

APS March 2017

Mechanical function of protein reduces the dimension of genotype space

a simple physical model

Jean-Pierre Eckmann (Geneva)

Albert Libchaber (Rockefeller)

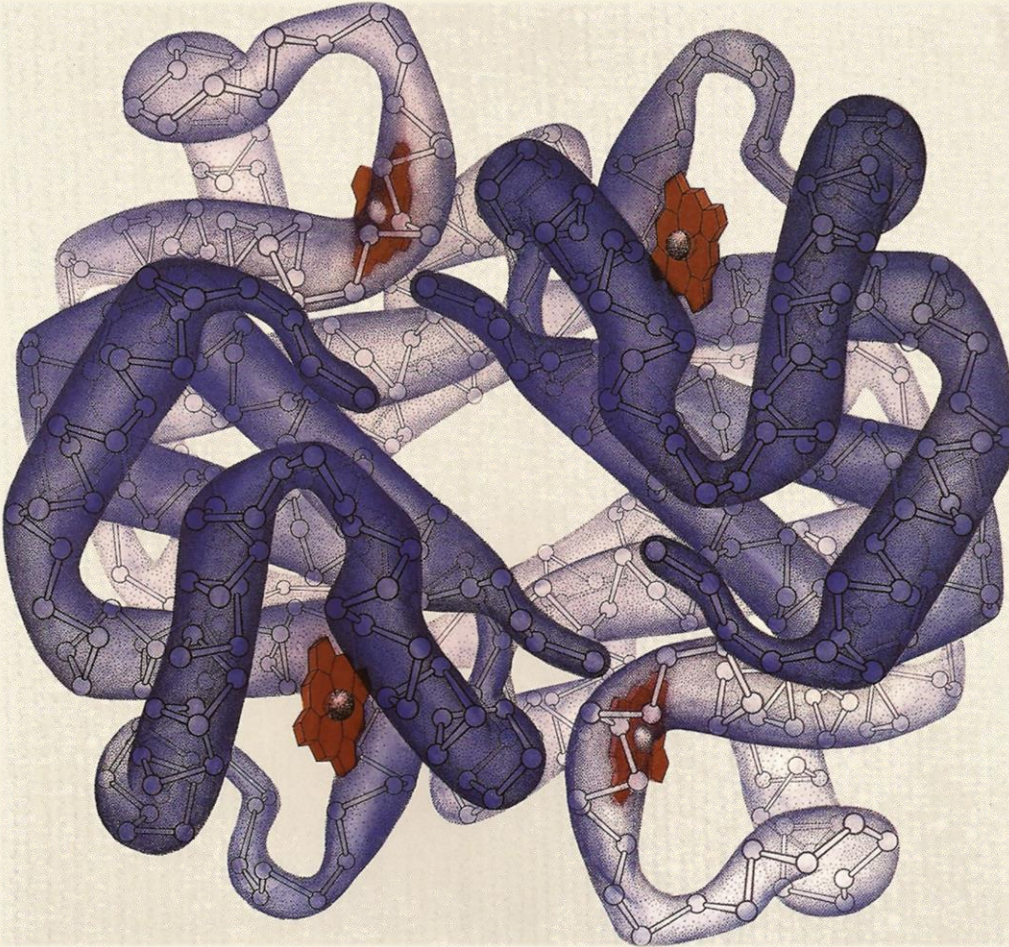
Stanislas Leibler and Michael Mitchell (IAS)

Sandipan Dutta & TT (IBS)



**Center for Soft
and Living matter**

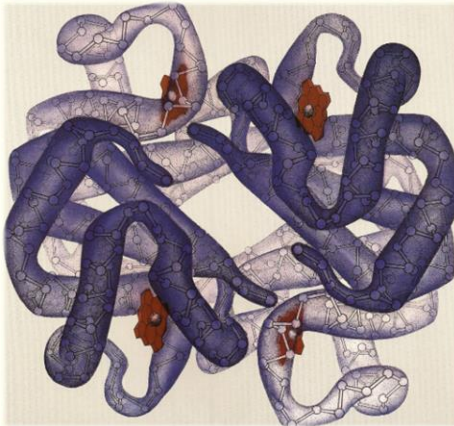
Proteins are complex systems made of hundreds to thousands amino acids arranged in a 3D structure



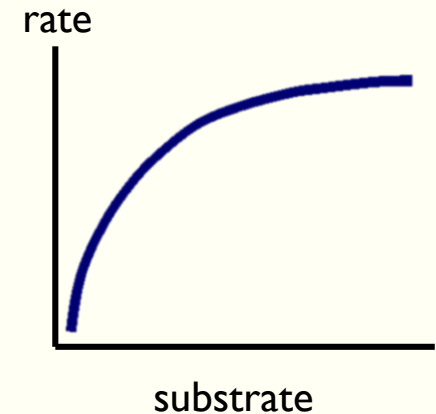
Hemoglobin

Irving Geiss, 1978

Genes and proteins have *many* degrees-of-freedom but protein function is described by a *few* of parameters



**dimensional
reduction**



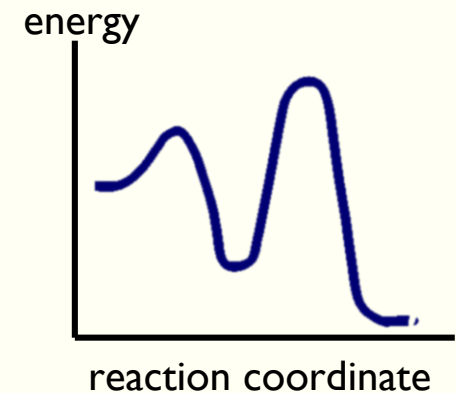
ATG CTG GAG AAC GGG CTG GCG CGG
Met Leu Glu Asn Gly Leu Ala Arg

TGG GAG CGG TTC CGC TGC AAC GTG
Trp Glu Arg Phe Arg Cys Asn Val

TGC ATC AGT GAG ATG CTC TTC ATG
Cys Ile Ser Glu Met Leu Phe Met

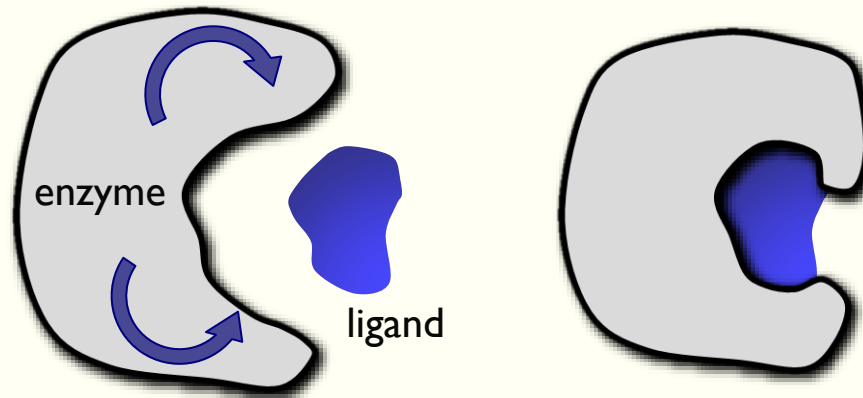
GAC GGC TGG AGG GAG CTG GGC TAC
Asp Gly Trp Arg Glu Leu Gly Tyr

TGG GCC GCC AAG CAG CGT GAC ACT
Trp Ala Ala Lys Gln Arg Asp Thr

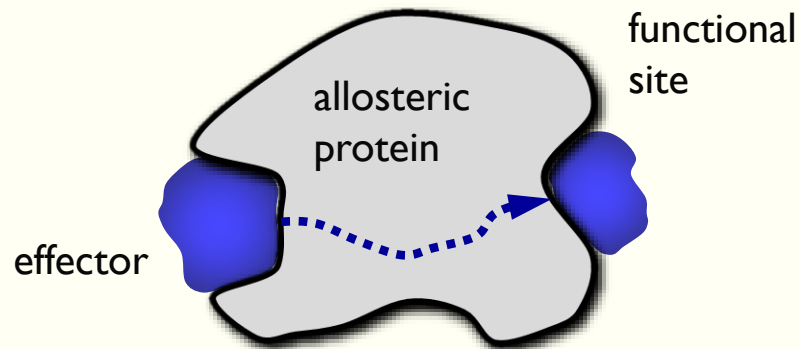


Many proteins use *large-scale, long-range* motions to function

Induced fit:



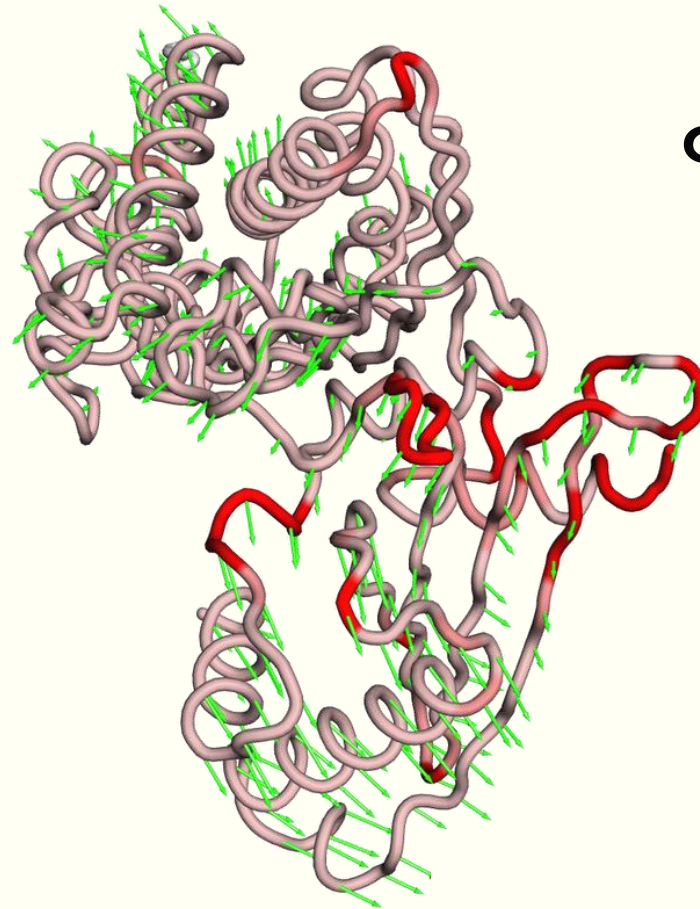
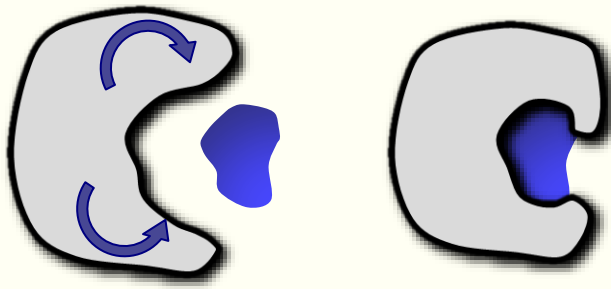
Allostery:



Looking at dynamics reveals simple patterns that help to understand the mechanical basis of protein function

Displacement:

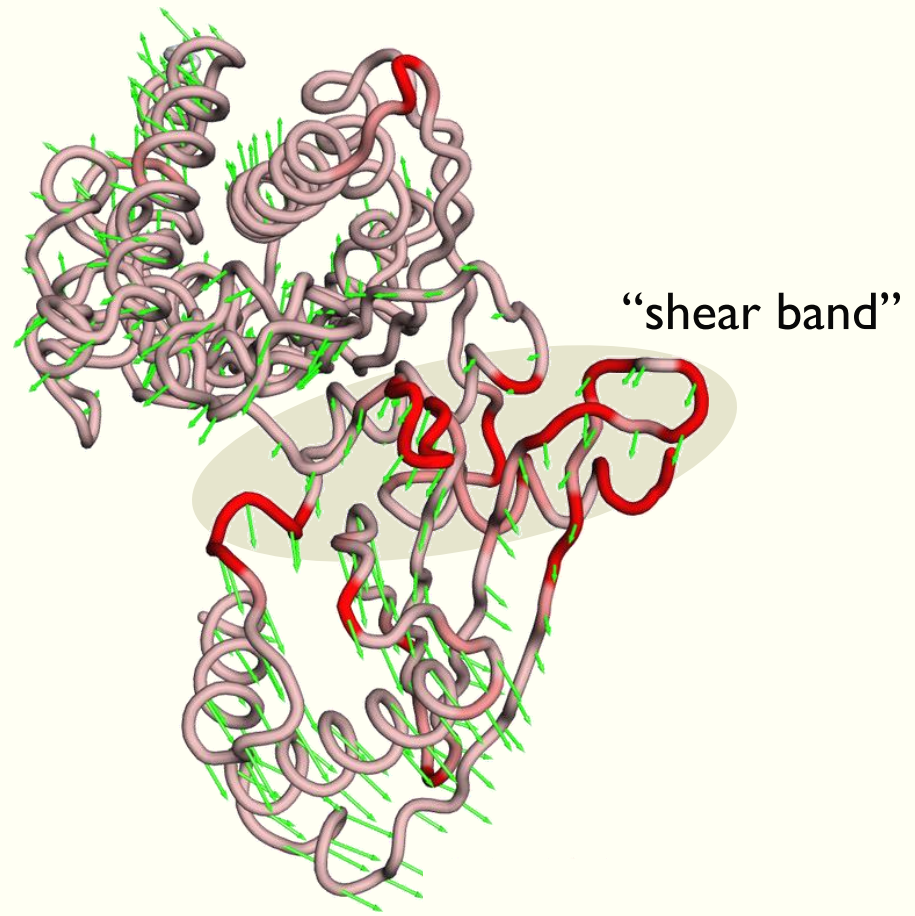
unbound → bound



Glucokinase

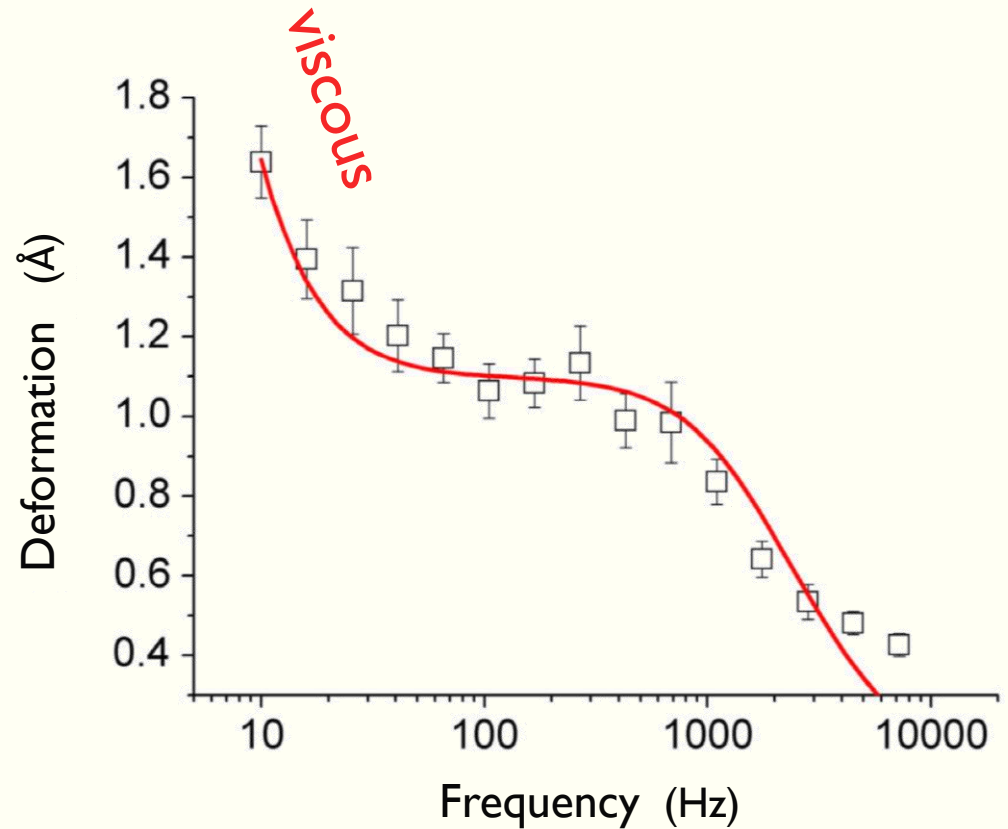
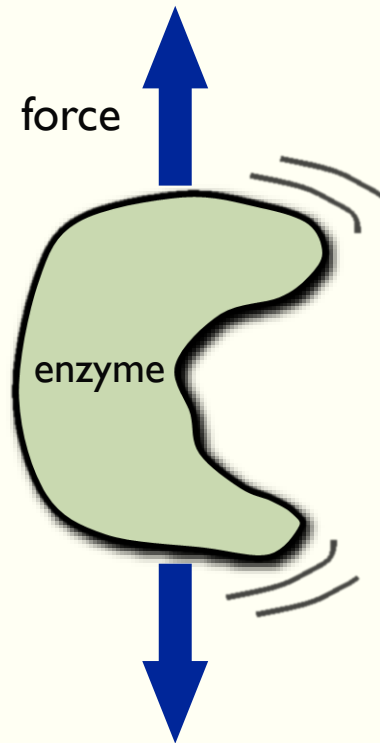
- Jean-Pierre Eckmann, TT
- Michael Mitchell, Stan Leibler & TT, PNAS 2016

The shear highlights the relative motion of amino acids within the protein

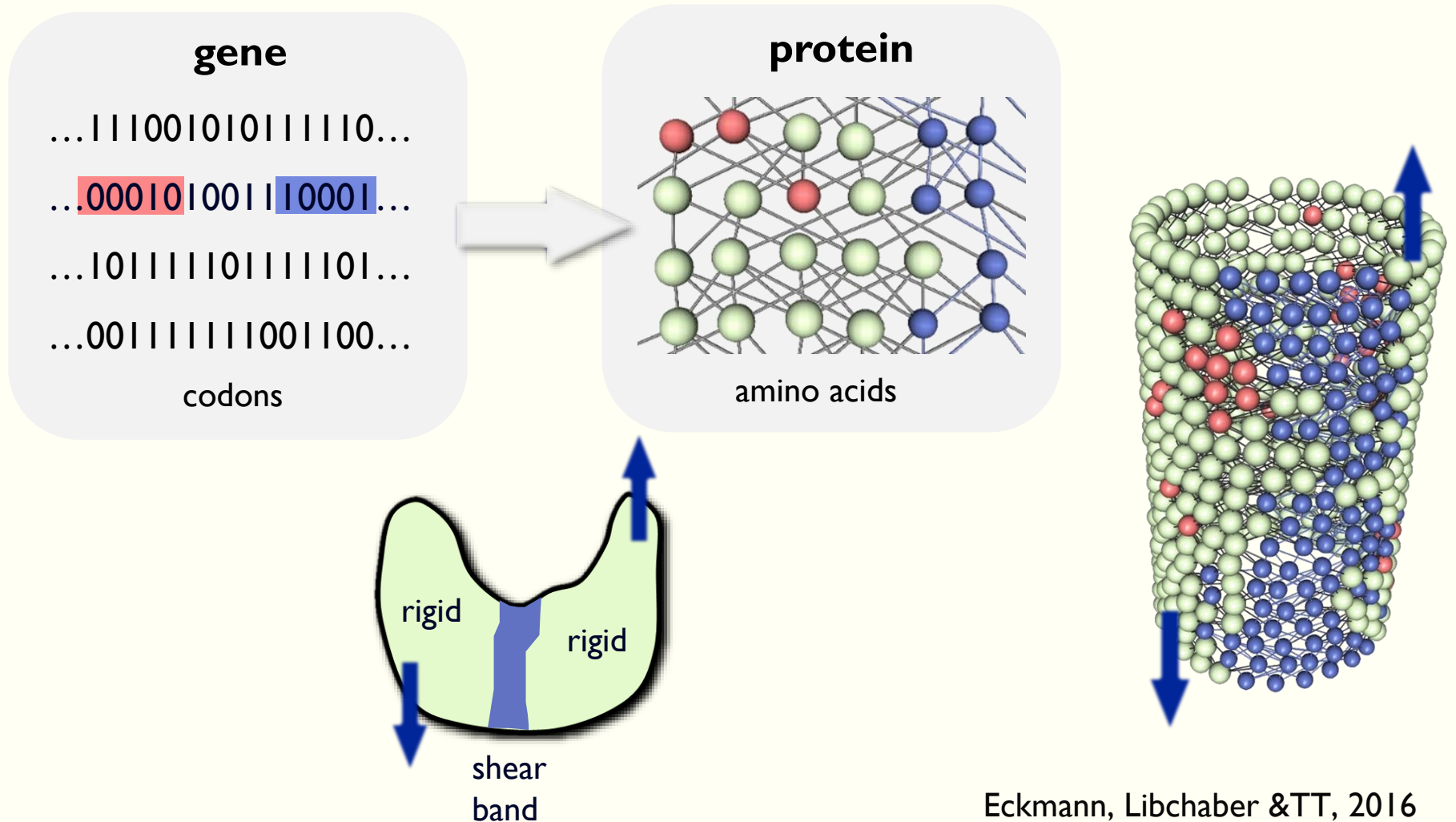


Shear = gradient (displacement)

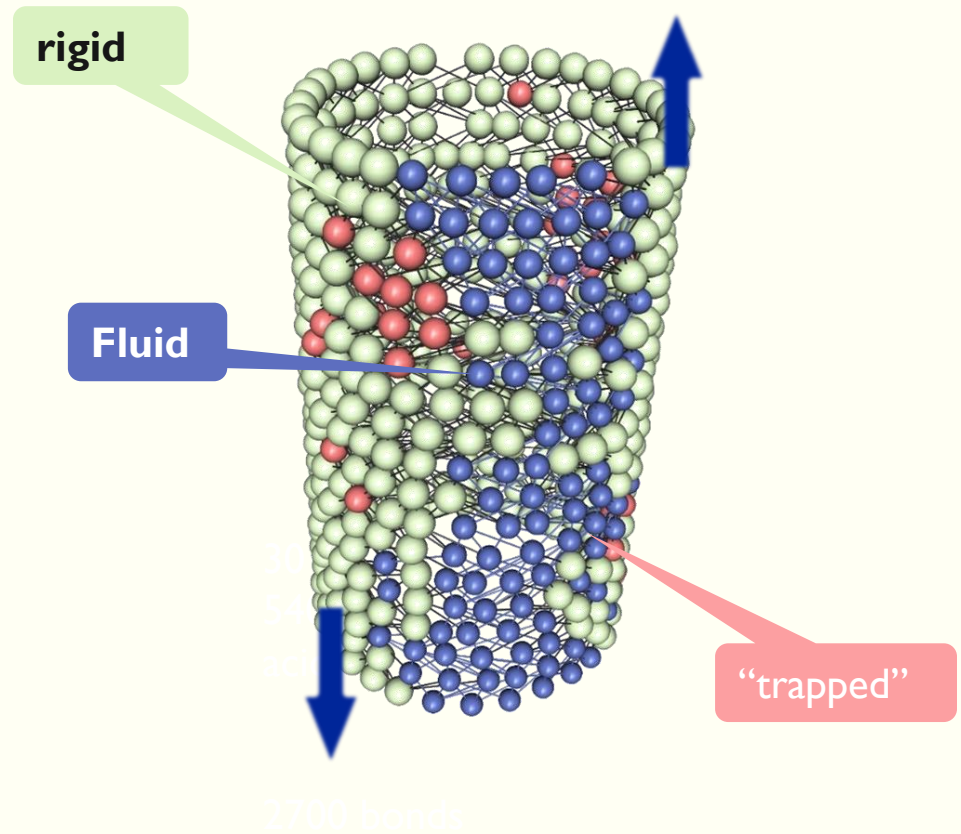
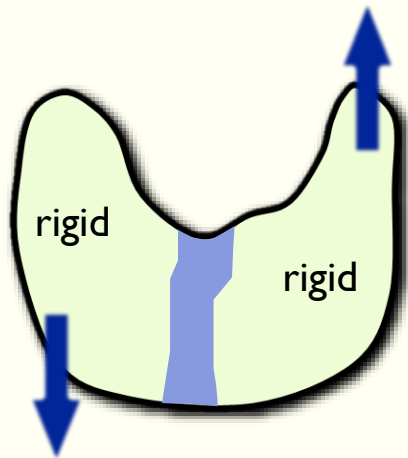
Proteins are soft and *viscous* at low frequency



A simple model captures the evolution of functional shear and hinge motions in the protein

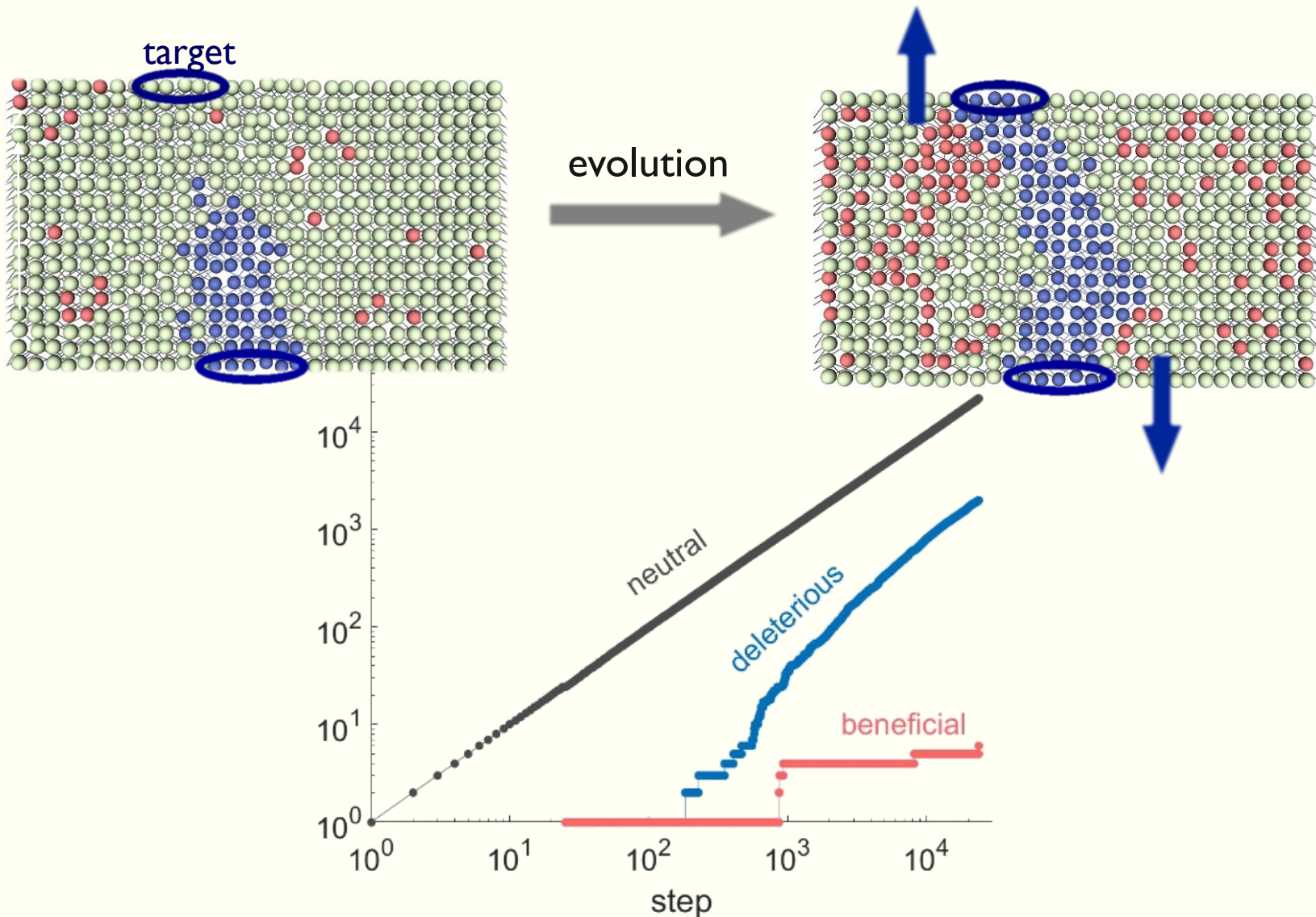


A concrete genotype-to-phenotype map shows how change in sequence affects the rigidity and the dynamics



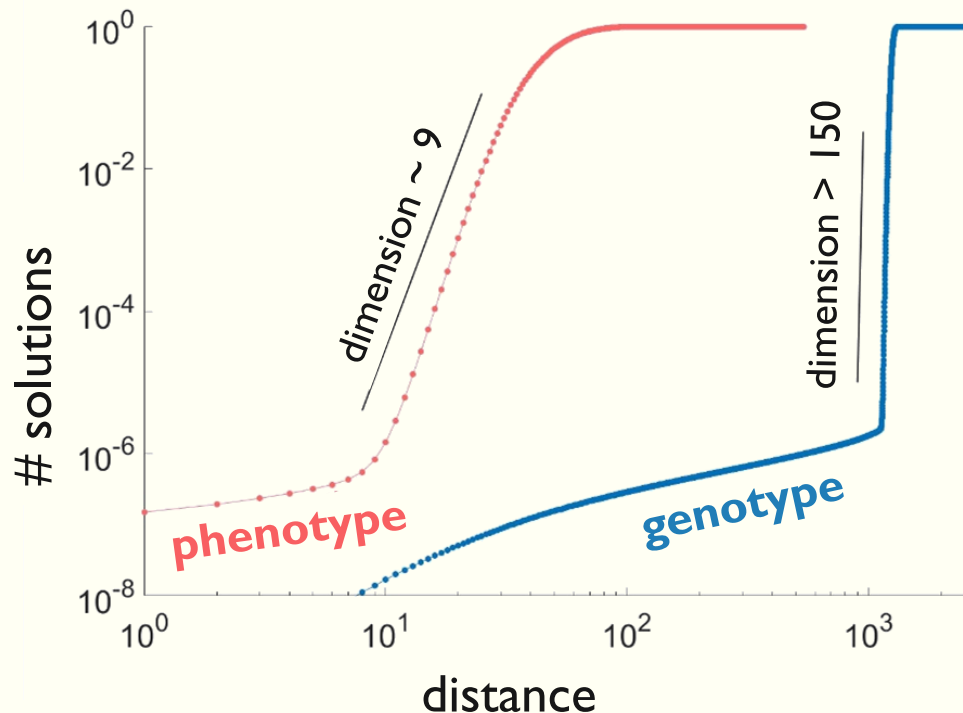
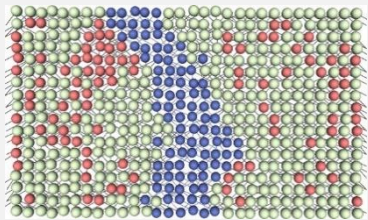
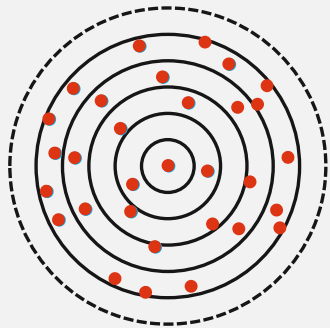
“Floppy”, soft mode (amorphous matter, glasses)

Evolution searches by mutating the gene for a fluid channel and slow, viscous modes

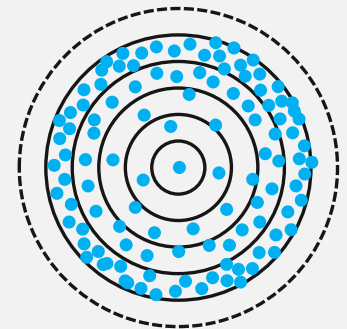


The genotype-to-phenotype map involves huge dimensional reduction

functional
phenotypes



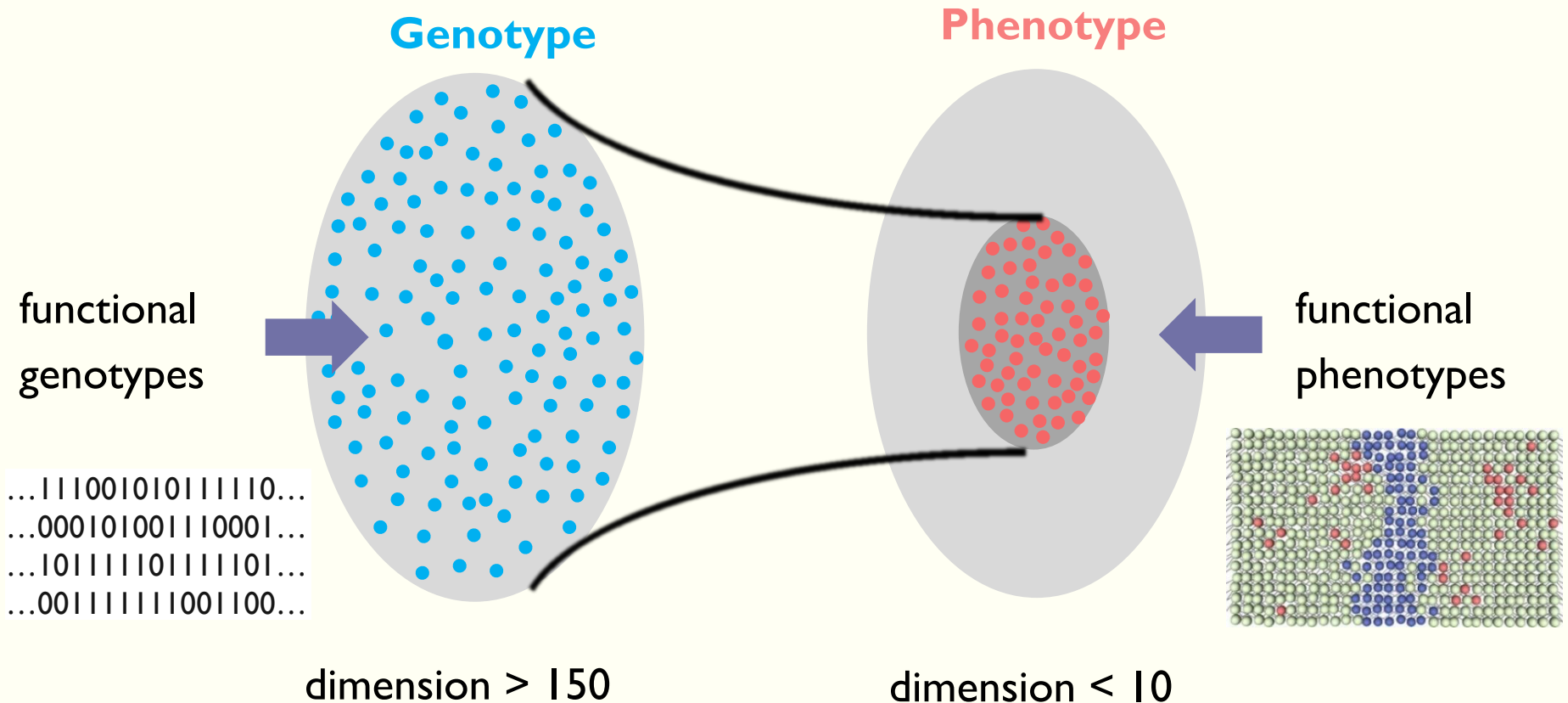
functional
genotypes



...111001010111110...
...000101001110001...
...101111101111101...
...00111111001100...

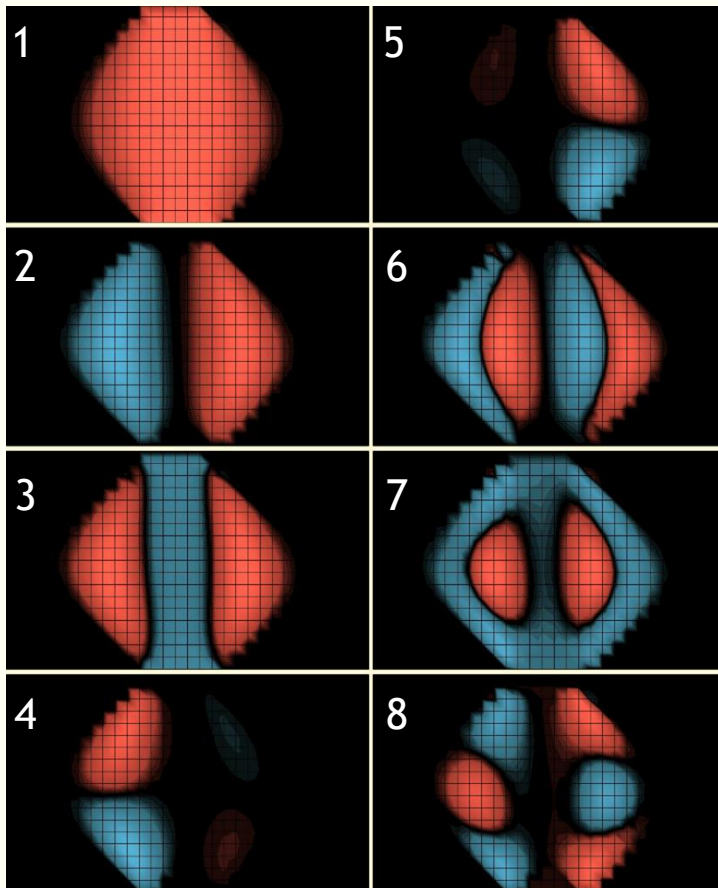
solutions $\sim$ (distance)^{dimension}

Phenotypes are constrained to low dimension by the continuity of force transmission along the protein

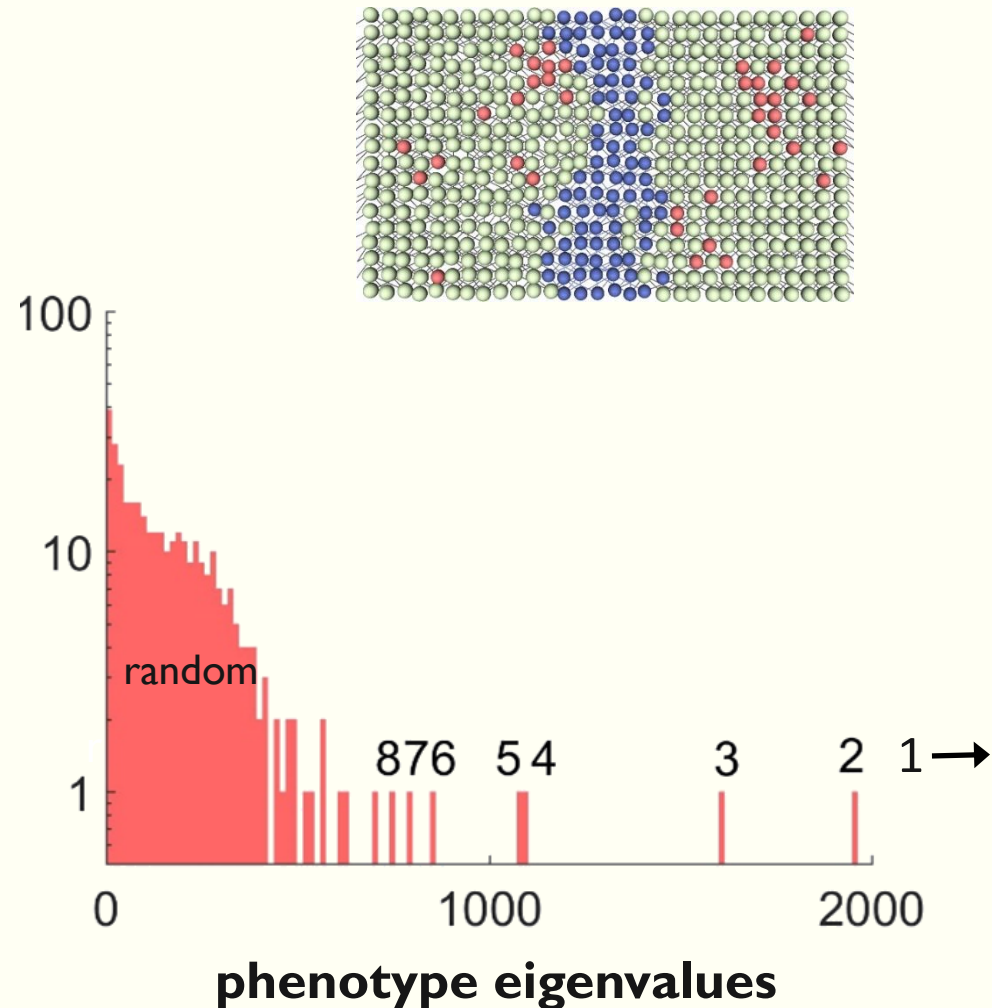


Phenotype modes reflect the few degrees-of-freedom that describe a continuous shear band

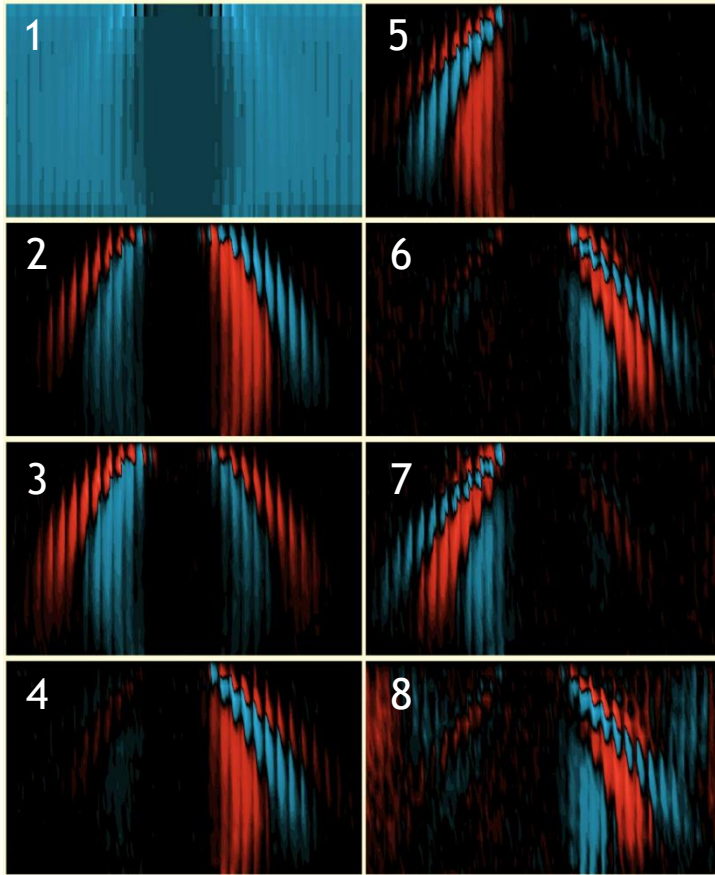
Principal components in spectrum of 10^6 phenotypes



phenotype modes

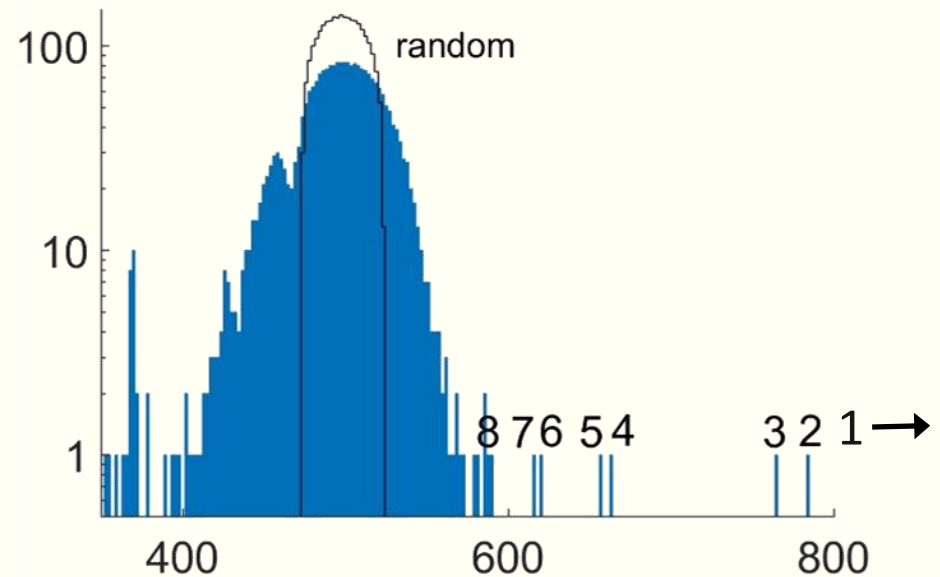


Genotype spectrum reveals that most variance is random with a few non-random sequence correlation modes



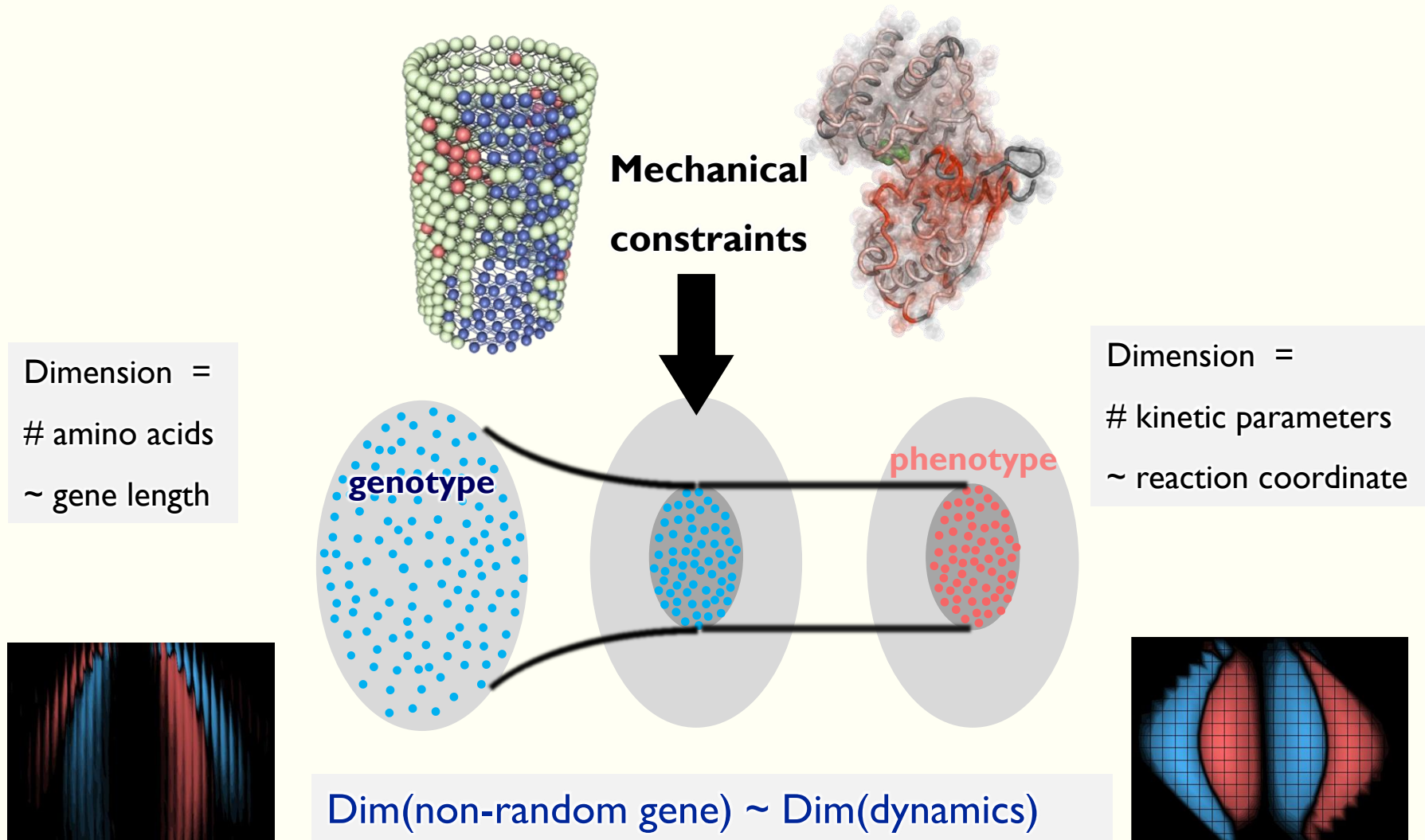
genotype modes

```
...11100101011110...  
...000101001110001...  
...101111101111101...  
...001111111001100...
```

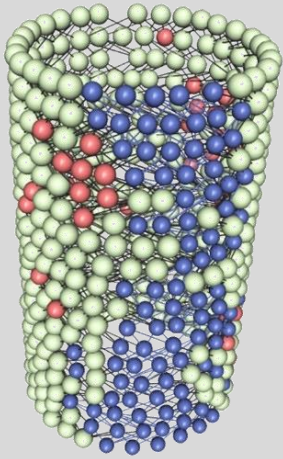


genotype eigenvalues

The evolution of collective motion shapes the geometry of phenotype-to-genotype map of protein



Looking at protein dynamics suggests simple physical understanding of protein as learning amorphous matter



- **Mechanical function:** “floppy” modes.
- **Dimensional reduction.**
- **Mapping:** mechanical modes $\leftrightarrow$ information “ripples”.



Thanks:

S. Dutta (IBS)

J.-P. Eckmann (Geneva)

A. Libchaber (Rockefeller)

S. Leibler & M. Mitchell (IAS)

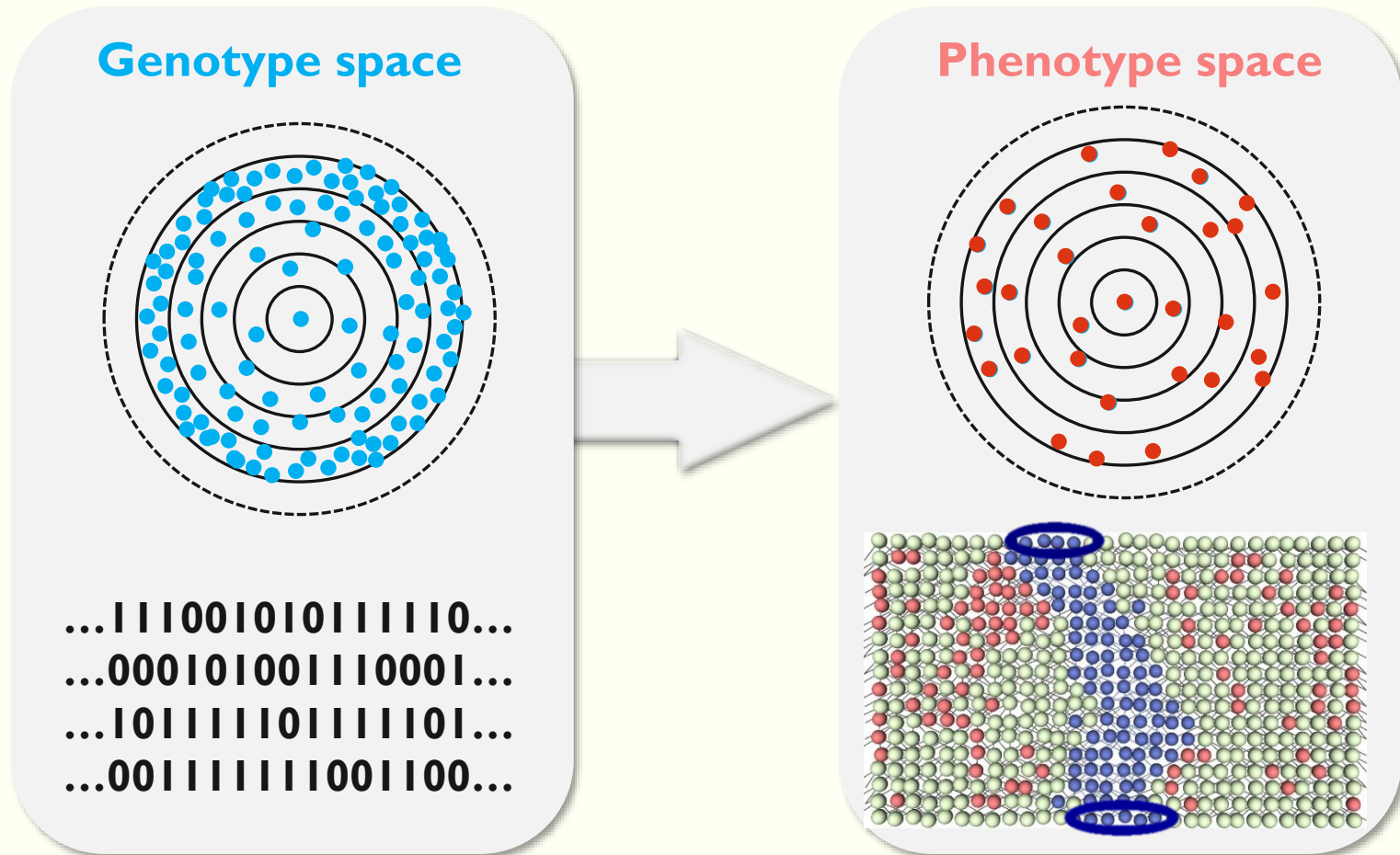


Center for Soft
and Living matter

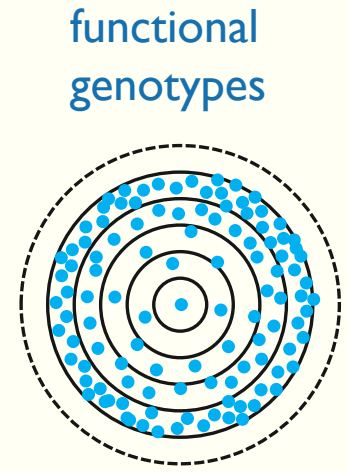
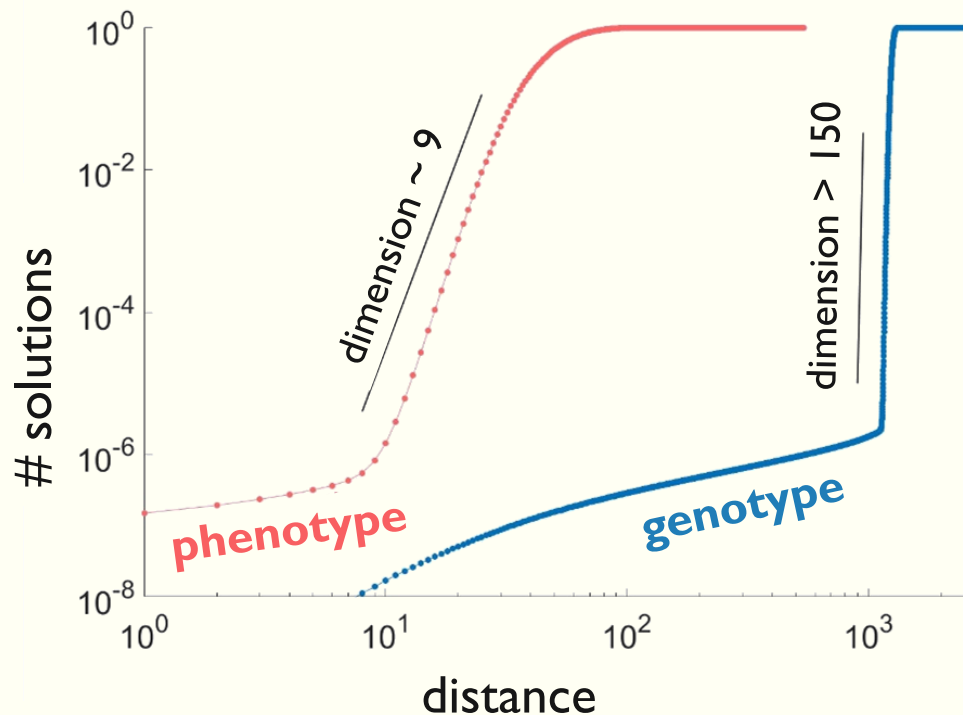
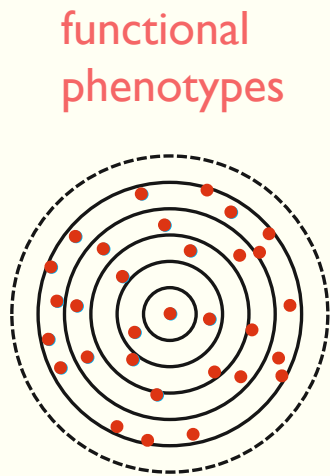
- Eckmann Libchaber & TT, arXiv: 2016
- Leibler Mitchell & TT, PNAS 2016
- TT, Proc Roy Soc A 2016

Additional slides

The simple model enables survey of solution space which reveals the geometry of the genotype-to-phenotype map

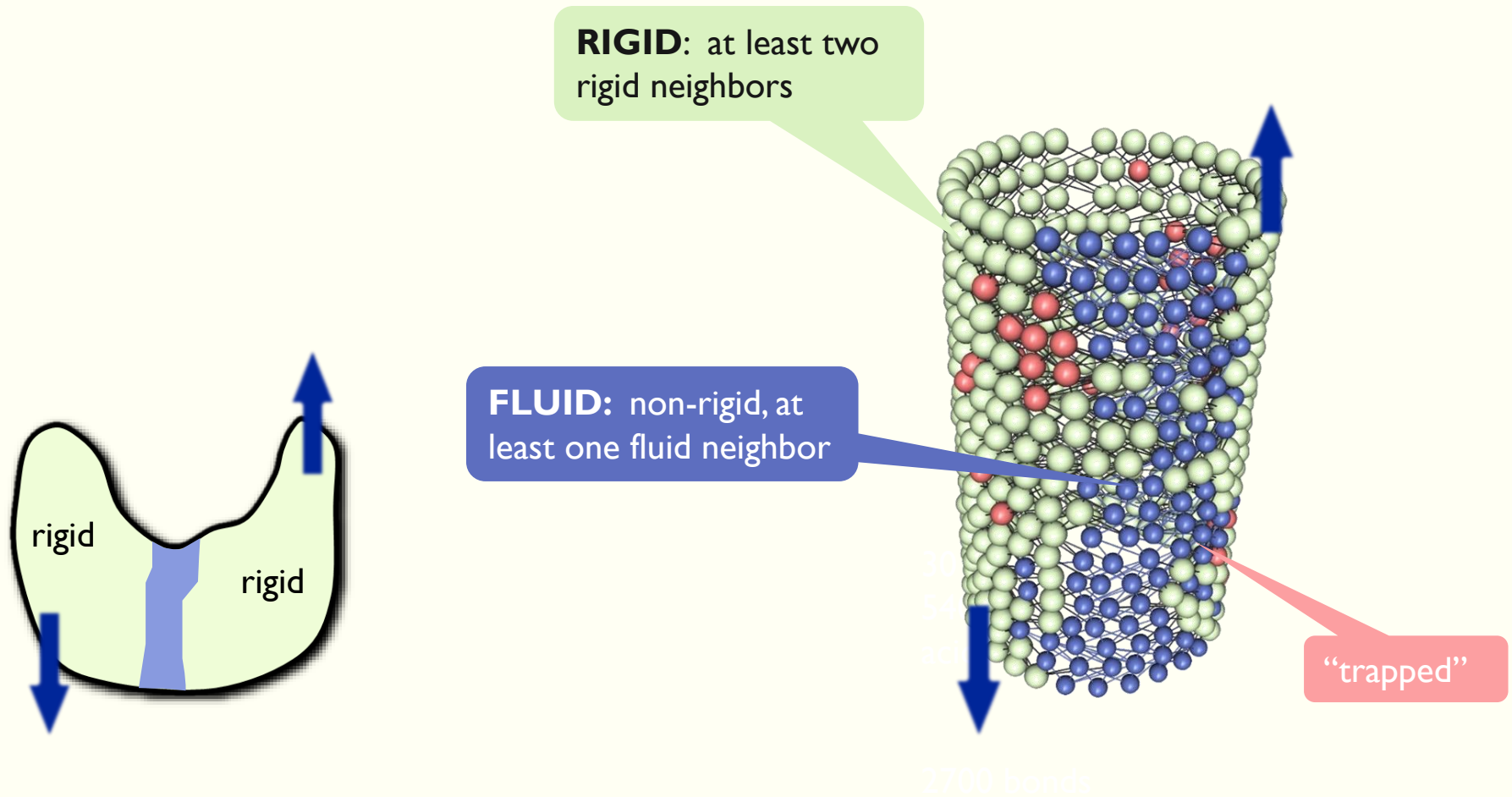


The genotype-to-phenotype map involves huge dimensional reduction



$$\# \text{ solutions} \sim (\text{distance})^{\text{dimension}}$$

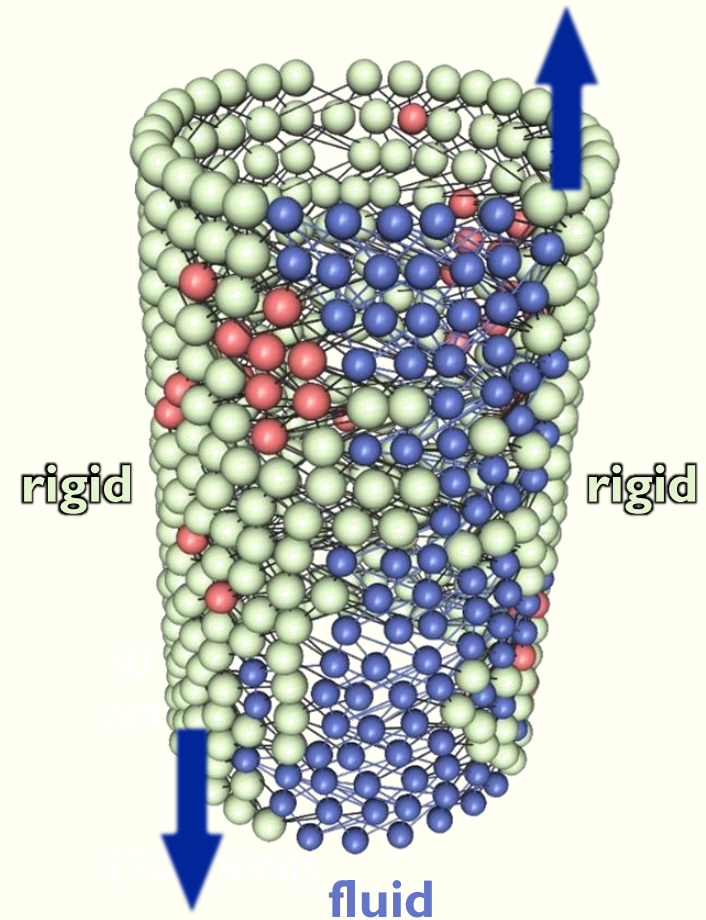
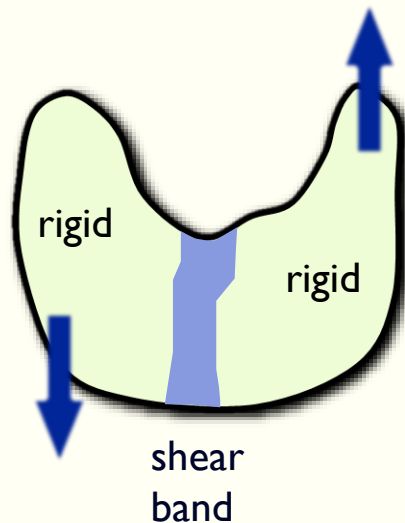
The collective rigidity and motion of the protein evolves by simple connectivity rules to induce a fluid channel



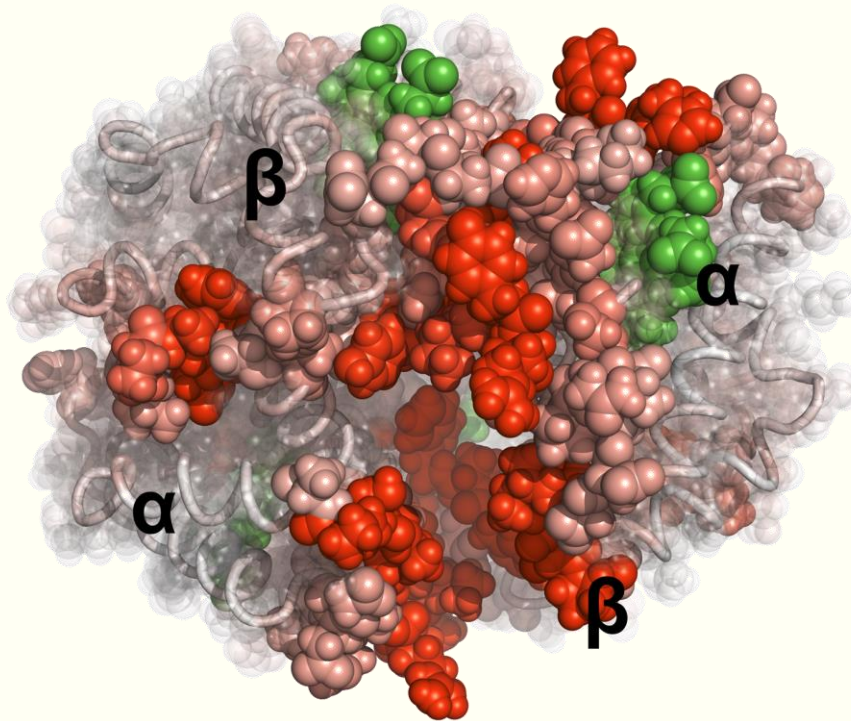
“Floppy”, soft mode (amorphous matter, glasses)

A concrete genotype-to-phenotype map shows how change in sequence affects the rigidity and the dynamics

“Floppy”, soft mode
(as in amorphous matter, glasses)

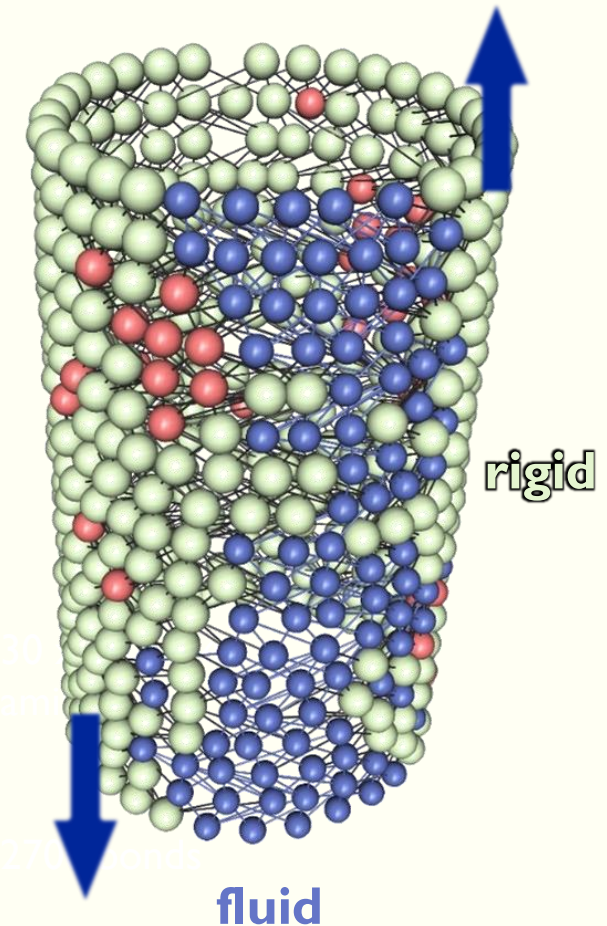
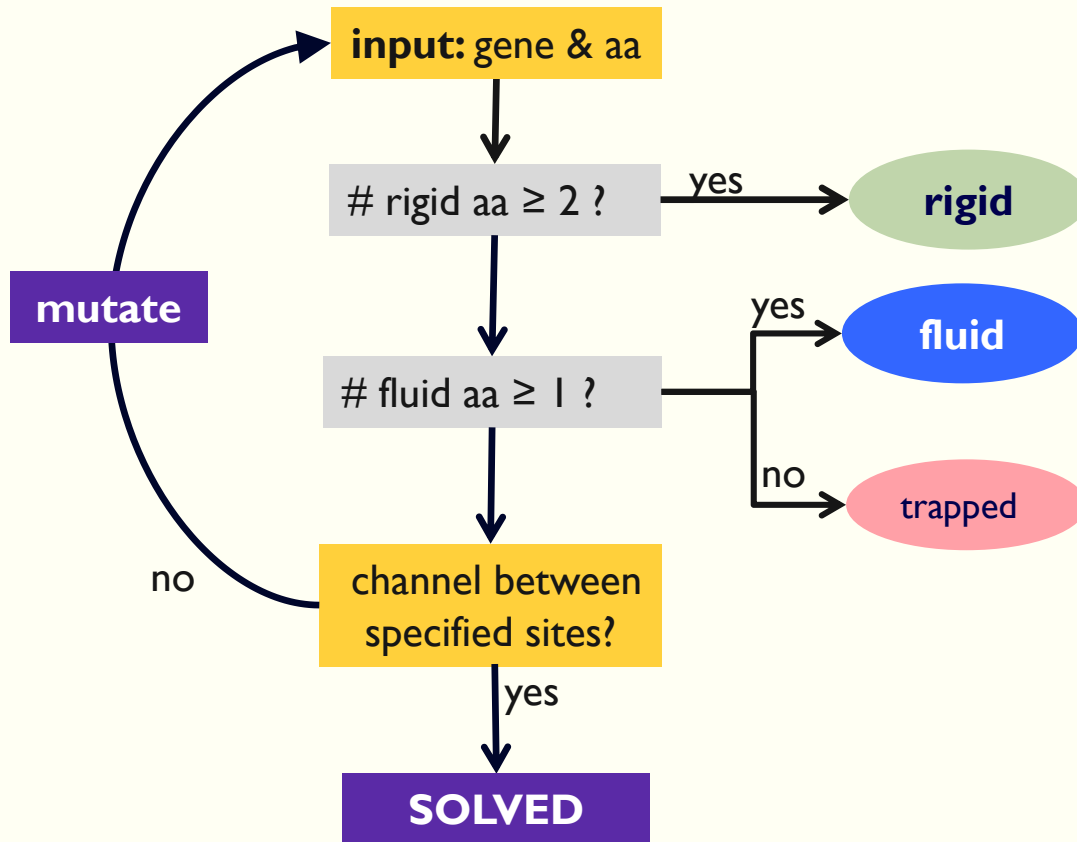


A two-dimensional “shear band” communicates mechanical signals between distant sites

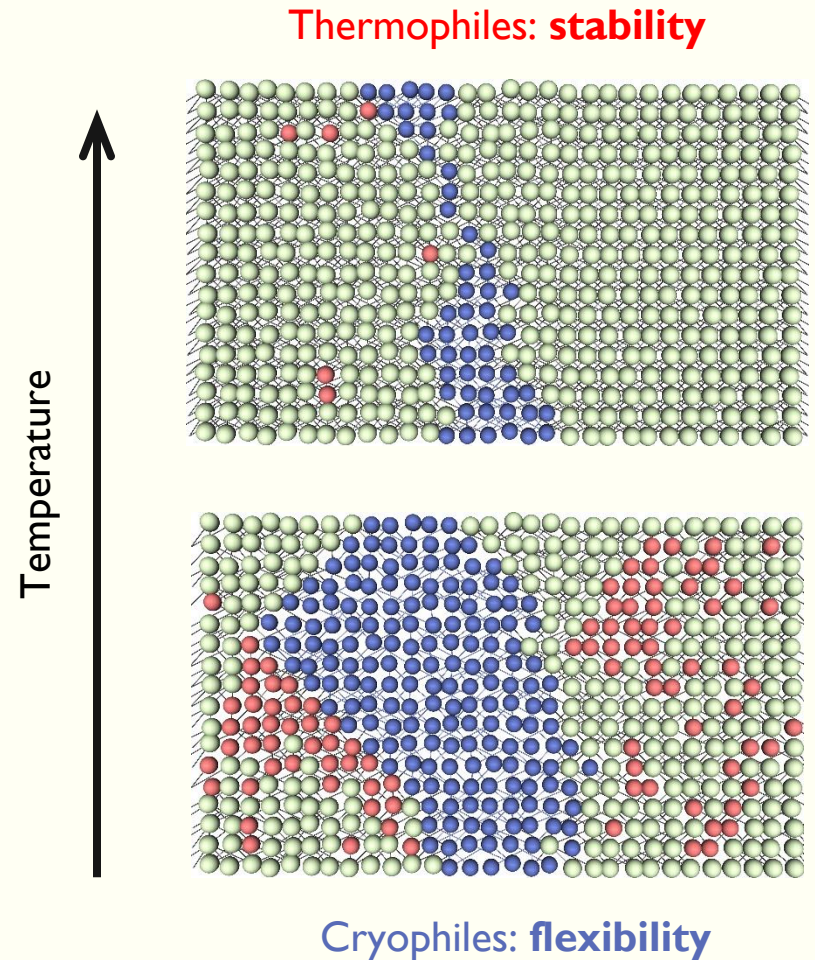
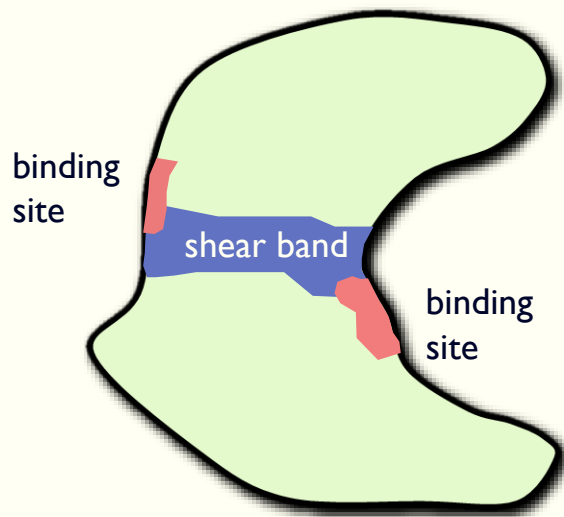


Hemoglobin

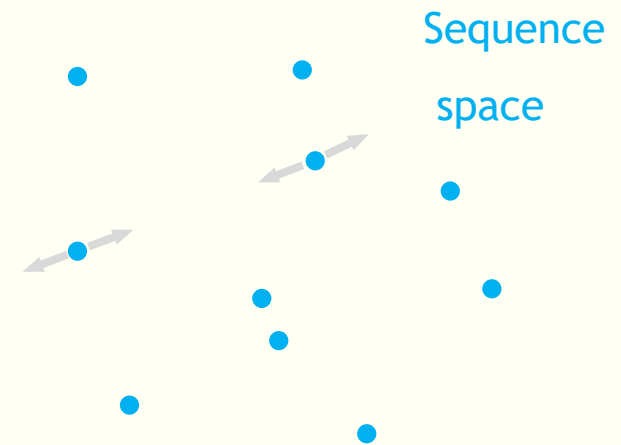
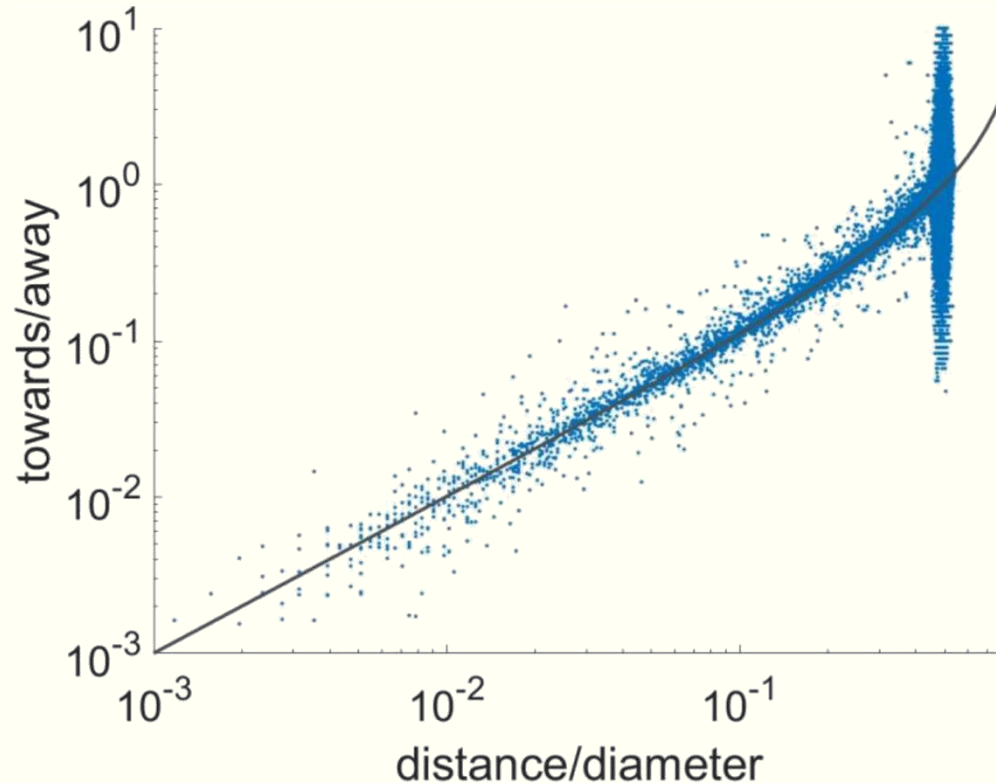
The collective rigidity and motion of the protein evolves by simple connectivity rules to induce a fluid channel



The geometry of the gene-to-function map allows for adaptation to multiple tasks, such as specificity and stability



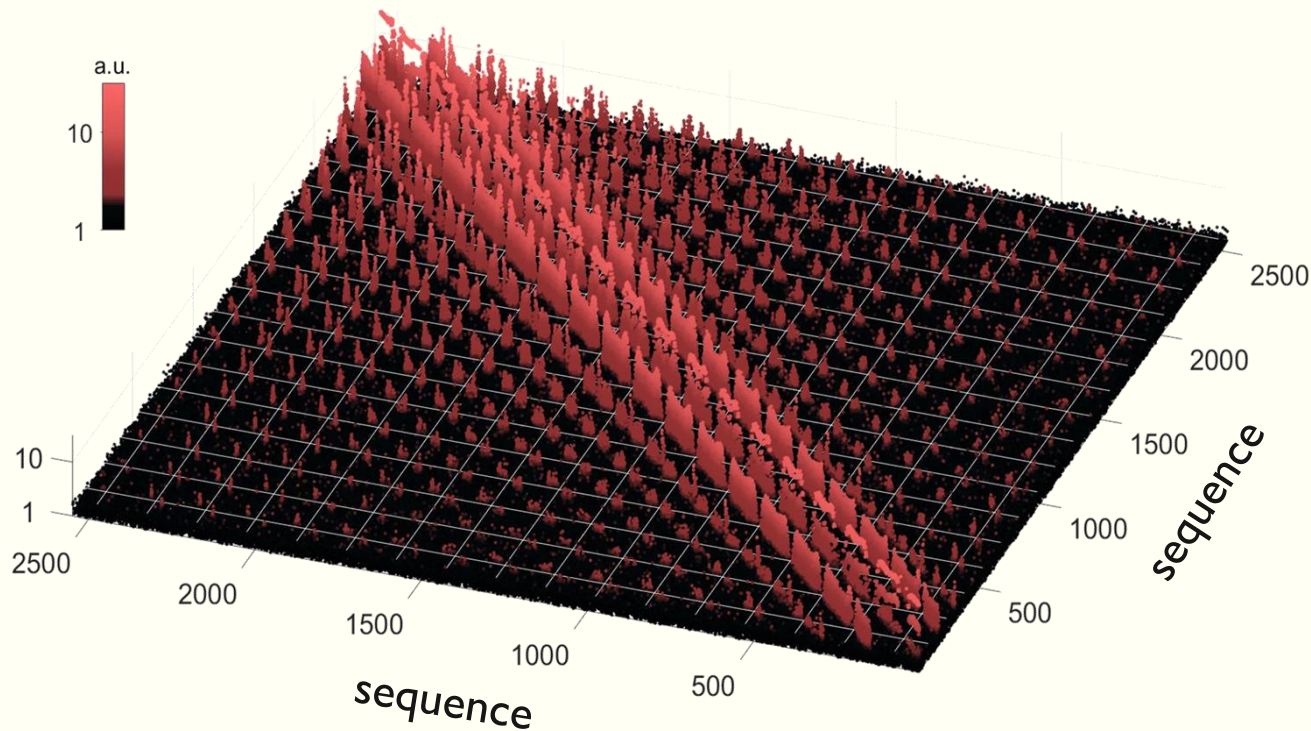
Geometry of protein universe: expansion via evolutionary search for functional genes



10^6 proteins in the model spread
through sequence universe

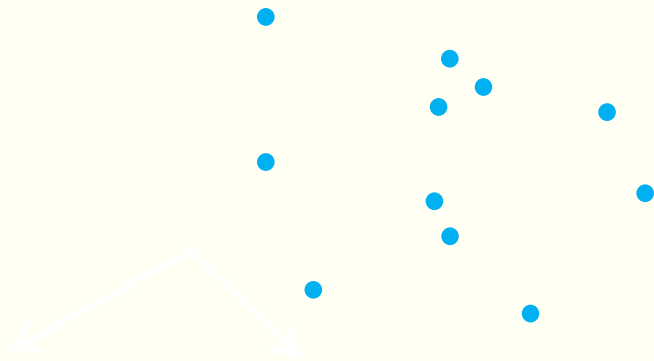
On the random background, functional genotypes show long-range correlations along the fluid channel

10^6 solutions {
11100101011111000100111010101...
10010001000101011100110101110...
00111110111110100110111111010...
100111111110011001100101111000...
...
...



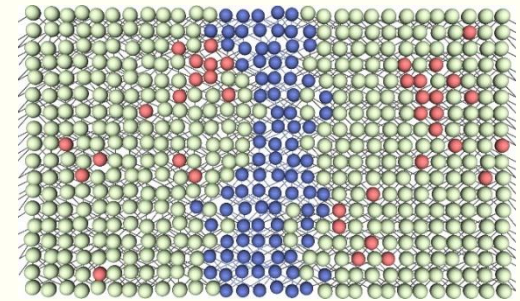
High-dimensional protein universe is extremely sparse, but evolution still easily solves for a functional mode

- Sequence space: $\text{dim} \sim \# \text{ amino acids}$
 $2^{2700} \sim 10^{900}$ sequences.



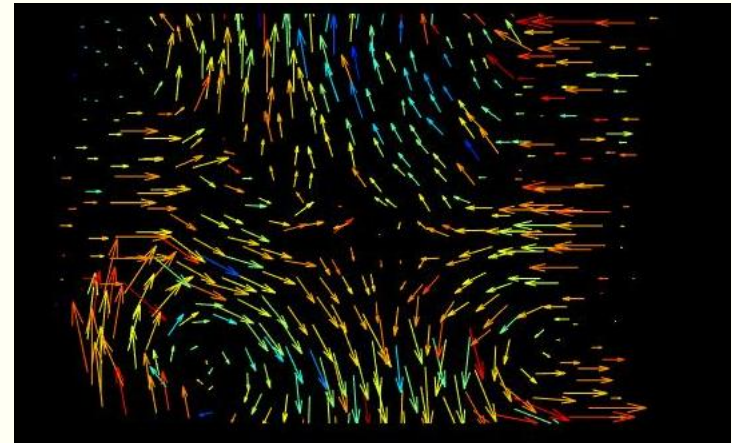
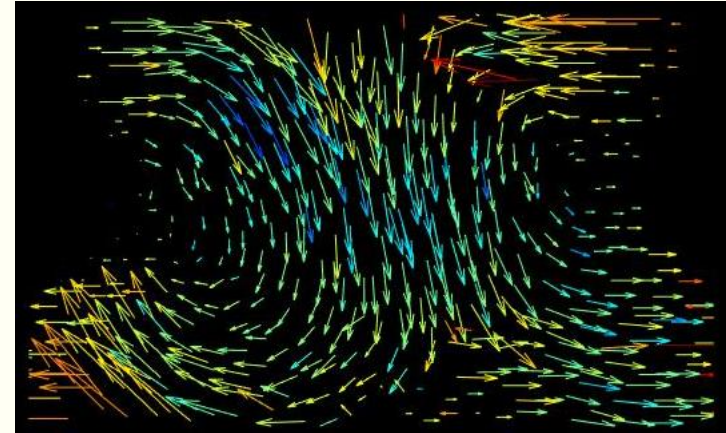
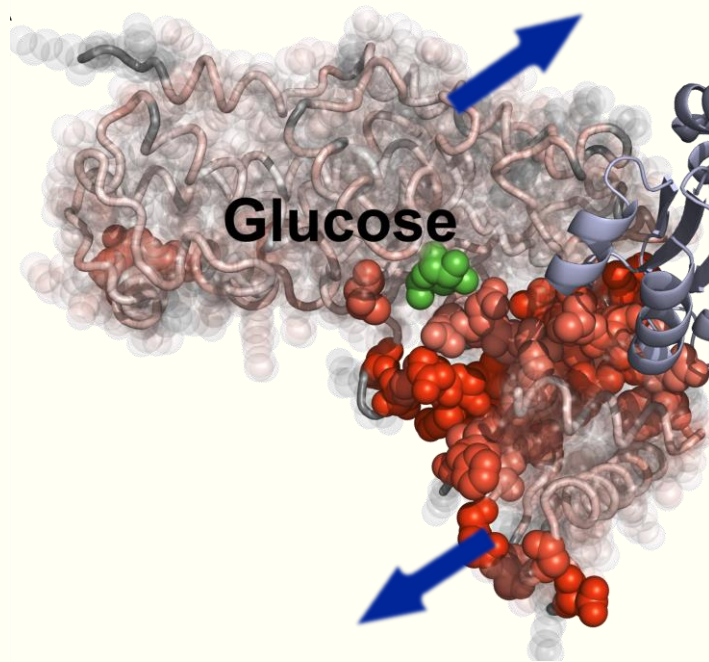
```
11100101011111000100111010101...
10010001000101011100110101110...
0011111011111010011011111010...
10011111110011001100101111000...
```

- Configuration space: $\text{dim} \sim \# \text{ amino acids}$
 $3^{540} \sim 10^{250}$ configurations

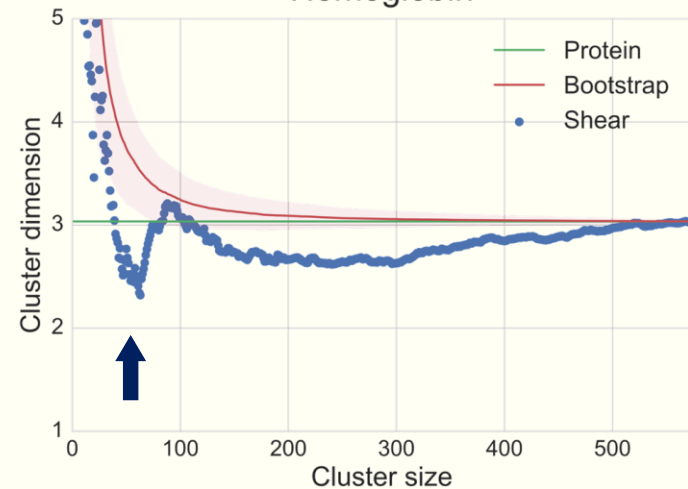
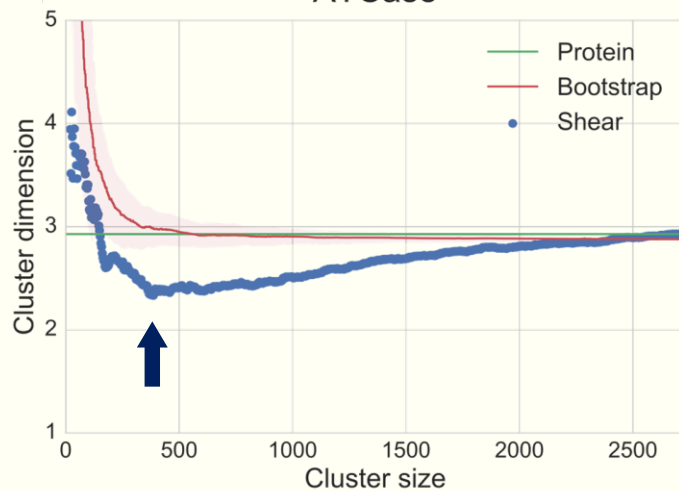
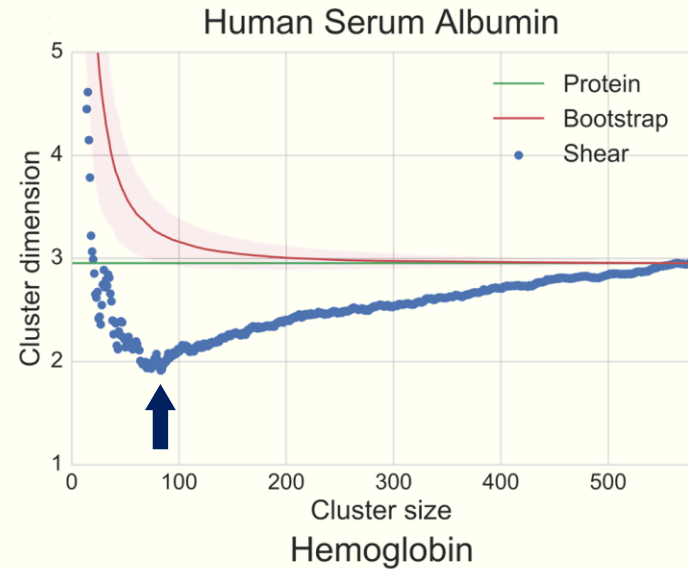
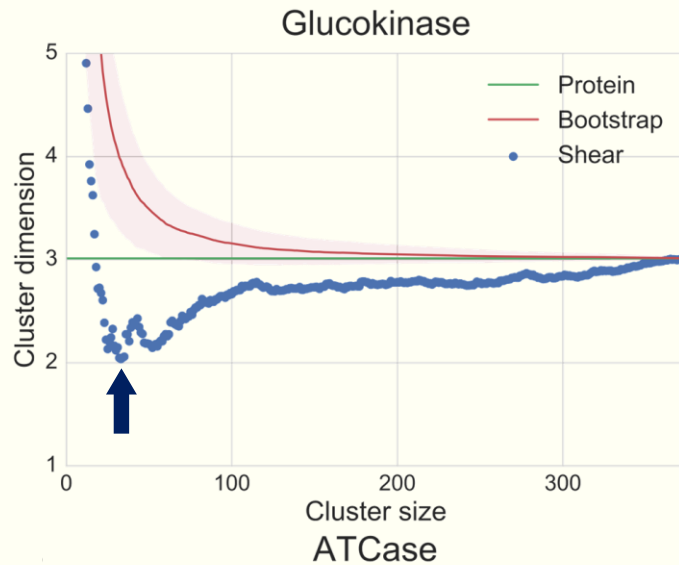


- Tiny fraction is functional: $\sim 10^{-240}$.

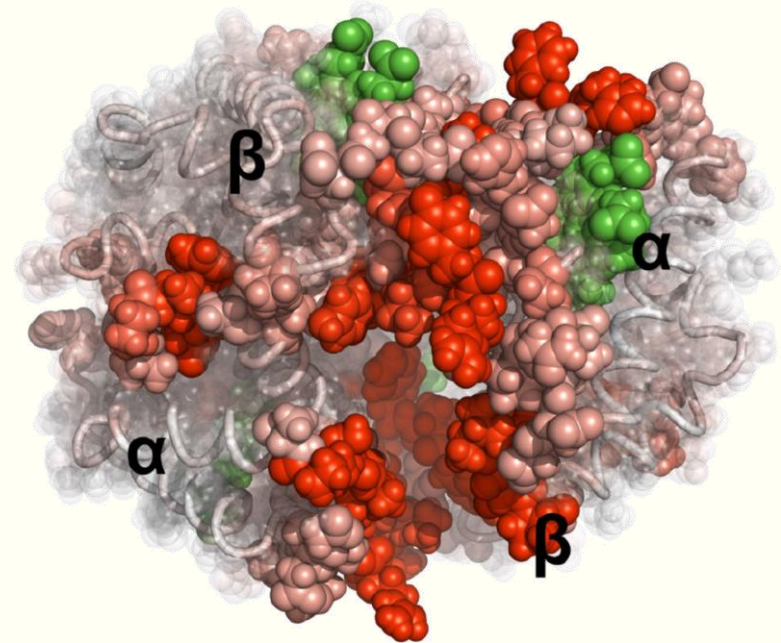
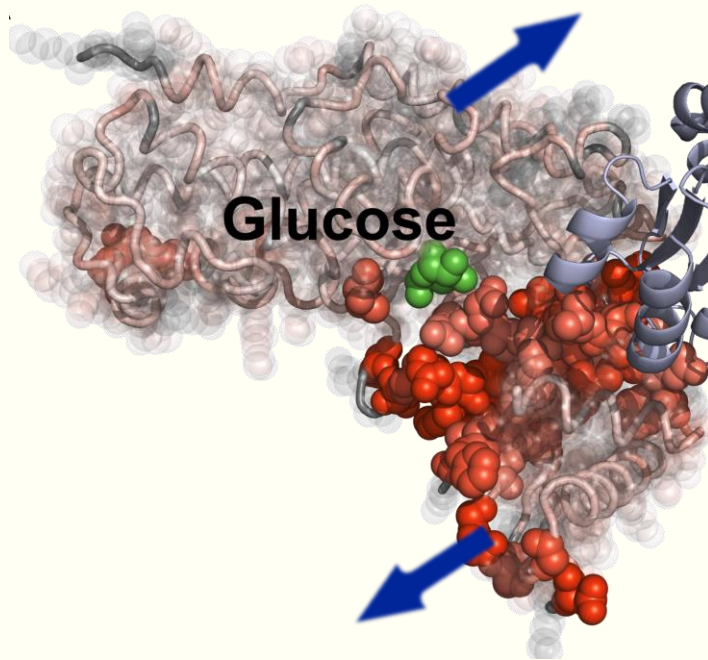
Formation of continuous shear band induces low frequency mechanical modes



Dimension of high strain regions is about two



Hint: proteins exhibit quasi-2D shear surfaces that couple active and allosteric sites



- Glucokinase shear region connects regulatory and active sites

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