNN. Two design cohorts were compared: a glycan-blind cohort (n=43) and a glycan-aware cohort incorporating NAG proximity bias (n=42). Candidate structures were predicted using AlphaFold3, and binder quality was assessed through Rosetta interface energy (ΔG_{cross}), interface contacts, glycan proximity, hydrogen bonding, and electrostatics.

Early findings show that glycan-aware binders achieved significantly stronger predicted binding energetics than glycan-blind binders (median G_{cross} = -57.7 vs. -41.7 REU; $p < 0.001$), with all

ten top-scoring designs by Rosetta energy belonging to the glycan-aware cohort. This effect was independent of binder length. Top candidates also showed a trend toward higher interface confidence in the glycan-aware group. These results support glycan-awareness as a valuable design constraint for targeting heavily glycosylated oncogenic receptors, with implications for improving computational strategies in glioblastoma-directed therapeutic design. Future work includes recombinant expression and surface plasmon resonance binding assays to experimentally validate top candidates, such as egfr_binder_56.

Keagan Hitt

Worcester Polytechnic Institute

REU: Samanta Lab, University of South Florida

Language Model Adapter for de novo Trans-Membrane Beta-Barrel Protein Design

An important group of proteins to design are membrane proteins, and a specific type of membrane protein is the transmembrane β -barrel (TMB), commonly used as nanopores. Nanopores are used for analysis and sequencing of DNA molecules. To analyze specific targets, the chemical properties of the TMB need to be compatible with the target, otherwise the nanopore will be unable to detect change in current. De novo design would allow for making these versatile and functional nanopores. With current methods, it is difficult to design these nanopores de novo due to the lack of PDB representations. Our goal is to computationally design sequences for membrane embedded nanopores by combining the structure-based protein design model ProteinMPNN with the membrane protein language model MemDLM. We expect that using a combination of embeddings from sequence-based language models and weights of structure-based design models will improve membrane protein design. Our hypothesis is that integrating explicit physical 3D-geometric principles with evolutionary language models produces proteins that are both highly stable and biologically functional, overcoming limitations of relying on just one approach. After training MemDLM, an adapter will be added to it. This adapter is a layer that can accept structural information from ProteinMPNN and feed it to the pre-trained MemDLM with frozen parameters. The system will then be tested on a blind dataset of both designed and naturally occurring beta barrels, and its ability to create effective TMBs will be tested through sequence identity, hydropathy of the barrel core, and propensity for beta sheet formation.

Francisco Jimenez

Queens College, City University of New York

REU: Khare Lab, Rutgers University

De Novo Design of a Type B Carbohydrate-Binding Modules for Targeted Alginate Recognition

Alginate is a highly valuable biopolymer derived from seaweed with extensive biomedical and industrial applications. Engineering proteins to interact with its irregular structure is a significant challenge. While Type A Carbohydrate-Binding Modules (CBMs) have an accessible aromatic plane, alginate requires a Type B CBM that has extended, positively charged concave grooves.

The de novo generation of these multiple-subsite interfaces is computationally demanding due to the "residue island problem." This project aims to engineer novel, highly specific Type B CBMs for targeted alginate binding using an iterative, high-throughput deep-learning pipeline. To navigate this challenge, our pipeline uses an inline loop of structural scaffolding and targeted sequence bias to sample the conformational design space, ensuring a strong pool of stable, chemically complementary candidate designs.

Rosemary Jurado

Bowie State University

REU: Gray Lab, Johns Hopkins University & Damo Lab, Bowie State University

Expanding Host Nutritional Immunity Through the De Novo Computational Design of Rare Metal-Binding Proteins

Metals are essential for life. Roughly 1/3 of the proteome is predicted to be represented by metalloproteins. Metals serve as enzymatic co-factors, structural stabilizers of proteins, and are frequently used to modulate signaling cascades. The importance of metals in biology is leveraged by the human immune system at host-pathogen interface. Several members of the S100 protein family sequester the essential transition metals zinc, manganese, and copper to starve invading pathogens, a process termed nutritional immunity. Metal sequestration is a potent antimicrobial mechanism, as essential enzymes requiring metal cofactors become inactivated. While there are S100 proteins that can bind cobalt, copper, iron, manganese, and zinc with high affinity, the full spectrum of metal chelation is incomplete. To engineer a de novo designed vanadium binding protein with the potential to expand the repertoire of nutritional immunity reagents, we utilized a computational approach. A multi-sequence alignment of S100 proteins was generated which was used to identify the canonical transition metal binding site comprising four amino acids: three histidines and one aspartic acid. This canonical site was evolved using dEVA, a multi-objective generative AI framework to create 100 evolutionary design iterations of natural sequence templates. Metal3d was used to evaluate how well these new designs would bind zinc (as metal3d is currently optimized to detect zinc binding only). We evaluated the top ~5 design structures using AlphaFold to assess whether the predicted folds preserved the intended His₃Asp geometry. This analysis revealed a key limitation: Metal3D, the scoring module underlying dEVA, is trained and validated exclusively on zinc. While its authors note the framework is theoretically extensible to other metals through retraining, it has no built-in parameter to specify a different target metal, and its accuracy for vanadium coordination specifically is unvalidated. Because vanadium is the metal this project aims to coordinate, and its favorability in the Irving-Williams selectivity series differs substantially from zinc's, scores from Metal3D could not be trusted to reflect true vanadium-binding potential. A baseline comparison against ProteinMPNN did confirm that dEVA's His/Asp sequence bias adds real value over naive sequence design, supporting the underlying design logic even as the scoring step needed to change. Based on this, we are transitioning to a new computational pipeline built around RFDiffusion, which generates protein backbones without being locked into a zinc-specific geometry and allows the metal-binding site to be specified directly. Designs will be scored using AllMetal3D, a more general metal-detection model, and the secondary coordination shell will be

optimized afterward using LigandMPNN to improve rigidity and selectivity around the bound vanadium. The immediate next step is a baseline validation: masking known vanadium-binding sites in natural proteins and testing whether LigandMPNN can recover the native sequence when given the vanadium ligand's position, which will confirm the model is learning real coordination geometry rather than memorizing training examples. From there, the plan proceeds through backbone generation, sequence design, scoring, and structural validation via AlphaFold3 or Boltz-2, with the goal of identifying a small set of designs for experimental testing in the Damo Lab.

Olivia Kong

University of California, Berkeley

REU: Oberdorfer Lab, TU Graz

Multi-state Design for a de novo Fluoroacetate Dehalogenase

The exceptional strength of the carbon-fluorine bond confers resistance to degradation and improved solubility. However, these same properties make fluorochemicals such as per- and poly- fluoroalkyl substances (PFAS) difficult to degrade, resulting in adverse environmental and health impacts. The natural enzyme fluoroacetate dehalogenase (FACD) defluorinates fluoroacetate to glycolate, making FACD a promising basis for designing fluorochemical-degrading enzymes. This project computationally designs a de novo fluoroacetate dehalogenase for the degradation of fluorocompounds. Our approach incorporates multiple active site geometries, an improvement from the single-state pipelines of current enzyme design methods. Key considerations for the active site geometry include bond angles, interatomic distances, and secondary structure. The design pipeline is constructed using ProtFlow, with core elements including backbone generation with RFdiffusion3, iterative backbone refinement, and reaction state prediction using Boltz-2. Throughout our refinement cycles we observe an increase in mean pLDDT and a decrease in motif RMSD. Our best designs show accurate reproduction of the active site geometry to $\sim 1\text{\AA}$ atom-wise RMSD. This method for designing defluorinating enzymes can be extended to accommodate various fluorochemical substrates to help recycle and degrade PFAS.

Shuaicheng Lai

Johns Hopkins University

REU: Merck Rahway

Accelerated Enzyme Engineering via Cell-Free Expression, Rosetta-guided design and High-Throughput screening

Enzymatic intermolecular C–C bond formation is an emerging strategy for pharmaceutical synthesis, offering a selective biocatalytic alternative for C–C bond formation under mild conditions. Thiamine pyrophosphate (TPP)-dependent enzymes are particularly attractive biocatalysts for this purpose, as they promote stereo-controlled aldehyde–ketone carbonylation under mild aqueous conditions. Bioinformatics-guided discovery identified YerE-2, a

TPP-dependent enzyme homolog, that catalyzes asymmetric intermolecular C–C bond formation between aldehyde and ketone substrates. However, YerE-2 is poorly soluble in *E. coli*, limiting its practical utility as a starting point for biocatalyst development. To address this challenge, we developed an accelerated protein engineering approach to enhance YerE-2 soluble expression. Traditional protein engineering workflows, relying on recombinant protein expression in microbial hosts followed by high-throughput screening to identify improved variants, are limited by the time required for cell growth and protein expression. Advances in cell-free protein expression, and structure-based computational design offer opportunities to accelerate enzyme engineering. In this study, we establish an integrated workflow that combines cell-free protein expression, Rosetta-guided rational protein design, and fluorescence-based high-throughput solubility assays to improve YerE-2 solubility and activity. This work provides an improved TPP-dependent biocatalyst for asymmetric C-C bond formation and lays the foundation for an accelerated protein engineering workflow.

Kelvin Lartey

Franklin and Marshall College

REU: Ovchinnikov Lab, MIT

Evaluating Organism-level Evolutionary information as Prior for Differentiable Phylogenetic Tree Inference

Understanding evolutionary history can reveal information relevant to understanding modern biological systems and the origins of protein functions. Ancestral Sequence Reconstruction (ASR) infers historical protein sequences from extant data, but its accuracy depends on both reconstructed ancestral sequences and the phylogenetic tree topology relating them. Conventional phylogenetic methods evaluate or sample discrete tree topologies using maximum-likelihood or Bayesian criteria, but the number of possible topologies grows rapidly as more sequences are added. A previously developed differentiable framework instead allows ancestral sequences and tree topology to be optimized jointly, while also creating a path for incorporating richer sequence-based potentials such as ProteinMPNN. However, weak alignments can still leave the tree topology underdetermined. To address this limitation, we incorporated organism proteome-sketch distances as a soft, low-weight, label-invariant prior that matches patterns of leaf-to-leaf tree distances throughout optimization rather than imposing a fixed guide tree. Because gene trees can genuinely disagree with organism trees, the prior must help when alignment evidence is weak, yield when alignment evidence is strong, and avoid harming reconstruction when the two histories differ. We hypothesize that an appropriate prior weight or calibration strategy can satisfy these constraints. We evaluate this hypothesis using synthetic datasets that vary in alignment strength and gene-organism agreement, together with hyperparameter sweeps, recovery thresholds, and null controls. We then examine curated DHFR sequences as a potential improvement case and PDZ domains as a no-harm case. This work seeks to determine when organism-level evolutionary information can improve differentiable phylogenetic inference and when sequence alignment should remain the dominant source of evidence.

Peter Le

University of California, Irvine

REU: Ljubetič Lab, Kemijski Inštitut

De Novo Design of a Quadrupedal Protein Walker

Naturally occurring motor proteins, such as kinesin and dynein, provide blueprints for the artificial engineering of cargo transport systems. Inspired by these biomolecular machines, a powered de novo protein walker has yet to be designed, but a random protein walker has recently been reported. The random walker uses flexible coiled-coil “feet” to bind and unbind to flexible footholds decorated at regular intervals on a fiber track. A powered walker, however, will require rigid feet to exhibit the asymmetry needed for directional motion. In this regard, a rigid variant of the random walker is being investigated. Because the morphological design space of these walkers is largely unexplored, there is still much room for creative experimentation. We propose a quadrupedal protein walker design, which reuses the coiled-coil interfaces from the flexible walker and a previously designed rigid track. The quadrupedal walker would only move two feet at once in repeatable “stepping” motions. The front pair of feet would step in unison, and likewise, the back pair. As this design is purely for proof of concept, it is intended as a foundation upon which further computational exploration may be conducted. The quadrupedal walker is put forth in hopes of pushing state-of-the-art protein walkers closer to prospective applications in drug delivery.

Celeste Litchfield

Brigham Young University

REU: IPD, University of Washington

Single-component Octahedral Nanoparticle Design for an Epstein-Barr Virus Vaccine

The Epstein-Barr virus (EBV) is ubiquitous in the population, causes both acute and chronic illnesses, and has no FDA approved treatments or vaccines. We aim to develop a vaccine candidate for EBV by scaffolding the gB glycoprotein trimer domain IV into a single component nanoparticle using computational protein design. Scaffolding the antigen into an octahedral nanoparticle optimizes exposed surface area for antibody binding and immunogenicity. We used diffusion to generate de novo trimeric nanoparticle backbones, docked them computationally, then used ProteinMPNN to design the sequences before assessing the sequences using AlphaFold 2. The designs were experimentally analyzed for complex formation using size exclusion chromatography and dynamic light scattering. The designs that matched the expected complex size in vitro were structurally validated by electron microscopy without the antigen. These successful design candidates will be tested for secretability in mammalian cell culture and for immunogenicity in murine models using both protein and mRNA administration methods.

Hannah Lott

University of Tulsa

REU: Yarov-Yarovoy Lab, University of California, Davis

Designing De Novo Mini-binders for Nav1.7 to Develop Chronic Pain Therapeutics

Chronic pain, defined as pain lasting for over three months, affects over 50 million adult Americans. Despite its prevalence, effective treatments without significant side effects remain elusive. Voltage-gated sodium channel 1.7 (Nav1.7) is known to be involved in nociception, making it an attractive target for chronic pain therapeutics. Based on the binding location of natural peptide toxins such as protoxin-II and huwentoxin-IV, mini-binders for voltage-sensing domain II (VSDII) could inhibit Nav1.7, reducing nociception signaling. Latent-X and RFdiffusion were used to generate over twelve thousand potential mini-binders for VSDII, the vast majority of which were between 50 and 70 residues in length. These potential mini-binders were cross-validated with Boltz-2 and extensively filtered by confidence metrics, stability, and developability. This pipeline demonstrated in silico proof of concept for the de novo design of Nav1.7 VSDII mini-binders using Latent-X. Surviving mini-binders from this pipeline are high-quality candidates for testing Nav1.7 inhibition and further therapeutic development.

Andy Luu

Swarthmore College

REU: Cooper Lab, Northeastern University

Threshold User Studies

Prior work on the Foldit lattice puzzle has explored how milestone-based reward structures, particularly three-star thresholds, influence player engagement and persistence. In these studies, participants complete a sequence of puzzles while earning stars upon reaching predefined performance thresholds. Researchers manipulate the placement of these thresholds (e.g., evenly spaced, front-loaded, or back-loaded) to examine how different reward schedules affect motivation, time spent on tasks, and progression through the puzzle sequence. The three-star system serves as an extrinsic feedback mechanism, signaling incremental achievement while encouraging continued optimization beyond mere task completion. By comparing different threshold configurations, user studies investigate whether specific reward distributions better sustain engagement, promote flow, or increase persistence in solving increasingly difficult lattice-folding challenges.

Martina Madrzak

Florida State University

REU: Strauch Lab, Washington University St Louis

Computational Design of a Universal H3 Influenza Binder Targeting the Conserved Anchor Region

Seasonal influenza vaccines require annual reformulation due to continual antigenic drift of the hemagglutinin (HA) protein which enables evasion of host immunity, limiting broad protection against emerging H3 variants and imposing significant financial burden. Although HA evolves

rapidly, the anchor region within the HA stem remains highly conserved, making it an attractive target for universal therapeutic design. This project aims to computationally design a de novo protein binder that stabilizes HA in its prefusion conformation, preventing the structural rearrangement required for viral membrane fusion and host cell entry. To identify a conserved binding interface, approximately 3,000 H3 HA sequences spanning multiple decades were analyzed using multiple sequence alignment, and residues with >89% consensus were selected as design hotspots. A high-resolution HA structure (PDB: 9D1U, 91.0% sequence identity to the consensus sequence) served as the design template. RFdiffusion generated approximately 50,000 candidate binders (55–64 residues) by systematically exploring combinations of conserved hotspots, which were subsequently filtered for a three-helix architecture. Candidate designs are being optimized with ProteinMPNN and structurally evaluated using AlphaFold2. Top candidates will be prioritized using Rosetta Relax based on three criteria: Rosetta interface binding energy ($dG \leq -30$), shape complementarity ($sc > 0.65$), and buried surface area ($SASA > 900 \text{ \AA}^2$), ensuring favorable binding thermodynamics and structural complementarity. This work presents a computational pipeline for designing broadly reactive H3 influenza binders targeting a conserved epitope. The resulting candidate binders will provide a foundation for experimental validation and may contribute to the development of universal influenza therapeutics and generalizable protein design strategies targeting conserved regions of rapidly evolving viral proteins.

Daniel Martinez

University of Maryland, Baltimore County

REU: Khmelinskaia Lab, Ludwig-Maximilians-Universität in Munich

Twist-Angle-Directed Design of Stimulus Responsive, Conformationally Switchable Protein Oligomers

Responsive protein assemblies are essential for cellular function, yet the de novo design of assemblies that controllably switch conformation in response to exogenous stimuli remains largely unexplored. Here I prototype an RFDiff3 protocol for extending our recently developed interface-seeded design strategy for functional oligomer design to include a functional core motif in addition to a protein-protein interface (PPI) seed. More specifically, I combine an experimentally validated, de novo metal-responsive twist motif with a known de novo PPI motif to construct protomers capable of adopting distinct conformations and thus control spatial arrangement of protein surfaces and protein oligomerization. I explore this combination to design protomers with two defined conformational states that promote oligomer disassembly by strain accumulation and, when the number of subunits is compatible with ring re-closure, enable the switch between a cyclicly symmetric ring and alternative topologies. Our multi-motif scaffolding approach establishes that twist motif's rotation angle can be used as a direct design parameter for simultaneously controlling oligomer assembly and topology, opening new possibilities for the design of biosynthetic systems with tunable, stimulus-gated regulatory mechanisms.

Noah Mitton*Whitworth University**REU: IPD, University of Washington*

Computational design of de novo protein binders that recognize and bind small molecule targets

Small molecules lack the library of natural binding scaffolds available to larger biomolecular targets, limiting the design of high-affinity binders for biosensing and therapeutic applications. Deep learning pipelines that combine RFDiffusion3, LigandMPNN, and a structure-prediction model now generate protein binders for small-molecule targets. Whether this pipeline generalizes across chemically distinct structures and different structure-prediction models remains untested. Here we propose to design de novo binders for three targets, two structurally similar steroids, cortisol and progesterone, and one distinct phenolic compound, BPA, using RFD3, LigandMPNN, and one of three structure-prediction models: AF3, ProteinX, or ESMFold. First RFD3 generates raw pocket backbones from precomputed center-of-mass tokens, then filters them for structural viability and steric clashes. LigandMPNN designs sequences for each backbone, and the structure-prediction model isolates the top 60% of variants by relative PAE score. This design-and-filter cycle repeats several times. The best candidates then undergo sequence expansion and a final validation fold, where they must clear strict thresholds for local confidence, target-specific hydrogen bonding, and correct ligand stereochemistry. Successful binders for these three targets will establish design rules transferable across the steroid and phenolic chemical space, supporting future biosensor and diagnostic tools for endocrine-disrupting compounds.

Marwa Naji*Johns Hopkins University**REU: Merck South San Francisco*

Property-aware de novo cyclic peptide design

Computational methods have expanded the ability to design de novo cyclic peptide sequences with favorable structural characteristics, yet many generated candidates lack the physiochemical properties required for therapeutic development. Specifically, existing approaches often focus on target binding, while overlooking the ADME (Absorption, Distribution, Metabolism, Excretion) properties that directly impact exposure, distribution and efficacy in vivo. We present a property-aware cyclic peptide design framework that integrates machine-learning based property prediction directly into sequence optimization. Using permeability as an example task, machine learning models for permeability prediction were trained and benchmarked using public data, and the highest-performing model was incorporated into the design framework to jointly optimize structural and permeability objectives. Generated cyclic peptide sequences were subsequently evaluated using internally developed ADME property prediction models and other scoring functions to compare property-aware and conventional design strategies. This workflow is expected to produce candidates with improved ADME properties while maintaining valid structural characteristics.

Tony Peonio*Colorado School of Mines**REU: Hosseinzadeh Lab, University of Oregon***Computational design of FPR2-specific cyclic peptides for resolution of chronic inflammation**

Current pharmacological strategies for managing chronic inflammation include specialized pro-resolving lipid mediators (SPMs), nonsteroidal anti-inflammatory drugs (NSAIDs), synthetic small molecules, and linear peptides. However, each of these has limitations: SPMs possess extremely short half-lives, NSAIDs temporarily reduce inflammation but fail to promote tissue healing, synthetic molecules often have toxic side effects, and linear peptides are highly susceptible to peptidase degradation. Consequently, there is an unfulfilled need for stable, long-lasting therapeutics that mimic endogenous SPMs. To address this, we developed multiple cyclic peptide computational pipelines, creating both head-to-tail cycles and thioether lariats to target formyl peptide receptor 2 (FPR2), a central regulator of inflammation resolution. In addition, we created a novel non-canonical amino acid based on an existing synthetic FPR2 agonist, which was then incorporated into a subset of our designs. Specifically, we focused on both de novo and motif-scaffolded designs, generating over 130,000 initial candidates with GenKIC and CyclicCAE before optimizing the top 12,000 using FastDesign. After filtering, over 1,000 peptides were docked into both FPR1 and FPR2 to model peptide-receptor complexes using either HighFold2 or FlexPepDock. We then evaluated the top candidates based on predicted binding affinities and P_{Near} calculations, with molecular dynamics simulations currently underway. Ultimately, we identified nine highly promising computational candidates and initiated in vitro validation on two lead binders, with plans to synthesize and evaluate the remaining candidates. Together, these pipelines represent a promising computational strategy for designing a new class of durable, pro-resolving therapeutics.

Alan Talavera*Saddleback College**REU: IPD, University of Washington***Transcriptional Activation-Based Readout for Cytosolic Delivery**

Recent advancements in cytosolic capture of extracellular degradation targets via protein EndoTags and pH-sensitive proteins have enabled new therapeutic modalities. This work seeks to extend these techniques to transport peptides and other therapeutic drug cargo into the cell via endocytosis, release them before lysosomal degradation, and accurately measure the degree of their escape. Current methodologies for measuring endosomal escape are resource intensive, difficult to scale, and computationally expensive. By combining specific dimerization domains, EndoTags, pH-sensitive proteins, and transcriptional activators, we created a new quantitative readout for endosomal escape events. Computational protein design employing PyRosetta, ProteinMPNN, and AlphaFold 3 was used to instill pH sensitivity in existing endosomal pore-forming designs in support of the endosomal escape goal. The establishment

of a faster and more scalable endosomal escape assay will accelerate the design-build-test cycle of peptide therapeutics, and therefore the discovery of highly specific and bioavailable drug delivery designs.

Sina Yarmohammadi

Stanford University

REU: Tandem AI

Enabling de novo heterochiral cyclic-peptide design at scale

Cyclic peptides are promising therapeutic molecules because their conformational rigidity and proteolytic stability can pair strong target binding with improved structural stability. Their computational design, however, requires both diverse structural training data and a reliable workflow that connects multiple prediction and design tools. This summer work addresses these needs through two complementary projects.

The first project expands the structural dataset available for CyclicCAE, a conformational autoencoder used to generate cyclic peptide backbones. Quantum-refined conformer libraries have now been generated for 10-, 11- and 12-residue macrocycles, roughly 2.5 , 3.5 and 7.5×10^5 conformers respectively, extending past the six- to nine-residue rings the released model was trained on. Training CyclicCAE on them is the next step.

The second project, PepGen-Auto, converts a previously manual and notebook-based peptide design workflow into an automated, restart-safe pipeline for cluster computing. Starting from a target protein sequence and a binding-site definition, twelve stages run unattended across four software environments: PepMLM binder generation, Boltz-2 structure prediction and quality control, Rosetta relaxation, anchor-motif identification, CyclicCAE scaffold generation around that anchor, heterochiral sequence design, multi-objective candidate selection, and optional PNear foldability validation. BoltzGen has been added as a parallel generative front end, scored on identical Rosetta metrics. Heavy computations are scheduled through Slurm, while a lightweight controller manages progression, throttling and recovery, and companion tools trace design provenance and computational yield.

Iris Zheng

Johns Hopkins University

REU: Merck Cambridge

Systematic Evaluation of Protein Folding Configurations and Confidence Metrics to Improve Experimental Binder Identification Success Rates

Advances in computational protein design have enabled the rapid generation of large libraries of protein binders. However, experimental validation rates are highly target dependent, ranging from $<1\%$ to $>50\%$. For challenging targets, large library screening is often still required to identify successful binders, resulting in long turnaround times for iterative design-make-test cycles. Continued advancements in deep learning-based structure prediction models (AlphaFold, Boltz, etc.) and the use of their confidence metrics for prioritizing designs are promising.

However, the subtle effects of folding-model configuration choices and confidence metric interpretation on binder identification remain incompletely understood in-part due to the lack of available data. In this work, we systematically investigate how protein binder-target complex folding settings influence the confidence metrics commonly used for computational binder prioritization and filtering. Using experimentally validated datasets, we retrospectively evaluated the impact of key modeling parameters on library selection. We further examine the predictive capabilities of multiple confidence metrics (pLDDT, ipTM, ipSAE) as well as alternative approaches for aggregating information across multiple model outputs. We report our findings of the key parameter choices that can impact the predictive ability of various confidence metrics as well as learning how filtering and prioritization criteria can be effectively tailored for specific design challenges and experimental screening constraints.

James Zhu

Pomona College

REU: Drew Lab, University of Illinois at Chicago

BindCraft guided de novo design of peptide binders targeting the bacterial quorum-sensing regulator Rgg3

Streptococcus species are major human pathogens responsible for diseases ranging from pharyngitis to life-threatening invasive infections. As antibiotic resistance continues to rise, anti-virulence strategies that suppress bacterial pathogenicity without inhibiting bacterial growth have emerged as attractive therapeutic alternatives. The quorum-sensing transcription factors Rgg2 and Rgg3 regulate virulence-associated gene expression through interactions with short hydrophobic peptide (SHP) pheromones, making them promising therapeutic targets [1]. In our previous work, we used BindCraft to design short helical peptide binders targeting the SHP-binding interface of Rgg3 and identified potent peptide inhibitors of Rgg3-mediated quorum sensing. The X-ray co-crystal structure of the highest-affinity peptide in complex with Rgg3 was subsequently determined [2], providing atomic-level insight into the structural factors which account for the high accuracy AI-guided peptide design. In the present study, we used this structure as the starting point for a second round of peptide binder design [3] to identify second-generation improved Rgg3 inhibitors. Thousands of independent design trajectories were generated and computationally evaluated using structural confidence metrics, interface quality, and sequence-based analyses to identify high-confidence candidates. We selected the top six peptide designs for experimental functional assays to evaluate their ability to inhibit Rgg3-mediated quorum sensing. Although one-shot computational design of protein-based inhibitors has become increasingly attainable, the design of short linear peptides remains challenging due to the difficulty in maintaining the stable folded conformations needed for potent activity. Our iterative computational-experimental workflow demonstrates the potential of AI-driven peptide design for developing highly specific anti-virulence peptide therapeutics targeting bacterial quorum-sensing regulators and provides a scalable strategy for engineering next-generation therapeutics.