Stowers Institute

Stowers.org

STOWERS RESEARCH CONFERENCES:

AI-Driven Approaches to Emergent Protein Behaviors

#SRCKC25

PROGRAM

Table of Contents

Speaker instructions

Speakers **must upload their presentations to the "Presenter" folder that was shared with them in advance of the meeting OR at the podium** at the designated times listed below, prior to their scheduled sessions.

Speaker file upload times:

Wednesday, October 22
1:00–2:00PM

Thursday, October 23
9:00–9:30 AM
1:00–1:30 PM

Friday, October 24
9:00–9:30 AM
12:30–1:00 PM

Speakers **must get equipped with a microphone at the podium** 15 minutes prior to their scheduled sessions.

Conference Program

All talks will be presented in the Stowers Auditorium
Food and beverages offered are complementary, please enjoy!

30 min Presentation + 10 min Q&A
15 min Presentation + 5 min Q&A

WEDNESDAY, OCTOBER 22

1:00 PM TO 2:00 PM	**CONFERENCE CHECK-IN & POSTER SET-UP** *Light snacks and refreshments are available outside of the Auditorium*

SESSION 1: GENERATIVE PROTEIN DESIGN & ENGINEERING

Session Chair: Arvind Pillai

2:15 PM	**OPENING REMARKS**
2:25 PM TO 2:45 PM	**ARVIND PILLAI, STOWERS INSTITUTE** *"Computational design of allosteric assemblies and enzymes"*
2:50 PM TO 3:10 PM	**CLAIRE LEBLANC, UNIVERSITY OF CALIFORNIA, BERKELEY** *"Interpretable biophysical neural networks of transcriptional activation domains separate roles of protein abundance and coactivator binding"*
3:15 PM TO 3:35 PM	**TIMO SCHNEIDER, ETH ZURICH** *"De novo design of peptides localizing at the interface of biomolecular condensates"*
3:35 TO 3:50 PM	**BREAK** – *refreshments available outside of the Auditorium*
3:55 PM TO 4:15 PM	**KATHERINE HATSTAT, UNIVERSITY OF CALIFORNIA, SAN FRANCISCO** *"Engineering bacterial signal transduction"*
4:20 PM TO 5:00 PM	**JOANNA SLUSKY AND SAMUEL LIM, UNIVERSITY OF KANSAS** *"Protein Design Insights from Large Datasets"*
5:30 PM TO 8:00 PM	**WELCOME DINNER at Char Bar** *Dinner is included in your registration fee.* *A shuttle is available to take you to the restaurant*

THURSDAY, OCTOBER 23

9:00 AM TO 9:30 AM	*Coffee and light breakfast options are available outside of the Auditorium*

SESSION 2: PROTEIN PLASTICITY

Session Chair: Jay Unruh

9:30 AM	**OPENING ANNOUNCEMENTS**
9:35 AM TO 10:15 AM	**JULIA ZEITLINGER AND MELANIE WEILERT, STOWERS INSTITUTE** *"Extracting mechanistic insights from AI sequence models – the intriguing case of low-affinity DNA binding motifs"*
10:20 AM TO 10:40 AM	**PETRAS KUNDROTAS, UNIVERSITY OF KANSAS** *"Conformational changes upon protein binding inferred from AlphaFold models"*
10:45 AM TO 11:05 AM	**RAJUN BURT, STOWERS INSTITUTE** *"A Quantitative 3D Proteome Reveals Tissue Dependent Structural Predictions & Interactions of Common Proteins"*
11:05 AM TO 11:20 AM	**BREAK** – *refreshments available outside of the Auditorium*
11:20 AM TO 11:40 AM	**MAJID BANI-YAGHOUB, UNIVERSITY OF MISSOURI KANSAS CITY** *"Identifying Time-Delayed Biological Feedback Loops Using Data-Driven Modeling: A New Approach with a Hes1 Case Study"*
11:45 AM TO 12:25 PM	**NEVILLE BETHEL AND ZINNIA MA, UNIVERSITY OF CALIFORNIA, SAN DIEGO** *"Assessing and Leveraging Structural Information Encoded in Protein Language Model Embeddings"*

12:30 PM TO 1:30 PM	**LUNCH BREAK** – *join us in the Stowers Library for lunch*

SESSION 3: DESIGNER ASSEMBLIES

Session Chair: Danielle Tullman-Ercek

1:40 PM TO 2:20 PM	**DANIELLE TULLMAN-ERCEK AND ANNIE GOMEZ, NORTHWESTERN UNIVERSITY** *"Towards designer protein assemblies using a combination of experiments, simulations, and machine learning"*
2:25 PM TO 2:45 PM	**SHIQI LIANG, NORTHWESTERN UNIVERSITY** *"Synthetic Rewiring of Virus-Like Particles via Circular Permutation Enables Modular Peptide Display and Protein Encapsulation"*
2:45 PM TO 3:00 PM	**BREAK** – *refreshments available outside of the Auditorium*
3:05 PM TO 3:45 PM	**CAROLYN MILLS AND NATASHA JONES, UNIVERSITY OF CALIFORNIA, SANTA BARBARA** *"Nature vs Nurture: Interrogating the intrinsic and extrinsic drivers of emergent protein behaviors using data-driven approaches"*
3:50 PM TO 4:30 PM	**ALLIE OBERMEYER AND SO YEON AHN, COLUMBIA UNIVERSITY** *"Self-Assembled Protein Particles for Cellular Delivery"*
4:45 PM TO 6:45 PM	**POSTER SESSION** proceed down the staircase to the Stowers Gallery Light snacks and refreshments are available at the base of the stairs ODD # POSTERS 4:45 PM TO 5:45PM EVEN # POSTERS 5:45 PM TO 6:45 PM
6:45 PM	**DINNER** – *join us in the Stowers Library for dinner*

FRIDAY, OCTOBER 24

9:00 AM TO 9:30 AM	*Coffee and light breakfast options are available outside of the Auditorium*

SESSION 4: AMYLOIDS

Session Chair: Randal Halfmann

9:30 AM	**OPENING ANNOUNCEMENTS**
9:35 AM TO 9:55 AM	**RANDAL HALFMANN, STOWERS INSTITUTE** *"Repurposing AI Tools to Address the Protein Misfolding Problem"*
10:00 AM TO 10:20 AM	**JAY UNRUH, STOWERS INSTITUTE** *"Engineering nucleation scaffolds to control polyQ amyloid assembly"*
10:25 AM TO 10:45 AM	**ALEX VON SCHULZE, STOWERS INSTITUTE** *"Integrating AI reasoning with experiments to prioritize mechanistic hypotheses"*
10:45 AM TO 11:05 AM	**BREAK** – *refreshments available outside of the Auditorium*
11:10 AM TO 11:30 AM	**YAN YAN, STOWERS INSTITUTE** *"Embracing Structural Mass Spectrometry for Protein Pleomorphism"*
11:35 AM TO 12:15 PM	**NIKOLAOS LOUROS AND MACY LOZEN, UNIVERSITY OF TEXAS SOUTHWESTERN** *"Machine-guided decoding and rebuilding of amyloids"*
12:15 PM TO 1:15 PM	**LUNCH BREAK** – *join us in the Stowers Library for lunch*

SESSION 5: BRIDGING AI AND EXPERIMENTAL WORKFLOWS

Session Chair: Jay Unruh

1:20 PM TO 2:00 PM	**STEFFEN LINDERT AND ELIJAH DAY, UNIVERSITY OF CALIFORNIA, LOS ANGELES** *"Learning from Sparse Data: Integrative Modeling in the Post-AlphaFold Era"*
2:05 PM TO 2:45 PM	**ANUM GLASGOW AND CHENLIN LU, COLUMBIA UNIVERSITY** *"A machine learning method to discover hidden protein conformational states"*

2:45 PM TO 3:00 PM	**BREAK** – *refreshments available outside of the Auditorium*
3:05 PM TO 3:25 PM	**YASHWANTH ASHOK, UNIVERSITY OF WASHINGTON** *"Dosa: A method to covalently barcode proteins for high throughput biochemistry"*
3:30 PM TO 4:10 PM	**POSSU HUANG AND BRAXTON BELL, STANFORD UNIVERSITY** *"Close the loop with macrocycles"*
4:10 PM TO 4:30 PM	**BREAK** – *refreshments available outside of the Auditorium*

SESSION 6: FUTURE DIRECTIONS & GRAND CHALLENGES

Session Chair: Randal Halfmann

4:35 PM TO 4:55 PM	**ALEX GARRUSS, STOWERS INSTITUTE** *"Using Protein Language Models for Mechanistic Insight"*
5:00 PM TO 5:20 PM	**CHARLES MCANANY, STOWERS INSTITUTE** *"PISA: a versatile interpretation tool for visualizing cis-regulatory rules in genomic data"*
5:25 PM TO 5:45 PM	**MOISES GUALAPURO, UNIVERSITY OF KANSAS** *"An AI-based framework for functional inference in uncharacterized Chlamydia trachomatis proteins"*
5:50 PM TO 6:30 PM	**ILYA VAKSER AND AMAR SINGH, UNIVERSITY OF KANSAS** *"AI-based atomic resolution modeling of cell-size systems at extra-long timescales"*
6:30 PM	**CLOSING REMARKS**
6:45 PM TO 8:00 PM	**CLOSING RECEPTION** – *join us in the Stowers Library for a reception. Small bites, desserts, and refreshments available.*

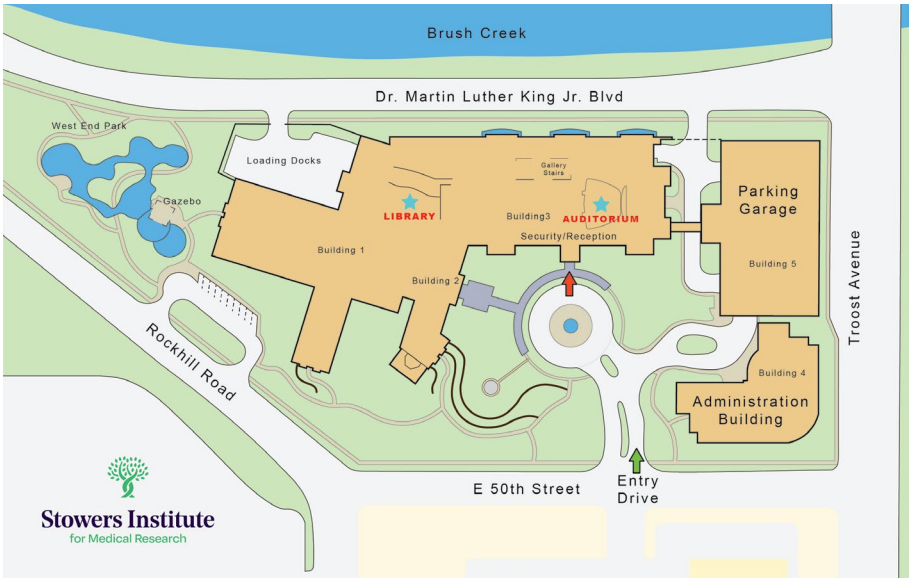
Poster Sessions

Poster Session

Thursday, October 23
Odd numbered posters: 4:45 PM – 5:45 PM
Even numbered posters: 5:45 PM – 6:45 PM

Author	Title	Poster #
Yan Yan Stowers Institute for Medical Research	Embracing Structural Mass Spectrometry for Protein Pleomorphism	1
Pushpa Kumar Balan University of Central Missouri	From Genomic Signals to Protein Behaviors: AI-Driven Feature Selection for Bioinformatics Modeling	2
Madeleine Clore National Institutes of Health		3
Chenlin Lu Columbia University	PFNet: a method to determine residue-level ensemble energies from HX/MS data	4
Yashwanth Ashok University of Washington	Dosa: A method to covalently barcode proteins for high throughput biochemistry	5
Rajan Burt Stowers Institute for Medical Research	A Quantitative 3D Proteome Reveals Tissue Dependent Structural Predictions & Interactions of Common Proteins	6
Timo Schneider ETH Zurich	De novo design of peptides localizing at the interface of biomolecular condensates	7
Alex Von Schulze Stowers Institute for Medical Research	Integrating AI reasoning with experiments to prioritize mechanistic hypotheses	8
Yusra Rahman University of Kansas Medical Center	Rheostat positions: A Missed Opportunity for Improving Computational Predictions	9
Sabrina Chen, Jessie Huh, Claire Sklar Johns Hopkins University	Predicting Supramolecular Self Assembly of Peptide Structures in AlphaFold3	10
Rajnish Kumar University of Kansas Medical Center	CNVs delineating in early development stages of human placenta	11
Tayla Miller Stowers Institute for Medical Research	Predicting the nucleating structure of amyloids through artificial evolution	12
Anastasiia Sivchenko University of Kansas Medical Center	A position-based substitution sensitivity framework reveals cold and hot spots for epistasis	13

Campus Map



All t**alks** will be presented in the Auditorium.
**Food and beverages are not allowed inside of the auditorium.*

Breaks will be in the hallway outside of the auditorium.

Lunch will be in the Stowers library.

Poster Session will be in the Gallery, down the staircase outside of the auditorium.

Conference Room 114 is reserved as a **workspace**.
** Food and beverages are allowed inside Conference Room 114.*

Shuttle Schedule

A shuttle will run between the Kansas City Marriott – Country Club Plaza Hotel (4445 Main St, Kansas City, MO 64111) and the Stowers Institute for Medical Research (1000 E. 50th St, Kansas City, MO 64110) at the times listed below:

Wednesday, October 22:

12:30 PM & 1:00 PM	Pick up at Marriott Hotel, drop off at Stowers
5:00 PM	Pick up at Stowers, drop off at Char Bar for welcome party
8:00PM	Pick up at Char Bar, drop off at Marriott Hotel and Stowers

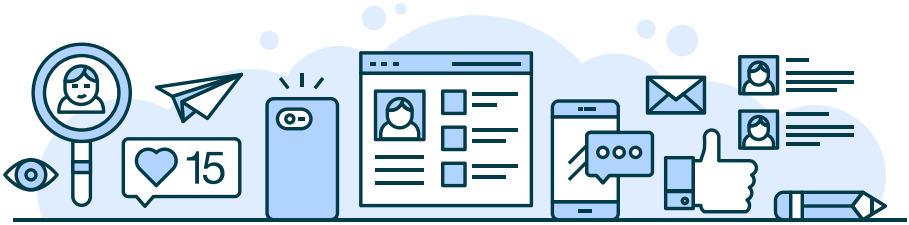
Thursday, October 23:

8:30 AM & 8:45 AM	Pick up at Marriott Hotel, drop off at Stowers
7:45 PM & 8:00 PM	Pick up at Stowers, drop off at Marriott Hotel

Friday, October 24:

8:30 AM & 8:45 AM	Pickup at Marriott Hotel, drop off at Stowers
7:45 pm & 8:00 PM	Pick up at Stowers, drop off at Marriott Hotel

SRC Social Media Guidelines



The Stowers Research Conference organizers **encourage the use of social media to share information and network with other attendees.**

We remind you to **remain courteous and respectful in your comments and posts.**

Avoid sharing speaker or poster presentation content that's not your own, copyrighted or trademarked, or material protected by other intellectual property rights.

Follow and tag **@stowersinstitute** on Instagram

Follow and tag **Stowers Institute for Medical Research** on LinkedIn

Follow and tag **@ScienceStowers** and **@Stowers_SRC** on X (Twitter)

Follow and tag **@stowersinstitute.bsky.social** on Bluesky

Use hashtag **#SRCKC25**

Invited Speakers and Trainees

JOANNA SLUSKY AND SAMUEL LIM, UNIVERSITY OF KANSAS
Lab Website

LAUREN PORTER AND MADELEINE CLORE, NATIONAL INSTITUTES OF HEALTH
Lab Website

NEVILLE BETHEL AND ZINNIA MA, UNIVERSITY OF CALIFORNIA, SAN DIEGO
Lab Website

DANIELLE TULLMAN-ERCEK AND ANNIE GOMEZ, NORTHWESTERN UNIVERSITY
Lab Website

CAROLYN MILLS AND NATASHA JONES, UNIVERSITY OF CALIFORNIA, SANTA BARBARA
Lab Website

ALLIE OBERMEYER AND SO YEON AHN, COLUMBIA UNIVERSITY
Lab Website

NIKOLAOS LOUROS AND MACY LOZEN, UNIVERSITY OF TEXAS SOUTHWESTERN
Lab Website

STEFFEN LINDERT AND ELIJAH DAY, UNIVERSITY OF CALIFORNIA, LOS ANGELES
Lab Website

ANUM GLASGOW AND CHENLIN LU, COLUMBIA UNIVERSITY
Lab Website

POSSU HUANG AND BRAXTON BELL, STANFORD UNIVERSITY
Lab Website

ILYA VAKSER AND AMAR SINGH, UNIVERSITY OF KANSAS
Lab Website

Dosa: A method to covalently barcode proteins for high throughput biochemistry

Yashwanth Ashok1, Christine Queitsch1, Stanley Fields1

1Department of Genome Sciences, University of Washington, Seattle, WA

Protein display technologies can assess in high throughput the effects of single amino acid substitutions via coupling of an activity to DNA sequencing. Here we show that protein variants can be covalently linked in vivo to RNA barcodes for analysis by in vitro biochemical experiments. Using fusion proteins composed of the E. coli tRNA (m5U54) methyltransferase TrmA (E358Q) and a protein of interest, we show that proteins can be crosslinked in vivo to a barcode-containing stem-loop derived from tRNA. We use this method, called Dosa, to show three different applications: to analyze a pool of 160 FLAG epitope variants for their binding to an anti- FLAG antibody, to profile substrate specificity of two proteases for their cleavage preference and to measure solubility of 836 variants of Aß(1-42). This method, applicable to both E. coli and mammalian cell expression, should be amenable to numerous biochemical activity assays.

From Genomic Signals to Protein Behaviors: AI-Driven Feature Selection for Bioinformatics Modeling

Pushpa Kumar Balan1, Aijing Feng1

1Department of Computer Science and Cybersecurity, University of Central Missouri

Bioinformatics, enabling models to extract the most informative signals from high-dimensional genomic and multi-omics datasets. By reducing noise and redundancy, feature selection enhances generalization and prevents overfitting, allowing AI systems to better capture the underlying biological principles of protein function and interaction. Recent advances in filter, wrapper, embedded, and hybrid feature selection methods have extended applications beyond classical gene prioritization to include cancer biomarker discovery, single- cell analysis, quantitative trait mapping, and integrative multi-omics studies.

In this poster, we review emerging approaches in feature selection with a particular focus on their role in uncovering determinants of complex protein behaviors. We further discuss how feature selection can be integrated with AI to identify disease-associated features and guide mechanistic insights. Finally, we outline future research directions where feature selection can inform the design and interpretation of high- throughput experiments, advancing our ability to predict, model, and control emergent protein behaviors and disease processes.

Identifying Time-Delayed Biological Feedback Loops Using Data-Driven Modeling: A New Approach with a Hes1 Case Study

Majid1,2

1Bani-Yaghoub, 2University of Missouri-Kansas City Missouri

Many biological processes such as gene regulation, cell signaling, and developmental patterning are driven by feedback mechanisms that include natural delays. These delays, often caused by transcription, translation, or transport, are critical for generating dynamic behaviors like oscillations and rhythms. However, identifying such delayed feedback from experimental data is difficult using traditional modeling techniques.

In this work, we introduce an innovative, data-driven method to detect time-delayed biological feedback loops from time-series data. The approach reformulates systems with time delays into equivalent non-delayed representations, enabling the use of sparse regression techniques to identify system dynamics, after which the original delay-based interactions are systematically reconstructed. We demonstrate the method using the well- known Hes1 gene regulatory network, where delayed negative feedback drives oscillations in gene expression. Our method reliably reconstructs the underlying feedback structure and is broadly applicable to other biological systems where time delays are known to play a role but remain challenging to model or parameterize.

The proposed approach opens new opportunities for studying complex cellular behaviors in areas such as developmental biology, circadian rhythms, and synthetic biology.

A Quantitative 3D Proteome Reveals Tissue Dependent Structural Predictions & Interactions of Common Proteins

Rajan Burt1,2 , Joe Varberg1 , Jay Unruh1 , Ying Zhang1 , Josefina Doffo1, Jonathan Russell1, Boris Rubinstein1, Jacob Yonke1, Laurence Florens1, Kausik Si1,2

1Stowers Institute for Medical Research, 2University of Kansas Medical Center

The functional diversity of biological systems cannot be fully explained by the number of genes or gene expression differences. Most of our analyses of tissue or organismal functional differences have thus far considered the amount and pattern of transcript expression. However, recent advances in proteomics provided an opportunity to ask how protein interactions and structure are associated with functional diversity. Here, we apply crosslinking mass spectrometry coupled with isobaric tag technology and deep learning protein structural prediction strategies to look at three domains of proteome variation: expression, interaction, and structure. Specifically, we integrate quantitative crosslinking both implicitly and explicitly: the former through ColabFold MSA subsampling and the latter through incorporation of crosslinks as covalently bound ligands in AlphaFold3. We measure these signatures across multiple murine tissues. Expectedly, we find unique protein expression patterns across tissues. Surprisingly, we also find unique interactions and predicted structures of commonly expressed proteins. We posit that in addition to tissue-specific expression, tissue-specific structure and protein- protein interactions allow commonly expressed proteins to serve tissue specific function.

Experimental characterization of the sequence space between two folds

Patrick S. Cantwell1,2, Ricardo Muñiz Trejo2, Joseph W. Thornton2

1University of Chicago, Biochemistry & Molecular Biophysics Program, 2University of Chicago, Department of Ecology & Evolution

There are thousands of unique protein folds in nature, but how new folds have evolved is unclear. Can new folds evolve from old folds by a stepwise process of amino acid substitutions? Historical evidence for this scenario is scarce, in part because sequence similarity would be lost quickly as folds diversify under distinct selective pressures. Previous protein engineering work by others produced two similar sequences that encode two different folds: G77 (3α fold) and G77 (α /ß-plait fold), which differ at 13 sites. A single stepwise mutational path connecting these two sequences was previously found, but whether this bridge is the only connection between the two folds is unknown. The distribution of folds across the binary sequence space between G77 and G77 determines the potential for one fold to evolve into the other, and this distribution in turn depends on how idiosyncratic are the rules by which amino acid sequence determines the G and Gfolds. To address this question, we are experimentally characterizing the free energy of folding into both folds of all 2 proteins using a novel high-throughput approach. We developed a yeast-display assay to quantify binding to the protein library using a set of designed minibinders that bind differentially to each fold. These measurements are then decomposed into the effects of each amino acid and combinations at each order on the free energy of folding, given the thermodynamic linkage between protein folding and binding. Analysis of simulated data shows that given a sufficiently diverse set of minibinders and coverage of the sequence space, folding energies and their sequence determinants can be inferred with high accuracy. When complete, this work will allow us to evaluate the mutational connectivity between the two folds, as well as the genetic and structural determinants that create or limit this connectivity.

Predicting Supramolecular Self Assembly of Peptide Structures in AlphaFold3

Sabrina Chen1, Jessie Huh1, Claire Sklar1

1Johns Hopkins University

Self assembled, peptide-based nanostructures have a wide range of applications in the pharmaceutical industry. A notable challenge in this field is the prediction of self-assembly behavior. This project assesses the utility of AlphaFold3 (AF3), a deep learning model for protein structure prediction, in predicting the interactions of peptide-based nanostructures, specifically in how the hydrophobic effect and other peptide-peptide interactions drive supramolecular self assembly.

We designed amphiphilic peptides composed of alternating hydrophobic (valine, leucine, isoleucine, phenylalanine) and hydrophilic (glutamic acid, aspartic acid, arginine) residues, with variations in sequence length, residue order, and glycine tails to mimic alkyl chain modifications. Utilizing AF3's multimer mode, we modeled assemblies at copy numbers ranging from 10 to 1000. AF3 generated a variety of morphologies, including micelles, vesicles, and nanotubes. We identified hydrophobic and hydrophilic regions and performed calculations of N-C terminus distances and packing scores using PyRosetta. Sequence order and residue chemistry strongly influenced produced outcomes: hydrophobic cores formed with solvent-exposed hydrophilic surfaces, while longer hydrophilic blocks favored ordered aggregates and glycine tails introduced curvature similar to vesicle-like assemblies. However, AF3 frequently produced structures with steric clashes, disordered packing, and aggregates inconsistent with experimentally observed morphologies, highlighting limitations when applied to solvent-dependent self-assembly.

Our findings demonstrate the emerging potential of AlphaFold as a tool for hypothesis generation in peptide material design, offering a pathway toward integrating sequence-level design with predictive structure-based assembly modeling. These methods could reduce reliance on costly and time-consuming experiments, improving the efficiency of developing functional peptide-based products.

An AI-based framework for functional inference in uncharacterized Chlamydia trachomatis proteins

Moises Gualapuro1, Kevin Hybiske2, Scott Hefty3

1Department of Computational Biology, University of Kansas, 2Department of Medicine, University of Washington, 3Department of Molecular Biology, University of Kansas

The obligate intracellular lifestyle of Chlamydia trachomatis complicates functional annotation of its genome, leaving approximately one-third of its proteome as hypothetical proteins (HPs). This incomplete picture limits the understanding of its pathogenesis, virulence, biology, and therapeutic development. To address this, we analyzed 209 HPs using AI-based algorithms. AlphaFold3 generated high-confidence 3D structural models, of which 37 (17.7%) showed very high global and local accuracy and 110 (52.6%) displayed reliable global folds and consistent local features. DeepFRI then inferred functions for these 147 high-quality models, assigning Gene Ontology (GO) terms with greater than 50% confidence. Biological processes, molecular functions, and cellular locations were determined for 112 of the HPs. Intrinsically disordered proteins were evaluated using the LoRA-DR algorithm, and annotations were assigned using MobiDB. Additional structural and functional insights were integrated from CATH, TMHMM, and FoldSeek, enhancing annotation robustness. Final annotations were mapped to the KEGG BRITE pathway vocabulary for improved clarity in Chlamydia research. This approach reduced the proportion of hypothetical proteins from ~33% to ~10%, revealing potential roles in electron transport, amino acid and nucleic acid metabolism, and energy production. These findings provide a structurally validated resource for future protein-protein interaction studies and the development of therapeutics against this important human pathogen

Engineering bacterial signal transduction

A. Katherine Hatstat1, William F. DeGrado1

1Department of Pharmaceutical Chemistry, University of California, San Francisco

Nature has evolved dynamic proteins that act as sensors, detecting changes in the environment and producing signals that allow cells to adapt and respond accordingly. For example, bacterial histidine kinases (HKs) detect environmental stimuli and transduce a signal across the membrane through multiple modular subdomains, initiating a phosphorelay and transcriptional response. Despite their ubiquity, HKs remain structurally elusive, HK specificity is still largely unknown, and our mechanistic understanding of HK signal transduction is still evolving. My research combines protein design and protein engineering to test and expand our fundamental knowledge of signal transduction from first principles. Here, I explore the modularity of HKs by generating large chimera libraries and correlating biophysical features of subdomains with receptor function. I also test the ability of emerging generative protein design tools to build functional HK subdomains de novo. Together, the combination of protein design and engineering expands our understanding of natural signaling mechanisms and enables us to build new-to-nature signaling systems.

Functional amyloid discovery with AlphaFold2

Jacob Jensen1, Wenhao Song1, Randal Halfmann1,2

1Stowers Institute for Medical Research, 2Department of Biochemistry and Molecular Biology, University of Kansas Medical Center

Many proteins can form semicrystalline aggregates, known as amyloids, that cause disease. A growing number of proteins are known to form amyloids as part of their normal physiological functions in, for example, programmed cell death, biofilm formation, and memory storage. How many of such amyloids exist and the breadth of their biological functions is not yet known. Answering this question could be accelerated by accurate prediction of amyloid structure from sequence. We have found that AlphaFold2 can accurately predict the structures of certain functional amyloids. We therefore developed a workflow, AmyloidSearching, to computationally screen proteomes for candidate functional amyloids. AmyloidSearching first models fragments of intrinsically disordered regions as multimers, and then uses a set of scoring metrics to evaluate whether multimer models resemble amyloids. We used this method to generate lists of candidate functional amyloids in diverse proteomes. We found that the top-scoring candidate in the human proteome has a conserved function and predicted amyloid structure across vertebrates. Testing it experimentally, we verified that it forms amyloid when exogenously expressed in yeast cells, and that a point mutation rationally designed to disrupt the predicted amyloid structure indeed prevents amyloid formation. Finally, we used gene editing to make the corresponding point mutation in mouse, and observed a loss-of-function phenotype consistent with the proteins having a native amyloid structure. Other candidates are being tested.

CNVs delineating in early development stages of human placenta

Rajnish Kumar, Purbasa Dasgupta, Soma Ray, Namrata Roy, Asef Jawad Niloy, Mounika Vallakati, Aniket Patankar & Soumen Paul1

1Paul Lab, Department of Pathology, University of Kansas medical center

Placenta ensures successful pregnancy. Progression of placental development is regulated and mediated mainly by trophoblast cells, a transient cell population exists only during pregnancy. Studies indicating that trophoblast cells, specially the invasive extravillous trophoblast cells (EVTs) boast similarity with tumor cells. Genomics copy number variation is is considered to be a source of genetic diversity and common features in human genome. Interestingly, EVT development is associated with polyploidy as well as and senescent features during early human gestation and has been implicated as a normal phenomena in case of adaptation of a developing human placenta. Interestingly, CNVs have also been implicated in diseases associated with multiple pregnancy complications. In this regard, CNVs profiling of trophoblast cells in early stages of human placenta will be interesting to gain insights of progression dynamics of CNVs with time points in these stages of placental development. In this study, we have performed CNV profiling using all cell types of placenta, trophoblast cells of early stages human placenta. The study outcome inline with the results of previous studies. Along with it our study indicate the presence of more tumor like cells of trophoblast in early states than latter stages of development of human placenta.

Conformational changes upon protein binding inferred from AlphaFold models

Petras Kundrotas1, Matthew Coppeland1, Ilya Vakser1

1The University of Kansas

With the recent arrival of Artificial Intelligence (AI)- based techniques (especially, AlphaFold), the field of structural biology underwent remarkable advancements in modeling of protein tertiary structures. The AlphaFold deep learning approach revolutionized protein structure prediction by achieving near-experimental accuracy on vast number of monomeric proteins. That resulted, among many things, in the AlphFoldDB database that contains modeled structures of nearly 200 million proteins. However, proteins in living organisms often function by interacting with other proteins. In the crowded cell environment, a protein finds intended functional partner by adjusting its structure to maximize complementarity with the partner's features (shape, charge, hydrophobic patches etc.). Thus, straightforward utilization of monomeric models directly from the AlphaFoldDB very often could not help to elucidate mechanism of protein binding/functioning. Besides, many models in the AlphaFoldDB contain large flexible regions which potentially can become ordered upon protein binding. In this study, we analyze models of proteins generated as monomers and within the homo-dimeric assembly based on our AI-based large-scale modeling efforts on experimentally verified binary interactions from the INTACT and BIOGRID databases. By comparing PAE (Predicted Aligned Error) matrix and pLDDT (predicted Local Distance Difference Test) distributions for the monomeric and homo-dimeric models, we found several distinct types of conformational changes upon protein binding that AlphaFold is able to capture. Remarkably, no defined structural changes upon binding can be detected if more than ~60 % protein residues are in the flexible protein regions. Our efforts are collected and summarized in the GWYRE database (beta.gwyre.org) with user-friendly interface for searching and extracting information.

De novo emergence of a heme-binding pocket in hemopexin's evolution

Gino La Bella1,2 , Arvind Pillai1

1Stowers Institute for Medical Research, 2National University of San Martin

The evolutionary history of hemopexin (Hpx) provides a powerful framework for understanding how novel ligand-binding sites arise in proteins. Hpx is a plasma glycoprotein of ~60 kDa, that binds free heme with sub- picomolar affinity, protecting tissues from oxidative stress and maintaining iron homeostasis. A central hypothesis in our work is that Hpx has evolved from an ancestral metalloprotease, with an internal duplication generating two ß -propeller domains together with an interdomain linker, forming a de novo high-affinity heme- binding pocket.

We combine phylogenetics and ancestral sequence reconstruction with AI-driven protein design (AlphaFold3, ProteinMPNN, RFdiffusion) to evaluate structural stability and develop strategies for improving solubility in heterologous expression systems such as E. coli. By using a curated and taxonomically balanced set of vertebrate orthologs from OrthoDB, we inferred a robust phylogeny and reconstructed ancestral proteins that represent key intermediates in the emergence of the Hpx fold and function.

Our combined approach of protein evolution and design aims not only to retrace the path that originated the heme-binding pocket but also to repurpose these principles for engineering contemporary Hpx variants. Our goal is to provide both biophysical insight into the evolution of binding sites and a platform for therapeutic, biosensor, and industrial applications.

Interpretable biophysical neural networks of transcriptional activation domains separate roles of protein abundance and coactivator binding

Claire LeBlanc1,2, Pooja Agarwal1,2,3 , Jack Demaray1, Gean Hu1,3,4 , Marissa Zintel1,2, Angelica Lam1,2 , Joel Enrique Castro Hernandez1,3, Max Staller1,2,5

1Center for Computational Biology, University of California Berkeley, 2Department of Molecular and Cell Biology, University of California Berkeley, 3Department of Electrical Engineering & Computer Science, University of California Berkeley, 4Department of Bioengineering, University of California Berkeley, 5Chan Zuckerberg Biohub– San Francisco

Deep neural networks have improved the accuracy of many difficult prediction tasks in biology, but it remains challenging to interpret these networks and learn molecular mechanisms. Here, we address the interpretability challenges associated with predicting transcriptional activation domains from protein sequence. Activation domains, regions within transcription factors that drive gene expression, were traditionally difficult to predict due to their sequence diversity and poor conservation. Multiple deep neural networks can now accurately predict activation domains, but these predictors are difficult to interpret. With the goal of interpretability, we designed simple neural networks that incorporated biophysical models of activation domains. The simplicity of these neural networks allowed us to visualize their parameters and directly interpret what the networks learned. The biophysical neural networks revealed two new ways that arrangement (i.e. the sequence grammar) of activation domain controlled function: 1) hydrophobic residues both increase activation domain strength and decrease protein abundance, and 2) acidic residues control both activation domain strength and protein abundance. Notably, the biophysical neural networks helped us to recognize the same signatures in complex interpreters of the deeper neural networks. We demonstrate how combining biophysical and deep neural networks maximizes both prediction accuracy and interpretability to yield insights into biological mechanisms.

Synthetic Rewiring of Virus-Like Particles via Circular Permutation Enables Modular Peptide Display and Protein Encapsulation

Shiqi Liang[1], Kaavya Butaney[2], Daniel de Castro Assumpção[2], James Jung[3], Nolan W. Kennedy[2], Danielle Tullman-Ercek[2,4]

[1]Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, Illinois, USA, [2]Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois, USA, [3]Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, USA, [4]Center for Synthetic Biology, Northwestern University, Evanston, Illinois, USA

Virus-like particles (VLPs) are self-assembling nanoparticles derived from viruses with potential as scaffolds for myriad applications. They are also excellent testbeds for engineering protein superstructures. Engineers often employ techniques such as amino acid substitutions and insertions/deletions. Yet evolution also utilizes circular permutation, a powerful natural strategy that has not been fully explored in engineering self-assembling protein nanoparticles. Here, we demonstrate this technique using the MS2 VLP as a model self-assembling, proteinaceous nanoparticle. We constructed, for the first time, a comprehensive circular permutation library of the fused MS2 coat protein dimer construct. The strategy revealed new terminal locations, validated via cryo- electron microscopy, that enabled C-terminal peptide tagging and led to a protein encapsulation strategy via covalent bonding – a feature the native coat protein does not permit. Our systematic study demonstrates the power of circular permutation for engineering new features as well as quantitatively and systematically exploring VLP structural determinants.

Annotation of protein sequences using some AI based techniques

Sukha Malladi1, Eric Ross1, Sofia Robb1, Jay Unruh1

1Stowers Institute of Medical Research

Annotation of protein sequences has become a challenge due the explosive growth of protein sequences available from growing number of species. The role of artifical intelligence (AI) techniques to annotate proteomes of a set of phylogenetically diverse species and observations are reported here. Two AI based annotation methods are tested and compared with a regular non AI based annotation method. The advantages, disadvantages and shortcomings of both AI and non-AI based methods are reported. The annotation process is set up to run on Stowers' slurm linux cluster as a pipeline with minimal manual intervention. Some the results of the execution times of the different methods and species are presented. It is attempted to examine how relatively less frequently sequenced and studied species are annotated when they are not already well characterised. Structural motifs in different types of proteins from the species are examined for trends.

PISA: a versatile interpretation tool for visualizing cis-regulatory rules in genomic data

Charles McAnany1, Melanie Weilert1, Grishma Mehta2, Fahad Kamulegeya1, Jennifer M Gardner1, Anshul Kundaje3, Julia Zeitlinger 1,4

1Stowers Institute for Medical Research, 2Indian Institute of Science Education and Research, 3StanfordUniversity, 4University of Kansas Medical Center

Sequence-to-function neural networks learn cis-regulatory sequence rules driving many types of genomic data.

However, interpreting these models to relate the sequence rules to underlying biological processes remains challenging, especially for complex genomic readouts such as MNase-seq, which maps nucleosome occupancy but is confounded by experimental bias.

We introduce pairwise influence by sequence attribution (PISA), an interpretation tool that combinatorially decodes which bases contributed to the readout at a specific genomic coordinate.

PISA visualizes the effects of transcription factor motifs, detects undiscovered motifs with complex contribution patterns, and reveals experimental biases.

By learning the bias for MNase-seq, PISA enables unprecedented nucleosome prediction models, allowing the de novo discovery of nucleosome-positioning motifs and their effects, as well as the design of sequences with altered nucleosome configurations.

These results show that PISA is a versatile tool that expands our ability to extract novel cis-regulatory sequence rules from genomics data, paving the way towards deciphering the cis-regulatory code.

Predicting the nucleating structure of amyloids through artificial evolution

Tayla Miller1, Randal Halfmann1,2

1Stowers Institute for Medical Research, Kansas City, MO, 2Department of Biochemistry, University of Kansas Medical Center

Some amyloids drive signal propagation in the innate immune system, while others cause aggregation- associated diseases like Alzheimer's. The rate-limiting structure on path to amyloid formation, known as the nucleus, is the ideal target for blocking aggregation of disease-causing amyloids. However, the nucleus structure cannot be determined experimentally due to its short lifetime on pathway to a more stable amyloid structure. We hypothesized that AlphaFold2 can be used to elucidate the nucleus structure. AlphaFold2 predictions are guided in part by multiple sequence alignments (MSA) of naturally occurring homologous proteins, with the assumption that they form the same structure dictated by a conserved function. We reasoned that the MSA for an amyloid-forming protein could be replaced by a synthetic pool of mutant sequences artificially selected for accelerated aggregation, yielding an MSA that uniquely informs on the amyloid's nucleus structure. This alignment could then be used by AlphaFold2 to predict the amyloid nucleus. As proof of principle, we generated a random mutant pool of human TICAM1. TICAM1 is an innate immune signaling protein containing a conserved amyloid-forming domain for functional signal propagation. The mutant pool was sorted based on the extent of self-assembly in yeast cells with the use of Distributed Amphifluoric FRET (DAmFRET) as the functional selection. Next-generation sequencing was used to determine the mutation's fold change in nucleation relative to wild- type. Mutations with an increased nucleation rate were used as the MSA for AlphaFold2, which yielded a high- confidence model that differed from the structure of mature amyloid. We then used RFdiffusion to design globular monomeric proteins stabilizing the hypothetical nucleus structure. As expected, they eliminated the nucleation barrier to TICAM1 amyloid formation in cells. This artificial evolution-based approach will now be extended for robust determination of the rate-limiting structure to amyloid formation for any amyloid-forming protein in living cells.

Rheostat positions: A Missed Opportunity for Improving Computational Predictions

Yusra Rahman1, Shwetha Sreenivasan1, Liskin Swint-Kruse1

1Dept. Biochem & Mol Biol, KUMC, Kansas City, KS 66160

Single amino acid substitutions play a critical role in shaping protein function, with broad implications for evolution, genetics, and bioengineering. Traditionally, studies have focused on conserved "toggle" positions, where substitutions are highly detrimental. In contrast, our lab studies "rheostat" positions, where substitutions result in a continuum of functional outcomes. Although these positions offer valuable insights into protein behavior, they are rarely highlighted in high-throughput, whole protein, site-saturating mutational datasets. We conducted a meta-analysis of approximately twenty published datasets of human proteins to identify a subset of studies with good experimental resolution (capable of detecting intermediate outcomes that occur at rheostat positions). We show that a common analysis metric used is the average substitution score for each position, which does not reliably identify rheostat positions. Next, we evaluated the predictive performance of Alphamissense, a state-of-the-art pathogenicity model, against these datasets. While it performs moderately overall, Alphamissense shows significantly reduced accuracy at rheostat positions compared to toggle positions. This reveals a major blind spot in current prediction tools and highlights the need for models and experiments that can detect nuanced functional changes.

Dynamic time warping for sequence alignments

Nicholas Razo1, Alex Holehouse1

1Washington University in St. Louis

Over the past thirty years, alignment-based methods have played a crucial role in elucidating protein evolutionary relationships. While alignment-based analysis has been successful for studying the conservation of folded proteins, it has fallen short in assessing the conservation of intrinsically disordered regions (IDRs). IDRs lack a fixed spatial structure, which reduces the need for absolute sequence conservation. This allows them to have different lengths and compositions while still conserving local or global sequence features. While traditional alignment tools struggle to capture these conserved features, recent advances in deep learning and physics-inspired bioinformatics allow us to encode sequences into information-rich vectorial representations, which in principle encode the conserved information. Unfortunately, these representations cannot be aligned using the traditional alignment tools. Here, we repurposed an approach from the signal processing literature called dynamic time warping, enabling direct alignment in embedded space. Our approach offers an alternative method to identify similar subregions. We show that even within disordered regions, local subregions can be highly conserved despite substantial sequence variation. In summary, our work offers a robust approach to assessing sequence similarity in arbitrary feature embeddings.

De novo design of peptides localizing at the interface of biomolecular condensates

Timo N. Schneider1, Marcos Gil-Garcia1, Marco A. Bühler1, Lucas F. Santos1, Lenka Faltova1, Gonzalo Guillén-Gosálbez1, Paolo Arosio1

1ETH Zurich

The interface of biomolecular condensates has been shown to play an important role in processes such as protein aggregation and biochemical reactions. Targeted modulation of these interfaces could, therefore, serve as an effective strategy for engineering condensates and modifying aberrant behaviors. However, the molecular grammar driving the preferential localization of molecules at condensate interfaces remains largely unknown. In this study, we developed a computational pipeline that combines coarse-grained simulations, machine learning, and mixed-integer linear programming to design peptides that partition at the interfaces of specific condensate targets. Using this workflow, we designed and synthesized peptides that localize at the interfaces of condensates formed by the intrinsically disordered protein regions (IDRs) of hnRNPA1, LAF-1, and DDX4. These peptides exhibit surfactant-like architectures, with one tail incorporated into the condensate and the other excluded from the dense phase. The distinct peptide sequences highlight the importance of the net charge of the scaffold protein as a key physicochemical parameter for designing peptides with preferential interfacial localization. Overall, our pipeline represents a promising strategy for the rational design of interface-localizing peptides and the identification of the corresponding molecular grammar.

A position-based substitution sensitivity framework reveals cold and hot spots for epistasis

Anastasiia Sivchenko1, Liskin Swint-Kruse1

1KUMC

In algorithms that predict protein mutational effects, a common input is the evolutionary information from multiple sequence alignments (MSA). MSAs provide a natural experiment to measure each protein position's tolerance to mutations during evolution. The common use of MSAs involves the assumption that conserved positions will not tolerate substitution in any homolog. This assumption is based on the fact that each protein family shares a common structural architecture and the expectation of position-specific roles. However, these assumptions do not apply to substitutions at non-conserved positions, as we have shown for the class of "rheostat" positions, where substitutions tune protein function over a wide range. Previous studies using families of existing natural and synthetic proteins led to the hypothesis that the locations of these positions are preserved during evolution. Here, we directly assessed the evolution of substitution sensitivities using mutagenesis datasets of extant and ancestrally reconstructed steroid receptor proteins reported by the Thornton lab. Indeed, more than half of protein sites preserved their position sensitivity classes, underlining the common architecture of protein positions' roles. In addition, our framework reveals that classifying positions by their substitution sensitivities provides information about the widespread epistasis observed for individual substitutions in these data. Our analyses showed that epistatic coldspots are associated with positions that are preserved as toggle, where epistatic hotspots directly correlate with a higher rheostatic character. These results suggest that a protein position's propensity to epistasis and ability to tune function are related phenomena with similar biophysical roots.

NCR Peptide–Protein Interactions in Legume–Rhizobium Symbiosis: AlphaFold Multimer Predictions andValidations

Vikram Venugopal1, Ketak Kapadnis2, Jay Unruh1, Jennifer Gardner1, Siva Sankari1

1Stowers Institute for Medical Research, 2Stowers Institute; Indian Institute of Science Education and Research

The symbiotic relationship between the legume Medicago sativa (alfalfa) and the bacterium Sinorhizobium meliloti underpins biological nitrogen fixation and agricultural productivity. Central to this interaction are nodule-specific cysteine-rich (NCR) peptides, a diverse family of plant-derived effectors. To explore emergent behaviors of NCR–protein interactions, AlphaFold Multimer was used to predict binding between ~700 NCR peptides and ~6,400 S. meliloti proteins, generating ~500 high-confidence candidates filtered by structural confidence and expression data.

Selected interactions were experimentally tested using yeast surface display, yeast two-hybrid, biolayer interferometry (BLI), isothermal titration calorimetry (ITC), and flow-induced dispersion analysis (FIDA). A novel symbiosis-associated protein, FixU, exhibited binding to three predicted NCR partners in yeast display, with validation underway through orthogonal assays. We are also developing functional assays for FixU and advancing high-throughput yeast display formats that enable pooled NCR–target screening.

These findings demonstrate how AI-driven structure prediction can prioritize complex peptide–protein interaction landscapes for systematic experimental validation. By combining AlphaFold Multimer predictions with scalable binding assays, we provide a framework to dissect NCR-mediated control of symbiosis, offering broader insights into the role of small peptides in host–microbe interactions.

Integrating AI reasoning with experiments to prioritize mechanistic hypotheses

Alex T. Von Schulze1, Justin Mehojah1, Lexie Berkowicz1,2, Xiaoqing Song1, Jay Unruh1, Randal Halfmann1,2

1Stowers Institute of Medical Research, Kansas City, MO, 64114, USA., 2Department of Biochemistry and Molecular Biology, University of Kansas School of Medicine, Kansas City, KS, 66160, USA.

Artificial intelligence (AI) promises to accelerate discovery, yet its capacity to generate mechanistic, testable hypotheses from experiments remains uncertain. Here we evaluate a novel reasoning model in alpha stage (creator undisclosed under confidentiality) that converts empirical observations and experimental goals into ranked mechanistic hypotheses. As a testbed, we supplied the model with a structured summary of our interpreted experimental observations indicating that proteasome inhibition with bortezomib (BTZ, 0.1 uM) reduces nucleation/assembly across amyloid-forming sequences in HEK293T cells (A ß 42 and poly-TA). These interpretations were derived from (i) Distributed Amphifluoric FRET (DAmFRET) profiles, (ii) semi-denaturing detergent–agarose gel electrophoresis (SDD-AGE), and (iii) genetically encoded multimer (GEM) diffusion measurements. The model returned 20 hypotheses, each scored for mechanistic plausibility, correctness review, novelty, experimental tractability, and overall feasibility. The top-ranked mechanisms implicated (i) O-GlcNAc– mediated kinetic arrest of amyloidogenesis and (ii) translational "braking" that lowers nascent-chain flux to assembly. Orthogonal data supported these AI-prioritized mechanisms. Unbiased proteomics after BTZ revealed significant suppression of aminoacyl-tRNA biosynthesis (A ß 42: FDR = 6.0×10^{-11}) and ribosome function (A ß 42: FDR = 2.4×10^{-5} ; poly-TA: FDR = 4.89×10^{-29}), consistent with translational slowing. Notably, O-GlcNAc transferase (OGT) was among 59 proteins downregulated by BTZ in both the A ß 42 and poly-TA datasets, aligning with the O- GlcNAc hypothesis. Causal perturbation tests further corroborated these predictions: thiamet-G (O-GlcNAcase inhibitor) increased A ß 42 and poly-TA assembly and partially rescued the BTZ phenotype, indicating that altered O-GlcNAc homeostasis modulates amyloid assembly. These findings demonstrate that reasoning models can narrow the experimental search space from many possibilities to a small set of rational, falsifiable mechanisms that withstand experimental validation. This observation → AI-ranking → validation loop provides a generalizable, rigorous strategy for integrating AI systems into hypothesis formation while maintaining scientific standards and testability.

Embracing Structural Mass Spectrometry for Protein Pleomorphism

Yan Yan1, Kai Zhang1, Gabrielle Pycior1, Pablo Adrian Guillen-Poza2, Ruben Hervas2, Paulo Leal1, Kexi Yi1, Zulin Yu1, Ying Zhang1, Laurence Florens1, Jay Unruh1, Kausik Si 1,3

1Stowers Institute for Medical Research, Kansas City, USA, 2School of Biomedical Sciences, The University of Hong Kong, Pokfulam, Hong Kong, 3Department Molecular and Integrative Physiology, KU Medical School, Kansas City, USA

Protein structure dictates protein function, and proteins often perform the same function by adopting similar structural basis in different cellular or tissue environments. Here, we report that in Drosophila melanogaster, the broadly expressed protein p62/SQSTM1 performs mutually exclusive functions in different tissues. In the midgut, it plays roles in autophagy, whereas in muscle, it participates in mitochondria-associated protein degradation. Using crosslinking mass spectrometry and structural modeling, we find that these functional differences are due to tissue-specific folding differences of the same p62 protein. The tissue-specific folded states were then validated by cryoEM and far-UV circular dichroism. We named this phenomenon as "Protein Polymorphism" to describe proteins that can adopt different structural and functional states depending on cell or tissue type. Further, we found deletion of a short intrinsically disordered region in p62 switches its folded state and function in muscle to resemble that in the midgut. However, this functional switch leads to muscle dysfunction, suggesting that tissue-specific folding and function are causally linked to tissue-specific physiology. We posit that tissue-specific folding of p62 represents a mechanism of tissue-specific function, where interactions between the inherent plasticity of protein folding and the physiochemical and/or molecular environment of a tissue allow a broadly expressed multifunctional protein to adopt tissue-specific folded states and functions.

Extracting mechanistic insights from AI sequence models – the intriguing case of low-affinity DNA binding motifs

Julia Zeitlinger1, Melanie Weilert1

1Stowers Institute of Medical Research, Kansas City, MO, 64114, USA.

Establishing Protein Sequence-Property Relationship via Experiments, Simulations and Interpretable Deep-Learning Model

A. Gomez, A. Pandey, B. Mehrafrooz , N. Kennedy, J. Miller, D. Tullman -Ercek, M. Olvera de la Cruz, W. Chen, S. Keten

Department of Material Science and Engineering, Northwestern University; Chemical and Biological Engineering, Northwestern University; Materials Research Science and Engineering Center, Northwestern University; Center for Synthetic Biology, Northwestern University

Bacterial microcompartments (MCPs) are protein -based organelles that self -assemble into complex architectures, yet their assembly mechanisms remain poorly understood. This project integrates experimental and computational approaches to map the apparent fit ness landscapes of PduA, a key shell protein of the 1,2 Propanediol Utilization (Pdu) MCPs using amino acid variant libraries. Favorable mutations were identified and used to train supervised machine learning models, including K -Nearest Neighbor, logistic regression, random forest, and XGBoost, to predict variant effects on assembly. Complementary all -atom molecular dynamics simulations that capture the nature and extent of self- assembled PduA nanostructures arising from specific intermolecular interactions . Together, these data -driven and molecular approaches will reveal how sequence- level changes shape higher -order assembly and stability in MCPs. The resulting framework aims to guide predictive mutational design of protein -based nanostructures.

PFNet: a method to determine residue-level ensemble energies from HX/MS data

Chenlin Lu, Kyle Weber, and Anum Glasgow

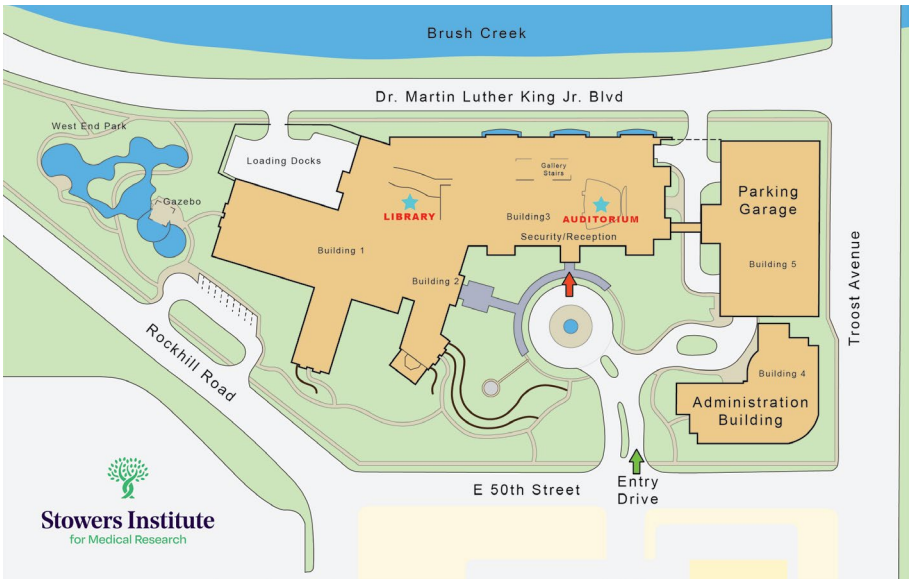
Protein function is encoded in sequence-specific energetic relationships that are invisible in the lowest-energy ground state protein structures (1). Even single mutations can perturb this network, resulting in unpredictable effects on a protein's local and global stability which may impact its function (2, 3). We commonly use hydrogen exchange/mass spectrometry (HX/MS) to qualitatively assess how molecular perturbations impact protein structure. However, in theory, HX/MS data contain all the information necessary to derive site-resolved energies of local unfolding (ΔGop) (4, 5). We introduce PFNet, a machine learning method to determine residue-level hydrogen exchange rates from conventional HX/MS datasets, which can then be converted to ΔGop. PFNet has a transformer architecture that takes peptide-level isotopic mass envelopes from HX/MS experiments as inputs, encodes the raw isotopic mass envelopes using a convolutional neural network, and outputs HX rates at the highest resolution that the data affords, alongside a confidence score and a predicted absolute error (pAE). We trained PFNet on billions of synthetic HX/MS isotopic envelopes with varied time points, back exchange, noise, and peptide coverage and developed a synthetic benchmark set to test its accuracy. We then benchmarked PFNet against HX/NMR to find that it has near-perfect accuracy. We will share examples of using PFNet to reveal the energetic effects of sequence variation on protein conformational ensembles, from point mutants to distantly related protein orthologs, to understand the molecular basis for their varied functions.

Attendee Resources

Location: Stowers Institute for Medical Research – 1000 E. 50th Street, Kansas City, MO 64110
General Phone Number: 816-926-4000
Driving directions from Kansas City International Airport:
(approximately 30 minutes under normal driving conditions)

1. Follow the signs on International Circle to the airport exit.
2. Take Cookingham Drive to I-29 South (right ramp), heading toward Kansas City.
3. Continue to follow I-29 South as it merges into I-29 South 71 (merge in left lane), then I-35South (merge in left lane again).
4. Cross the Paseo Bridge and get in the lane for I-70 South 71 (Exit 3)
5. Merge right and follow sign to South 71 Highway (Exit 2M).
6. Take the Emmanuel Cleaver II Boulevard exit and turn right onto Emmanuel Cleaver II Boulevard.
7. Follow Cleaver II to Troost Avenue and turn left on Troost.
8. Turn right on 50th Street and take another immediate right at the Stowers Institute's entrance (1000 E. 50th Street, Kansas City, MO 64110).

Campus Map



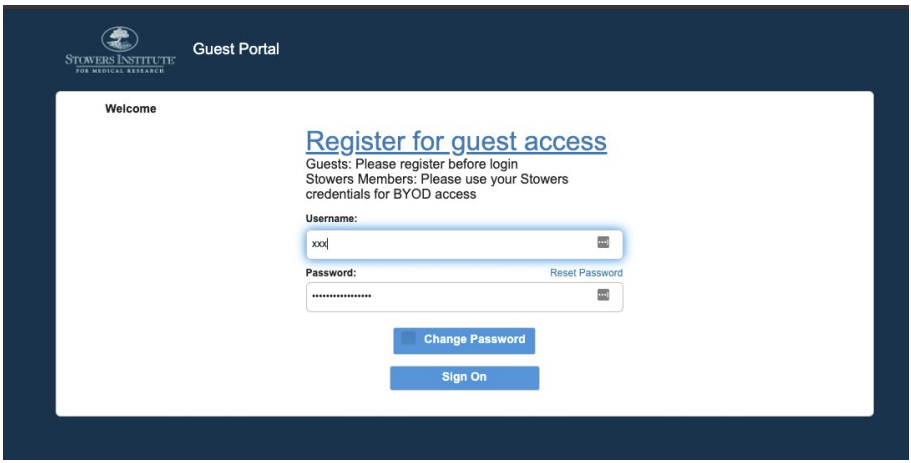
Parking: Visitor parking is available in the parking structure located between the Administration Building (to your right as you enter the campus) and the Research Building (to your left). Please park on the 5th floor of the parking garage and take the elevator down to the 1st floor to enter the Research Building.

Restrooms: Located by the two large planters on the 1st floor of the Research Building. All-gender restrooms are available at the base of the stairs on the B1 level.

Mother's Room: At the base of the stairs on the B1 level, located next to the all-gender restrooms. An entry key will be provided by security to guests that request to use it.

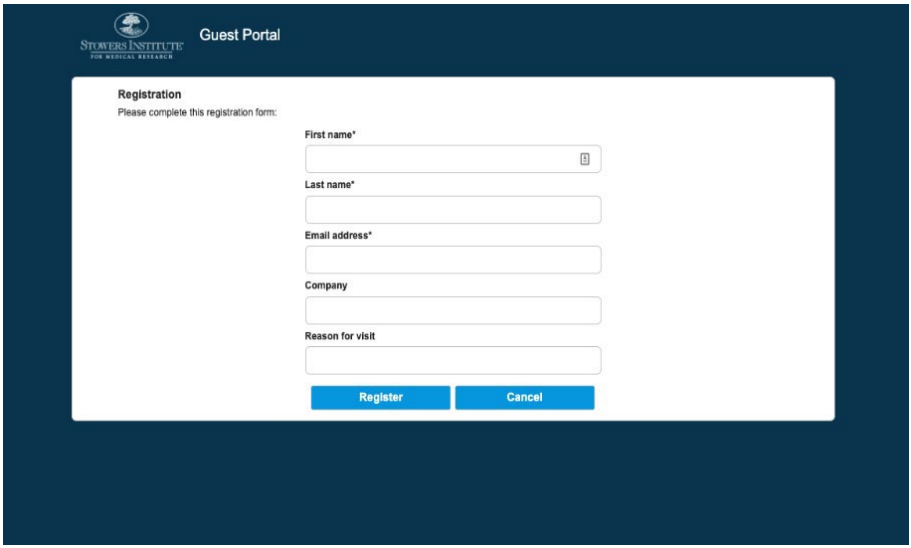
Tobacco Free Campus: Tobacco of any kind is prohibited in all the Institute facilities and on the Institute grounds. The Institute does not offer any designated smoking areas.

Wireless Access: Stowers Guests may connect to the **stowers_guest SSID**
When connected, you will see a captive portal. Click on the link at the top of the page to register for guest access.



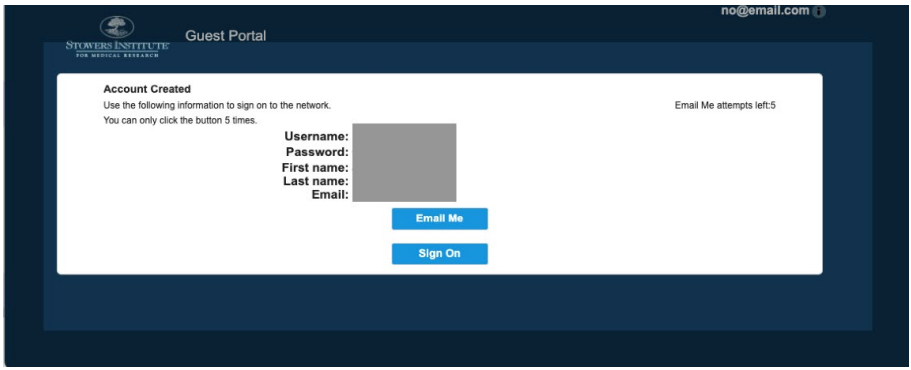
- You will be prompted to enter their information. Name and email are the only required fields.

- After entering the required information click Register.



- You will see a page like this with sign on information. You can use the userid and password to connect other devices without going through the registration, by entering the userid and password in the first page of the captive portal.

- Click the "Sign On" button.



After clicking sign on, you will be prompted to accept the "Acceptable use Policy". After clicking accept you should now have internet access.

If you have any questions, please contact the Stowers Help Desk at 816-926-4150.

Kansas City Fun: The Kansas City area offers a wealth of cultural, educational, and entertaining opportunities to explore, many of which are inexpensive or free of charge. The following is a brief listing of local attractions, event calendars, and resources.

Useful Resources
Information about Kansas City and the surrounding area can be found at kcmo.gov or at visitkc.com. Two publications to help you discover things to do in the area are: "Insider's Guide to Kansas City" by Katie van Luchene and "Day Trips from Kansas City" by Shifra Stein. Both can be found in local bookstores.

Linda Hall Library
5109 Cherry Street, 816.363.4600
The Linda Hall Library in Kansas City is one of the world's foremost independent science research libraries. Founded in 1946 through a philanthropic bequest, it houses vast collections spanning science, engineering, and technology, serving researchers, students, and the public. Known for its international research programs and collaborations, the Library promotes lifelong learning with innovative programming, digital access, and educational resources. The Library also maintains its grounds as a public urban arboretum, offering both scholarly and community engagement opportunities.
www.lindahall.org

Nelson-Atkins Museum of Art
4525 Oak Street, 816-751-1278
Opened in 1933, the Nelson-Atkins Museum has more than 50 galleries and several period rooms. The museum's outstanding feature is its collection of Asian art. The collection of Chinese landscape paintings is one of the finest in the West, and the museum's holdings of Chinese ceramics and decorative arts are also noteworthy. Besides European paintings from the Renaissance on, the museum also has notable collections of ancient Egyptian sculpture, Japanese porcelains and lacquer, and English pottery. The E.F. Pierson Sculpture Garden was dedicated

in 1972, and the Henry Moore Sculpture Garden opened in 1989. Admission is free.
nelson-atkins.org

Kemper Museum of Contemporary Art
4420 Warwick, 816-753-5784
Founded in 1994, the Kemper Museum of Contemporary Art presents modern and contemporary art of the highest quality and significance. It collects, preserves, documents, interprets, and exhibits a growing permanent collection; develops and presents special exhibitions; and offers a variety of educational programs. Admission is always free, and the Museum serves a diverse and inclusive public population.
kemperart.org

Union Station
30 West Pershing Road, 816-460-2020
This fully restored 1914 landmark is Kansas City's most prominent destination for entertainment and cultural activities. The Station is home to a permanent rail exhibit with vintage rail cars, an interactive science center, a vibrant Theater District featuring giant screen movies and live theater, fine restaurants, unique shops, spaces for meetings and events and much more. Of course, you can still catch the train at Union Station, once again among Amtrak's busiest stops.
unionstation.org

Loose Park
Intersection: Wornall Road and 51st Street, 816-784-5300
Loose Park is one of Kansas City's most beautiful parks. The park is home to a lake, a walking path, a shelter house, Civil War markers, tennis courts, a wading pool, picnic areas and the famous Rose Garden. The Rose Garden is popular for all types of outdoor special events including theatrical performances and wedding ceremonies.

Westport
Westport is one of Kansas City's premier destinations for dining, shopping, site seeing and is the heart of the city's nightlife. Located in the midtown, Westport is just north of the Country Club Plaza and a few miles south of downtown Kansas City. Historically, Westport was built along the Santa Fe Trail as an outfitting center for wagon trains heading west. Today the area is filled with renovated and new buildings housing trendy shops, restaurants, and nightspots.

SRC Meeting Policies

The Stowers Research Conference (SRC) series supports an environment for the exchange of scientific ideas that is grounded in dignity and respect for all program participants. SRC believes that a diverse, inclusive and collegial community culture promotes scientific creativity and progress. The conference code of conduct outlined here has been adapted from the policies outlined by the Society for Developmental Biology, sponsor of the first SRC meeting.

No Harassment Policy
Program participants are expected to conduct themselves in a professional manner and to treat each other with dignity and respect. This expectation applies to the organizers, event attendees, volunteers, employees, consultants, vendors, and others while on Stowers premises, while representing SRC elsewhere, and while attending events organized by SRC.
SRC will not tolerate any discrimination in the form of sexual harassment and other forms of harassment. Program participants shall not engage in any conduct that could reasonably be construed as unlawful harassment against an individual. Program participants shall not make unwelcome sexual advances, make requests for sexual favors, or engage in other verbal or physical conduct of a sexual or offensive nature.

No Violence Policy
SRC does not tolerate any type of violence or threats of such violence while on Stowers premises, while representing SRC elsewhere, and while attending events organized by SRC. SRC prohibits acts or threats of violence by or against any program participants. In addition, SRC does not permit the possession or the concealed or open carrying of weapons anywhere on the Stowers premises.

Reporting
If a program participant discovers any conduct which they believe violates this policy or is otherwise detrimental to the organization, they are asked to promptly report it to the organizers.

Recording, Photography, and Session Etiquette
While in sessions, please mute all cell phones and other electronic devices. Photography or the electronic capture of scientific sessions and posters is not permitted without the expressed consent of the presenting author(s). Respect other individuals' and organizations' intellectual property and confidential information.

Photo Release
SRC has an official photographer for the meeting. Photos taken at the meeting may be used in future SRC publications, on the SRC website, or in other materials.

By registering for the meeting, you agree to allow SRC to use your photo in and SRC related publication or website.

Smoking Policy

The use of tobacco, in any form, or the use of any smoking device is prohibited in all Stowers facilities and on the Stowers campus. Stowers facilities are defined as common work areas, auditoriums, classrooms, conference and meeting rooms, private offices, elevators, hallways, food service facilities, employee lounges, stairs, restrooms, vehicles, the parking garage, and all facilities owned by the Stowers Group of Companies.

Stowers Institute

Stowers.org

Creative, innovative, fearless science

At the Stowers Institute, we aim to attract scientists who are visionary risk-takers, who seek to solve the most challenging and intractable problems in biology.

Meet the Scientists at the Stowers Institute for Medical Research
stowers.org/scientists