

STATPHYS28
TOKYO UNIVERSITY, AUGUST 2023

DECIPHERING PROTEINS IN THEORY AND EXPERIMENT

E Weinreb, E Moses, Y Peleg, T Unger, S Albeck, JL Sussman (Weizmann)

J-P Eckmann, J Rougemont (Geneva)

G Zocchi (UCLA)

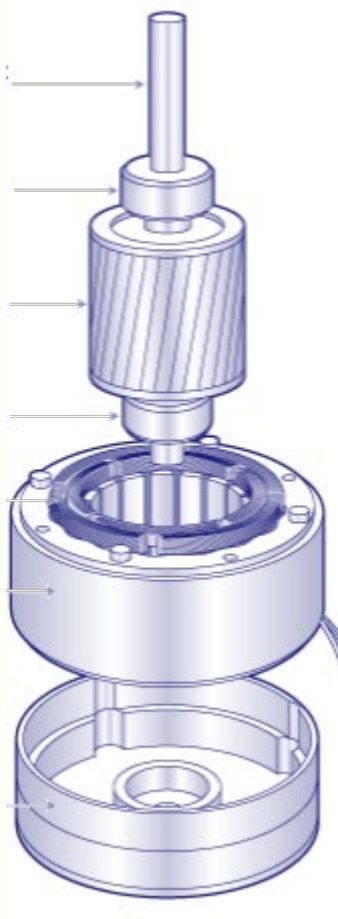
A Libchaber, M Mitchell, S Leibler (Rockefeller)

J McBride, S Dutta, M Siek, BA Grzybowski, TT (IBS, UNIST)



WE ARE RECRUITING
AT ALL LEVELS

What is the logic of protein machines?



Rotor of electric motor
aliexpresss.com

BLUEPRINT



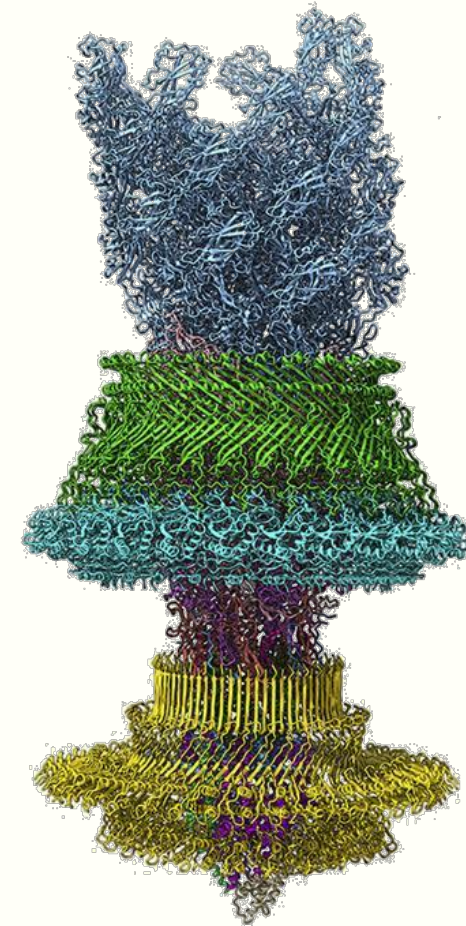
STRUCTURE



DYNAMICS



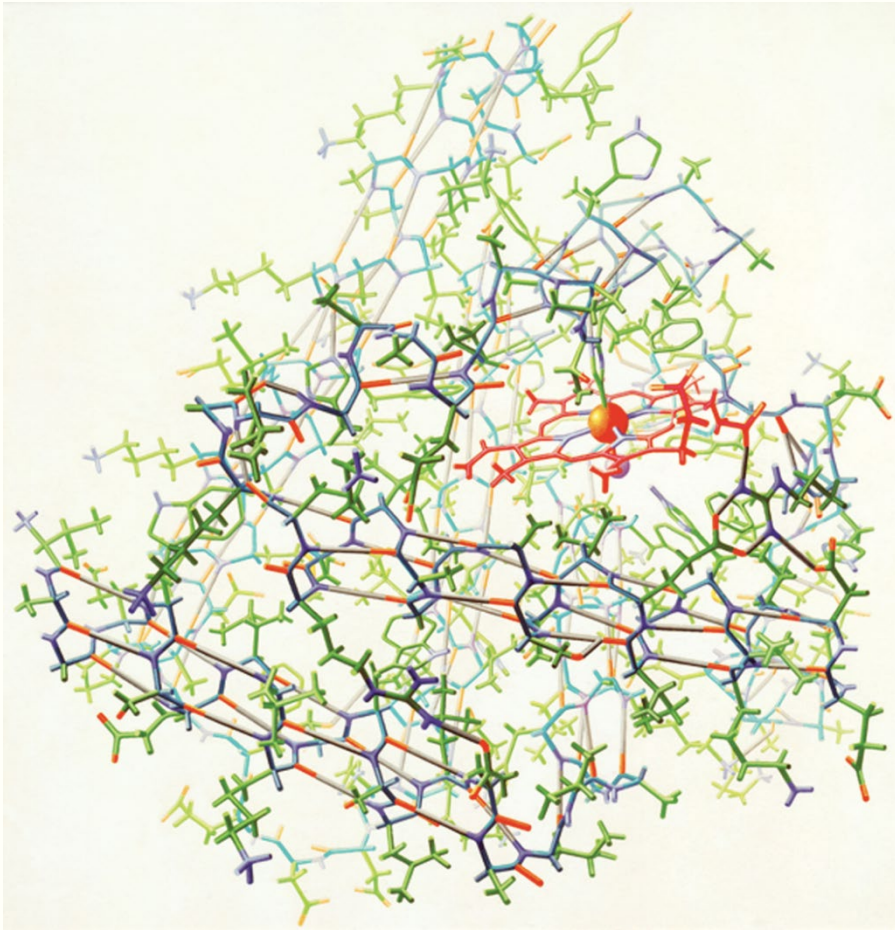
FUNCTION



flagellar motor:

P. Ball, Chem World 2021

However, despite the explosion of data and recent breakthroughs in understanding folding (AI),
fundamental questions about the logic of protein machines remain open

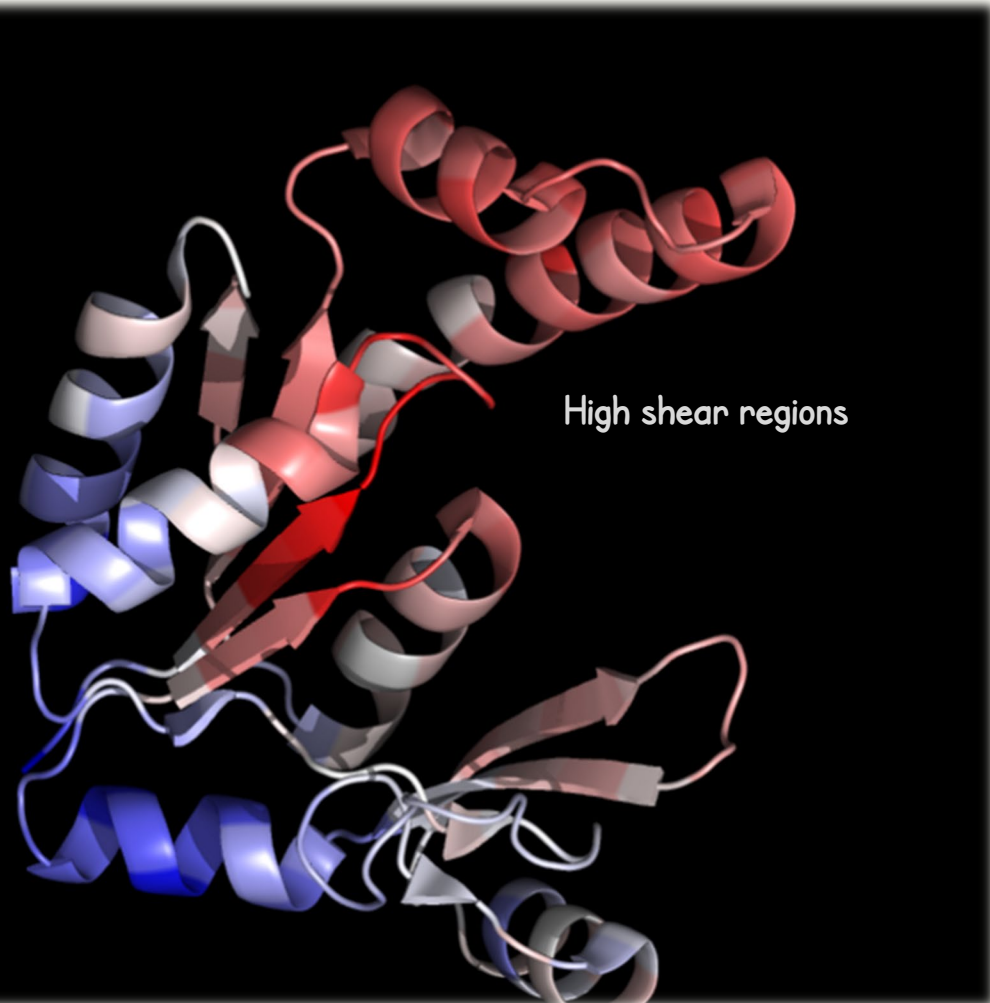


Kendrew (1961) – 3D structure of myoglobin (art by I. Geis)

information-rich matter (evolutionary memory)
made of **heterogeneous** building blocks (20 aa)
that interact **collectively**
and function within a **multiscale** living system

How the collective interactions that determine
protein function are encoded in the gene?

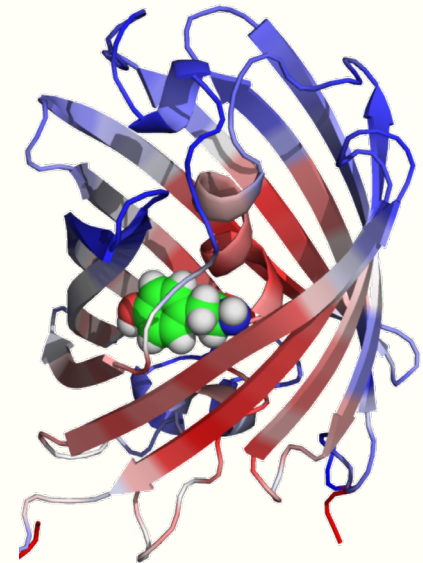
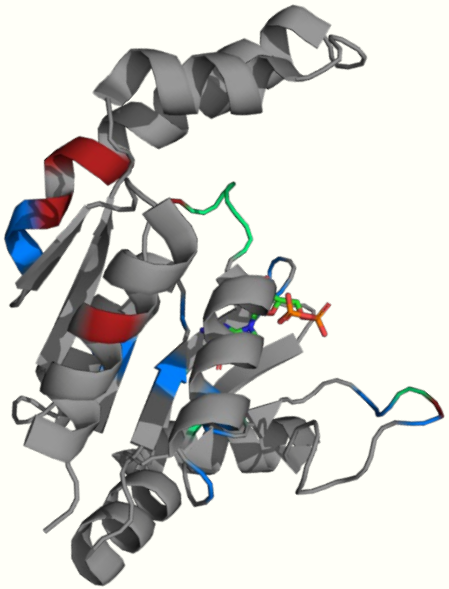
Proteins as sequence-encoded viscoelastic machines



- Proteins evolve a network of flexible viscoelastic elements that facilitate high-shear deformation essential for their function (e.g., allostery).
- Existence of these high-shear regions and their effect on function can be predicted and tested against experiments.

Proteins are viscoelastic machines with biochemical functions

1. **Theory** of protein as **viscoelastic** matter predicts the evolution of functionally-relevant collective motion.
2. **Experimental test – guanylate kinase:** long-range high-shear effects measured and tested vs. predictions.
3. **Generalizing with AI (AlphaFold)** by predicting the effect of **single mutations** on structure and phenotype.

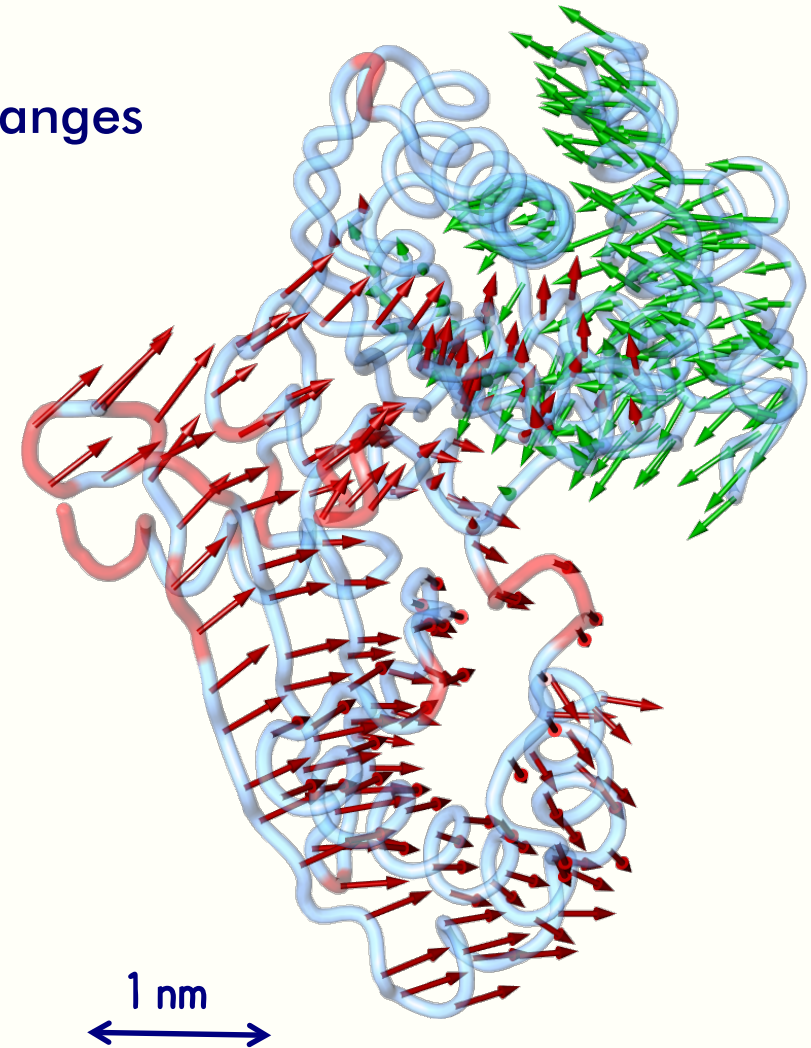
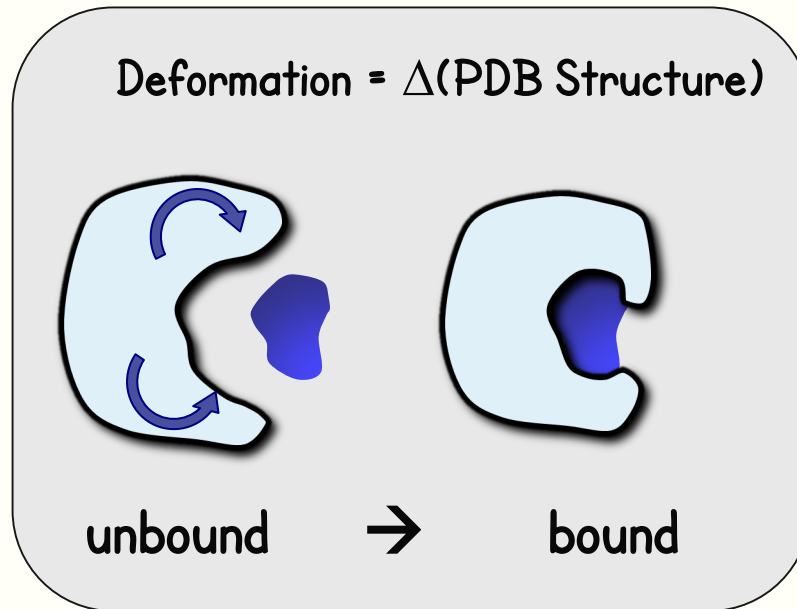


McBride et al. (2023) *AlphaFold2 can predict structural and phenotypic effects of single mutations* (in review)

Weinreb et al. (2023) *High-shear regions in an enzyme impact viscoelastic mechanics and activity* (in review)

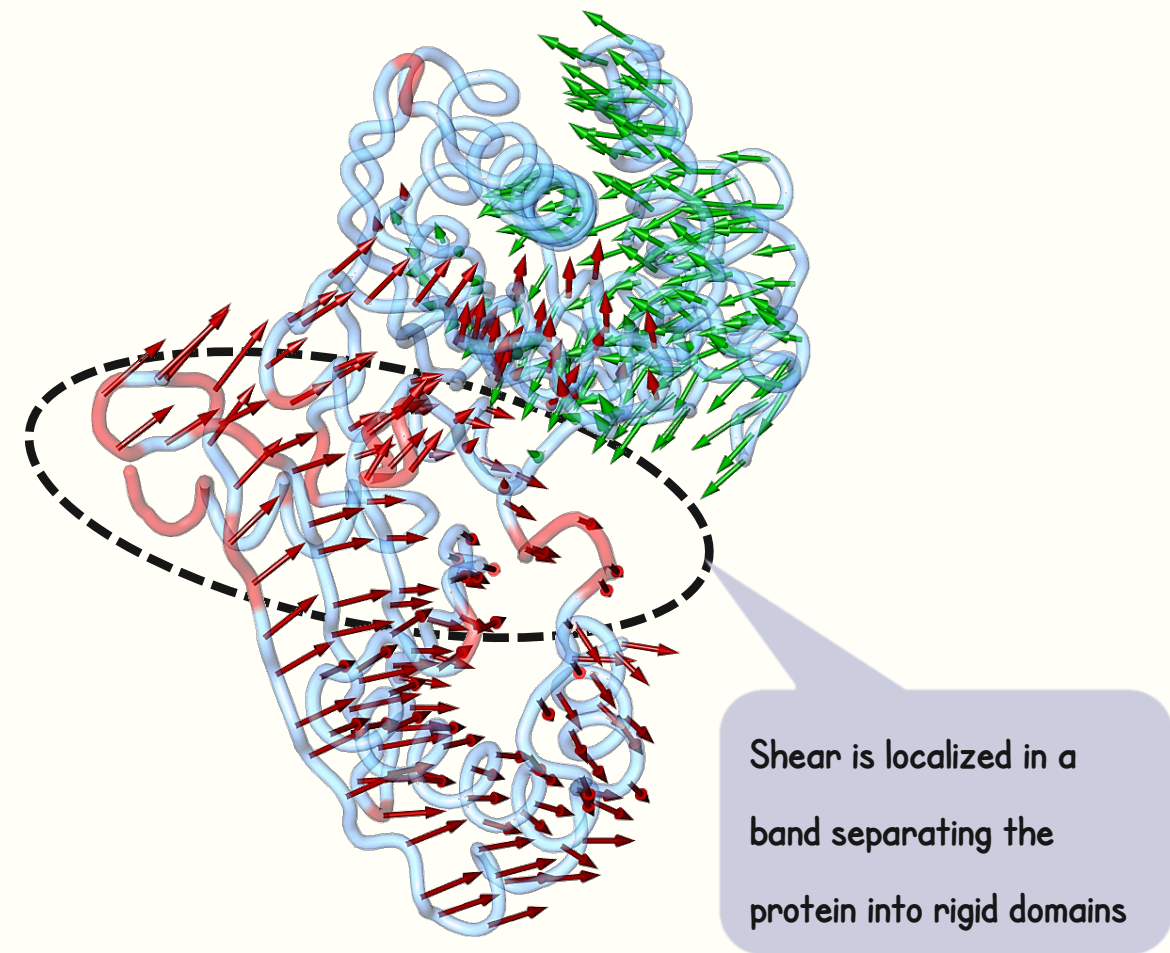
1. THEORETICAL APPROACH:

Looking at **motion** and **deformations** identifies structural changes relevant to function and how they are encoded in the gene



Glucokinase
(PDB: 1v4s → 1v4t)

Shear measures the internal motion of amino acids in the protein



SHEAR = how much amino acids move w.r.t. to neighbors

$$\text{affine shear} = \varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \sum_k \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} \right)$$

Geometric meaning: **metric** of deformation $\varepsilon = \frac{1}{2}(\mathbf{g} - \mathbf{1})$

Discrete version by tracing motion of neighboring atoms

Glucokinase (PDB: 1v4s → 1v4t)

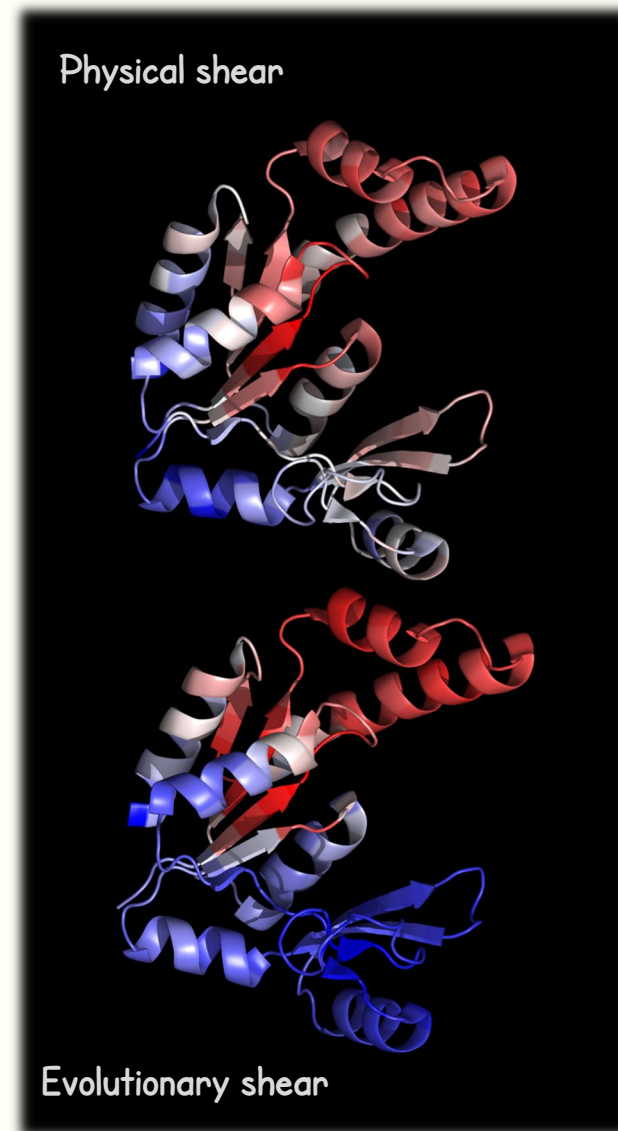
Localized rearrangement (= shear) can be induced by physical forces or mutation

PHYSICAL FORCES

Deformation during binding
(induced fit, allostery...)

EVOLUTION

Mutations induce structural changes.

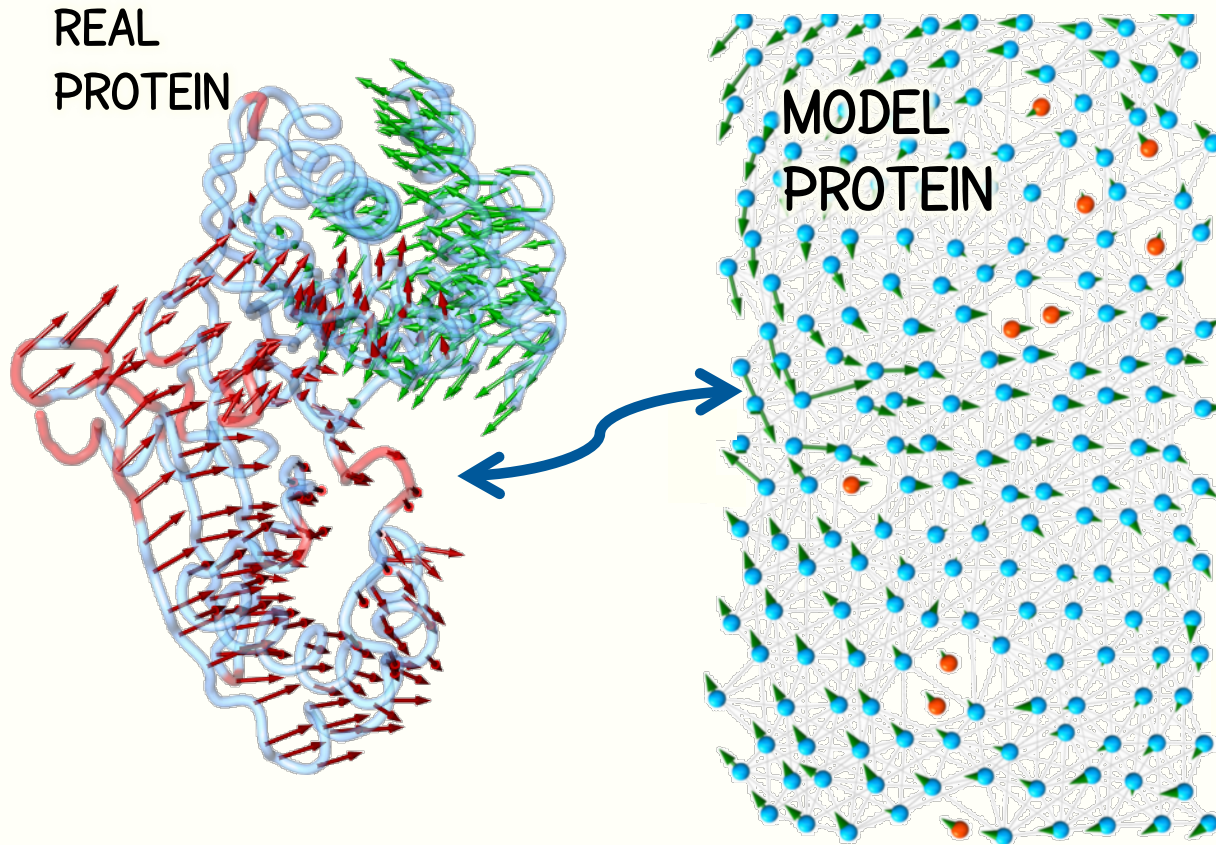


structure(**bound**) – structure(**unbound**)
→ **physical Shear**

structure(**WT**) – structure(**mutant**)
→ “**Evolutionary Shear**”

Physical theory of protein as evolving matter (Green functions)

links collective physical interactions among amino acids and genetic correlations



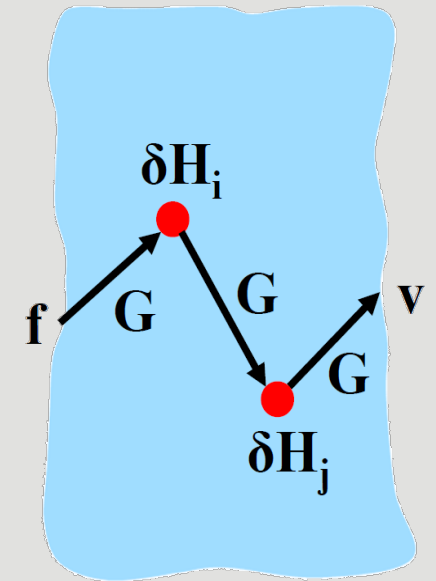
- "gene" = 0111010100...

Two types of amino acids
connected by "springs"

amino acids	H	P
H	strong	weak
P	weak	weak



GREEN FUNCTION FORMALISM

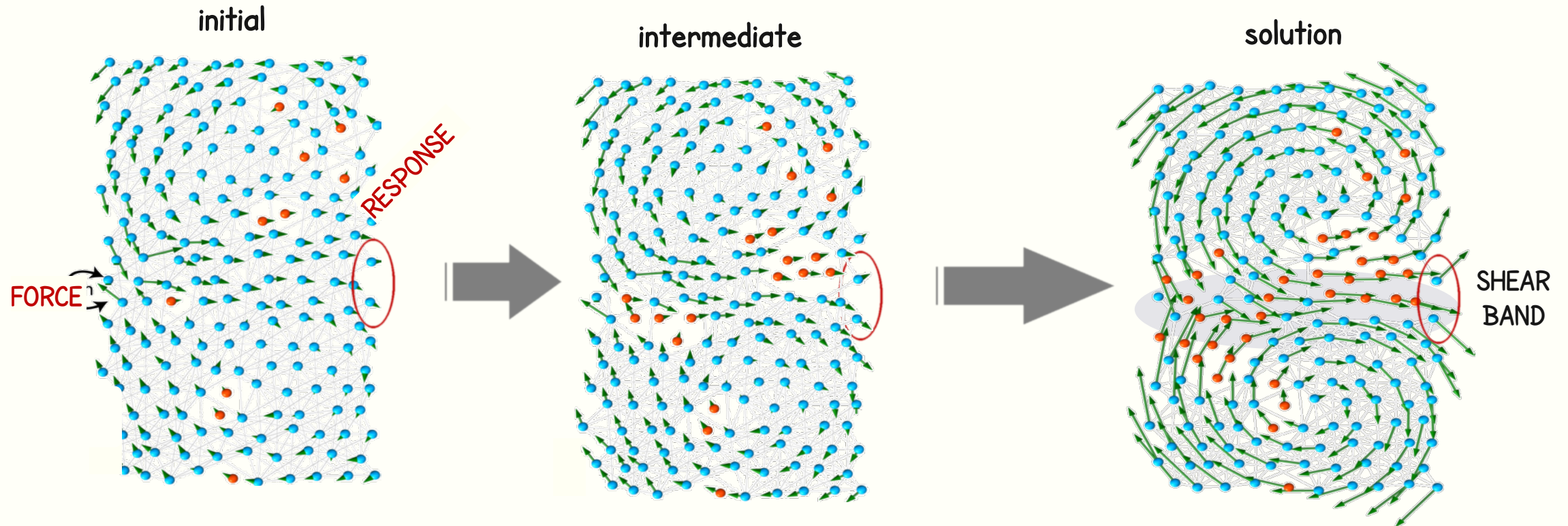


Mutations = local perturbations

Interactions $\rightarrow$ Dyson series,
multiple scattering...

$$\delta G = -G \delta H G + G \delta H G \delta H G - \dots$$

Green function theory explains how functional proteins with shear bands emerge during evolution

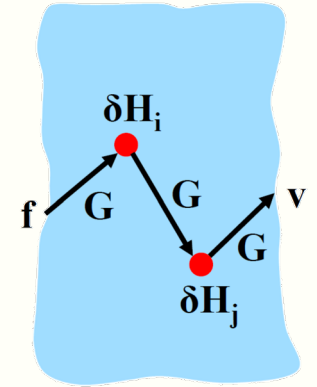


High-shear regions facilitate **versatile, sophisticated functions**: specific recognition, induced fit, allostery...

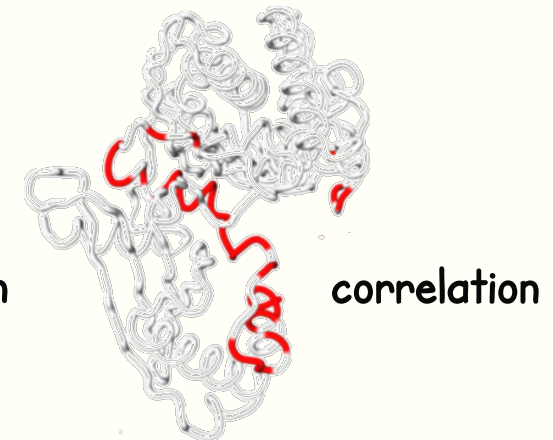
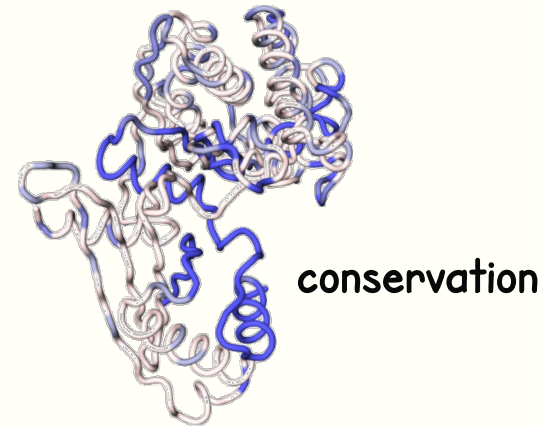
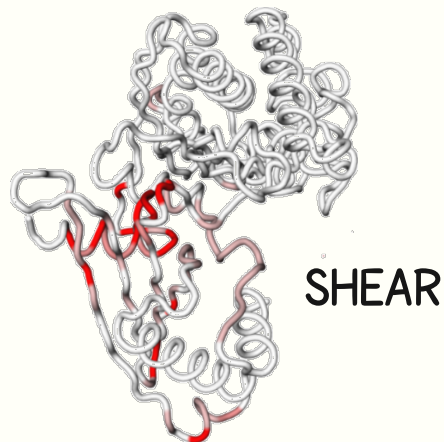
Summary: Theory predicts the evolution of high-shear regions in the protein

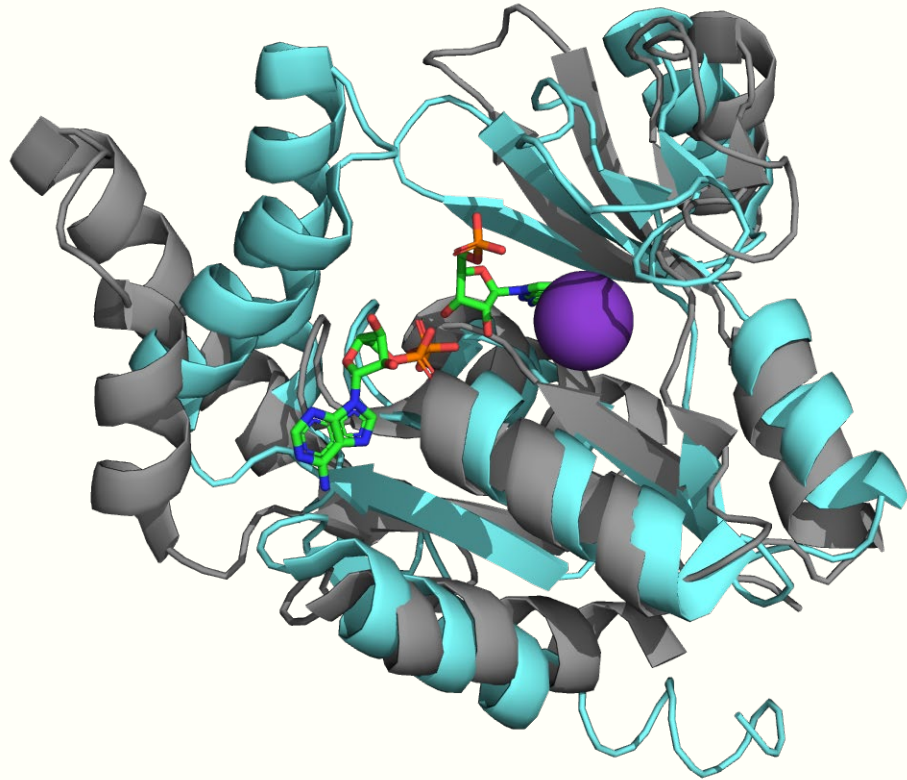
High-shear regions evolve that are:

- essential for **biochemical function**
- through **large-scale motion** and long-range mechanical interactions,
- and therefore, genetically **conserved** and **correlated**.



Relevant amino acids can be **predicted** by computing the physical shear during binding or the "evolutionary shear" induced by mutations.





2. Measurements of Guanylate Kinase mutants link dynamics – function – structure

We tested predictions by introducing **mutations** and, in parallel,

Computing: (A) **Physical shear** - bound\unbound

(B) **Evolutionary shear** - deformation AlphaFold

Measuring: (C) **Biochemical function** - enzymatic kinetics

(D) **Viscoelastic dynamics** - nano-rheology

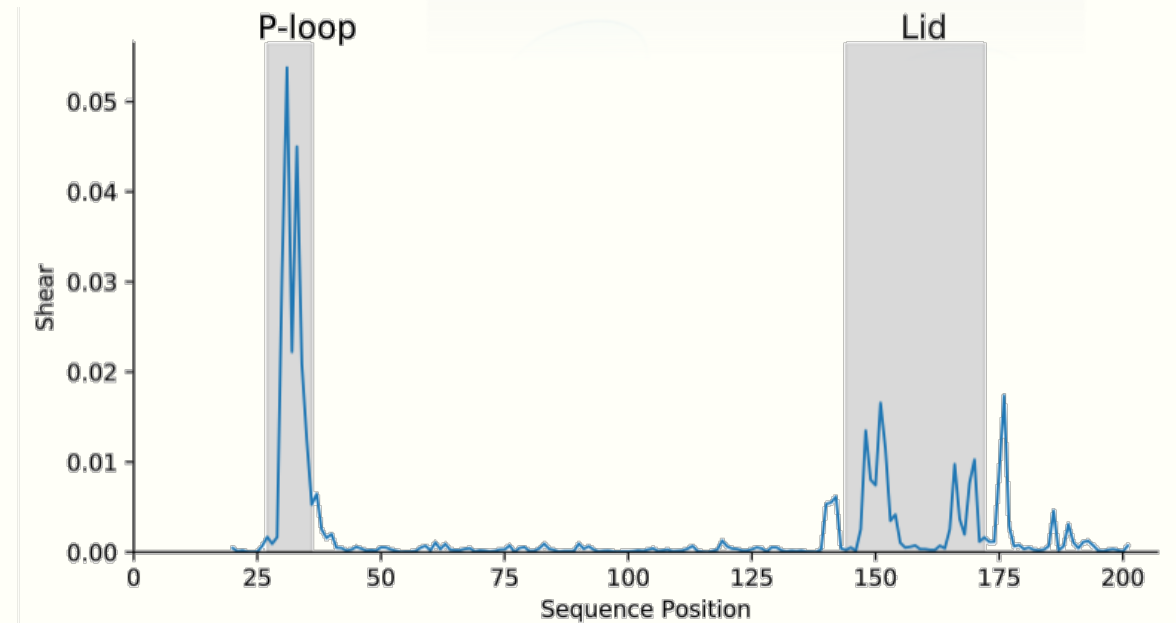
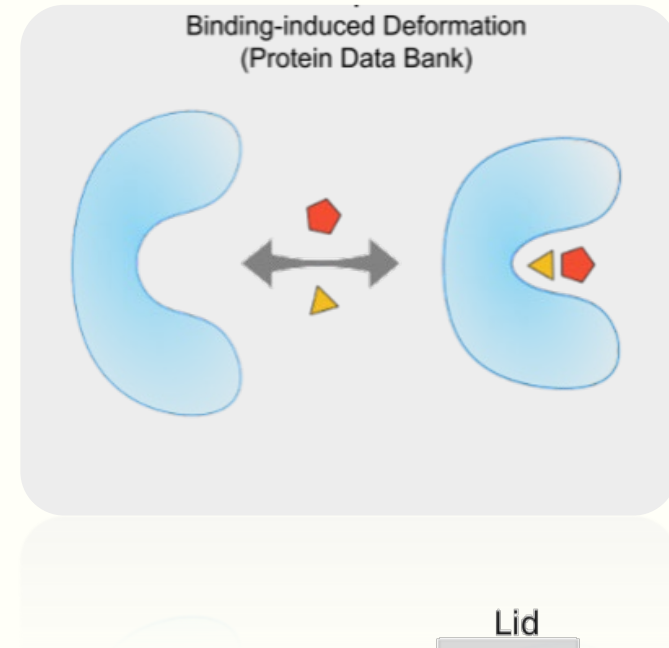
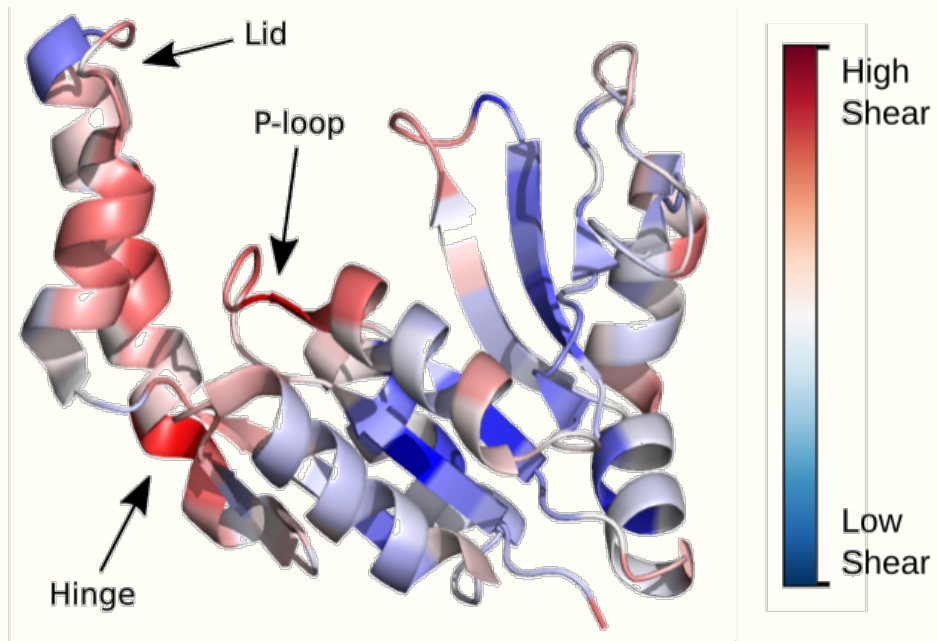
High-shear regions in guanylate kinase impact viscoelastic mechanics and enzymatic function

Eyal Weinreb, John M. McBride, Marta Siek, Jacques Rougemont, Renaud Renault, Yoav Peleg, Tamar Unger, Shira Albeck, Joel L. Sussman, Bartosz A. Grzybowski, Giovanni Zocchi, Jean-Pierre Eckmann, Elisha Moses, TT

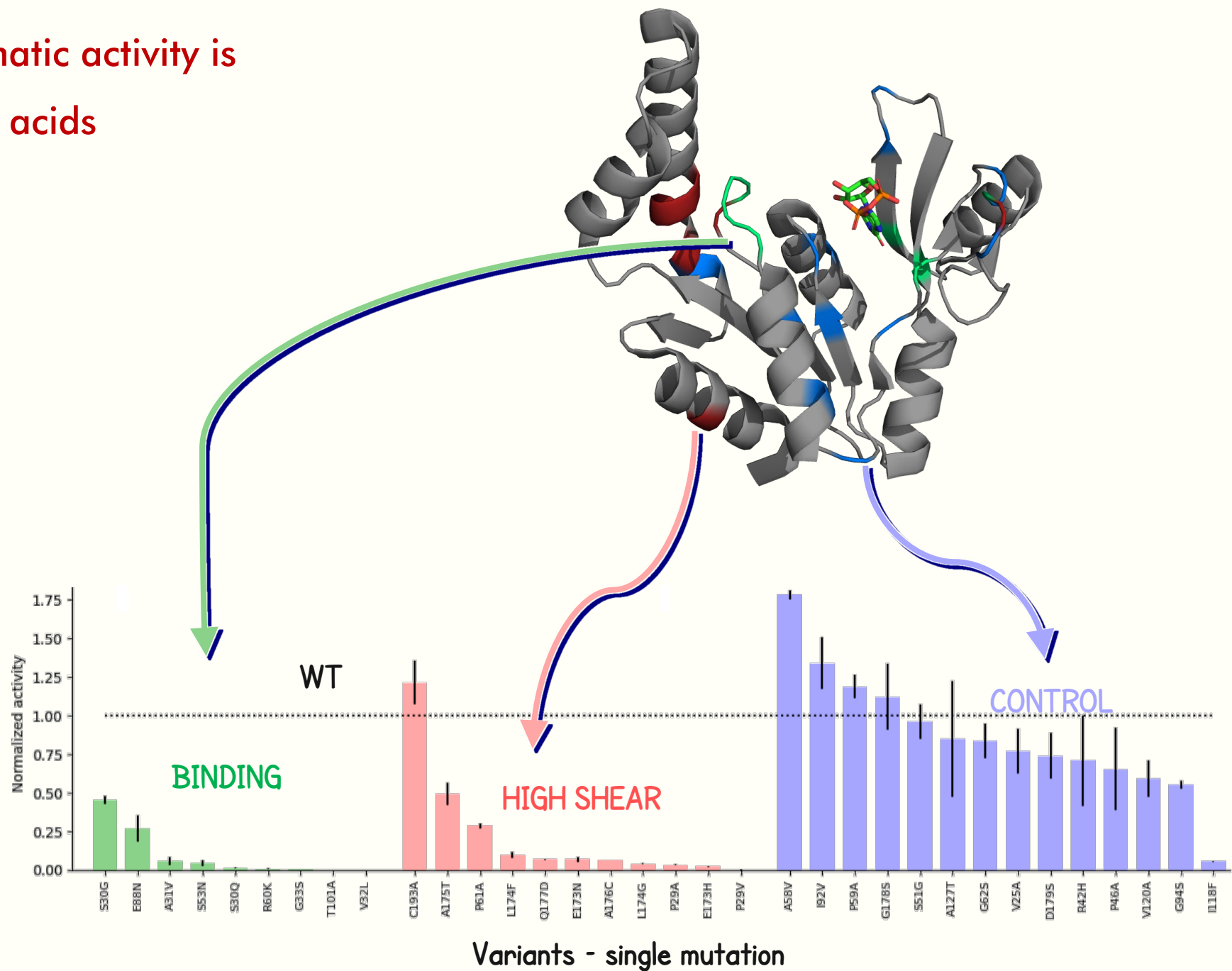
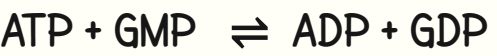
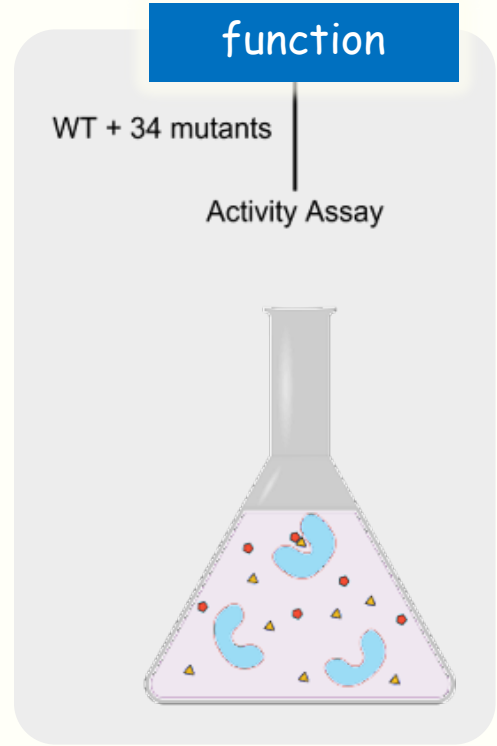
(in review)

We calculate the shear and introduce 34 mutations in three regions:

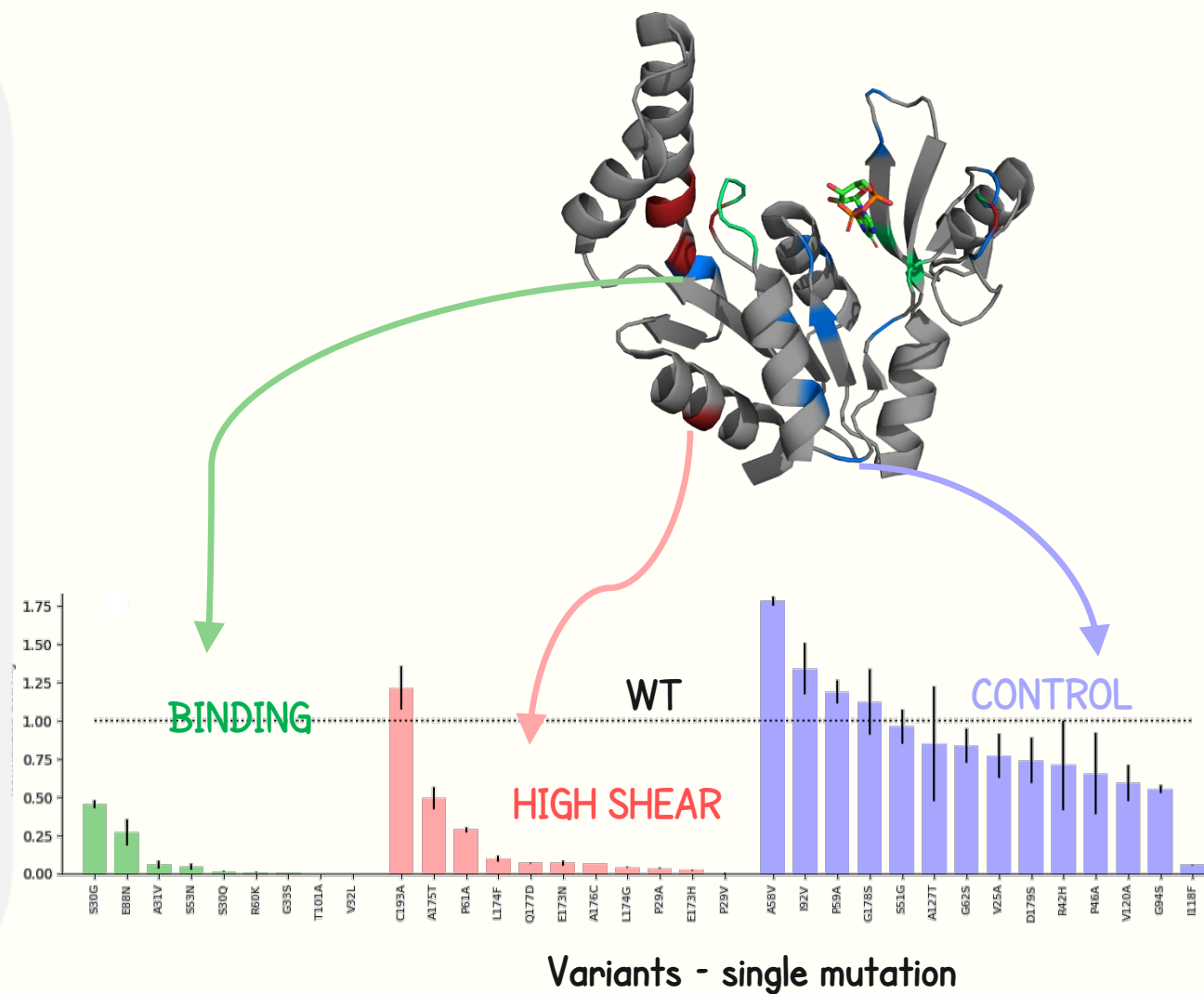
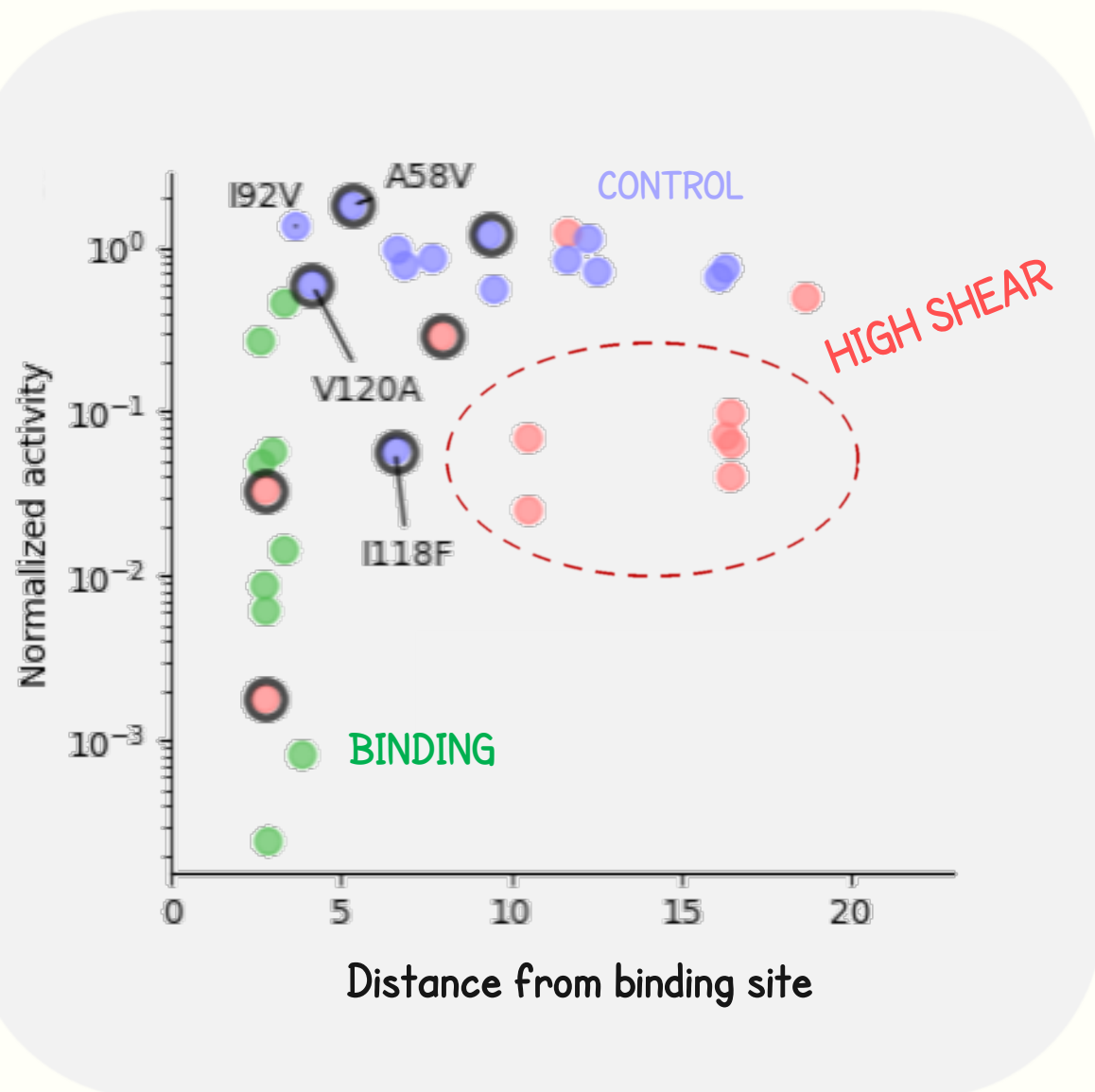
- (i) high-shear
- (ii) binding site
- (iii) control



Effect of mutation on enzymatic activity is strong at high-shear amino acids

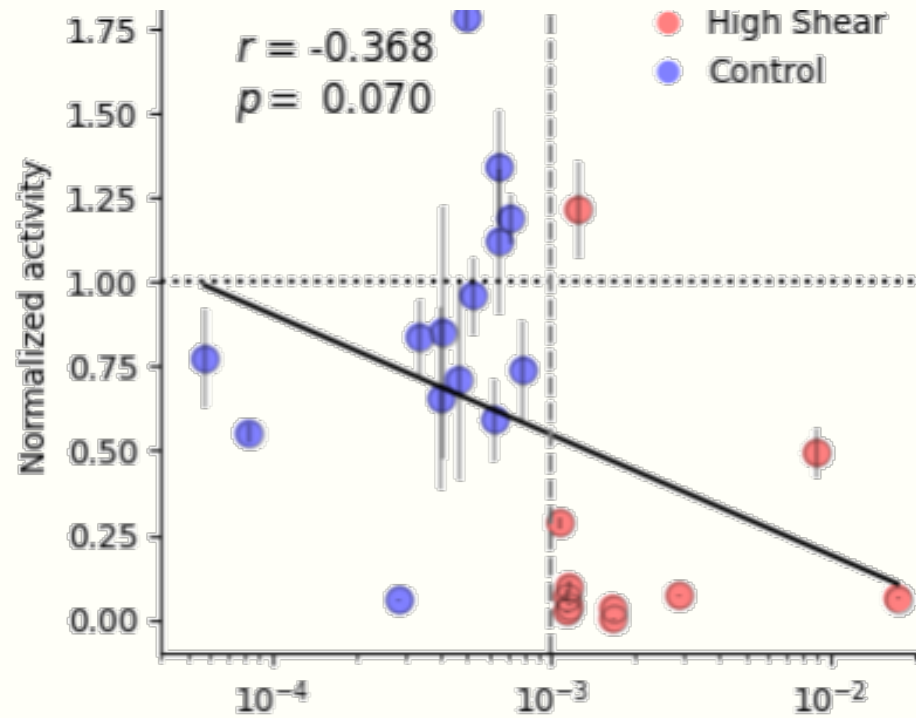


High-shear mutation have significant effect even far from binding site

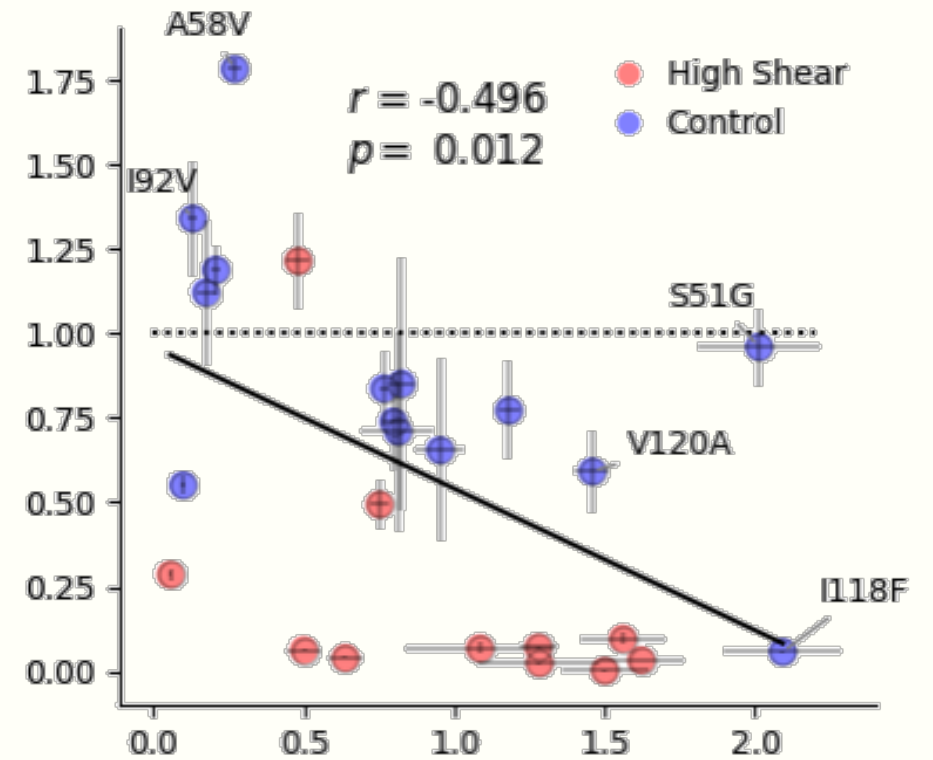


Mutations at high-shear regions have a strong effect on enzymatic function

- of both PDB (binding shear) and AlphaFold (mutation deformation)

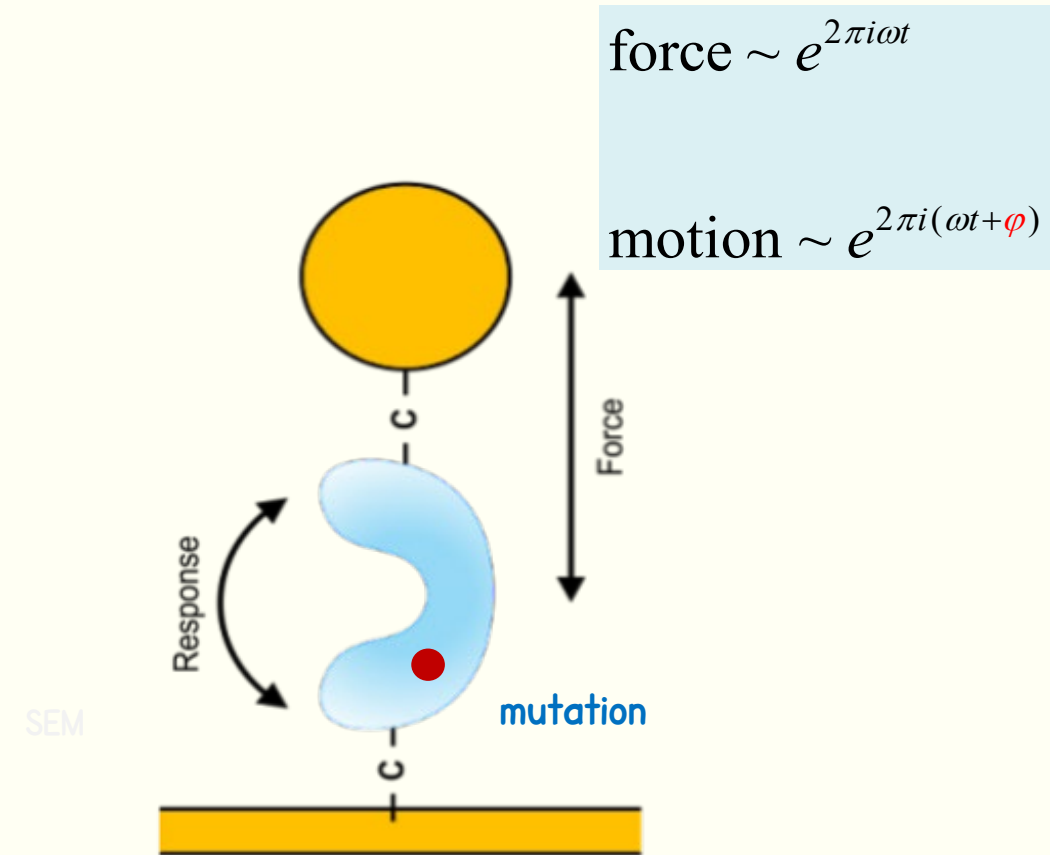
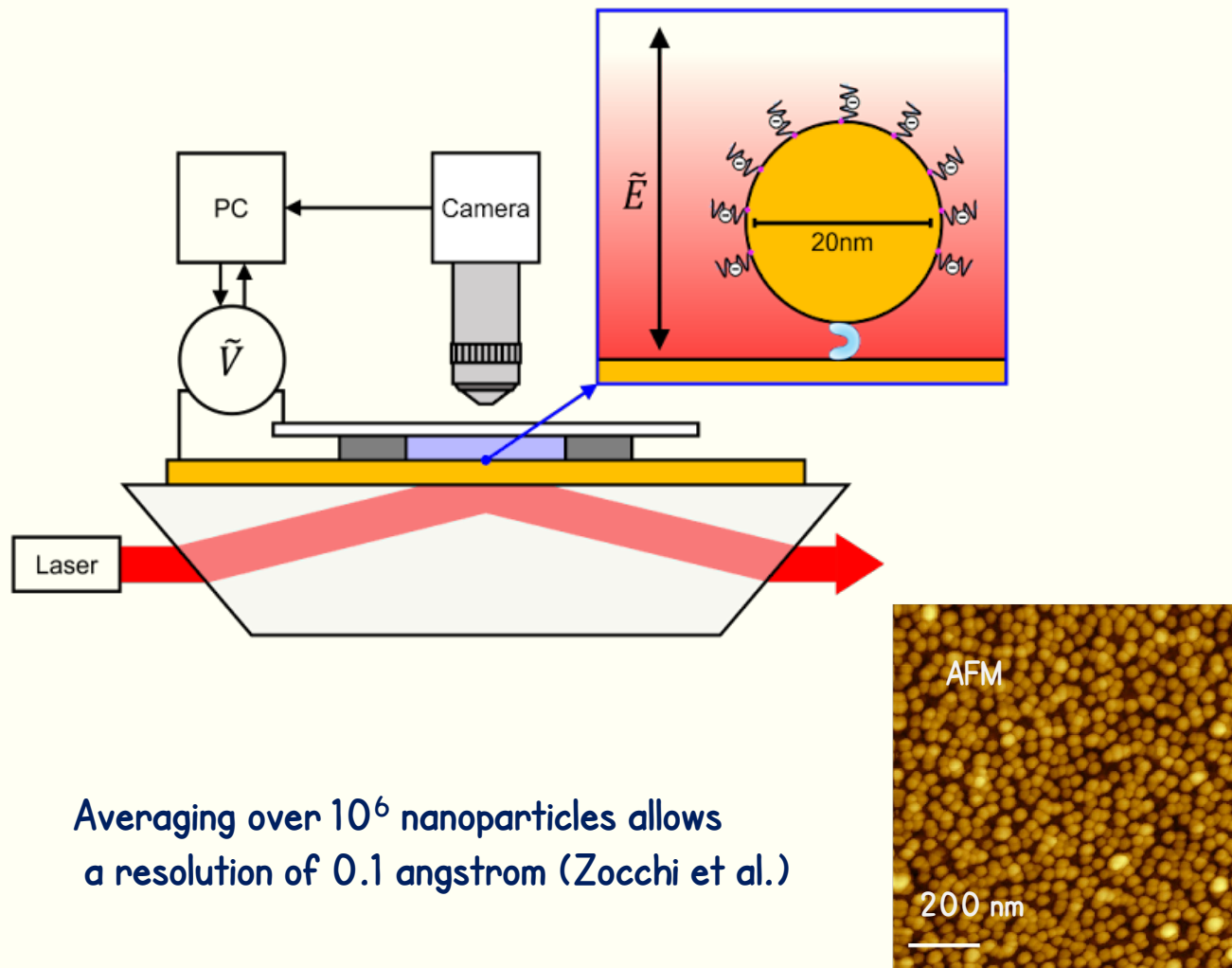


physical SHEAR (PDB) at mutated aa

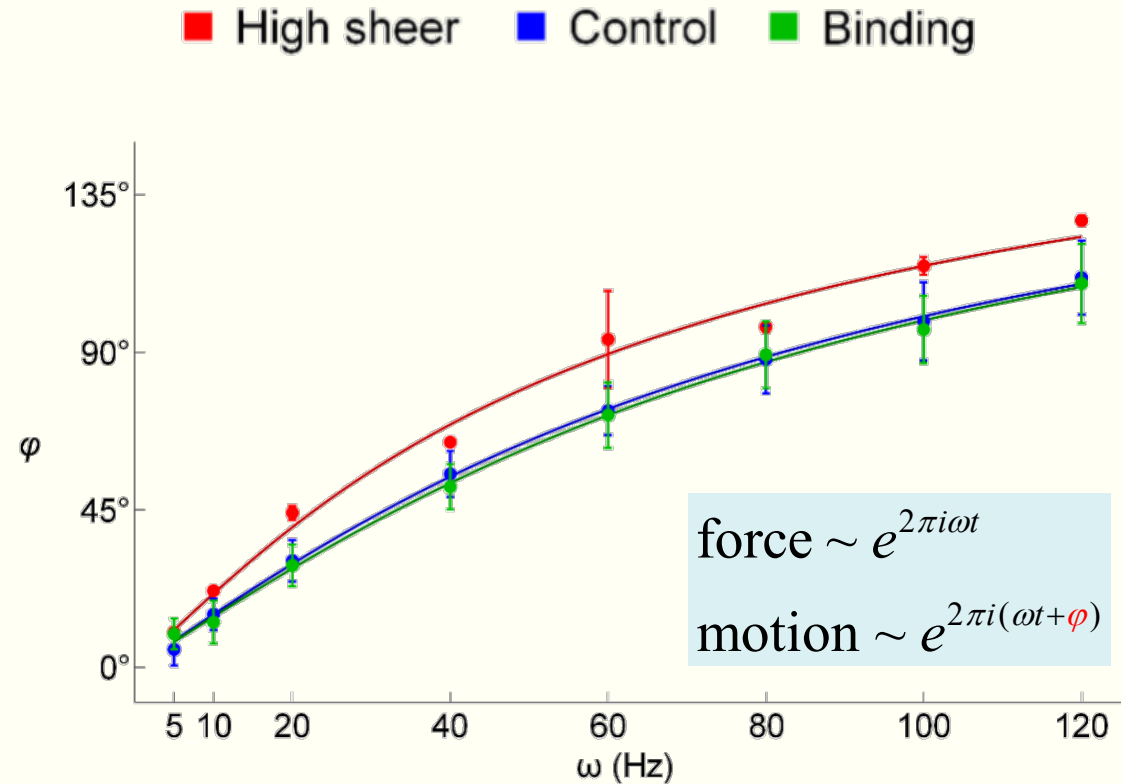
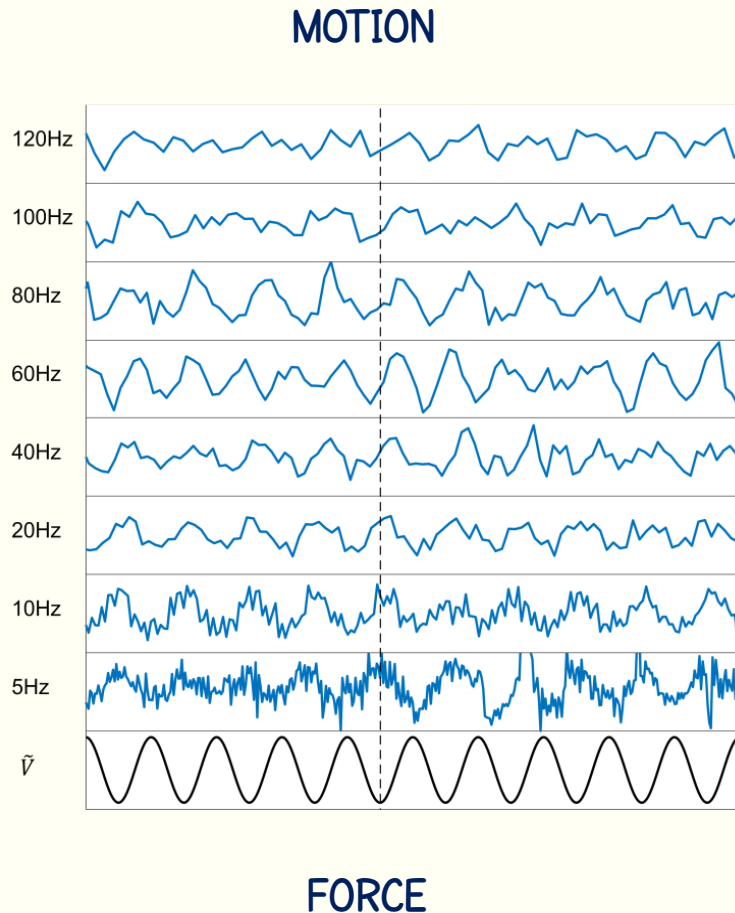


Evolutionary SHEAR at mutated aa

Nano-rheology of guanylate kinase exhibits viscoelastic dynamics in dependence of mechanical response on frequency

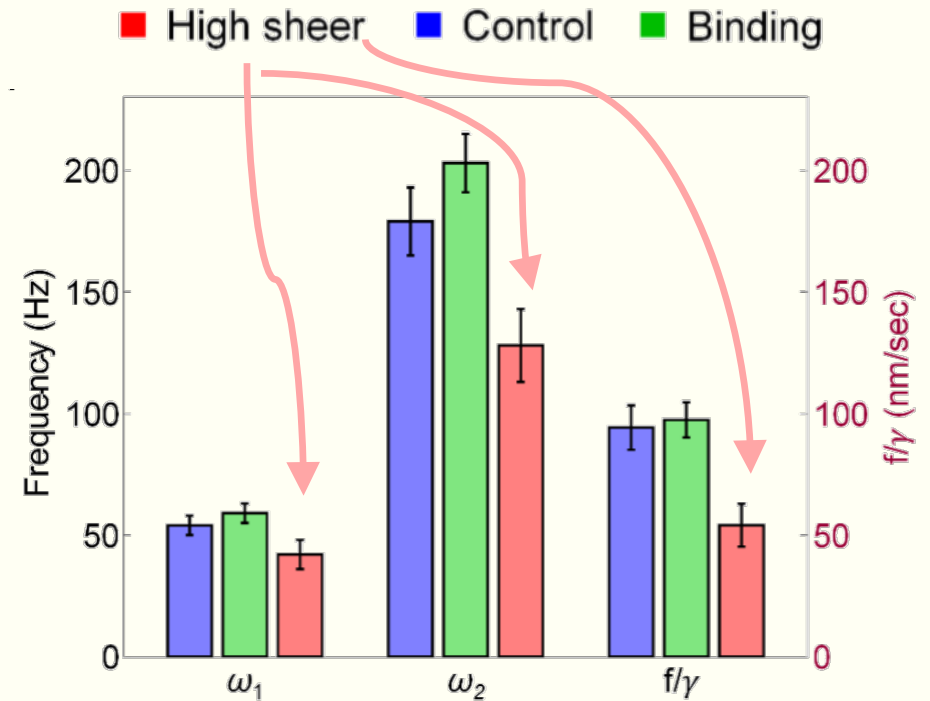
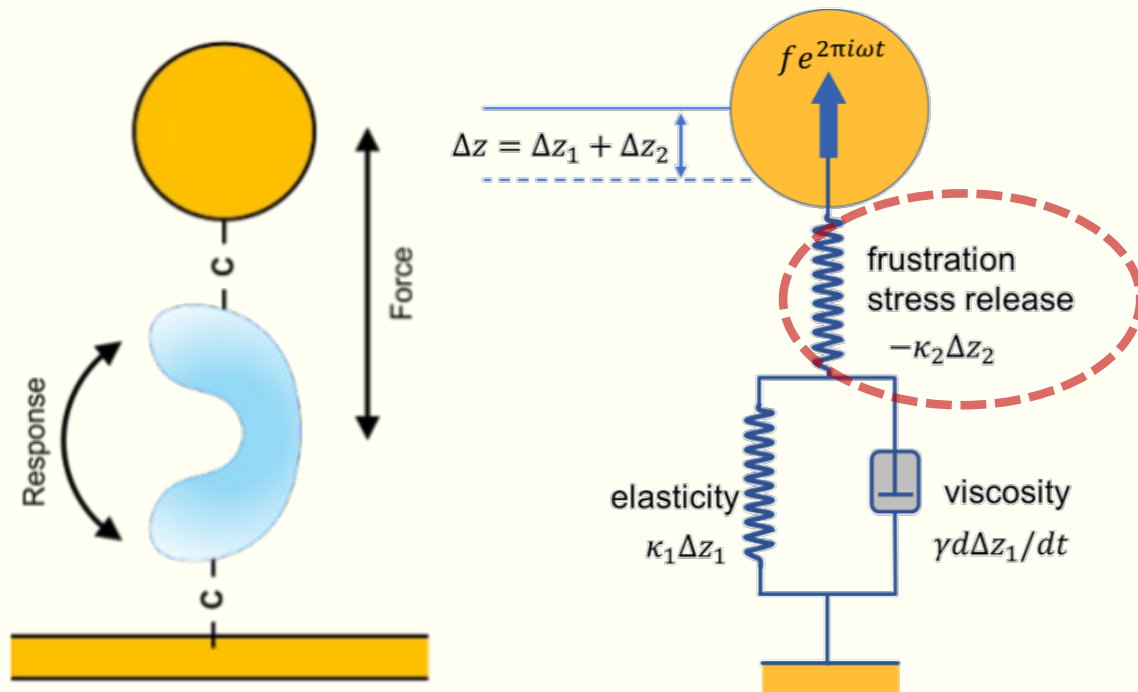


Nano-rheology exhibits phase lag between applied electric force and internal motion and its frequency dependence is different for high shear



HIGH-SHEAR MUTATIONS AFFECT THE VISCOELASTIC RESPONSE OF THE PROTEIN, MAKING IT STIFFER AND MORE VISCOUS

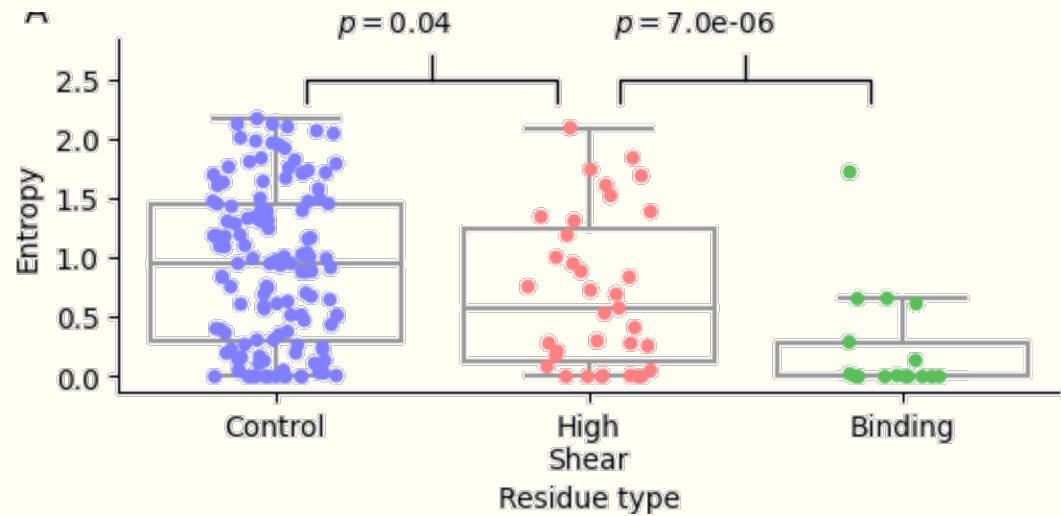
Protein viscoelasticity captures stress released from frustrated bonds



$$\Delta z(\omega) = \Delta z_1 + \Delta z_2 = \frac{f}{\gamma} \left(\frac{1}{\omega_1 + i\omega} - \frac{1}{\omega_2} \right)$$

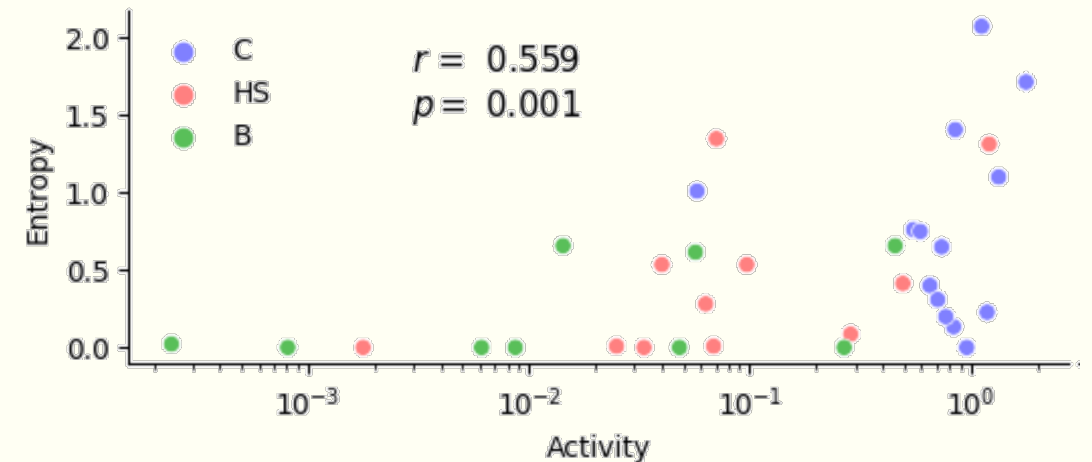
$\omega_1 = \frac{\kappa_1}{\gamma}$ viscoelastic relaxation frequency
 $\omega_2 = \frac{\kappa_2}{\gamma}$ stress release frequency (frustration)

entropy per residue from alignment (MSA) of 956 sequences



High-shear regions are conserved through evolution

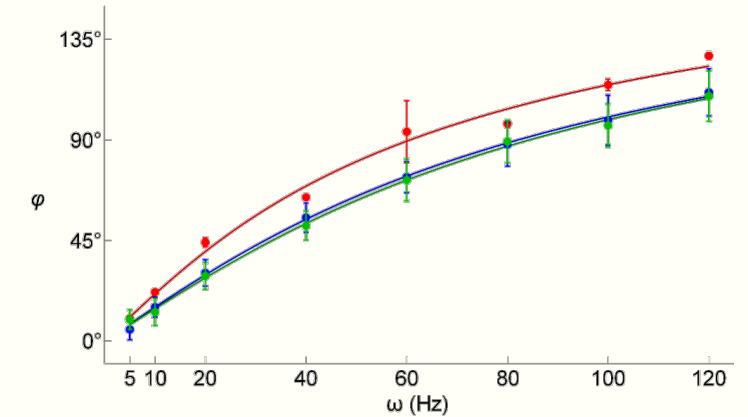
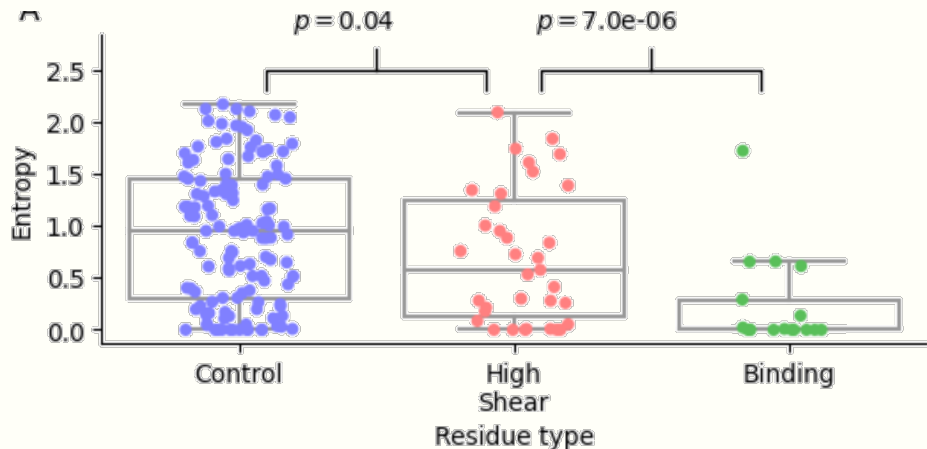
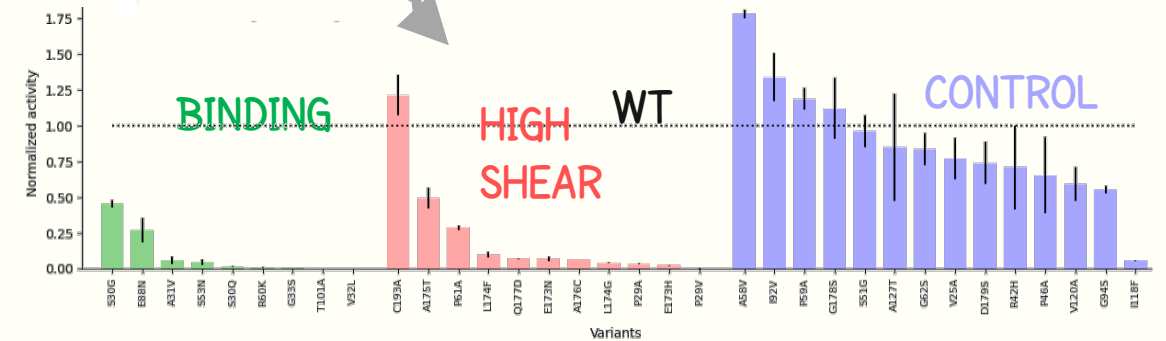
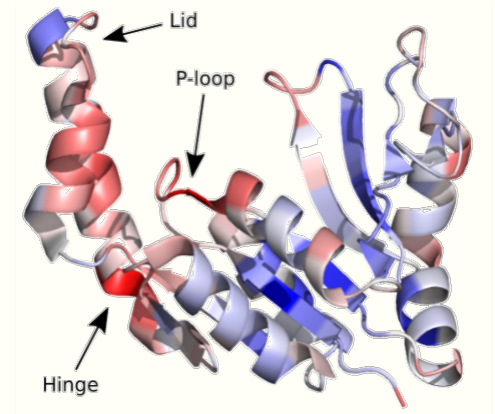
Sequence conservation correlates with effect on function



2. Summary: experimental test

Theory predicts high-shear regions from structural data and AI which are:

- essential for biochemical function
- through large-scale motion and viscoelastic interactions,
- and therefore, genetically conserved and correlated.



3. Using AI (Alphafold) to predict single mutation effects on structure and phenotype

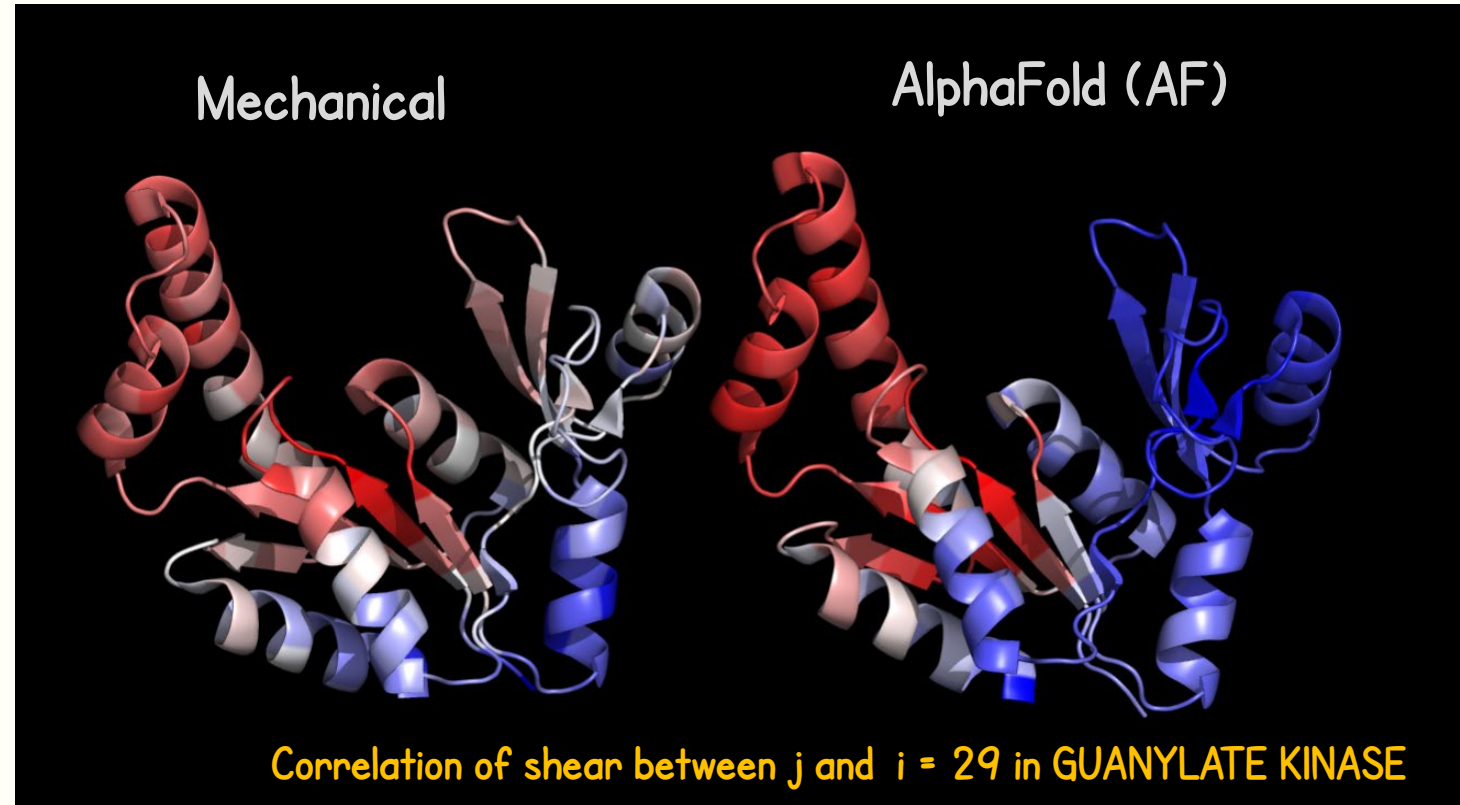
AlphaFold predicts protein structure at unprecedented accuracy



CAN WE GENERALIZE ALPHAFOLD'S PREDICTIONS TO OTHER PROTEINS?

- AF not validated or expected to predict single mutation effects.

TEST: using shear to predict mutation effect on **structure and phenotype** in a large data set of proteins



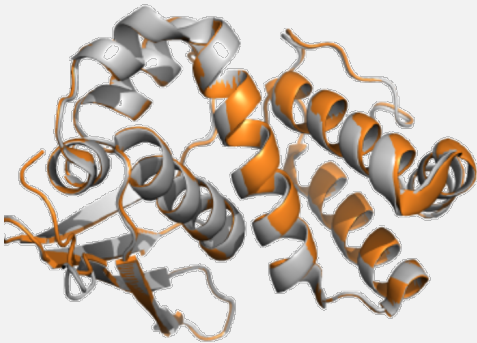
AlphaFold2 can predict single-mutation effects on structure and phenotype

John M. McBride, K Polev, A Abdirasulov, V Reinharz, BA Grzybowski, TT (in review)

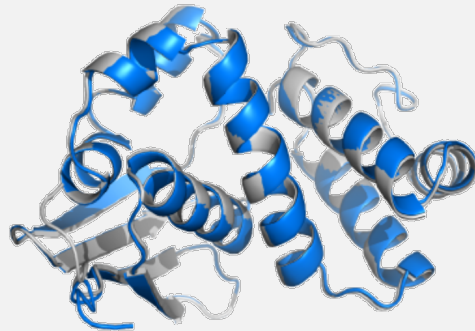
We compared shear from PDB and AF in $\sim 5,000$ pairs differing by 1-3 positions

Structural change

PDB



AF

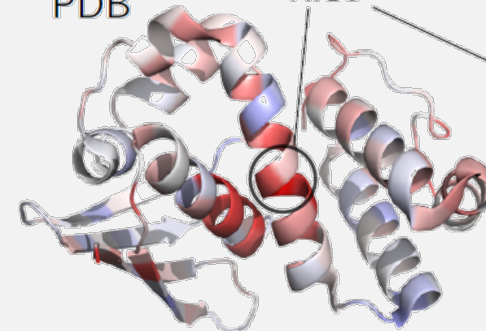


Example: H-NOX (Heme-Nitric oxide/Oxygen binding)

Mutation A71G

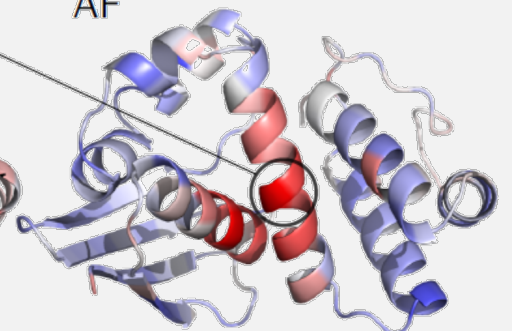
Shear

PDB



A71G

AF

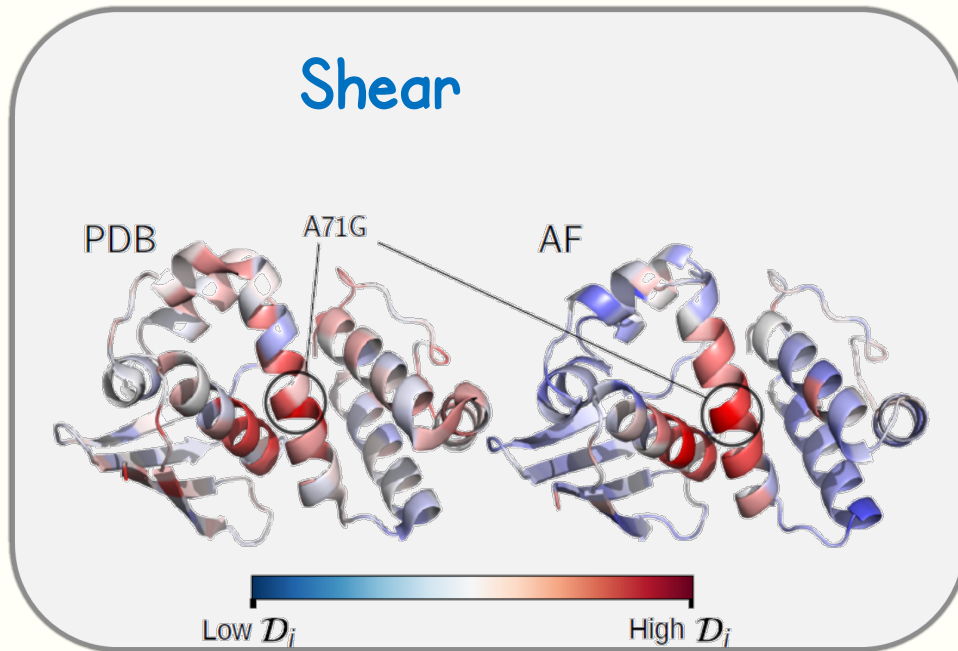


Low D_i

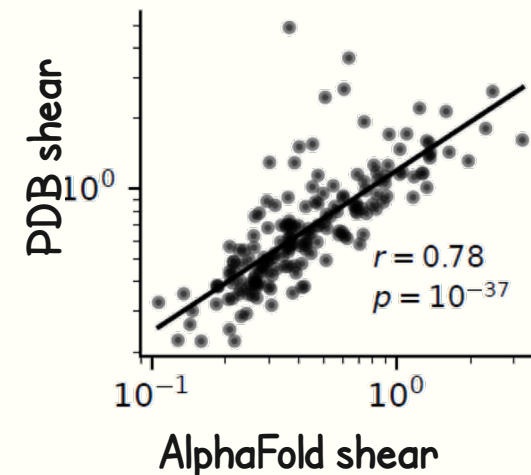
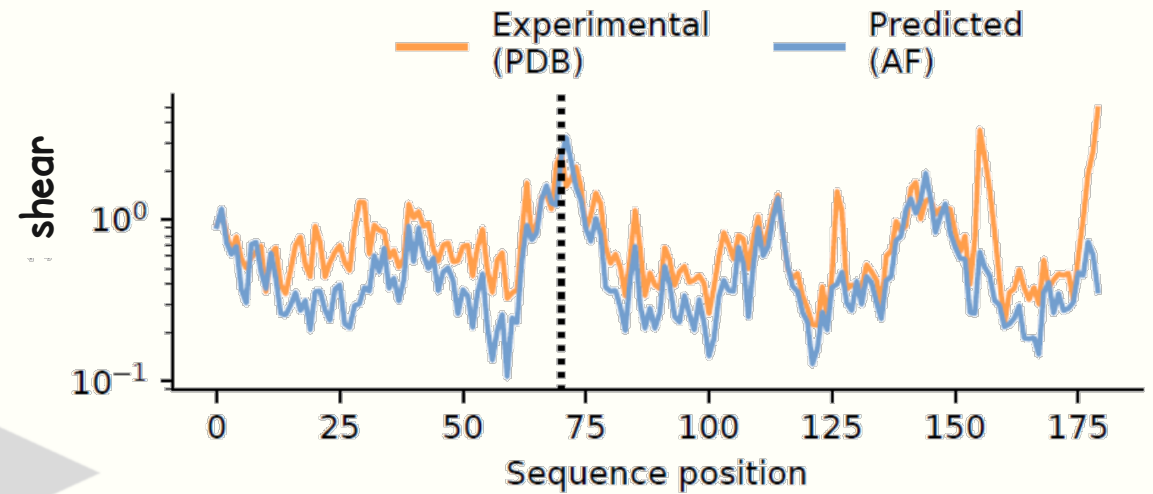
High D_i

* Same crystallographic groups in PDB

Shear predicted from AF correlates with shear computed from PDB



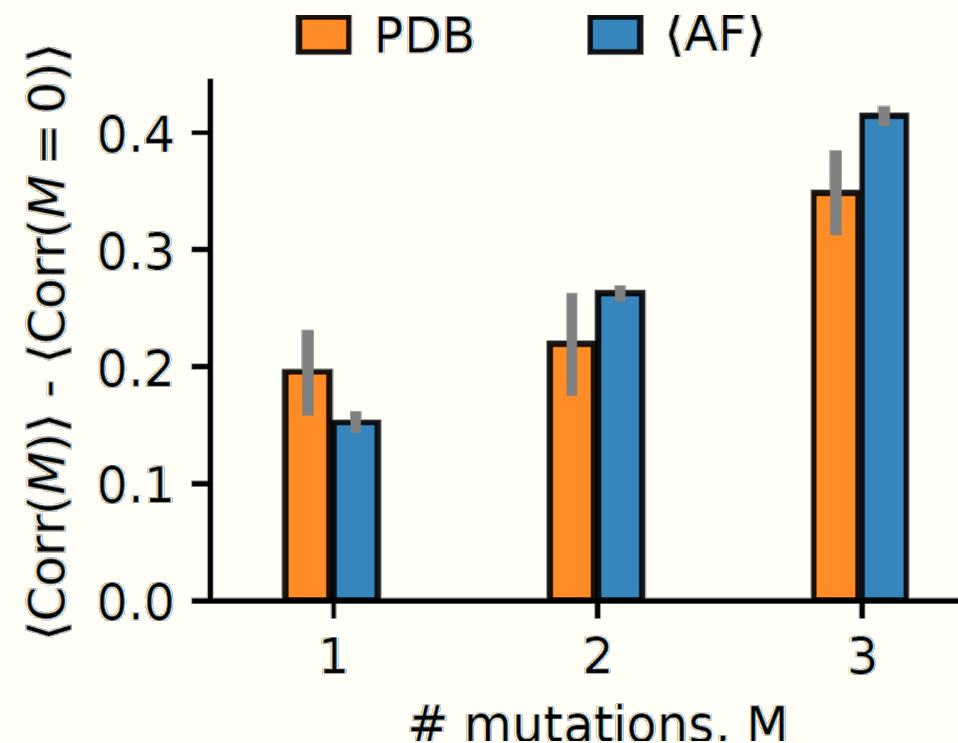
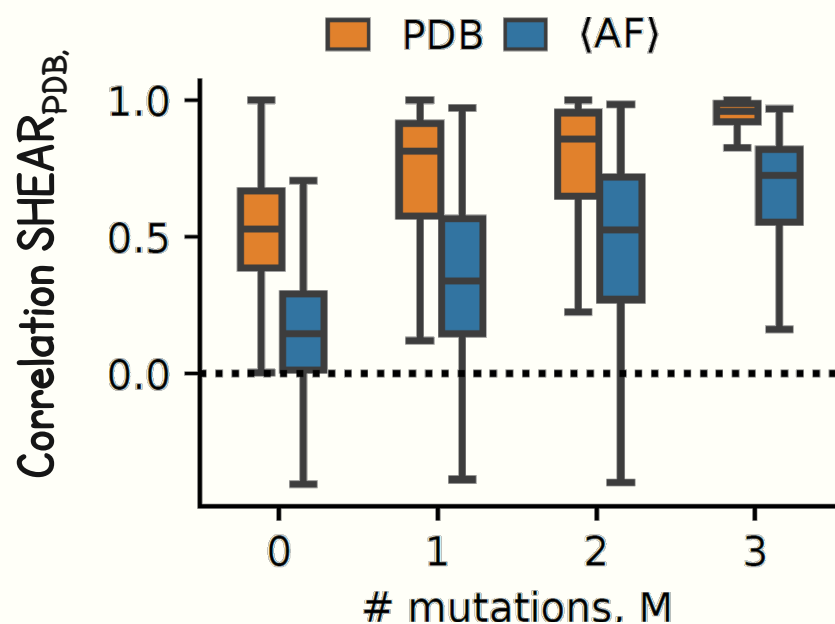
H-NOX = HEME-NITRIC OXIDE/OXYGEN BINDING



Systematic in $\sim 5,000$ mutated pairs:

Alphafold captures $\sim$ whole experimental signal above the noise

Fluctuations in experimental measurements
hinder the evaluation of mutation effects



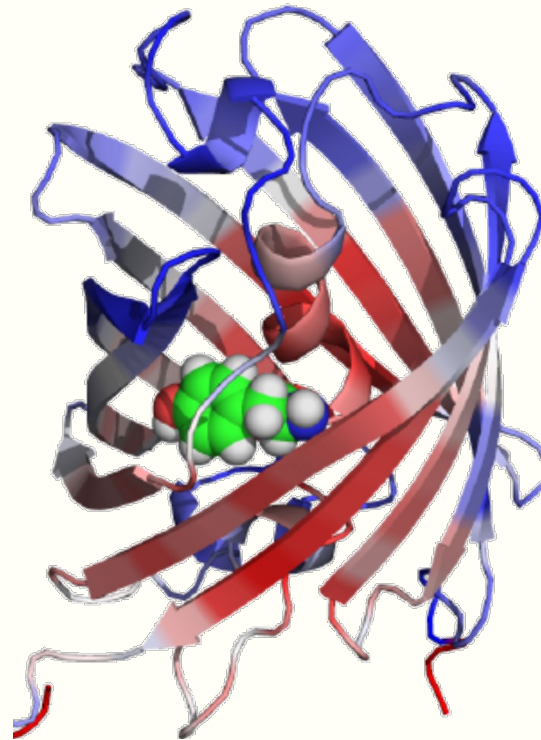
But AF predictions of mutation effects are (on average)
as reliable as repeat exp. measurements

Shear correlates with phenotypic changes in fluorescence in BFP

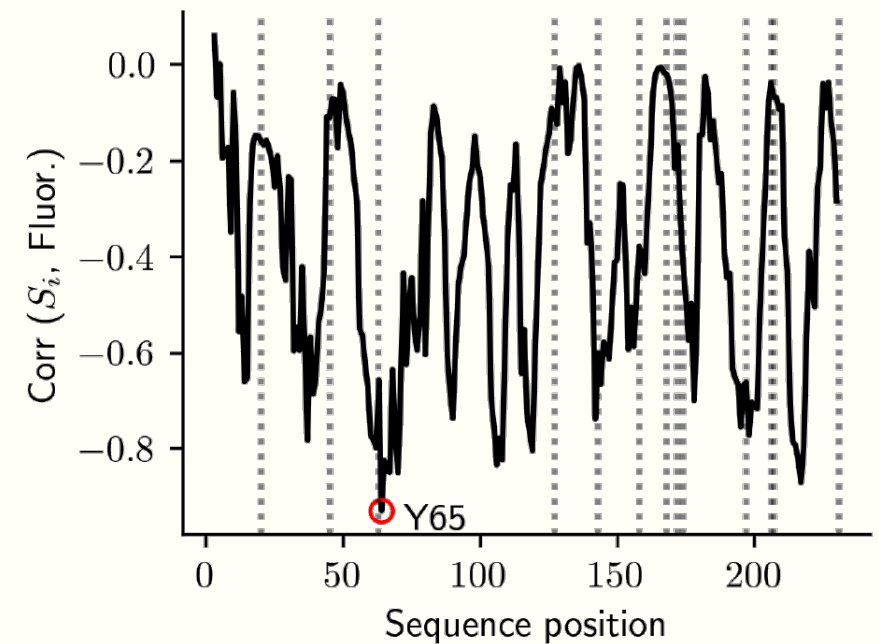
Example: Blue fluorescent protein (BFP)

Fluorescence measured for 2^{13}
= 8,192 mutations
(Poelwijk et al. Nat. Comm. 2019).

Red region: shear in amino acids
strongly correlated with phenotype

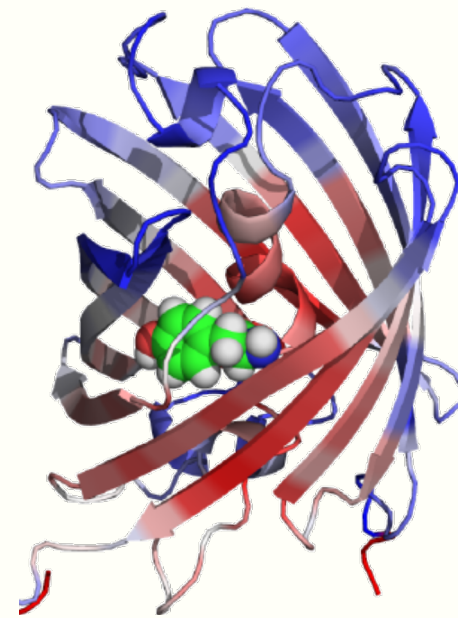
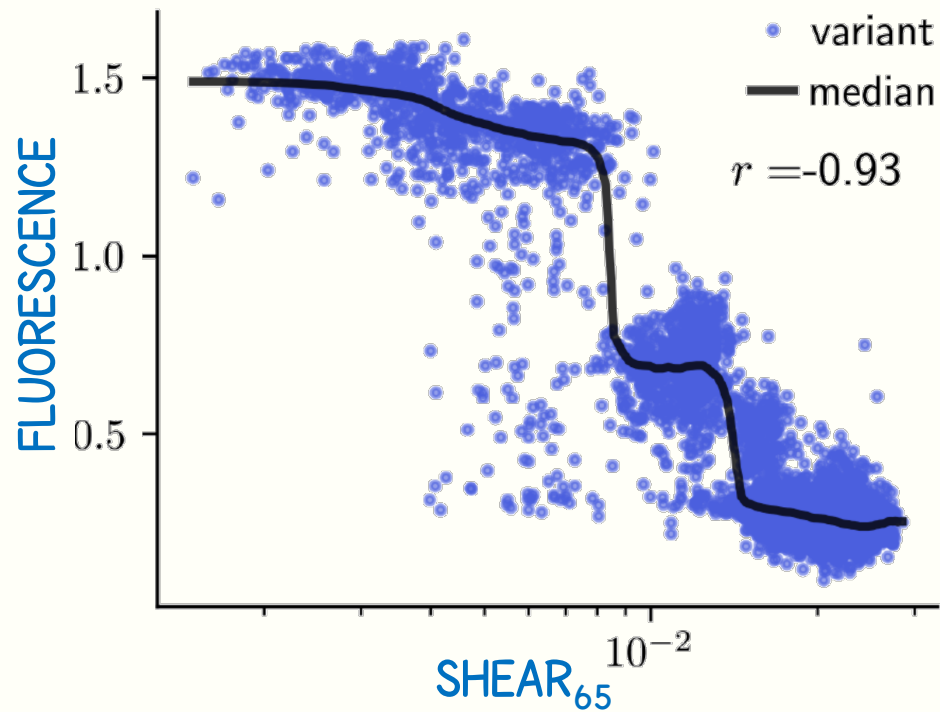


correlation of AF-predicted shear
and fluorescence change



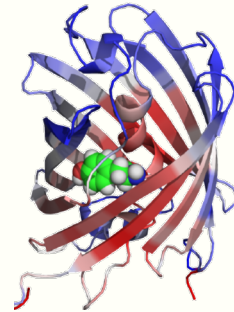
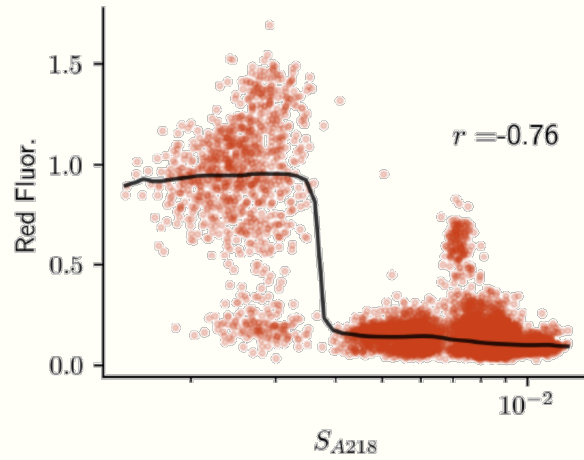
Dash - 13 mutations

Phenotype Shear is a ~binary predictor of fluorescence level in BFP

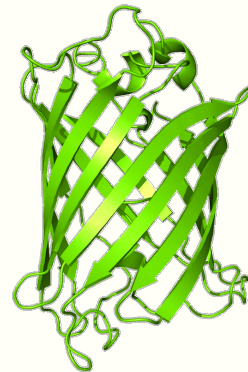
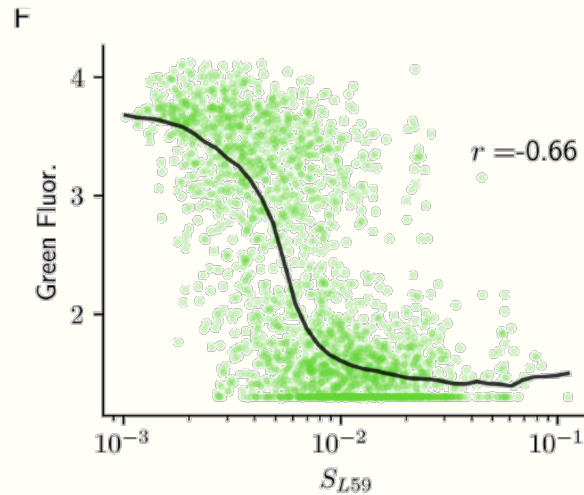
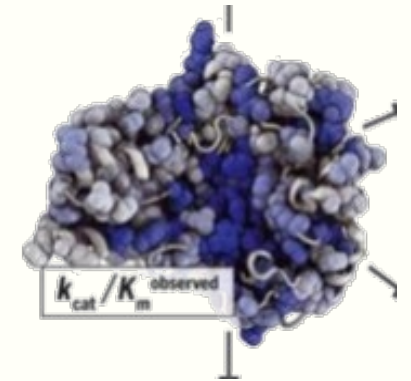
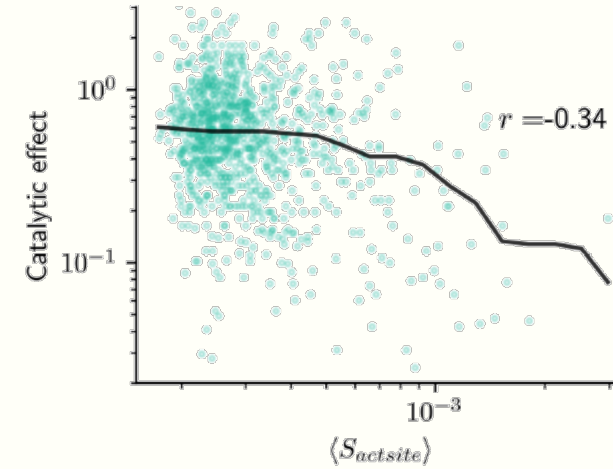


...on average, AF can predict mutation effects

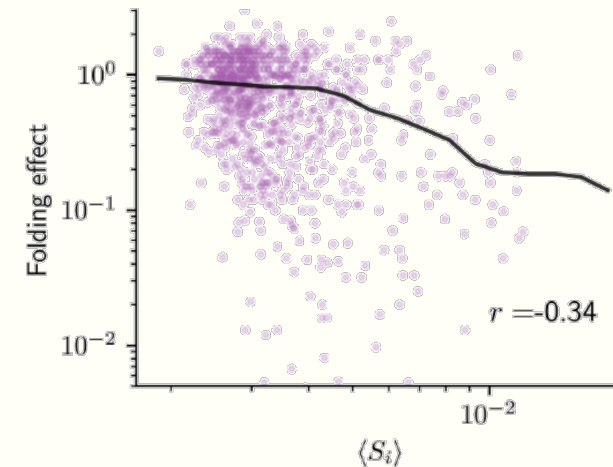
Shear-phenotype correlation is also found in other high-throughput datasets



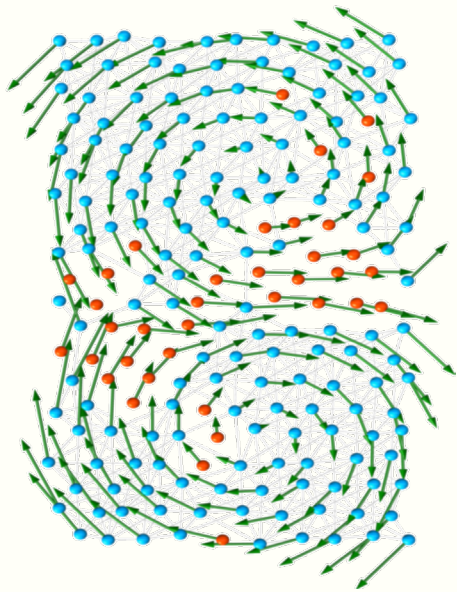
Poelwijk et al. Nat. Commun. 10, 4213 (2019).



GFP: Sarkisyan et al. Nature (2016).

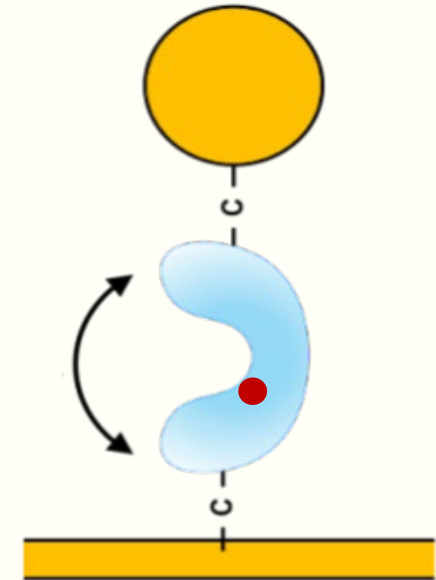
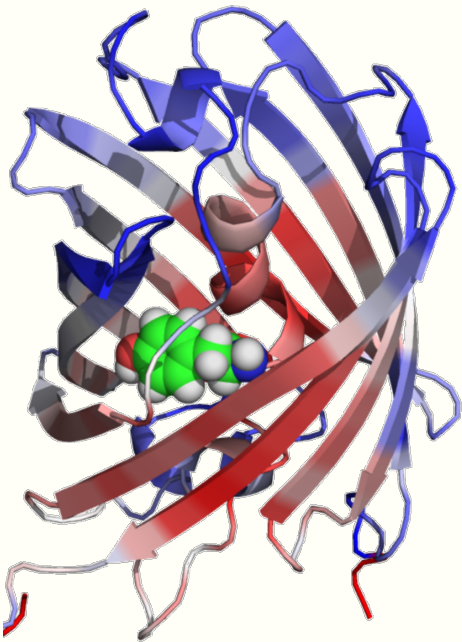


PafA (kinase) Markin et al. Science (2021).



Summary: proteins as sequence-encoded viscoelastic machines

1. **THEORY** of protein as **viscoelastic** matter predicts the evolution of **high-shear** regions essential to functionally relevant large-scale motion.
2. **EXPERIMENT** (**guanylate kinase**) links enzymatic function to **predicted** high-shear regions and their effect on viscoelastic motion.
3. **Generalizing with AI (AlphaFold)**: a method for predicting the effect of **single mutations** on **structure and phenotype** through computation of induced shear.



McBride et al. 2023 *AlphaFold2 can predict structural and phenotypic effects of single mutations* (in review, arXiv)

Weinreb et al. (2023) *High-shear regions in an enzyme impact viscoelastic mechanics and activity* (in review)

E Weinreb, E Moses, Y Peleg, T Unger, S Albeck, JL Sussman (Weizmann)

J-P Eckmann, J Rougemont (Geneva)

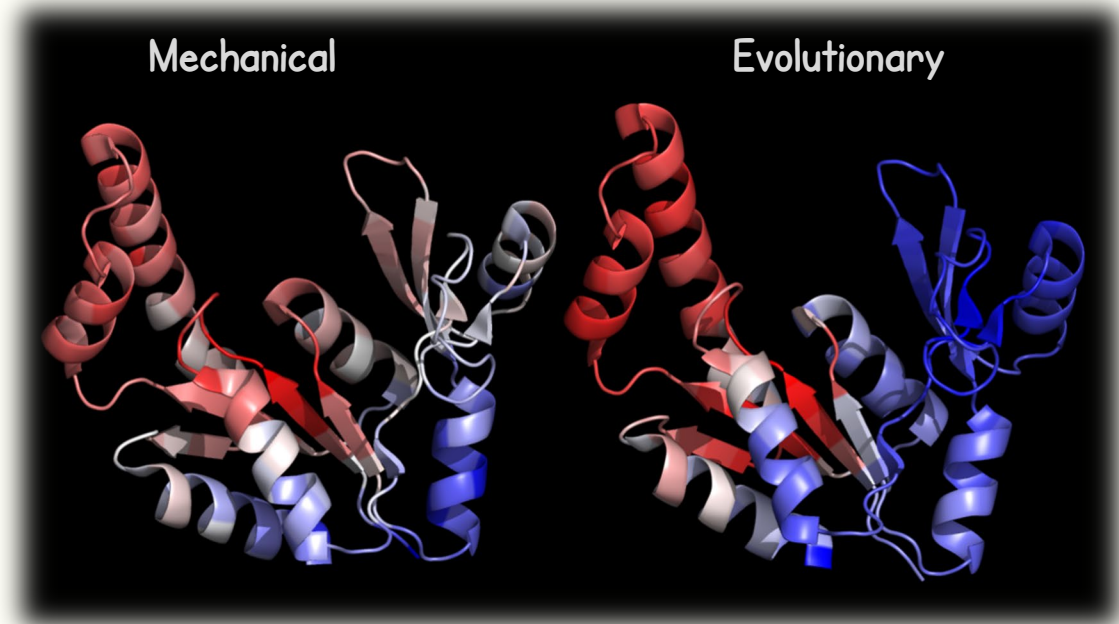
G Zocchi (UCLA)

A Libchaber, M Mitchell, S Leibler (Rockefeller)

J McBride, S Dutta, M Siek, BA Grzybowski, TT (IBS)



**WE ARE RECRUITING
AT ALL LEVELS**



- **High-shear regions in guanylate kinase impact viscoelastic mechanics and enzymatic function** Weinreb, McBride, Siek, Rougemont, Renault, Peleg, Unger, Albeck, Sussman, Grzybowski, Zocchi, Eckmann, Moses, TT (in review)
- **AlphaFold2 can predict structural and phenotypic effects of single mutations** McBride, Polev, Reinharz, Grzybowski, TT (in review)
- **Work on protein theory:** McBride et al. *Mol. Biol. Evo* 2022; Eckmann et al., *Bioessays* 2021; Eckmann et al., *Rev Mod Phys* 2019; Eckmann et al. *PRX* 2017; Dutta et al., *PNAS* 2018; Mitchell et al. *PNAS* 2016

END