

What Is Protein?

A geometric view

J.-P. Eckmann (U. Geneva) & T.T.



Solvay 2016

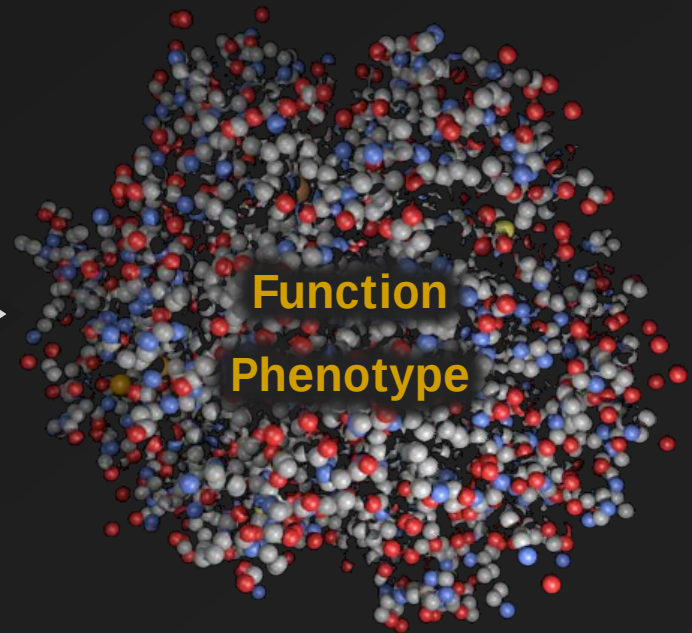
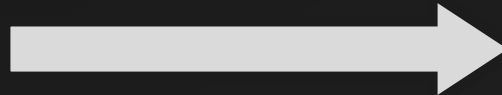


Protein =

1D DNA gene → 3D *functional* aggregate of amino acids

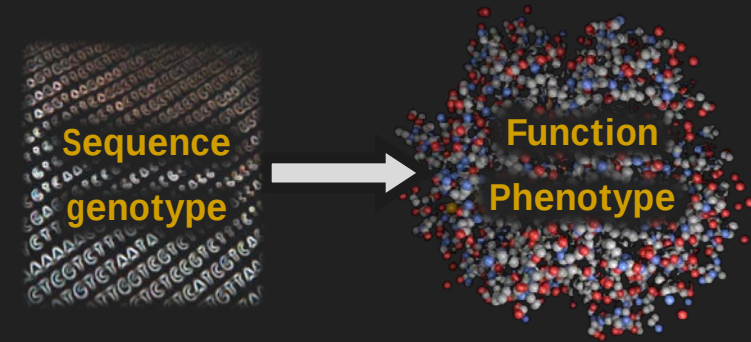


?



Deciphering sequence-to-function map is a hard problem

- **Collective** , long-range, many-body interactions.
- **Sparse** functional sequences
 << high-dim sequence universe:

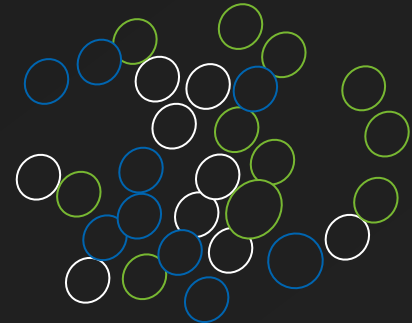


- protein is **non-generic**, information-rich phase (stat-mech?)
- Slow evolution: not enough statistics.



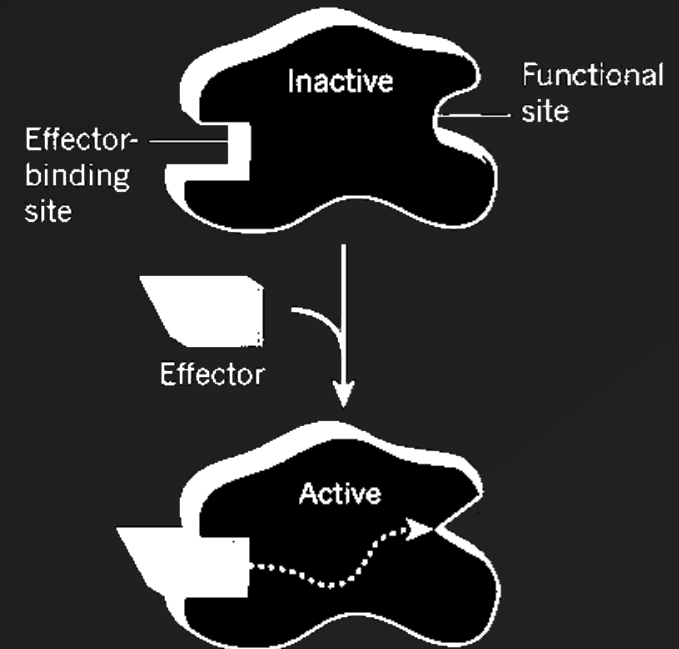
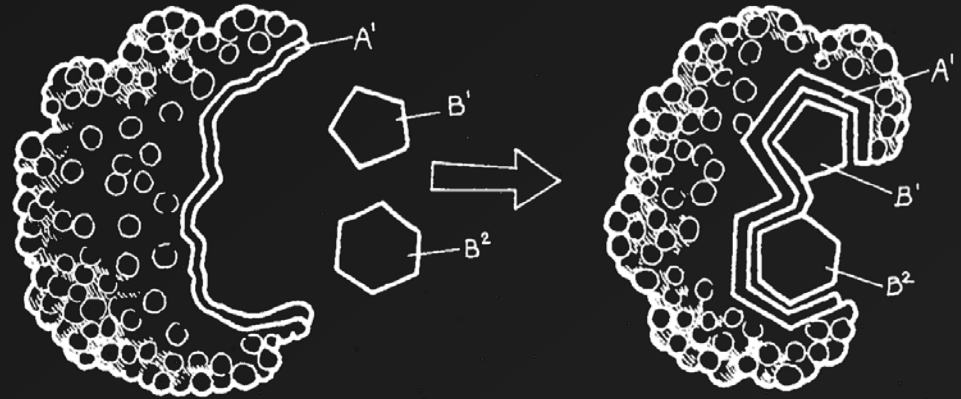
Outline: protein as learning amorphous material

- **Model:** Gene evolves to express *mechanical* phenotype (“floppy” modes).
- Explicit gene-to-function map: survey of protein universe geometry.
- Huge **dimensional reduction** from gene to function.
- Spectral correspondence: mechano-genetic duality of protein.



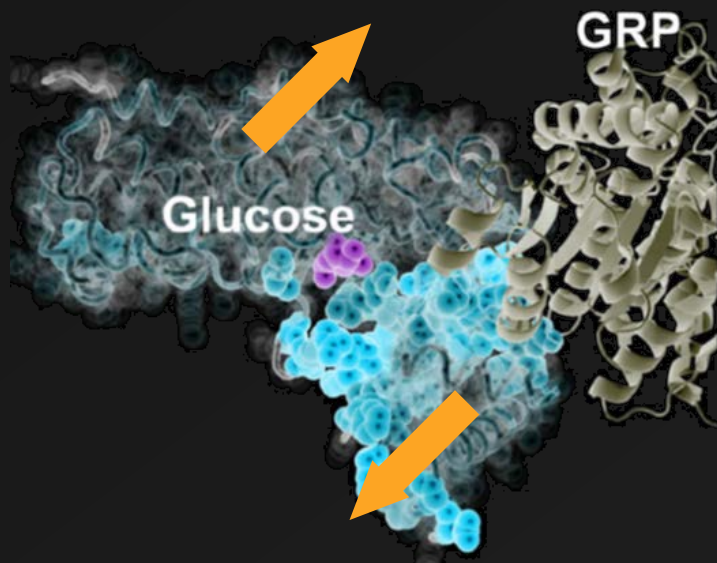
Hint 1: large-scale long-range motions are central to function

- Induced fit = **large-scale** deformation of enzymes.
- Allostery = **long-range** regulatory signals between distant sites.
- Ion channels, ribosomes, many others...



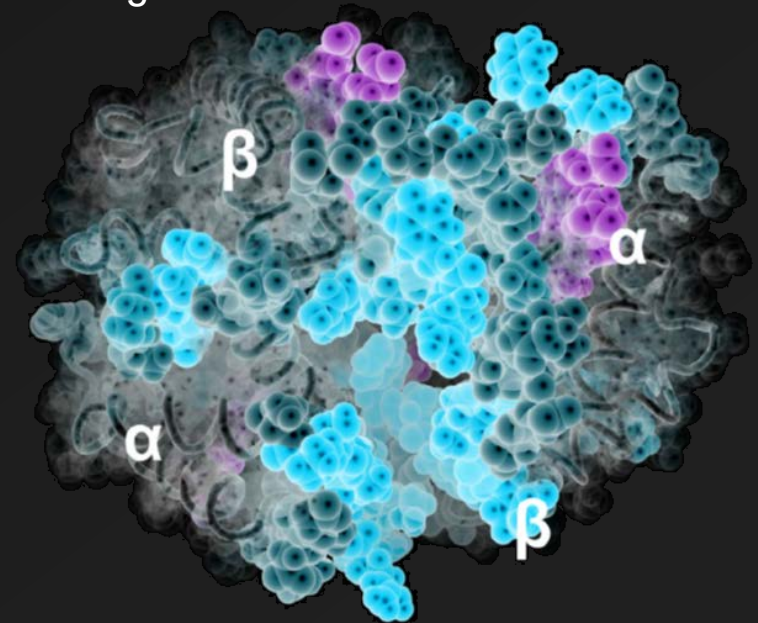
Hint 2: proteins exhibit ~2D “shear bands” that couple active and allosteric sites

- Glucokinase



shear region connects
regulatory and active sites.

- Hemoglobin



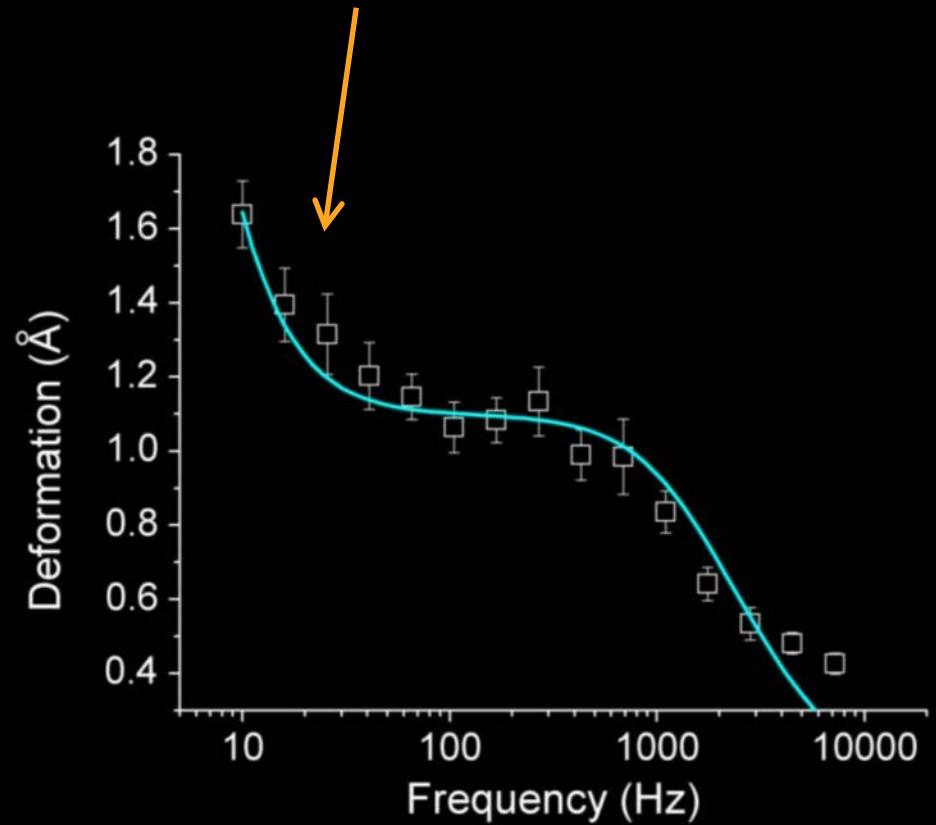
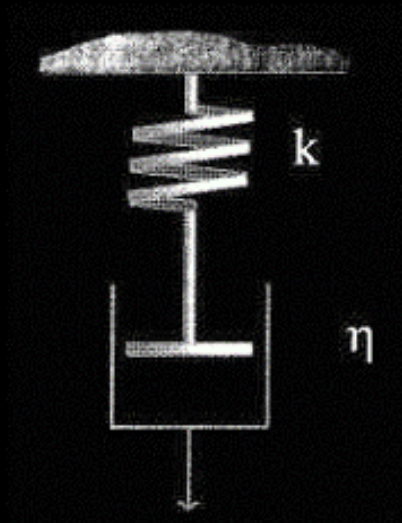
$$\varepsilon = \frac{d\vec{u}}{dx}$$

shear communicates mechanical
signal between subdomains.

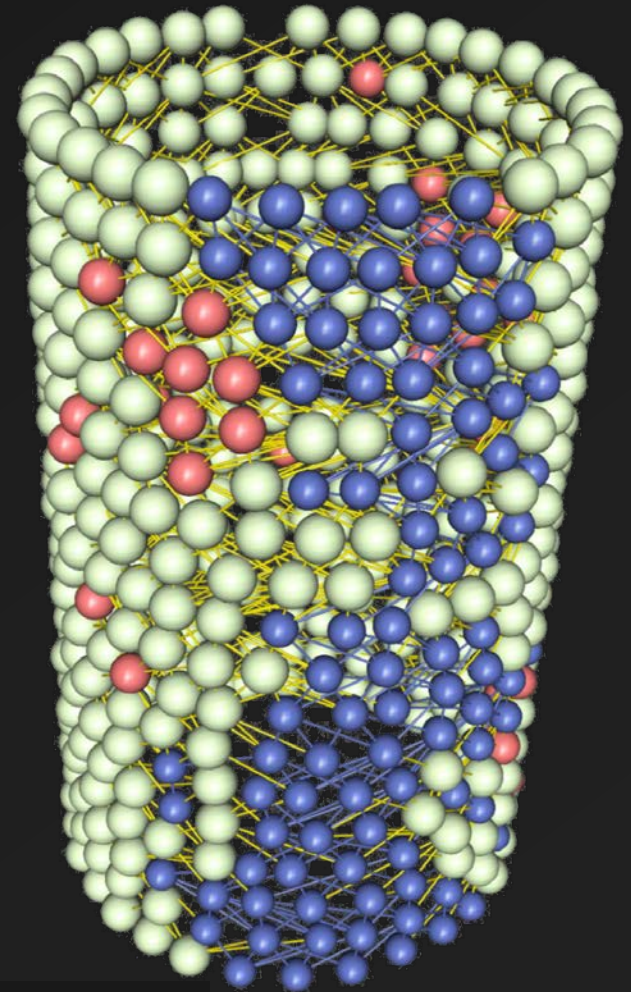
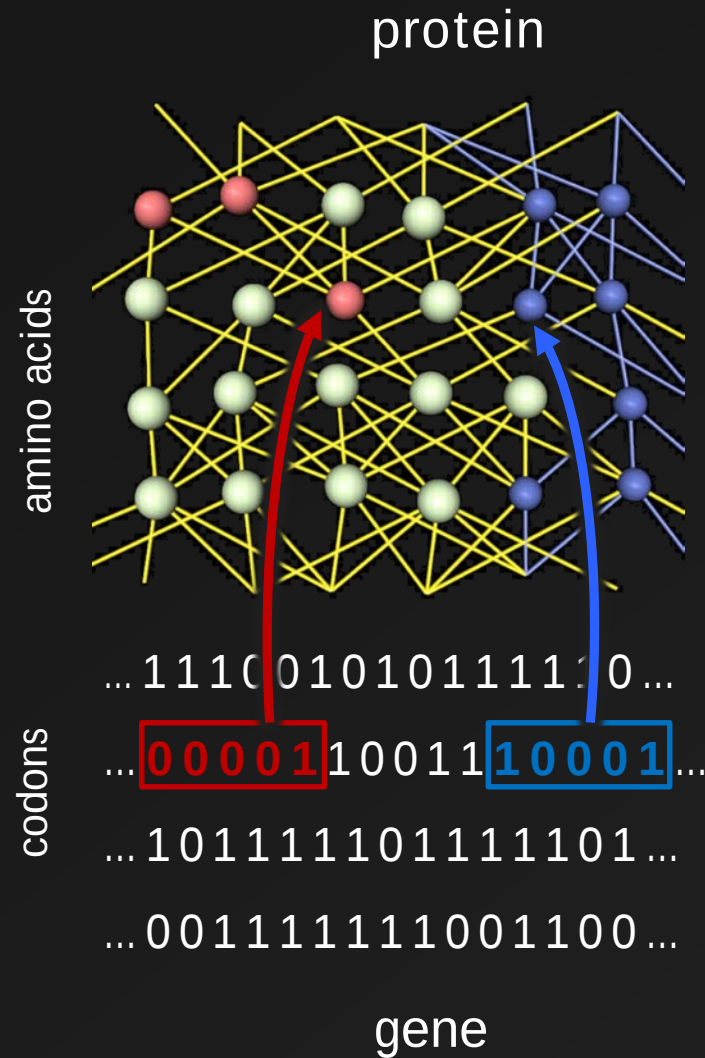
(Mitchell, Leibler & T.T., in review)

Hint 3: low-energy viscous modes in protein

Viscous response $\omega \leq 10^2$ Hz
Zocchi et al. (2012)



Protein as learning amorphous matter



$30 \times 18 = 540$
amino acids (aa)

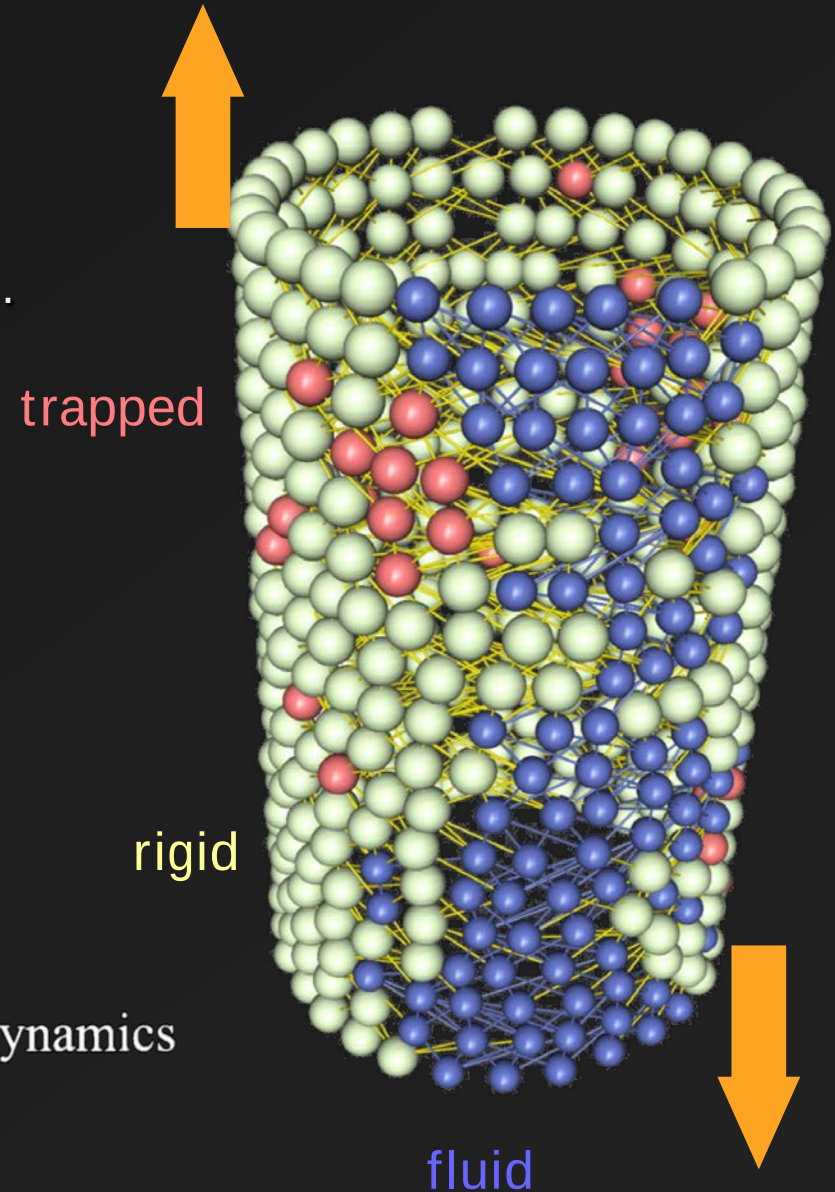
Protein as learning amorphous matter

- Protein adapts AA network to exhibit specific 0 energy mode via “shear band”.
- “Floppy” modes:

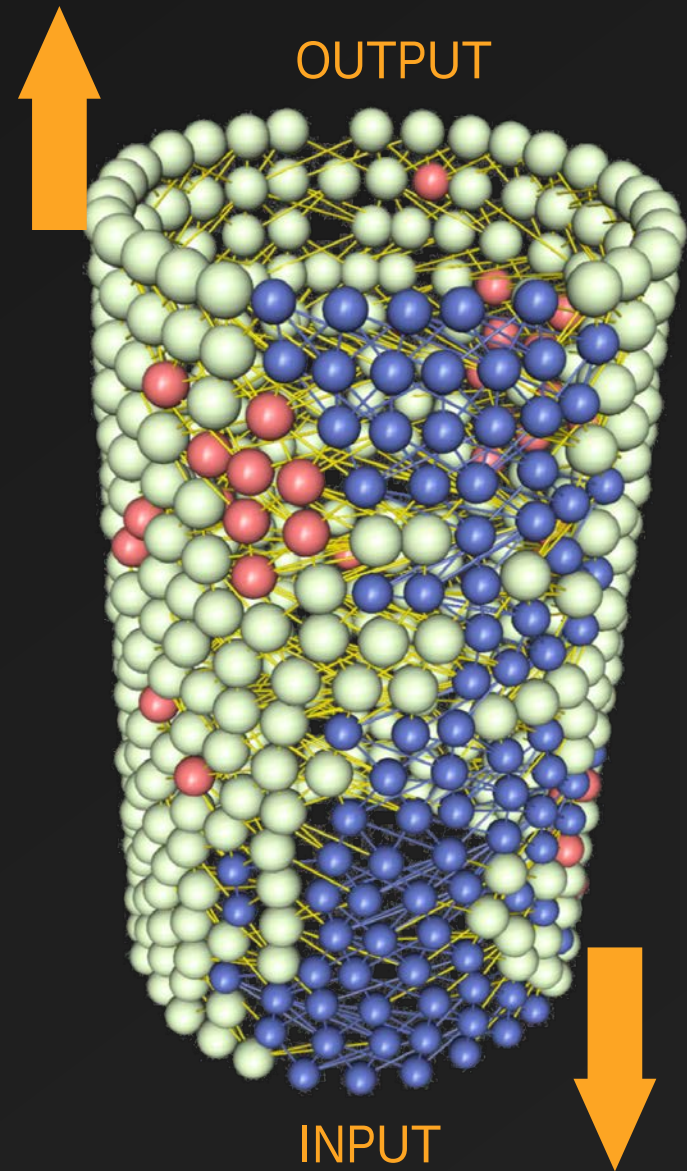
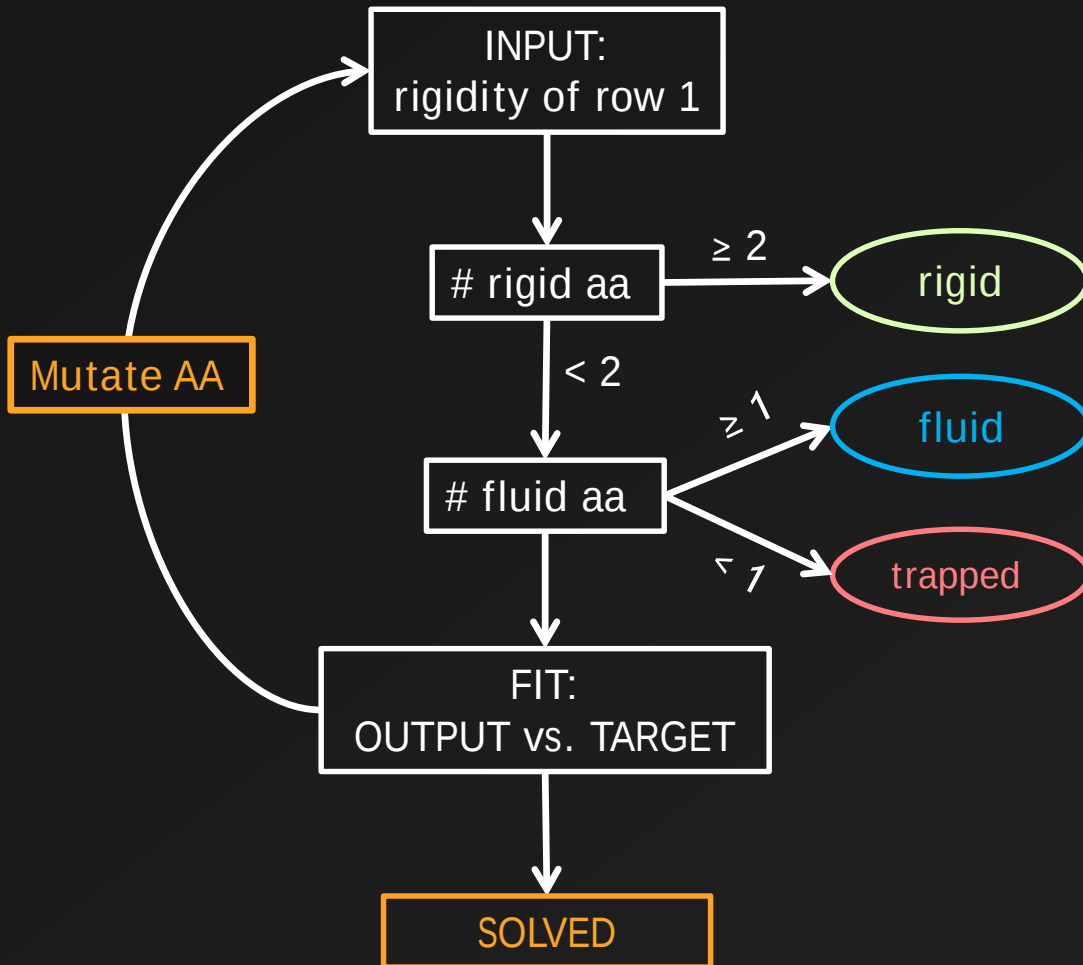
Physics Reports 296 (1998) 65–236

Amorphous solids: their structure, lattice dynamics
and elasticity

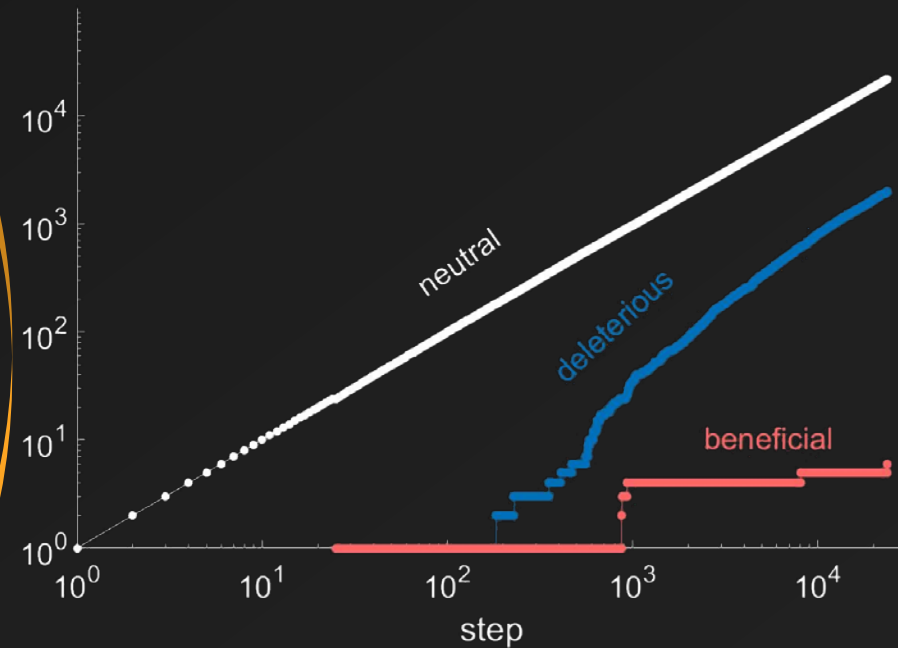
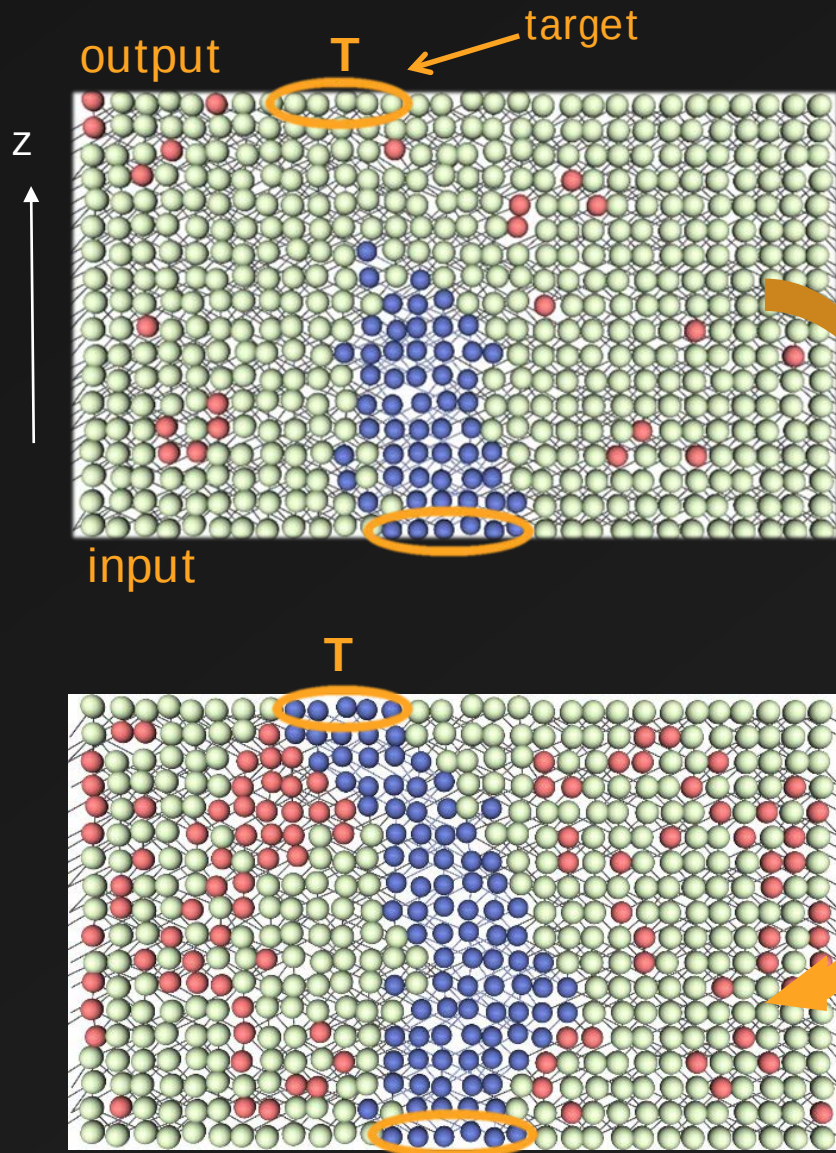
Shlomo Alexander



Rules of the protein game

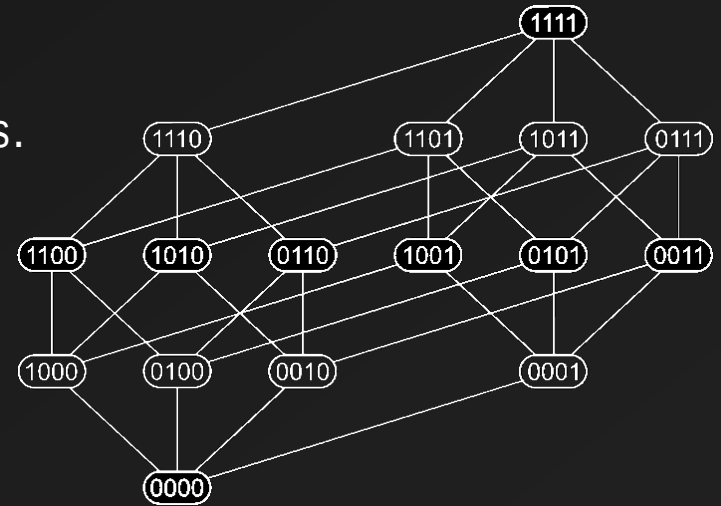


Evolution converges to solution by trial and error



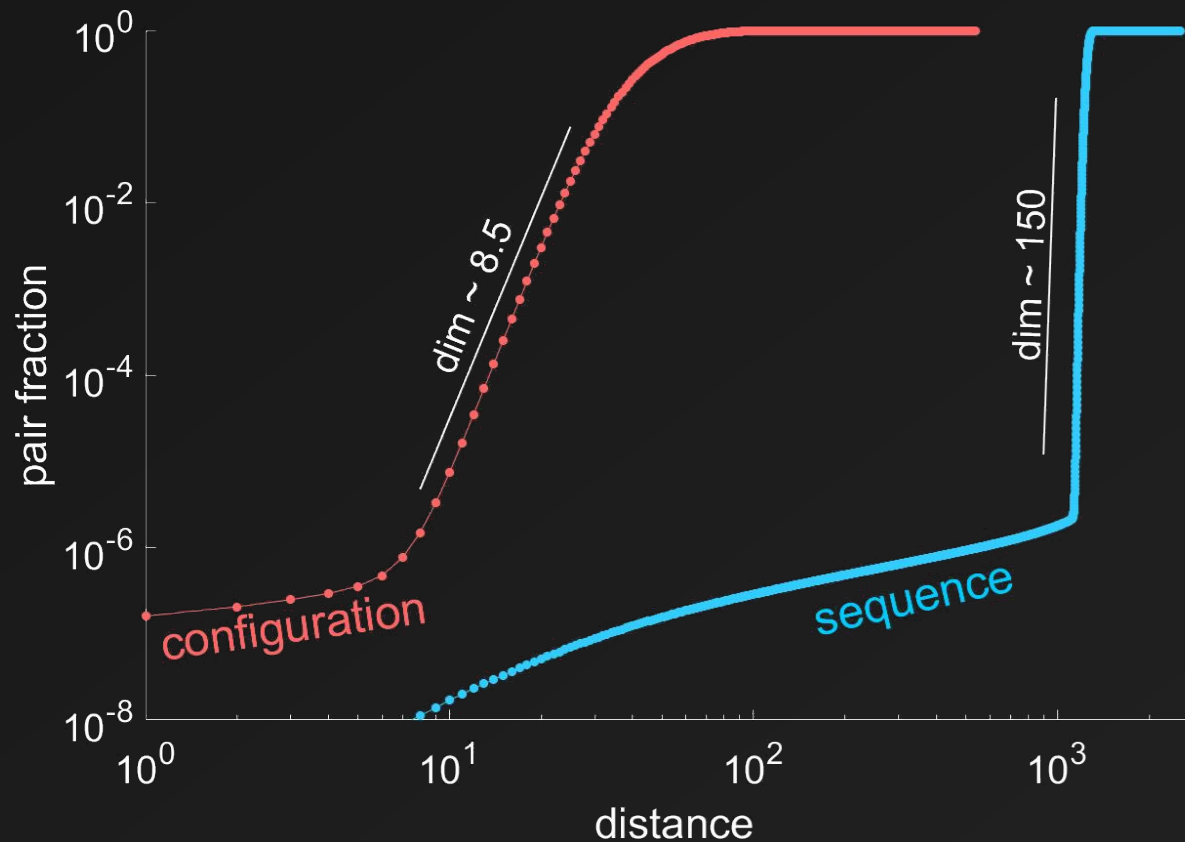
High-dimensional protein universe is extremely sparse

- Sequence space: $\text{dim} \sim \# \text{ amino acids}$
 $32^{540} \sim 10^{810}$ sequences.



- Configuration space: $\text{dim} \sim \# \text{ amino acids}$
 $3^{540} \sim 10^{250}$ configurations.
- Tiny fraction is functional: $\sim 10^{-240}$.

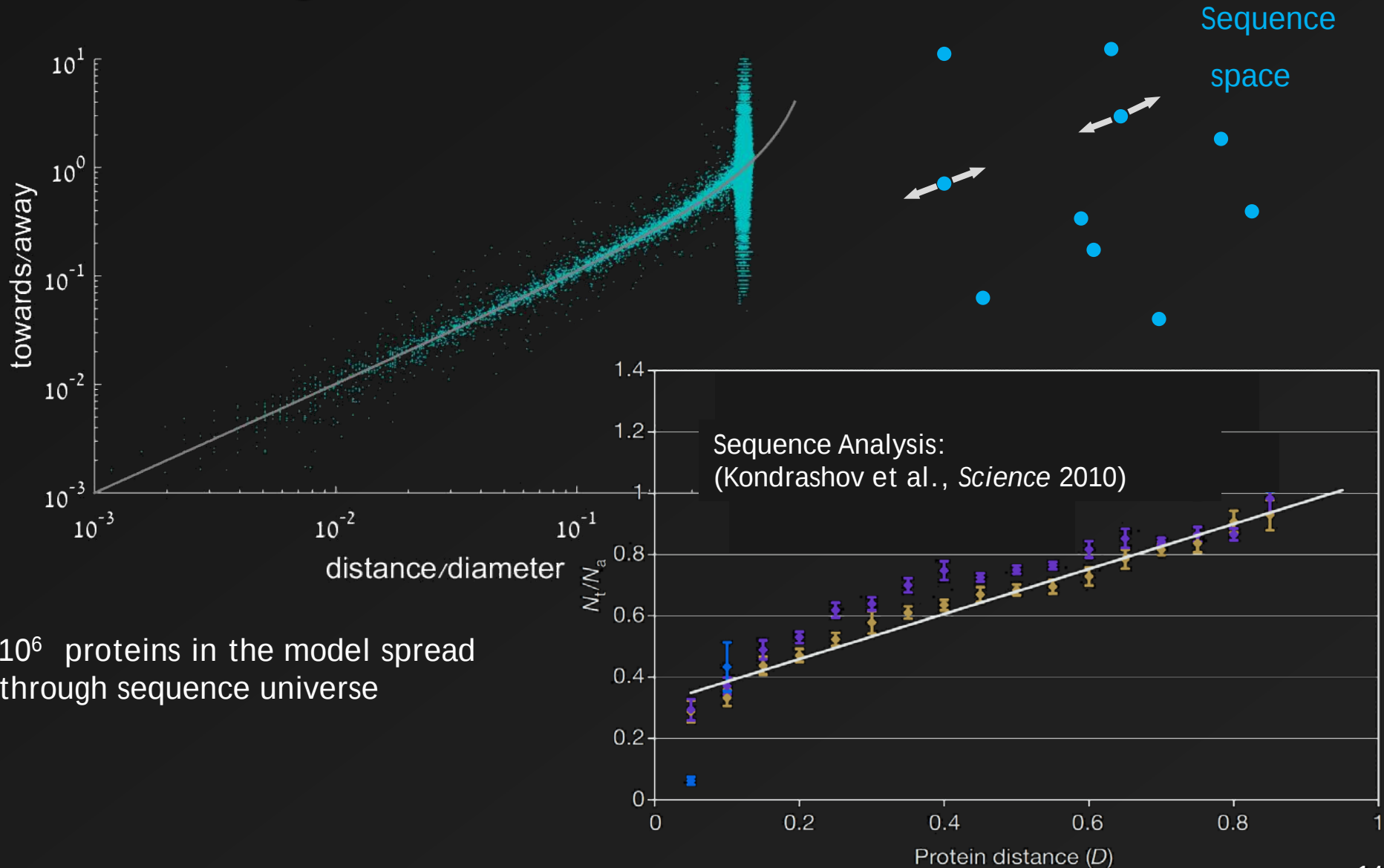
Model enables massive survey of protein universe: map involves huge dimensional reduction



$$\text{dim} = \frac{d\log(\text{pair \#})}{d\log(\text{distance})}$$

Correlation dimension (Procaccia & Grassberger)

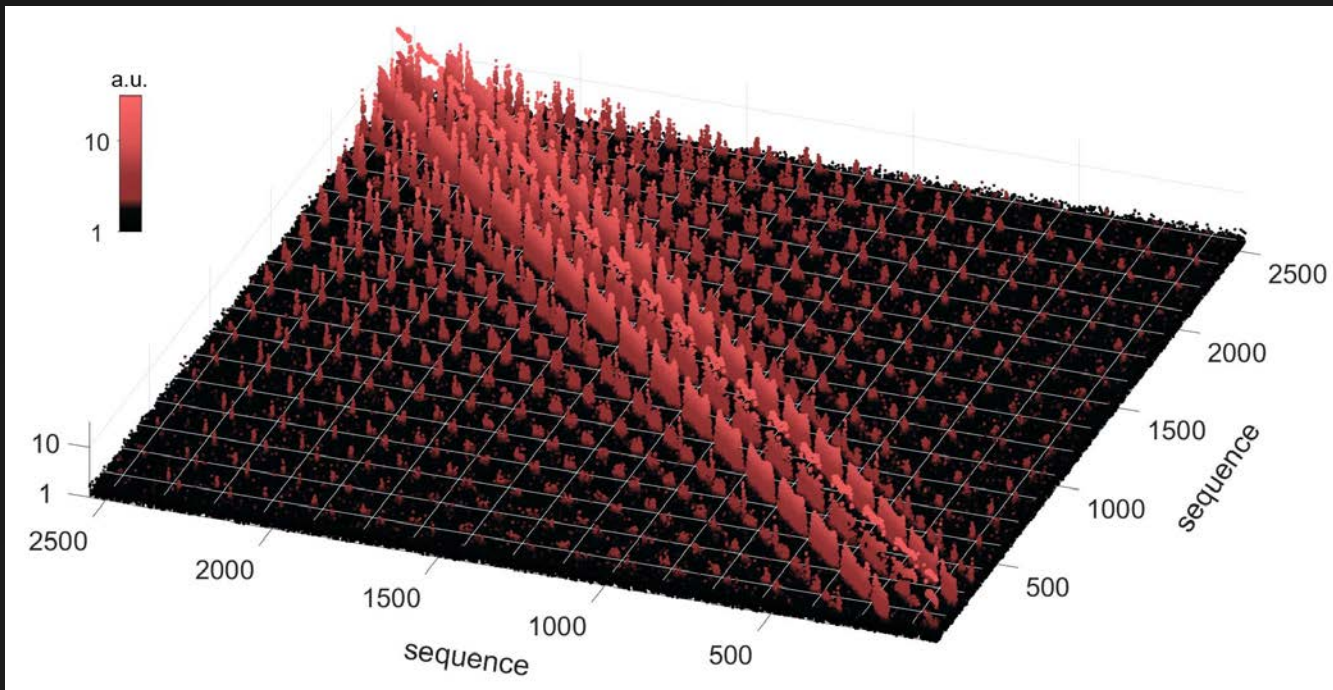
protein evolution surveys the sequence universe for functional genes



Gene sequences exhibit long-range correlations among solutions

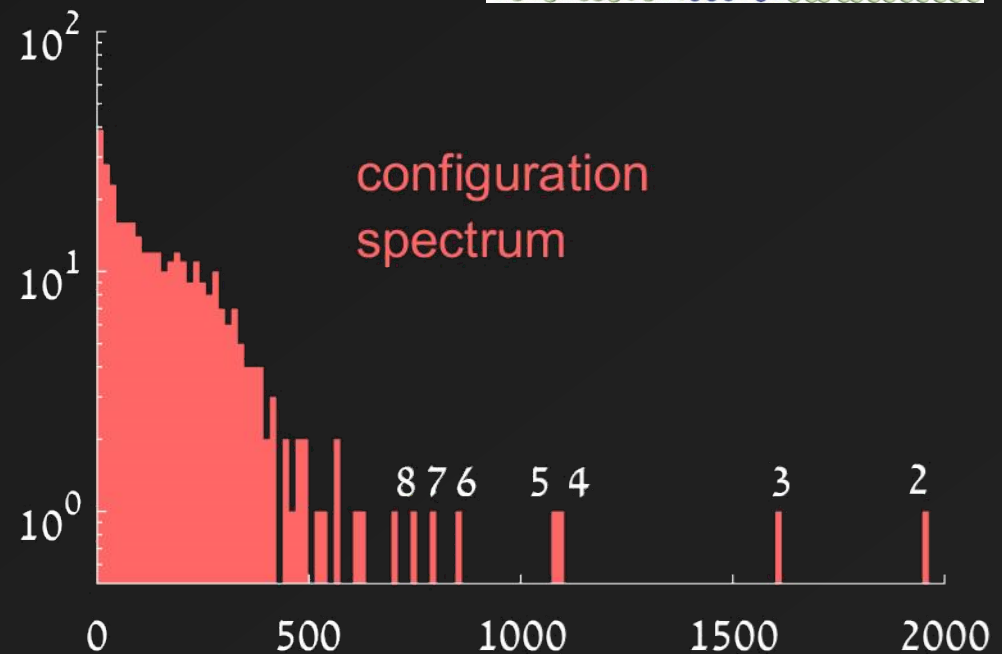
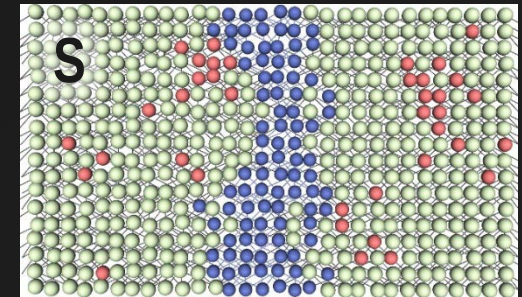
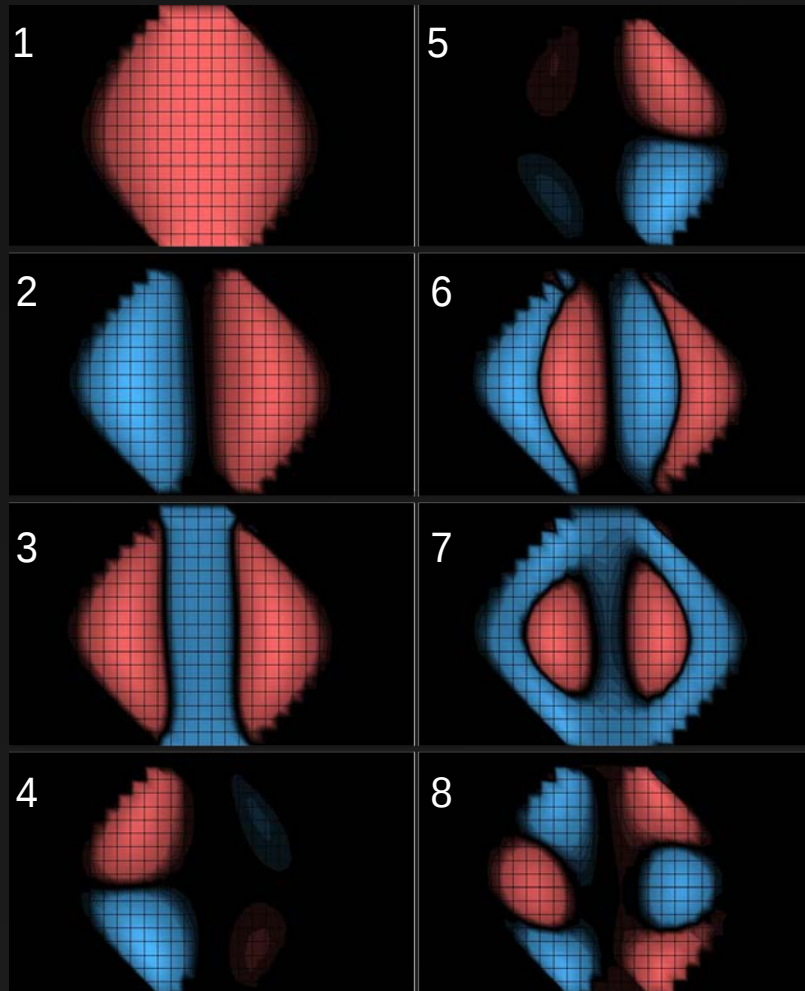
10^6 solutions

1 1 1 0 0 1 0 1 0 1 1 1 1 1 0 0 0 1 0 0 1 1 1 0 1 0 1 0 1 ...
1 0 0 1 0 0 0 1 0 0 0 1 0 1 0 1 1 1 0 0 1 1 0 1 0 1 1 1 0 ...
0 0 1 1 1 1 1 1 0 1 1 1 1 1 0 1 0 0 1 1 0 1 1 1 1 1 1 0 1 0 ...
1 0 0 1 1 1 1 1 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 1 1 1 1 0 0 0 ...



Configuration modes reflect continuity of mechanical forces

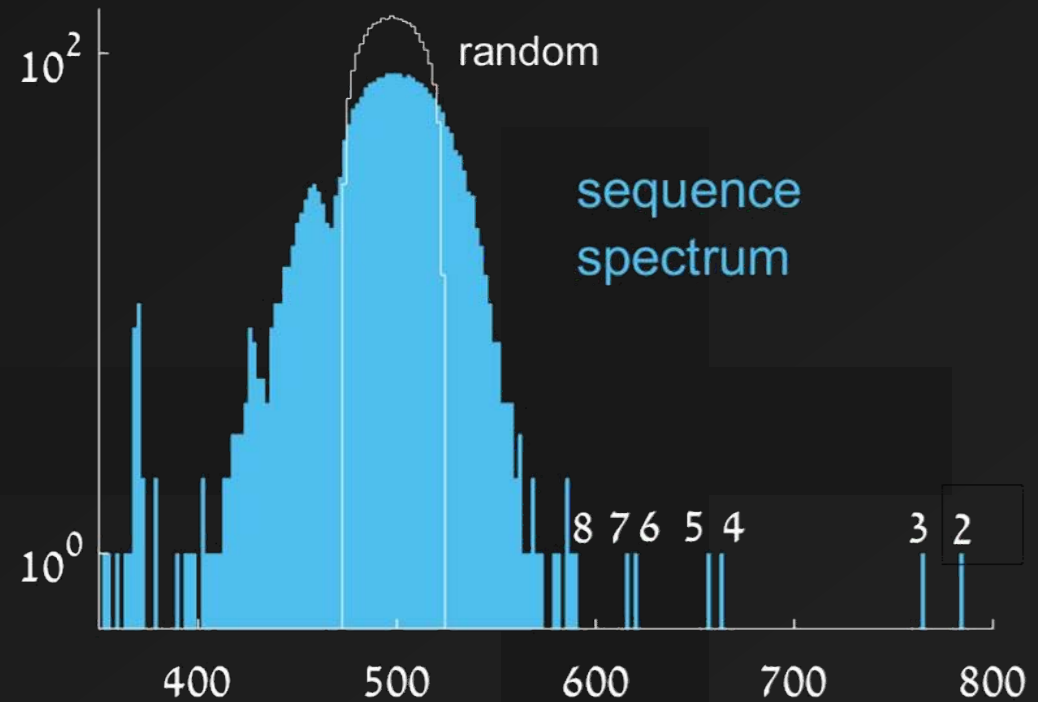
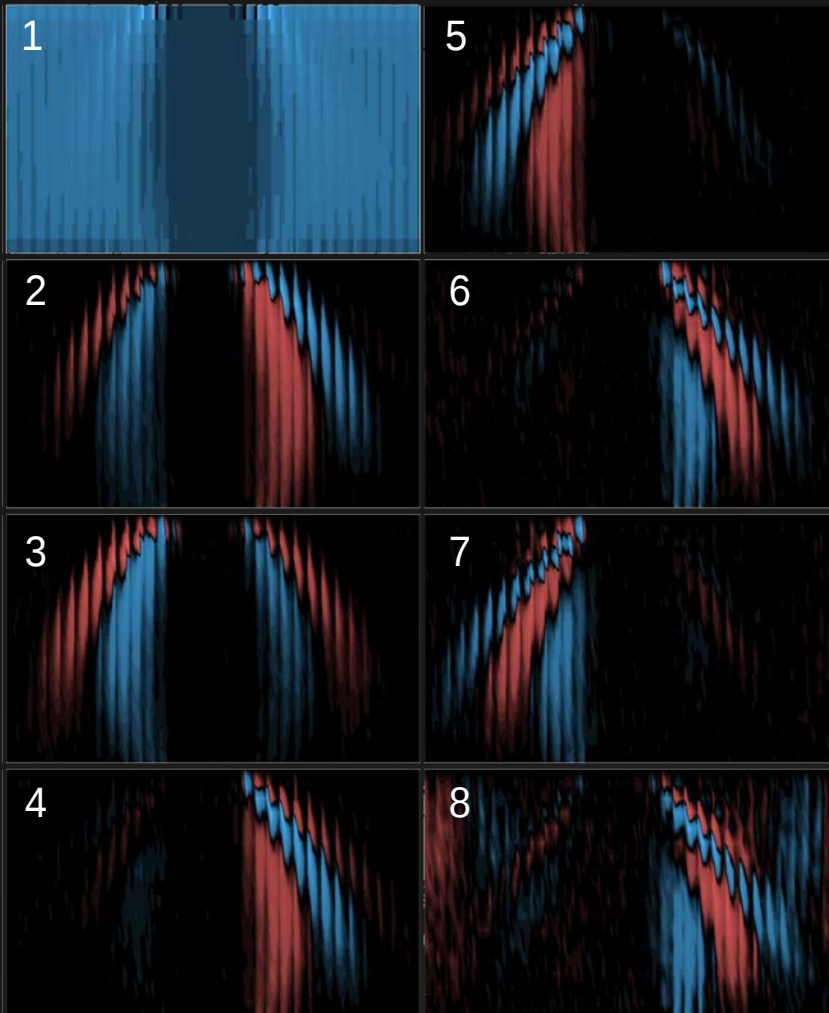
Singular value decomposition (SVD) spectrum



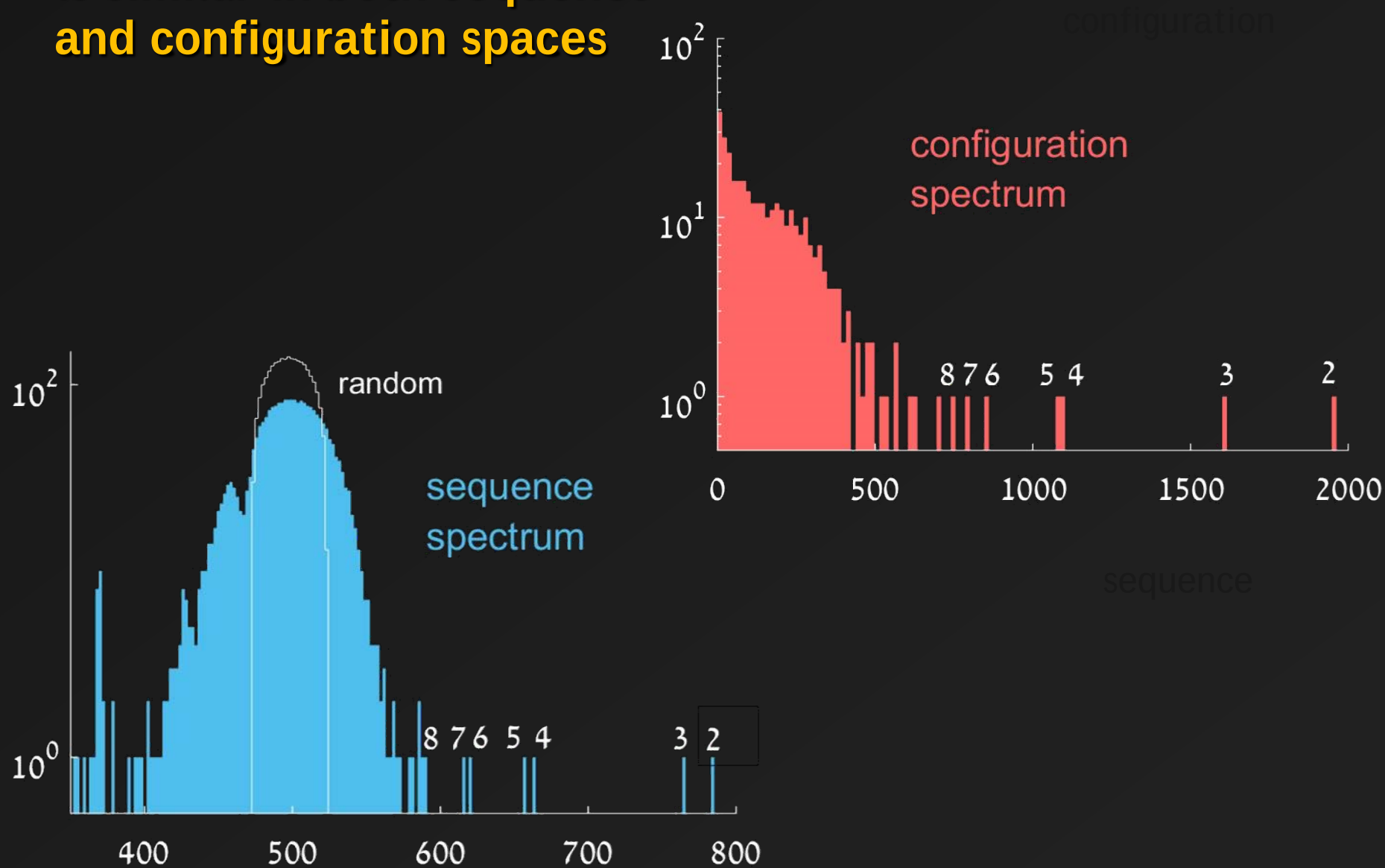
- Geometric constraints ~ number of independent degrees-of-freedom.

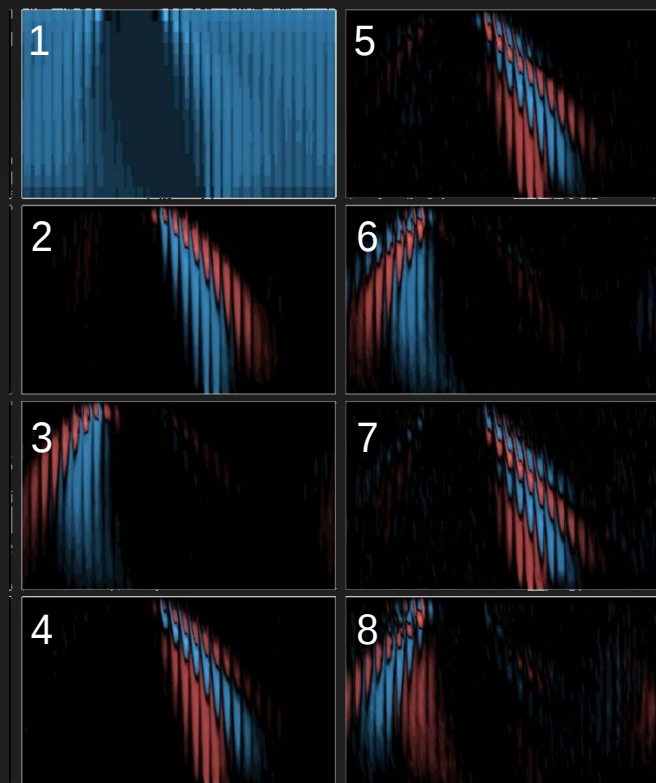
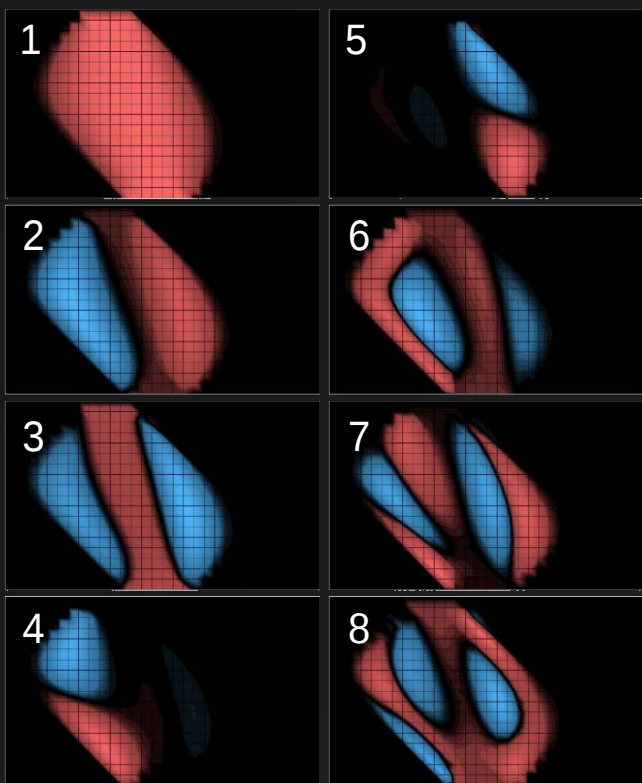
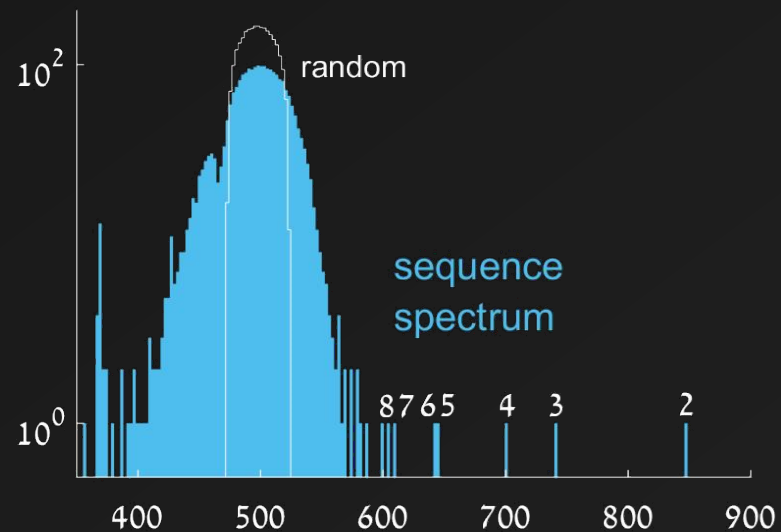
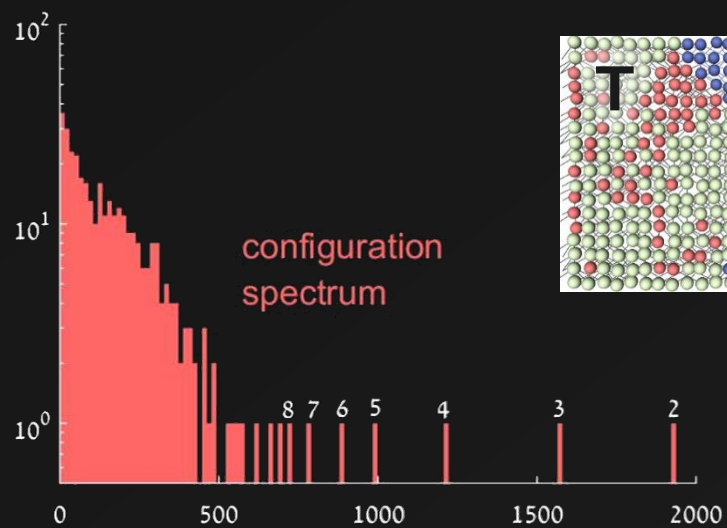
Sequence modes correspond to configuration modes

SVD spectrum:



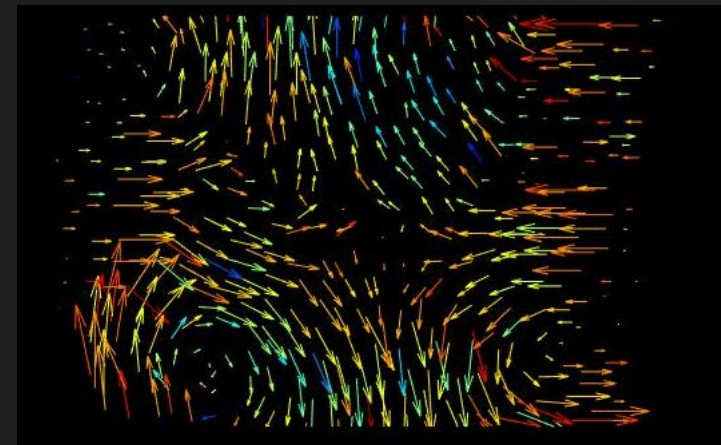
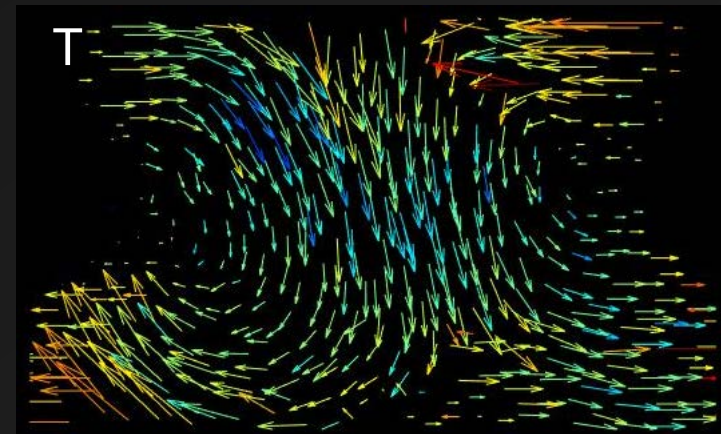
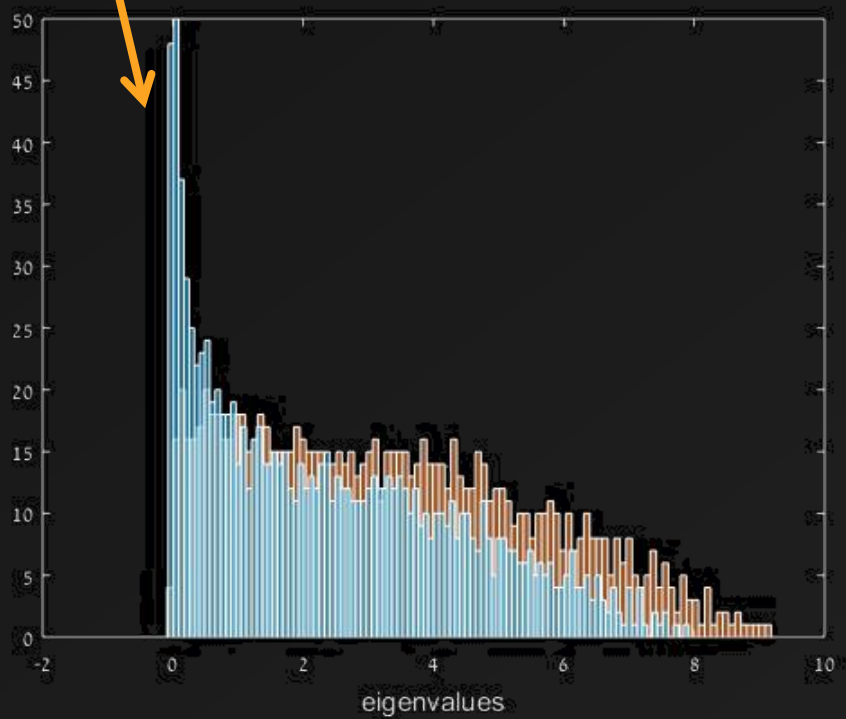
The information dimension is similar in both sequence and configuration spaces





Evolution induces low frequency mechanical modes

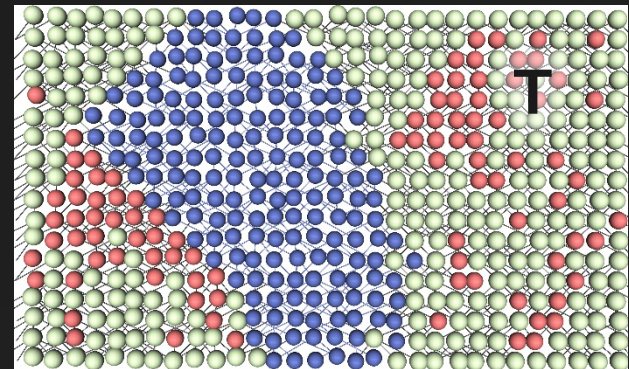
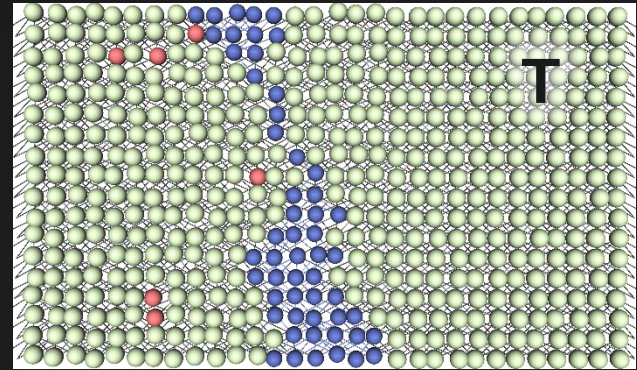
Low modes



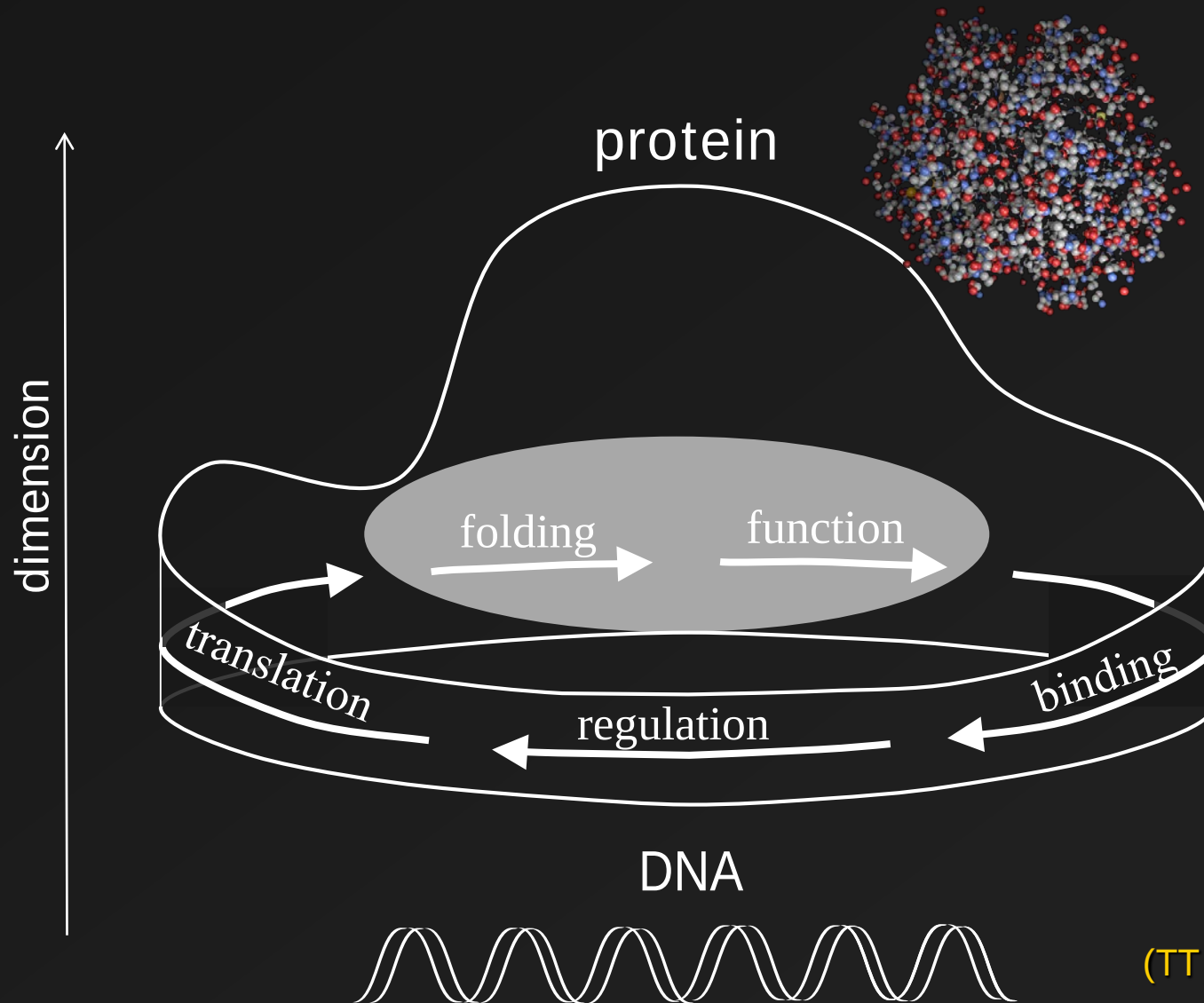
Solution space allows adaptation to multiple tasks

- Thermophiles: stability
- Cryophiles: flexibility

Temp

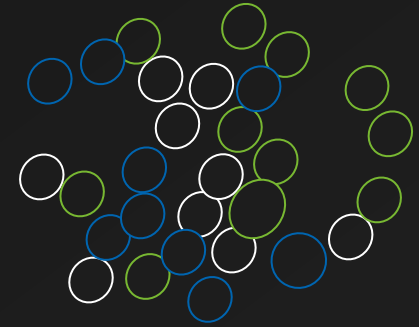


Protein synthesis and function involves huge dimensional expansion and reduction



(TT, Phil Trans 2016)

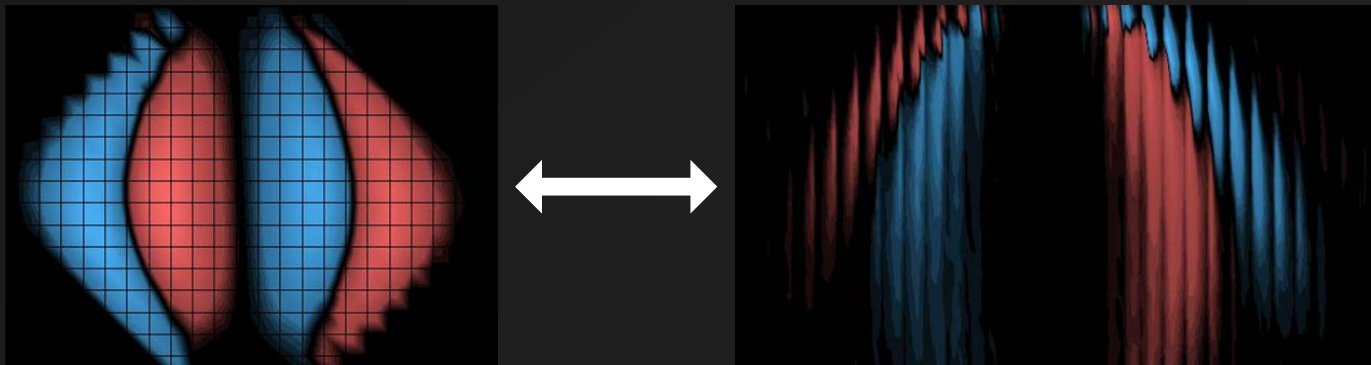
What is protein?

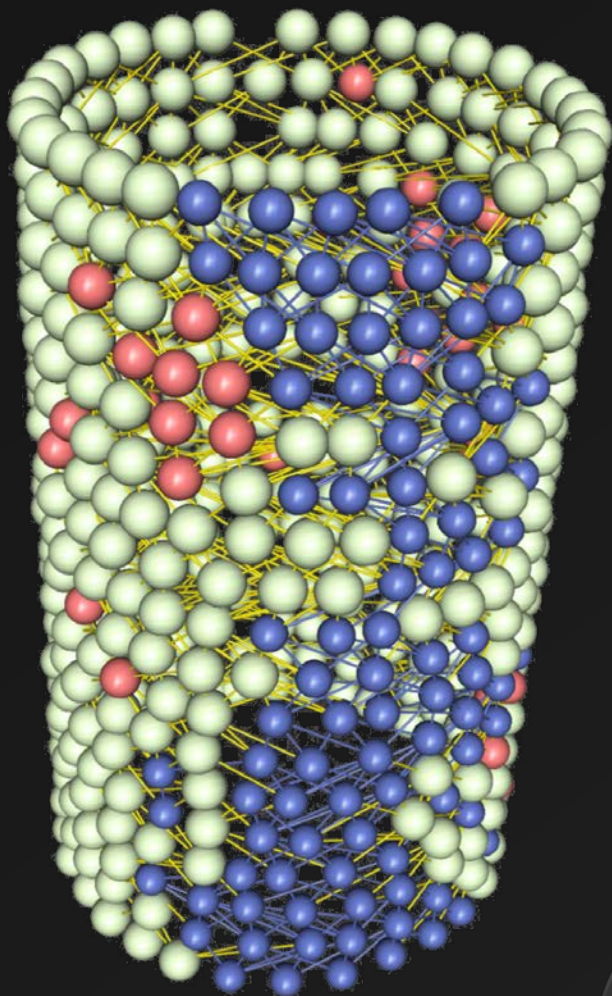


- *Learning, information-rich amorphous matter.*
- Mechanical function: *slow, “floppy” modes.*
- Huge dimensional reduction: # amino acids $\gg$ ~10 slow modes.
- **Mechanogenetic duality:**

Slow mechanical modes $\leftrightarrow$ information “*ripples*”

in almost-random sequence space.





THANK YOU

J.-P. Eckmann (U. Geneva)

S. Leibler & M. Mitchell (IAS)

