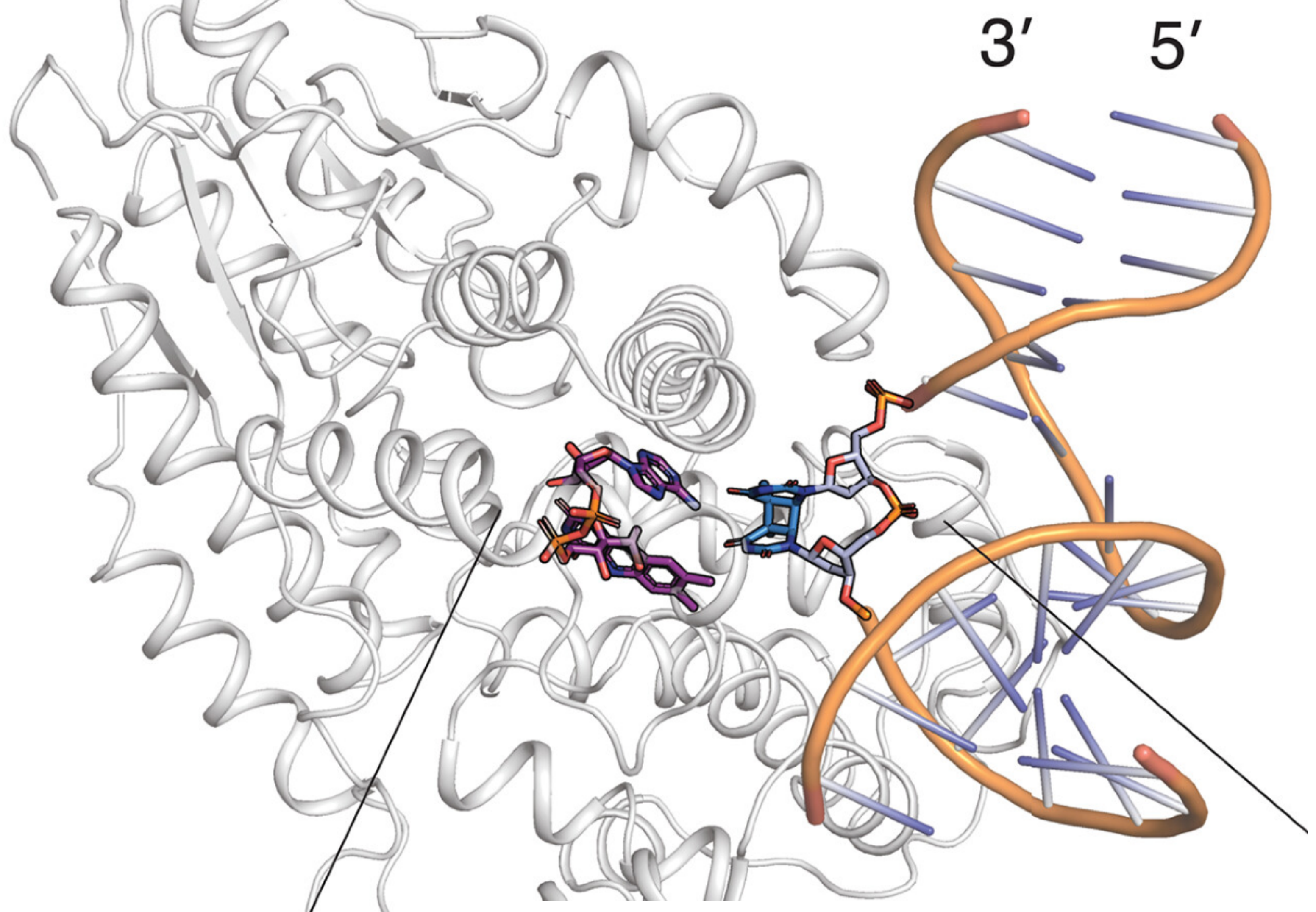
Stanford | Bioengineering
SCHOOLS OF ENGINEERING & MEDICINE

Frontiers of Materials That Learn workshop

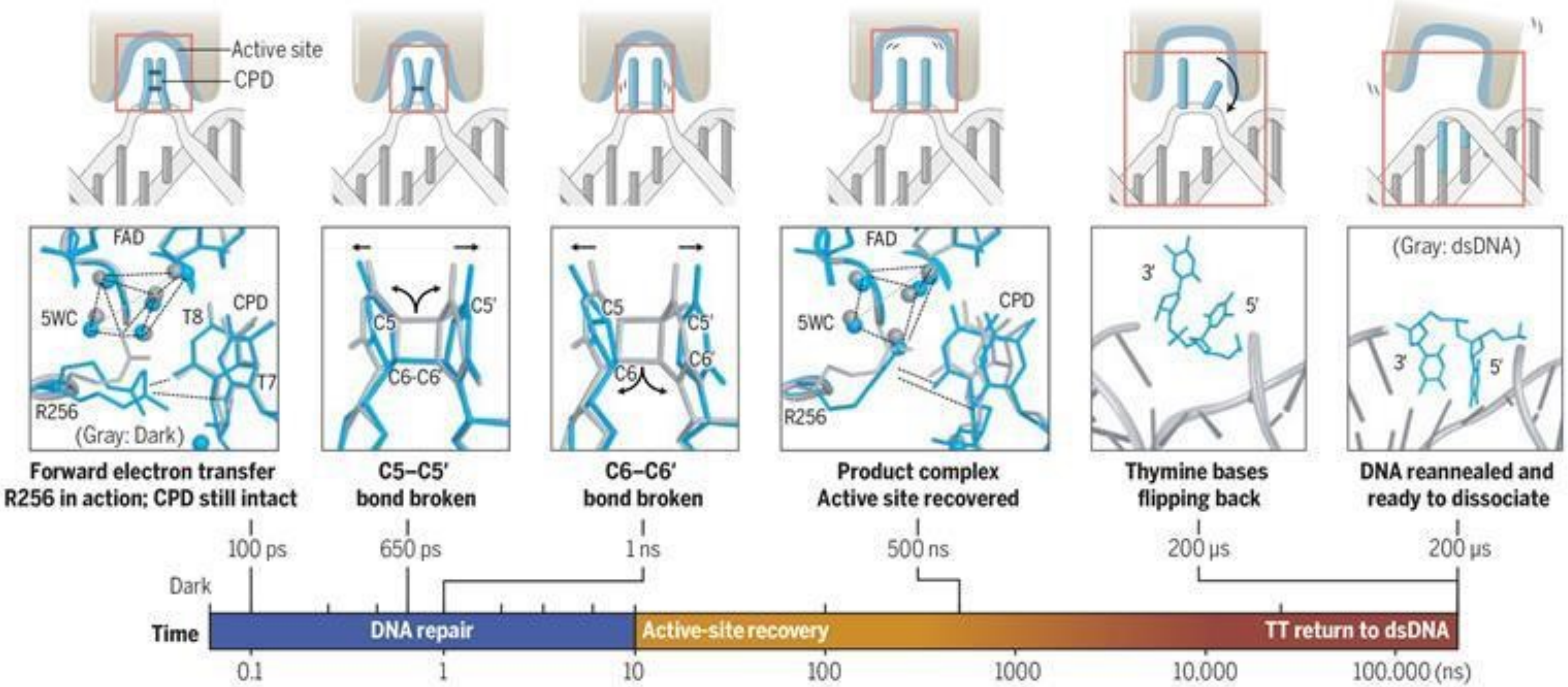
# Designed protein interactions and their emergent behaviors

Possu Huang
Assistant Professor
Stanford Bioengineering

Christou, et al. Science, 2023

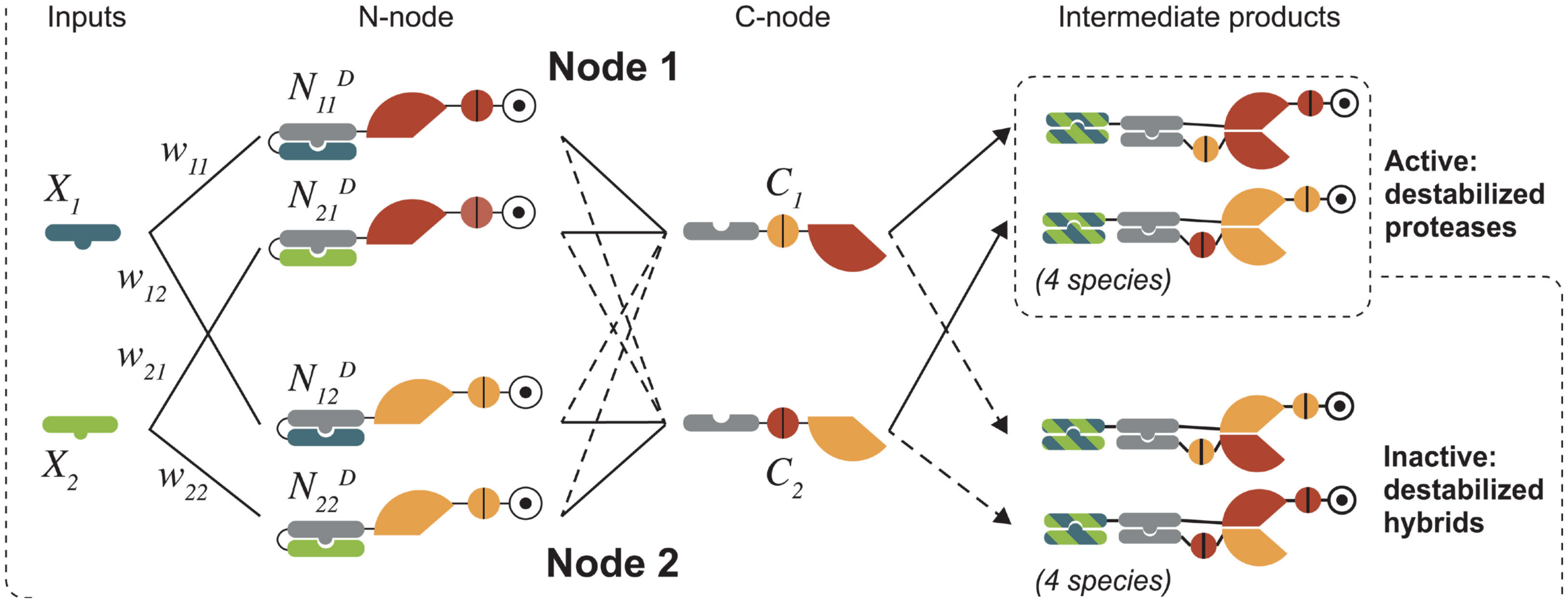


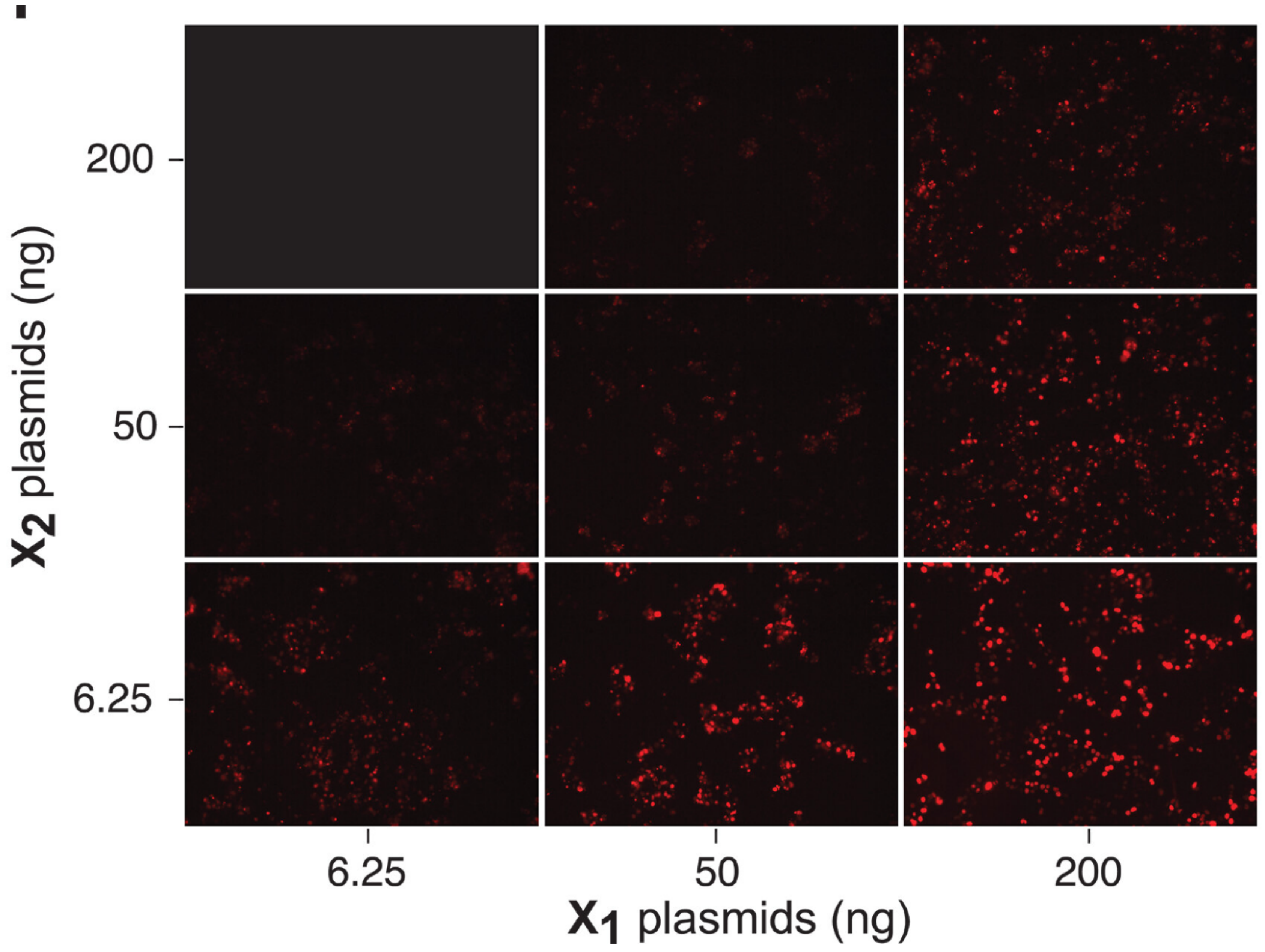
# Exhibit A: Protein responding to its environment



maestre-reyna, et al. Science, 2023

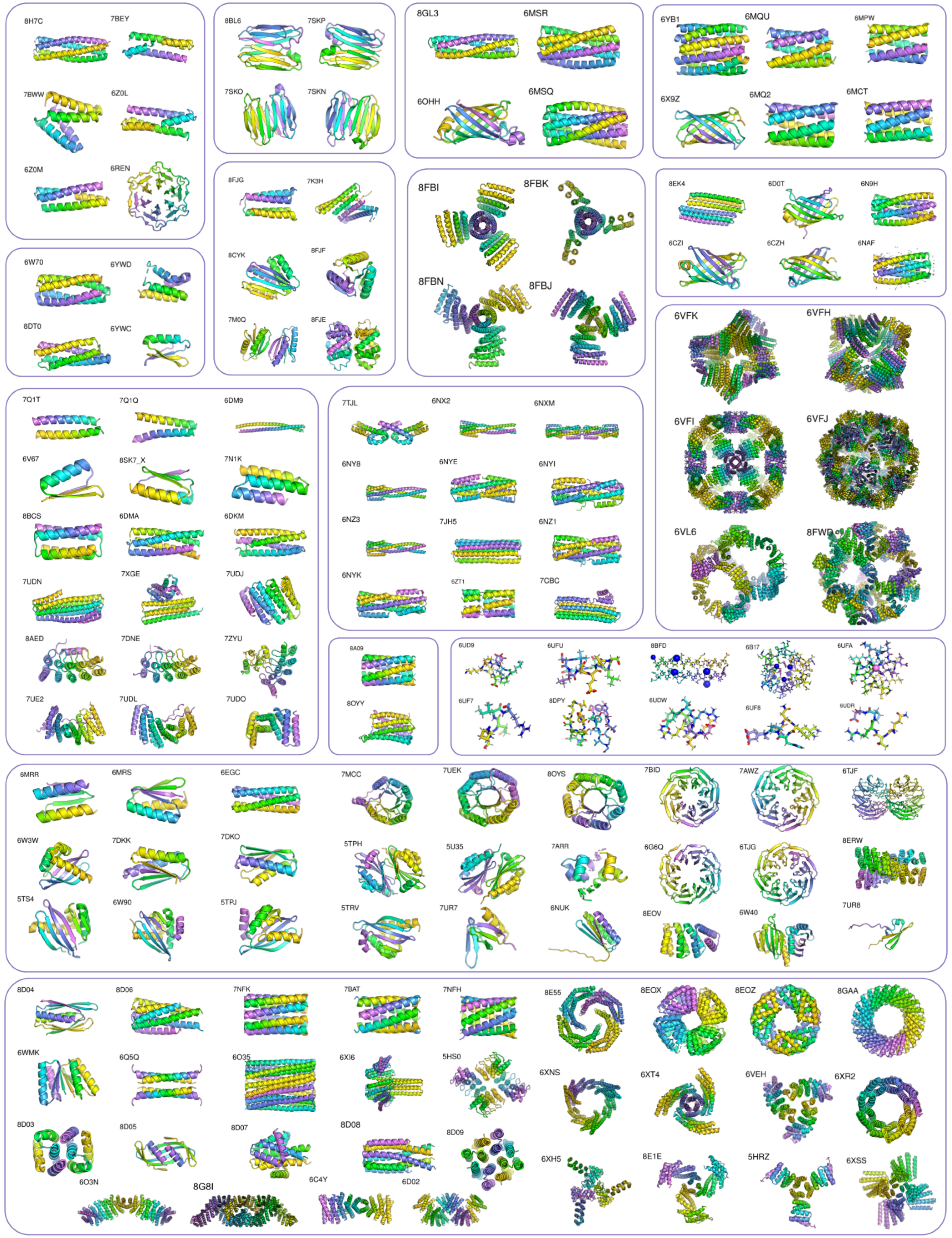
# Exhibit B: Protein logic through assembly and protease cleavage





Chen, et al. Science, 2024

# The current state of de novo protein design

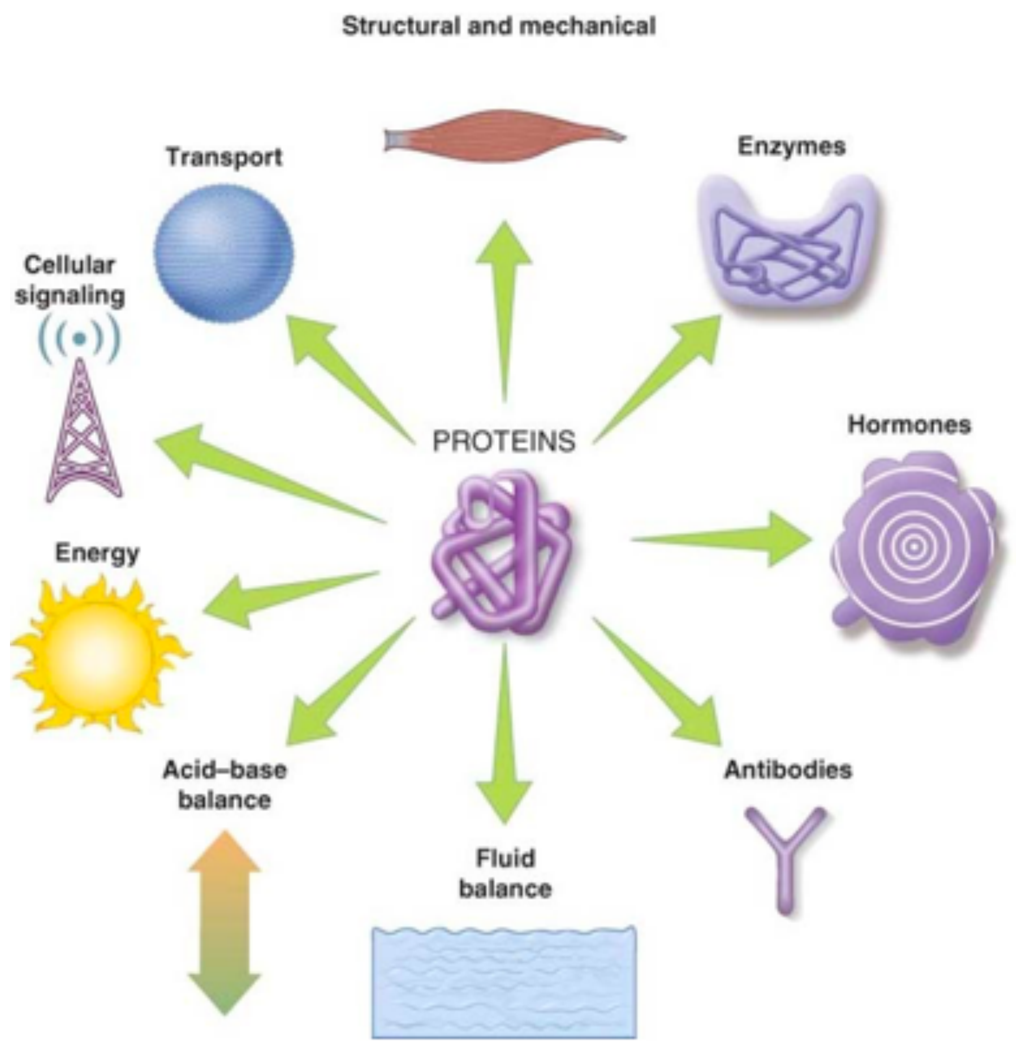


Chu, Lu and Huang, “Sparks of function by de novo protein design”, Nat.Biotech 2024

# The “central dogma” of de novo protein design

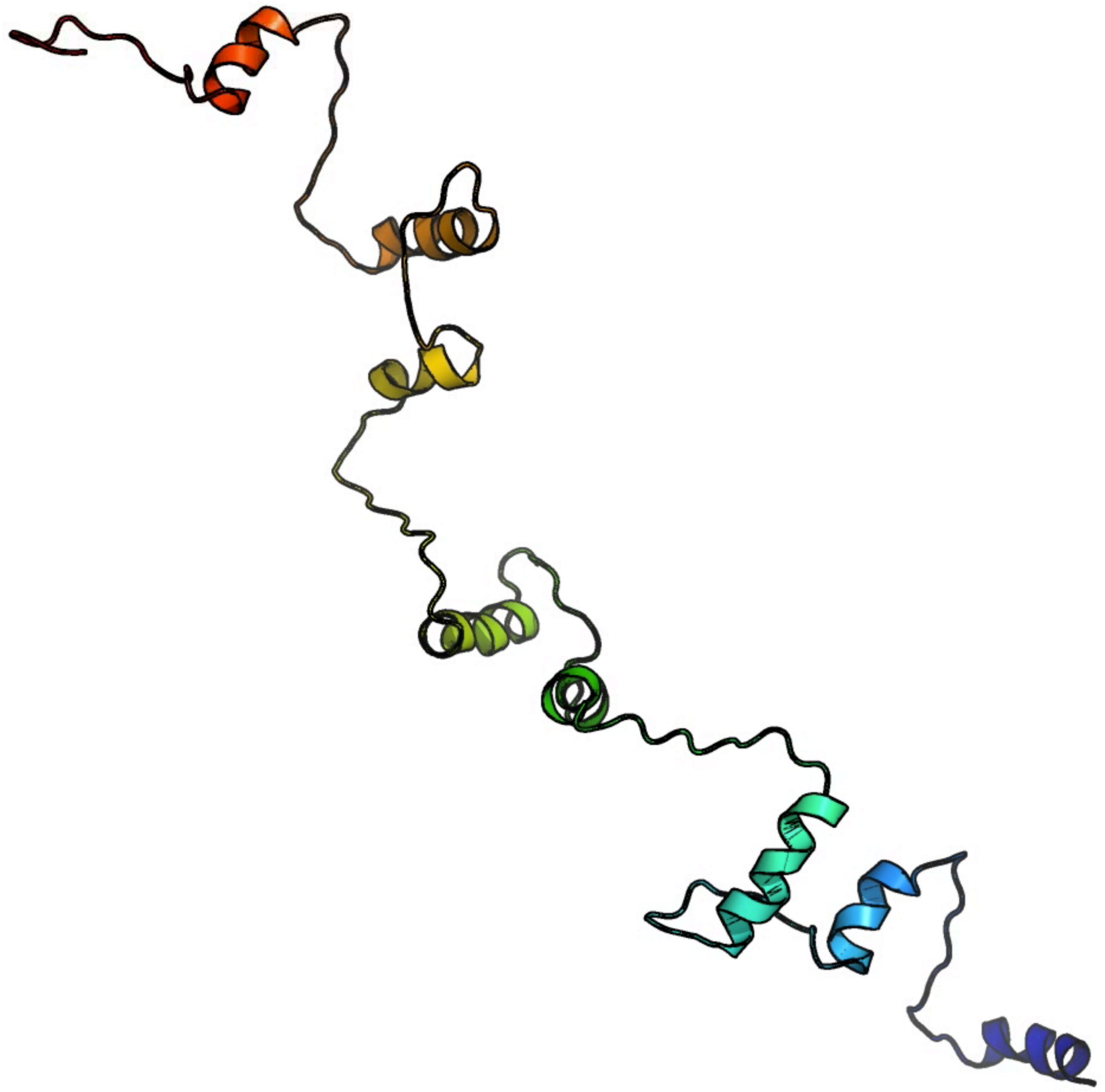
## Function

Antibodies
Enzymes
Transporters



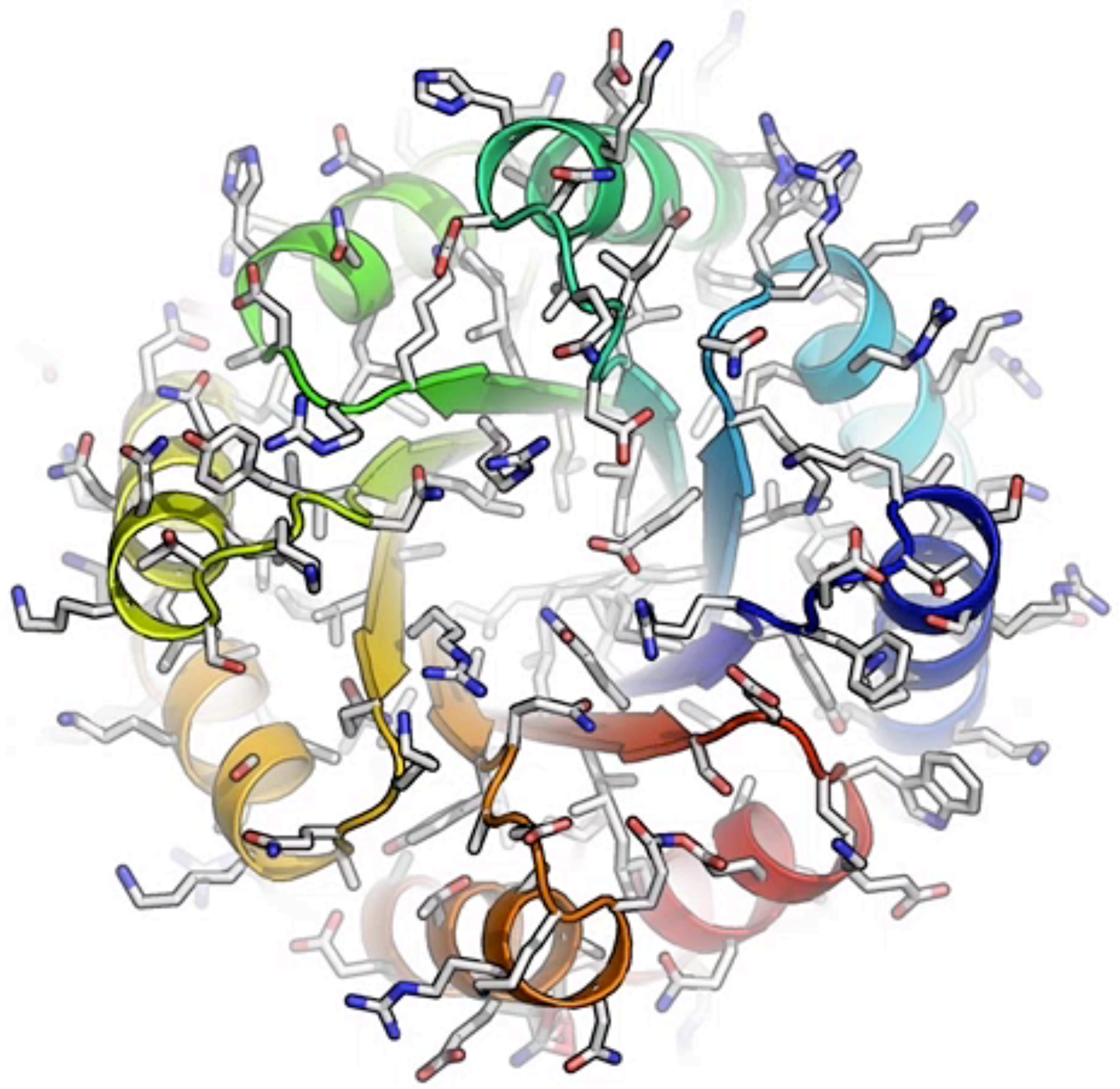
## Structure

Binding sites
Active sites
Allostery

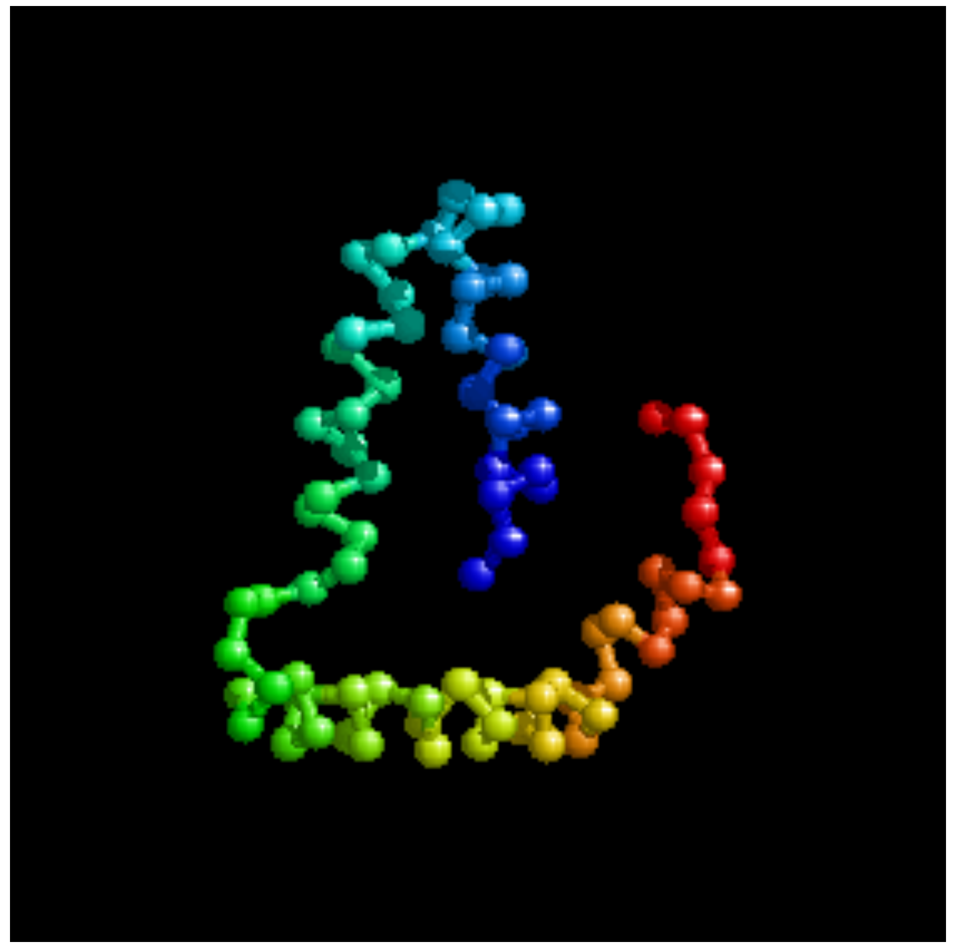


## Sequence

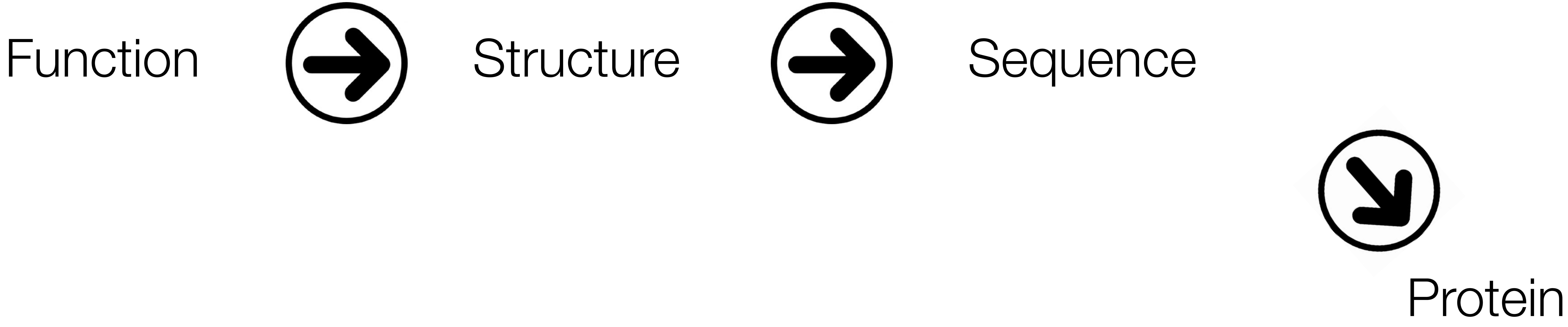
MCMC
Autoregressive
Direct prediction



## Protein!



# The “central dogma” of de novo protein design



Function

Structure

Sequence

Protein

…why do we need to break this down?

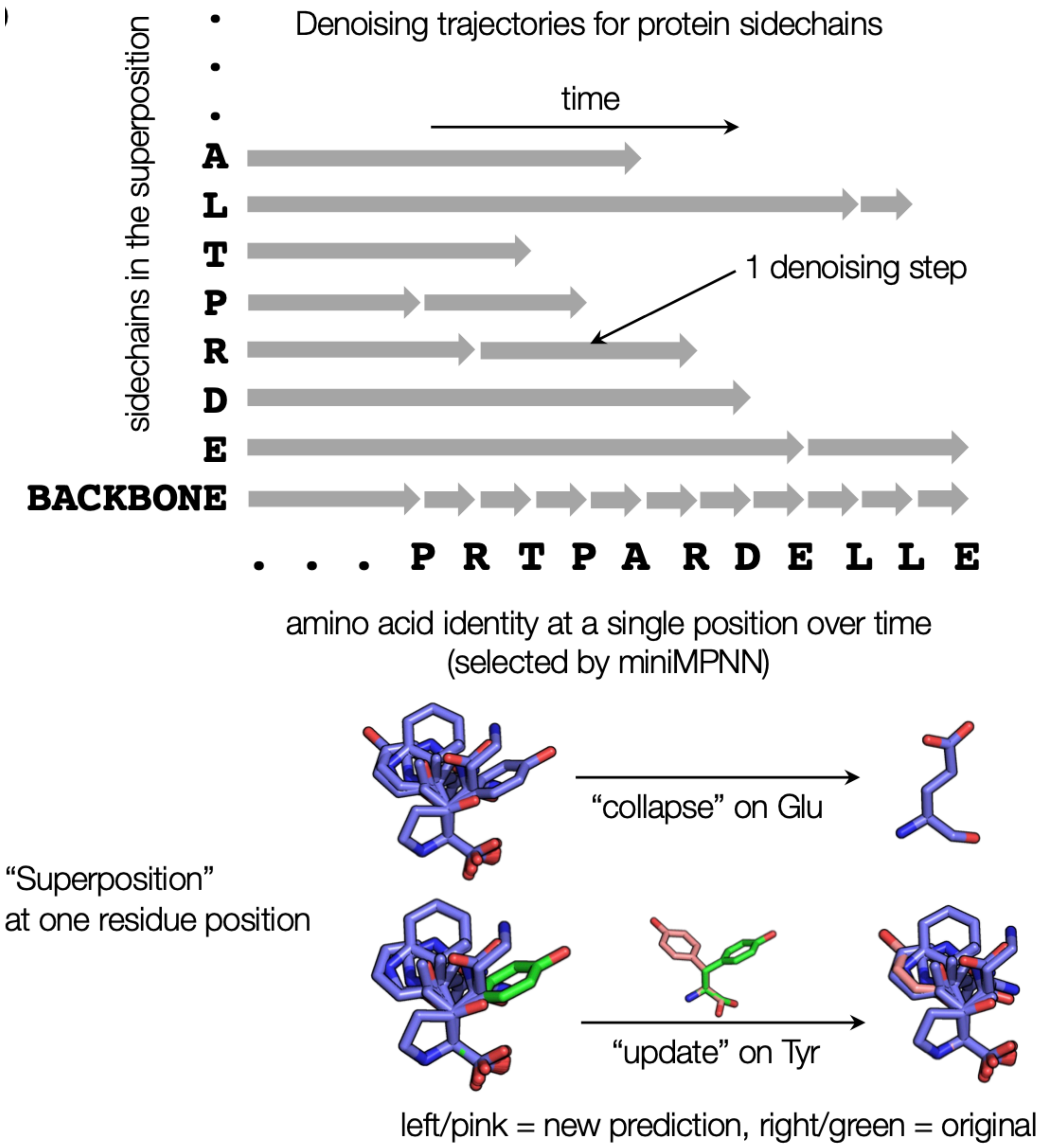
Proteins have no sense of sequence or structure:
they are “one”

# The “central dogma” of de novo protein design

Function → Structure → Sequence → Protein!

Function → Structure and Sequence → Protein

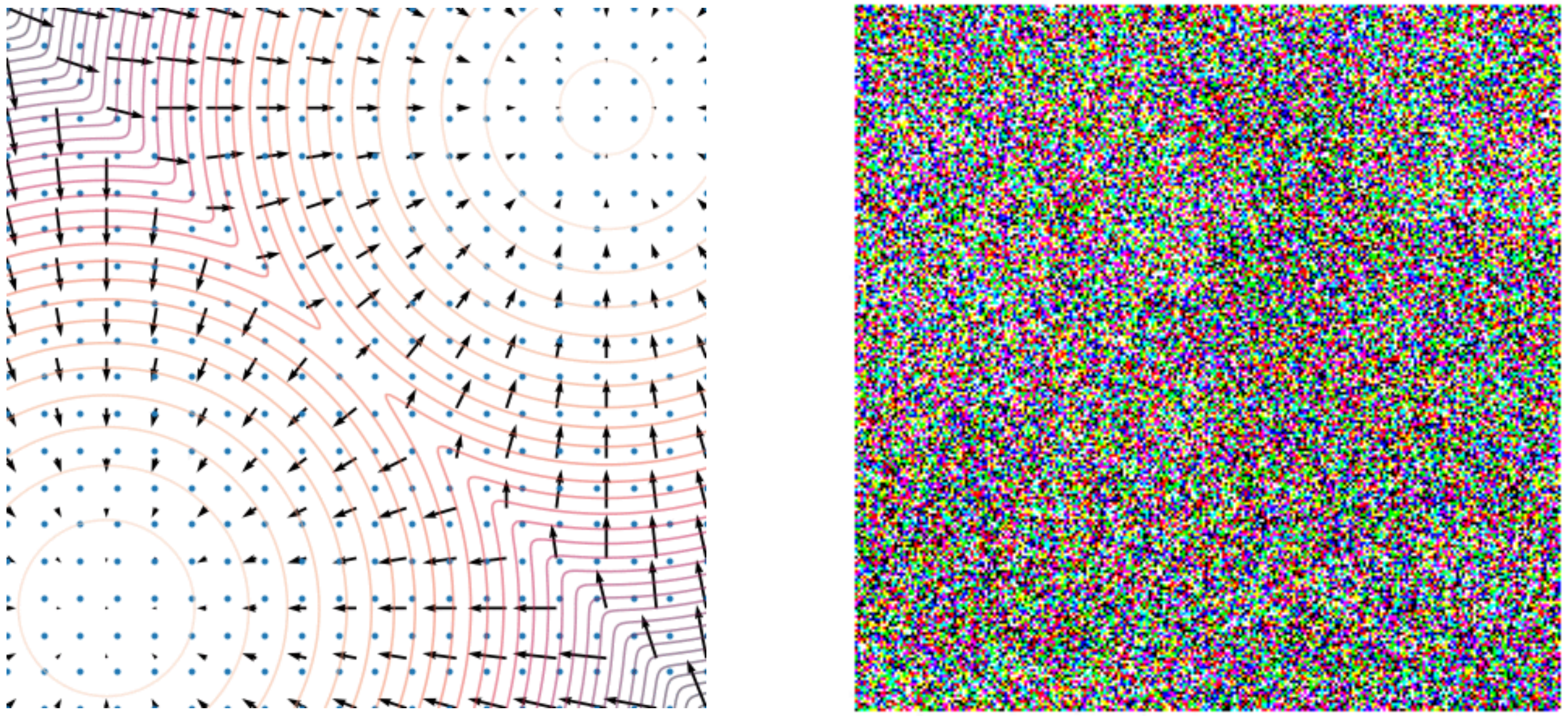
# Sequence-structure co-design through sidechain “superposition”



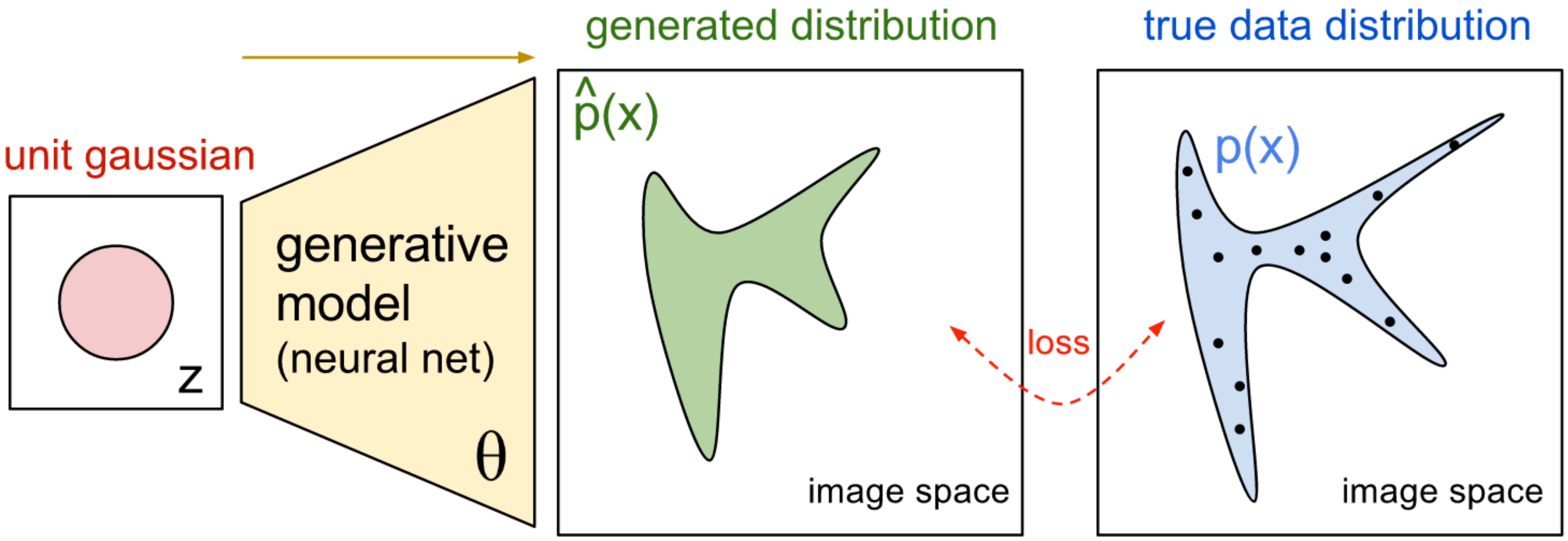
Chu, et al. “An all atom protein generative model”, PNAS, in press

# How? Modeling proteins with diffusion

- Diffusion models are a powerful new class of generative models which work via an iterative generation process
- Improves quality, and also allows for information to interact
- Interleaving sequence and structure generation allows both to feedback to each other



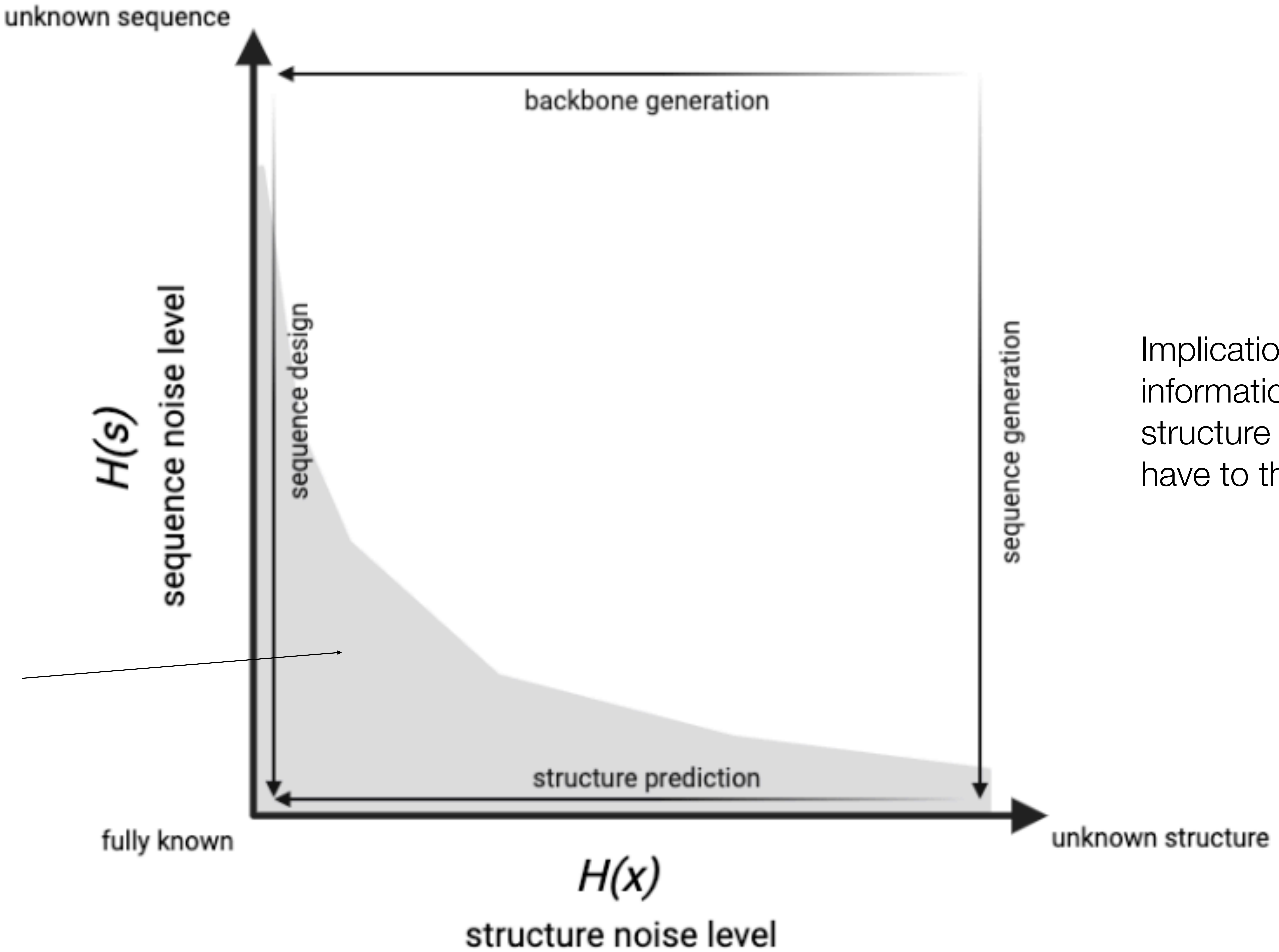
# Generative model should learn the data distribution



openAI.com

How do we make sure that the underlying data distribution is captured?

# Visualizing the model



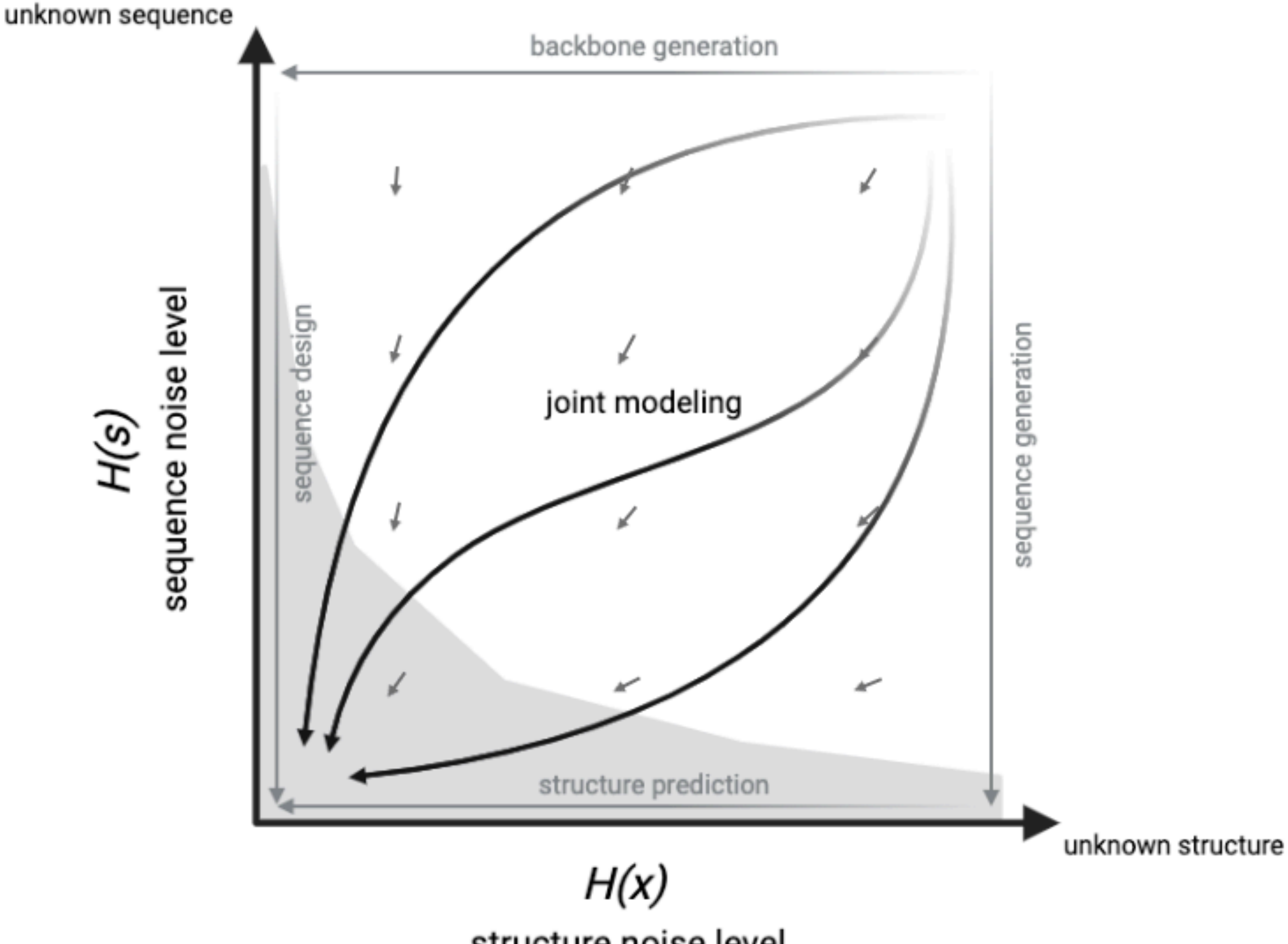
When we get into this area, the protein is basically determined – there is not much “designing” left to do

Implication: even if we have information about BOTH the structure and sequence, we have to throw one out

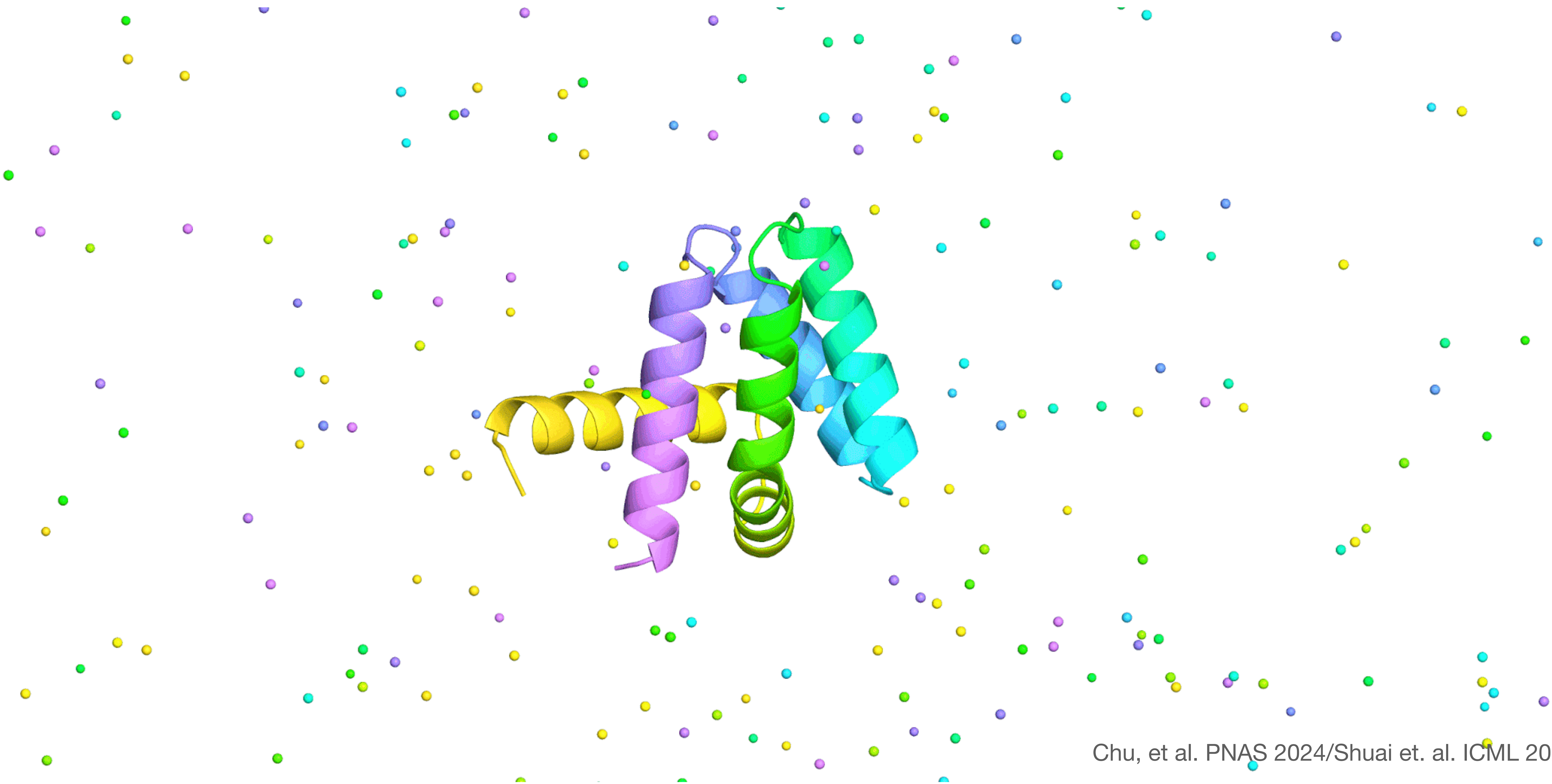
# Towards a joint model of protein structure and sequence

Describe a model which
parameterizes a gradier
field on the entire space

us to integrate
ths through this
ures existing
special cases



# Protein structure/sequence joint optimization



Chu, et al. PNAS 2024/Shuai et. al. ICML 2025

# Conditional generation of active sites

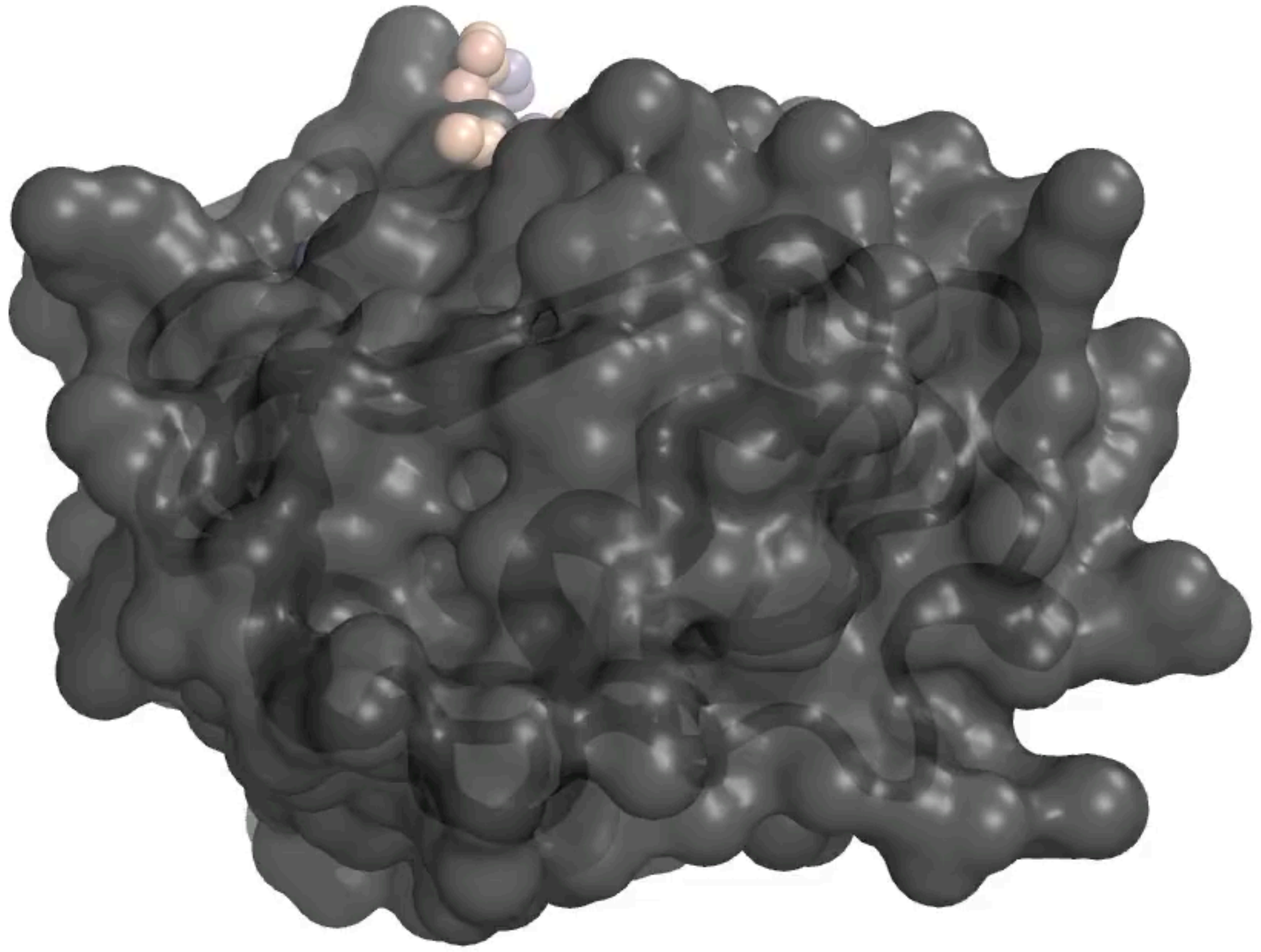
Defining the objectives
to optimize, usually
through spatial arrangements
of amino acids

Finding a structure or defining
a topology to create a new
structure that can host the amino
acid residues that define function

Resolving the amino
acid identities that
can maintain the designed
functional protein

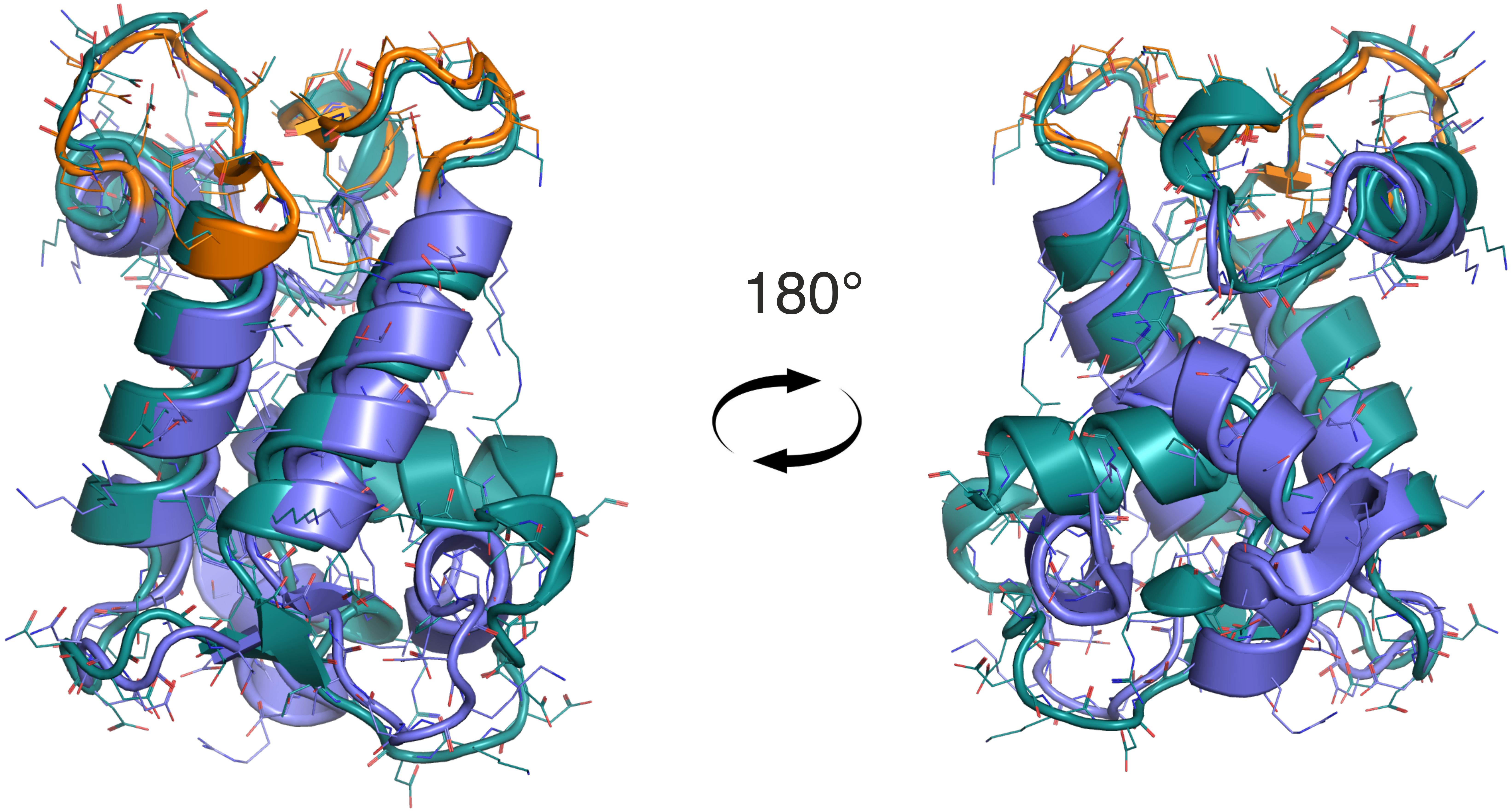
Chu, Lu, & Huang, Nature Biotecholgy, 2024

# Conditional generation of active sites



RosettaCommons.org

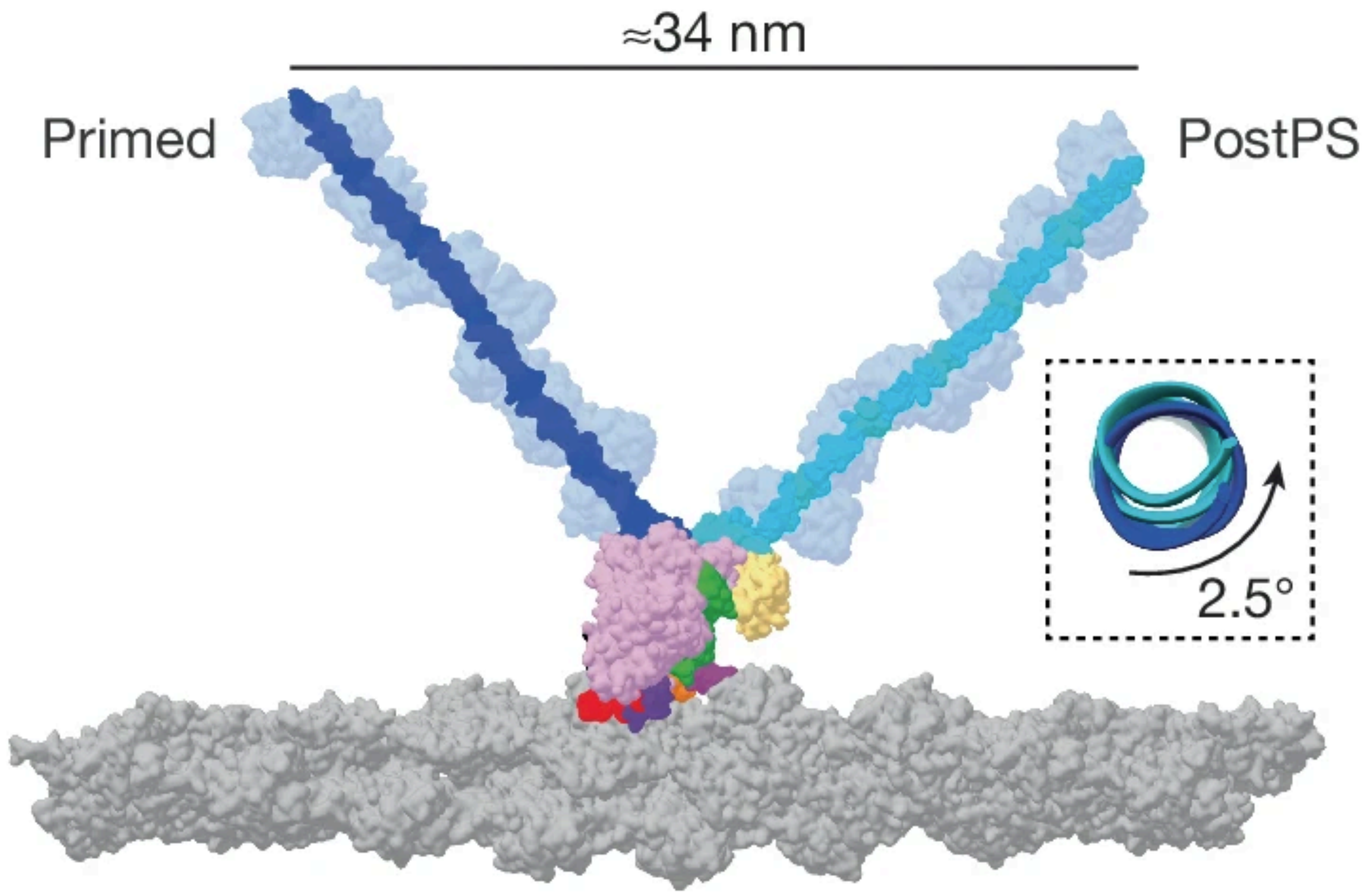
# Conditional generation of active sites

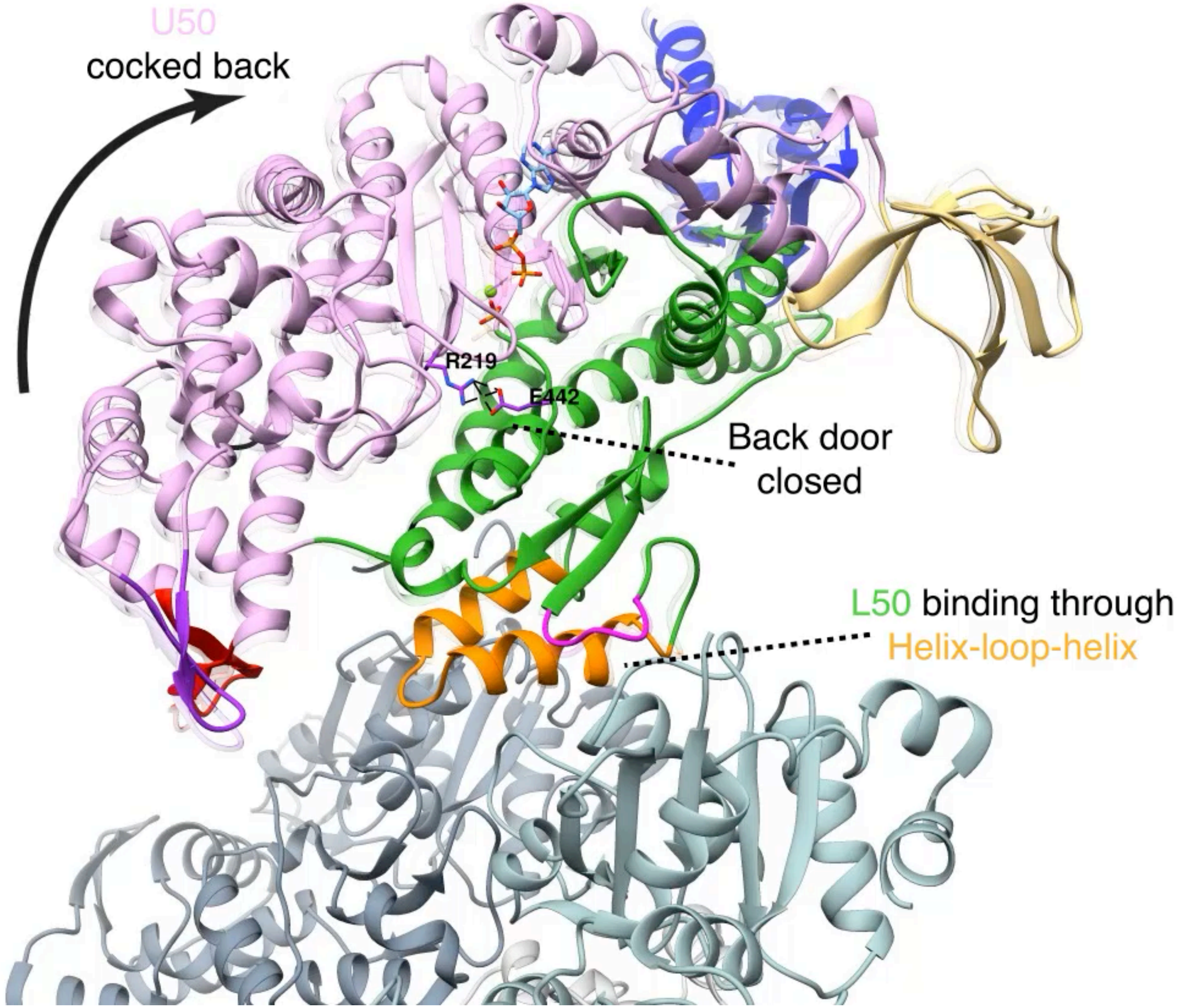


Orange: motif
Teal: generated structure
Purple: crystal structure

Lu, et al. Protpardelle-1c, bioRxiv 2025

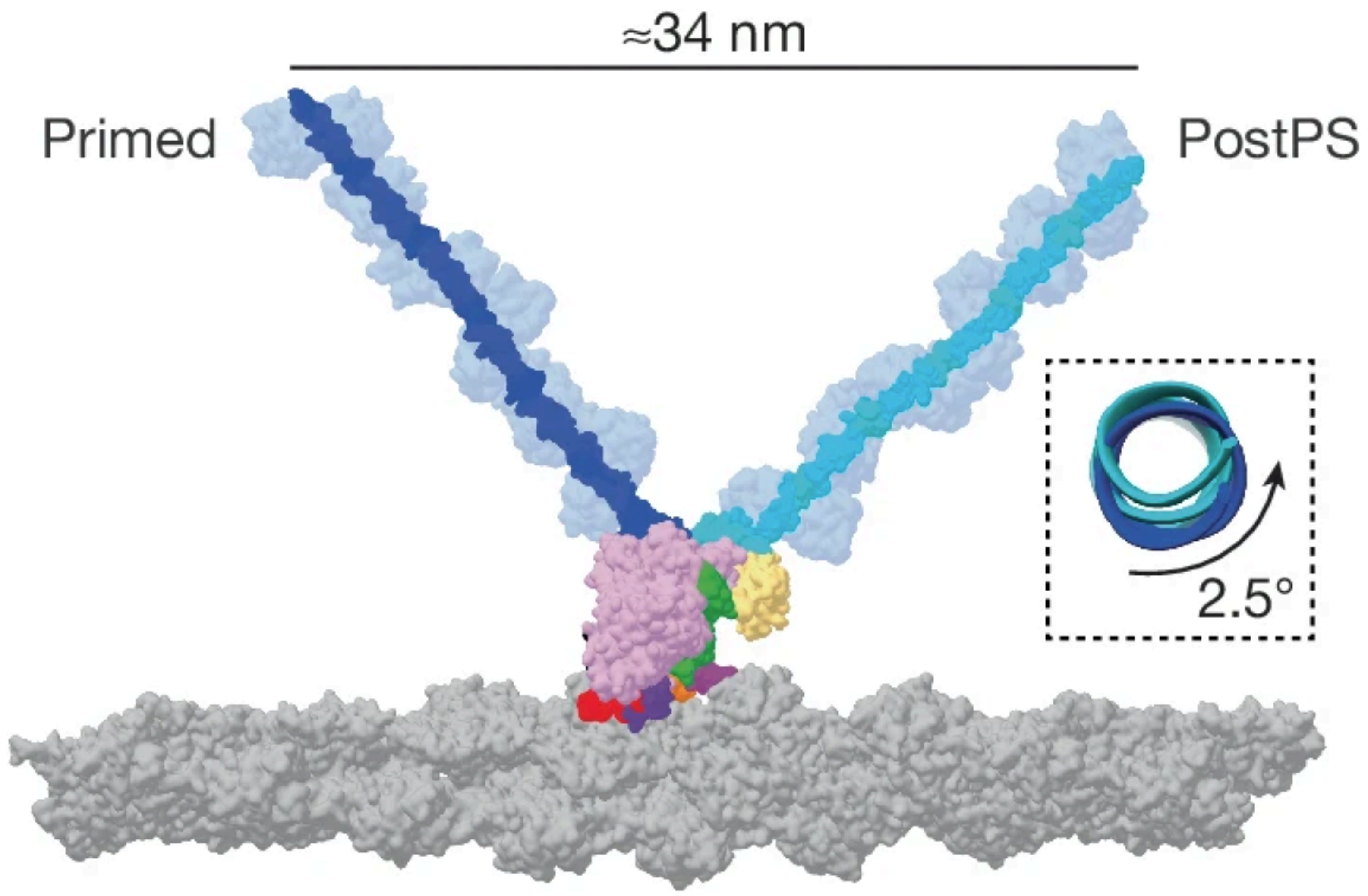
# Protein that senses and respond to environment: molecular motor



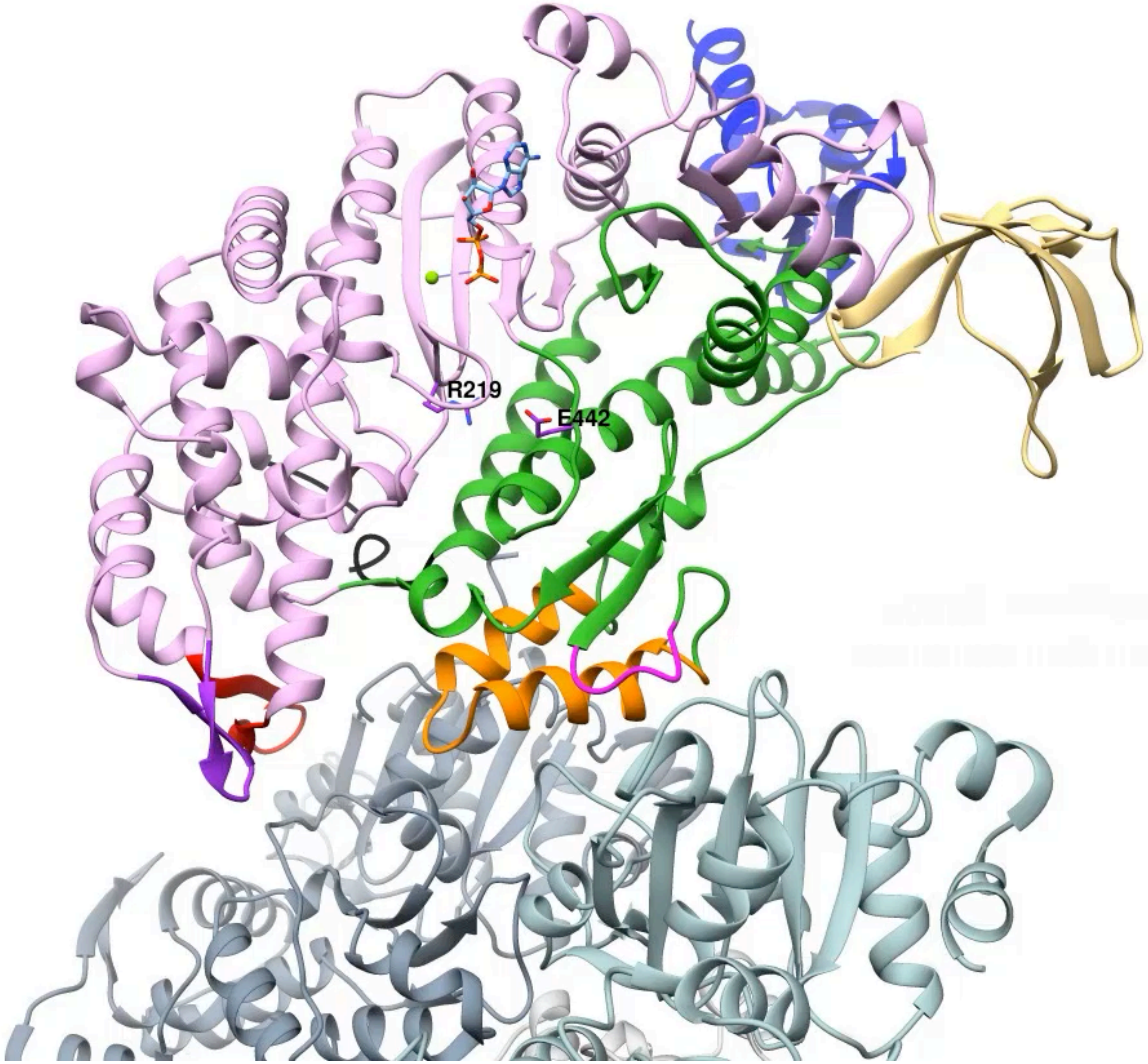


Klebl, et al. Nature 2025

# Protein that senses and respond to environment: molecular motor

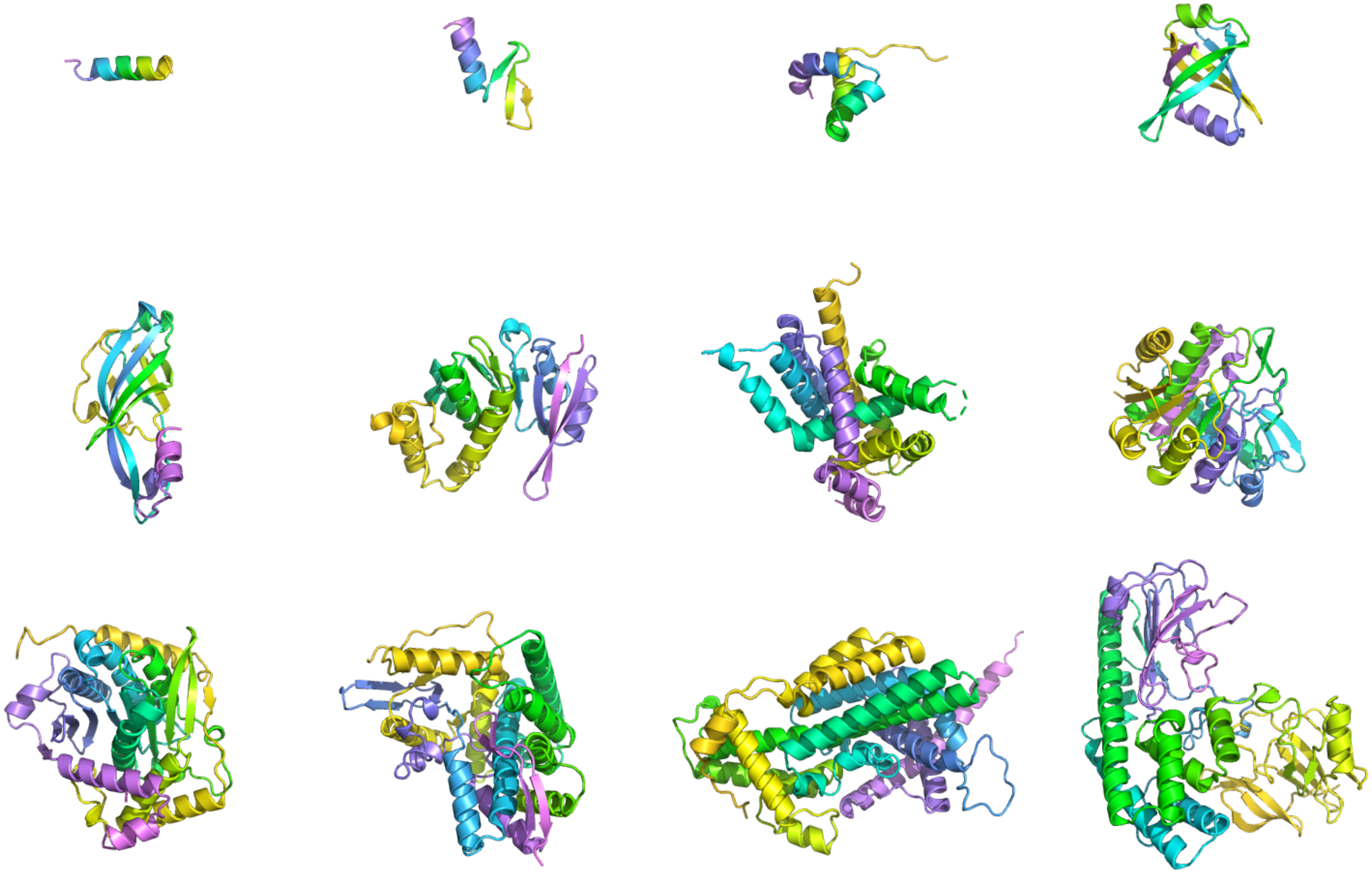


Productive powerstroke

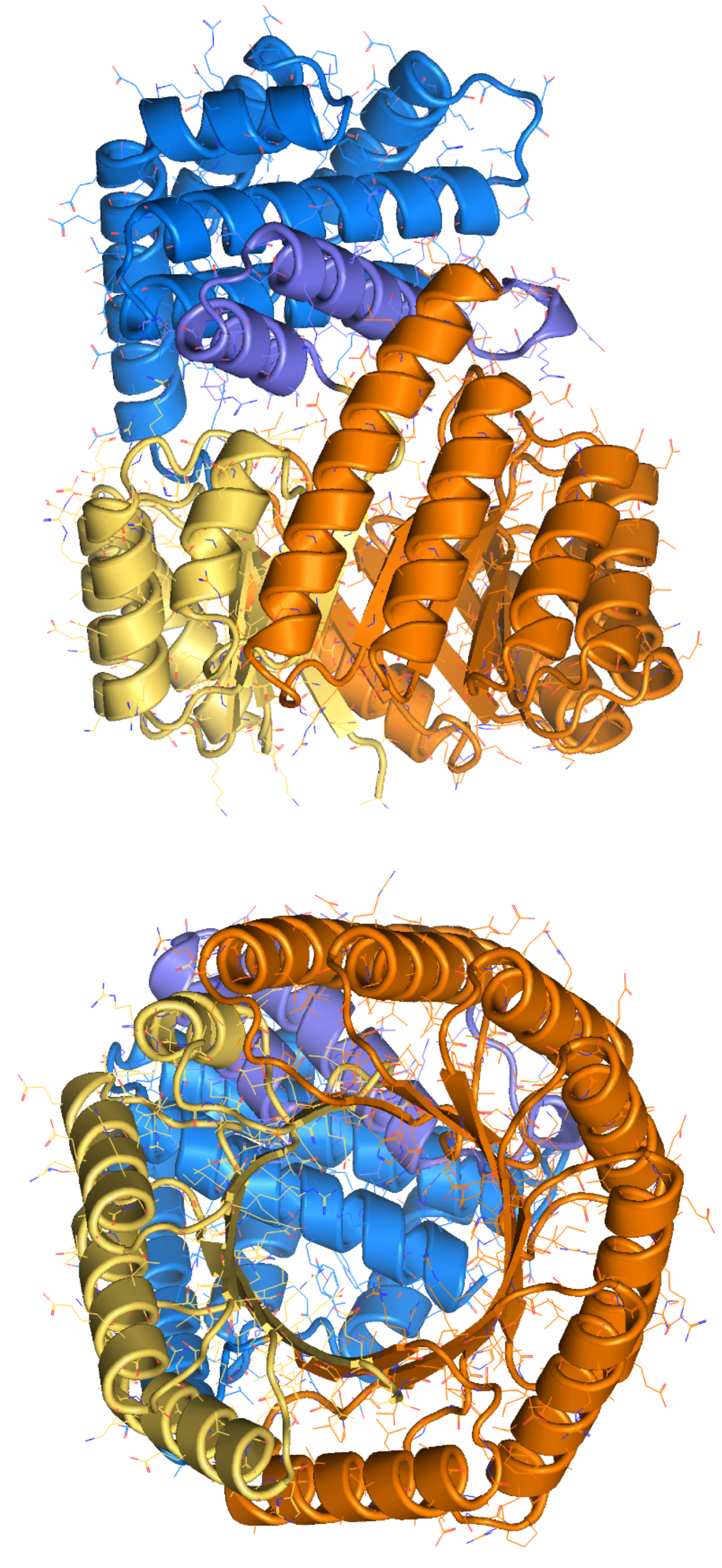


Klebl, et al. Nature 2025

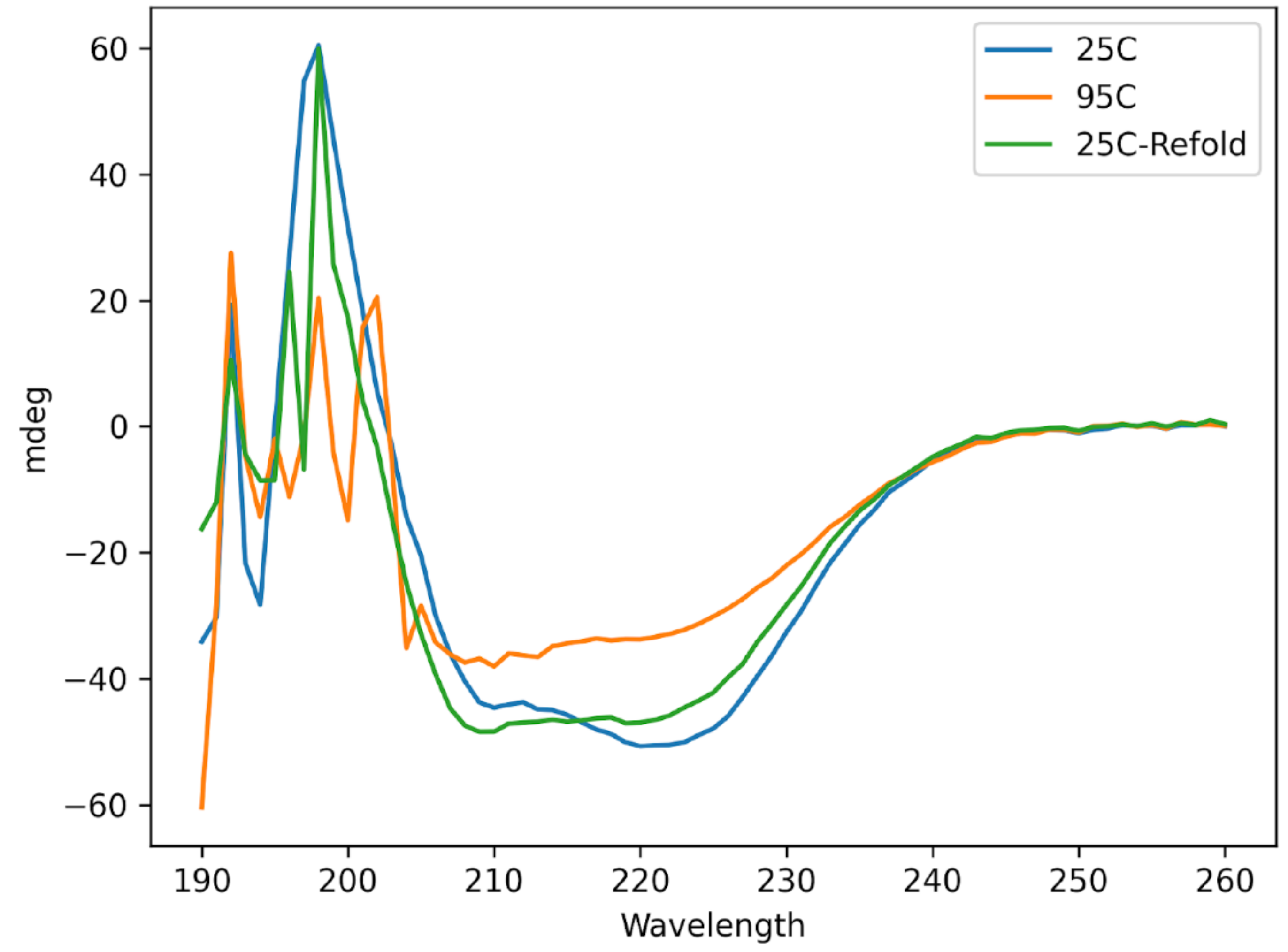
# Expressive structural samples from Protpardelle-1c

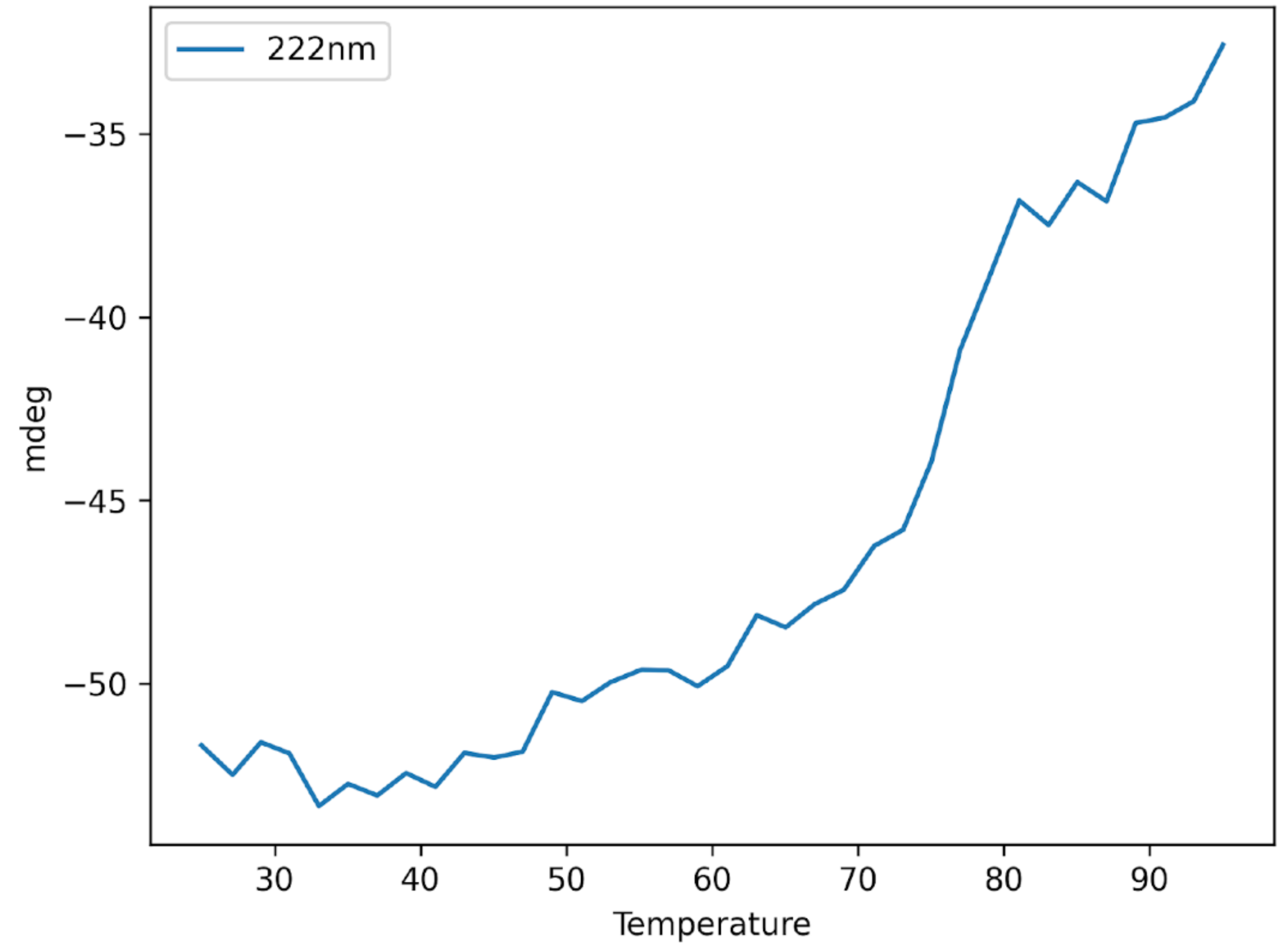


# De novo creation of a stably folded “multi-domain” barrel

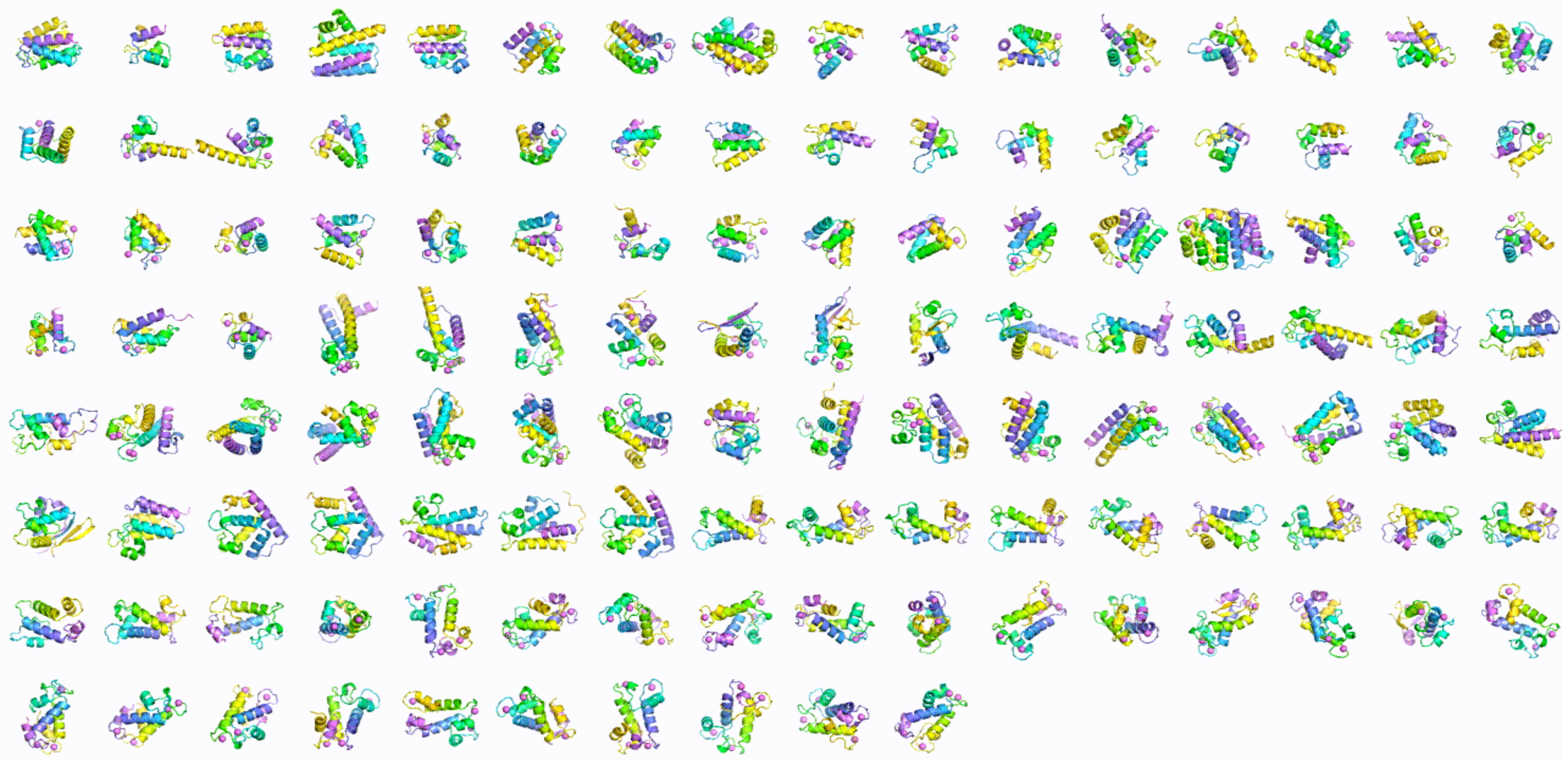


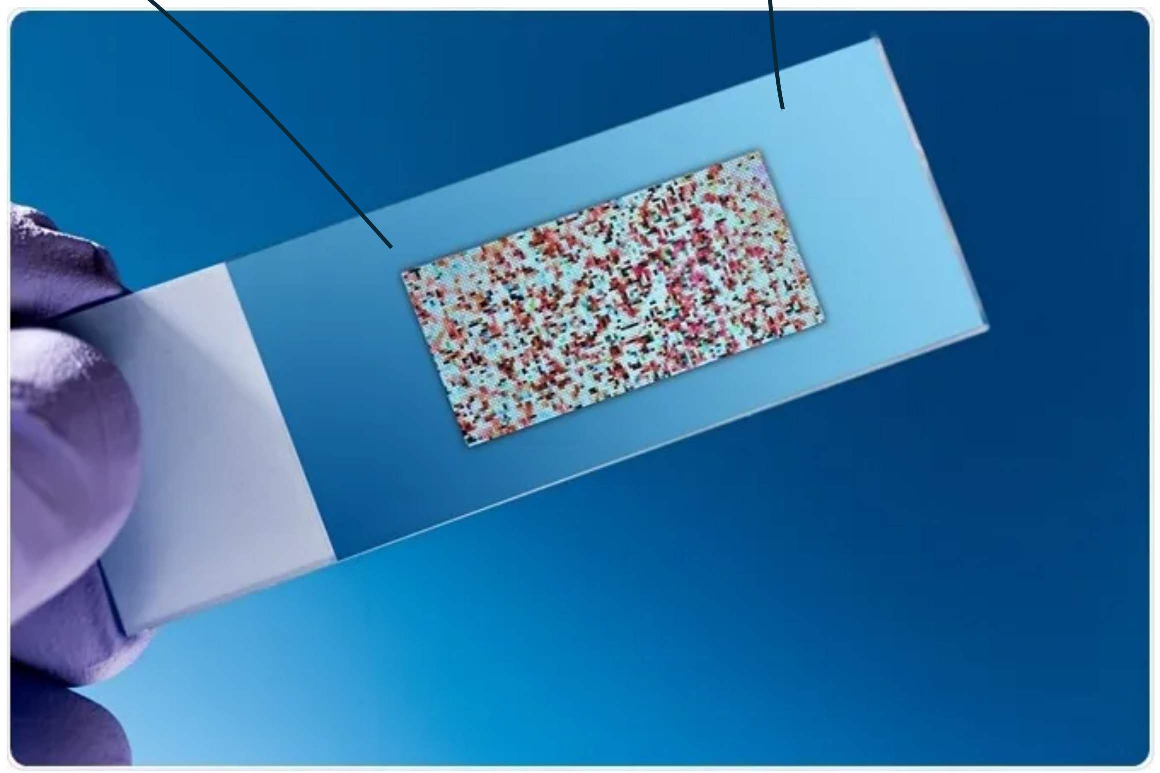
AF2 prediction





# Advancing beyond the state-of-the-art requires new high-throughput data generation system





Rapid feedback and iteration cycles to improve molecules and design model