



FROM INFORMATION TO CELLS - I2CELL

February 20-24, 2018

Eynsham Hall, Oxfordshire, UK

Computation and Mechanics in Living Matter: The Protein Problem

Sandipan Dutta & TT (IBS)

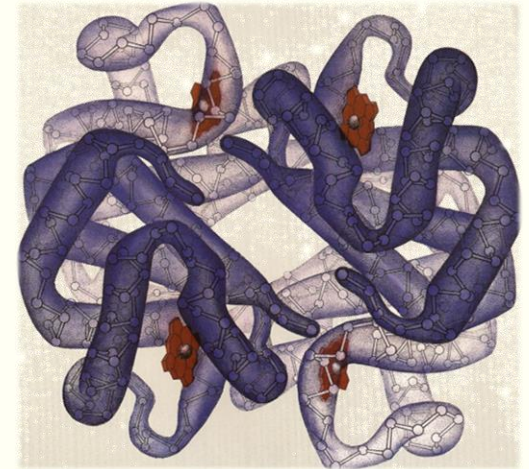
Albert Libchaber (Rockefeller)

Jean-Pierre Eckmann (Geneva)

In living information-processing systems, computational and physical aspects are entangled

[This was discussed in detail by Dowe and Danchin this morning]

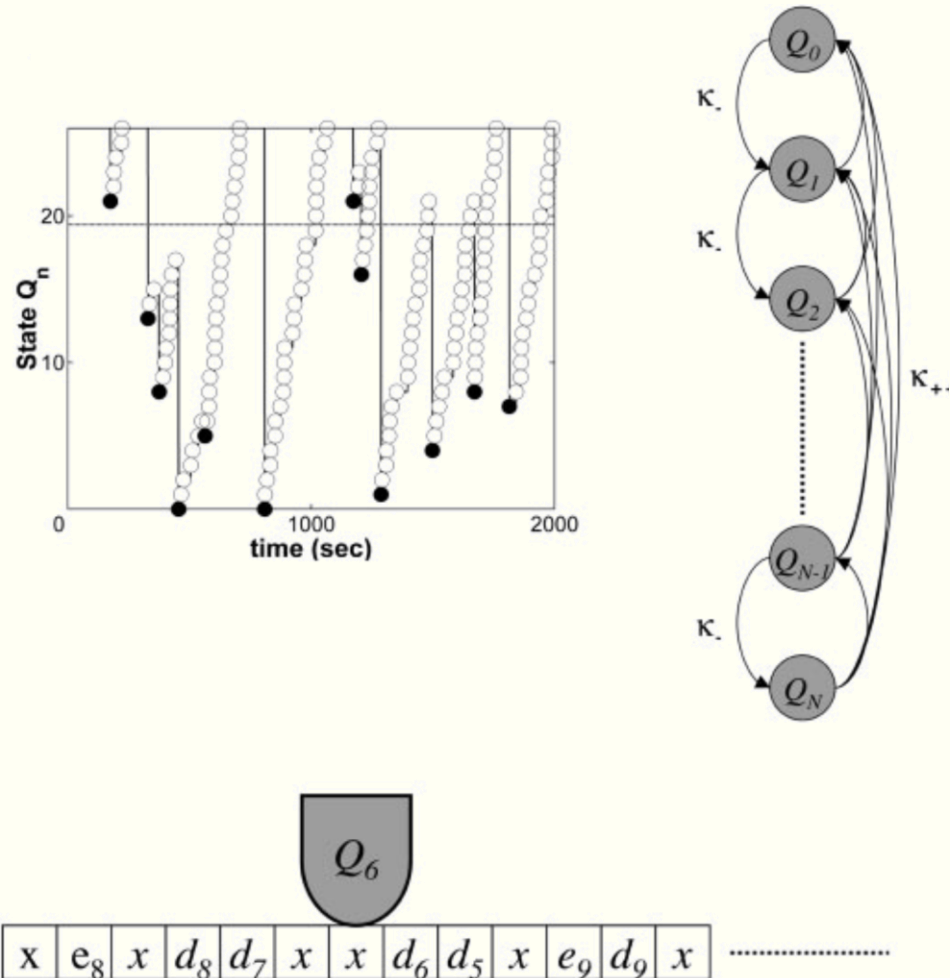
- Implemented in molecules – inherently **stochastic**, uses biochemical **recognition**.
- **Geometric space-time** considerations are essential **(Marshall)**
- Outcome of **evolutionary** process with long-term **memory**,
- Wherein **self-reproduction** is a major determinant of fitness **(Libchahber)**
- **Representation** and decoding by molecules and chemical process (especially compression)



Together with DNA and RNA,
protein is the smallest physical object where living and inanimate become discernable.

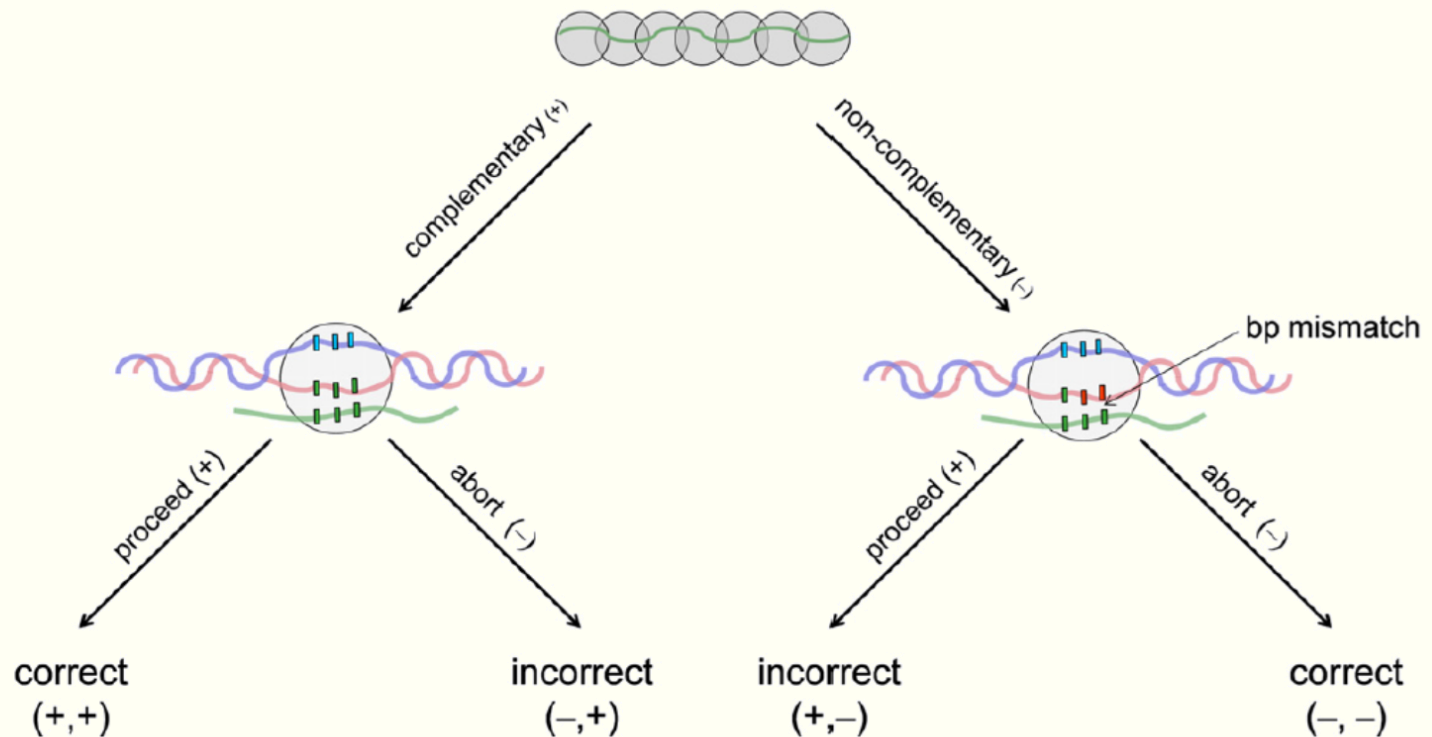
Looking at molecular information systems in computation terms may provide insight on function and design

- Protein–DNA computation utilizes **stochastic RecA assembly** cascade
= probabilistic finite-state machine.



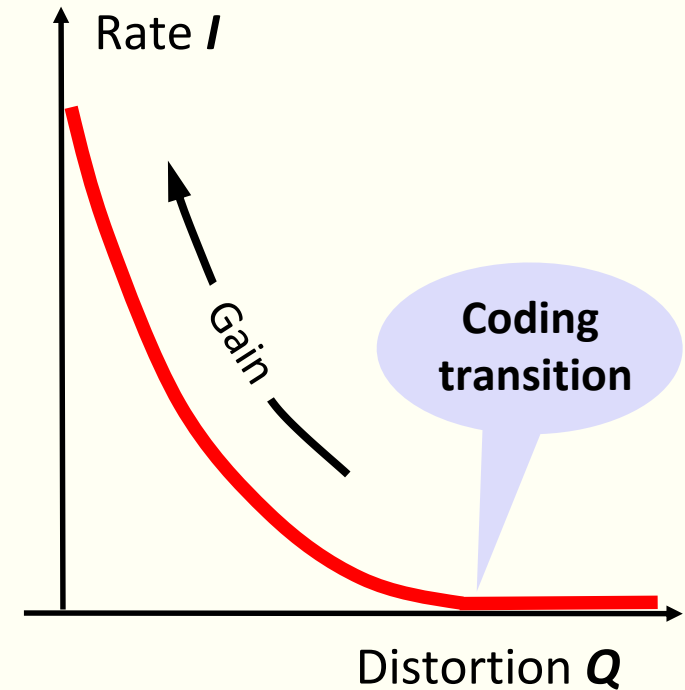
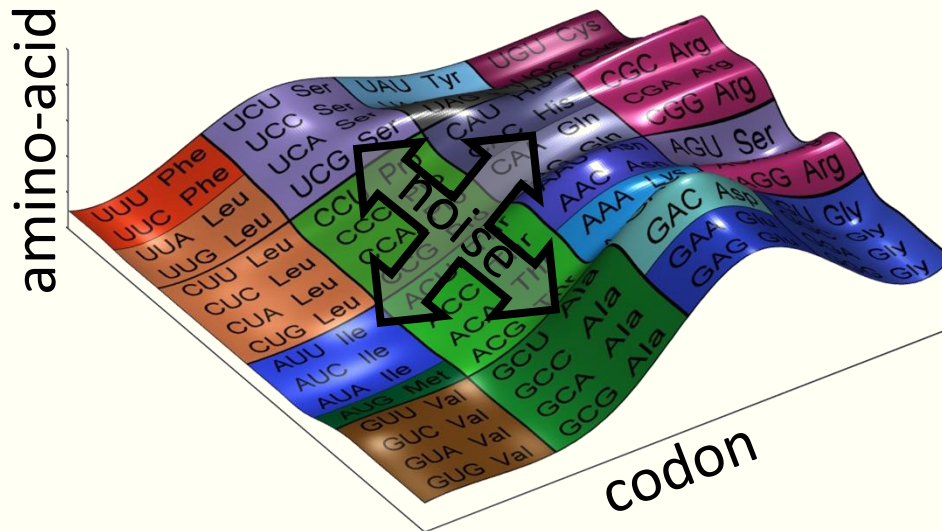
Mechanical deformation is used to tune the DNA homology detection system to maximal fitness

- RecA-mediated homology search as a nearly optimal signal detection system



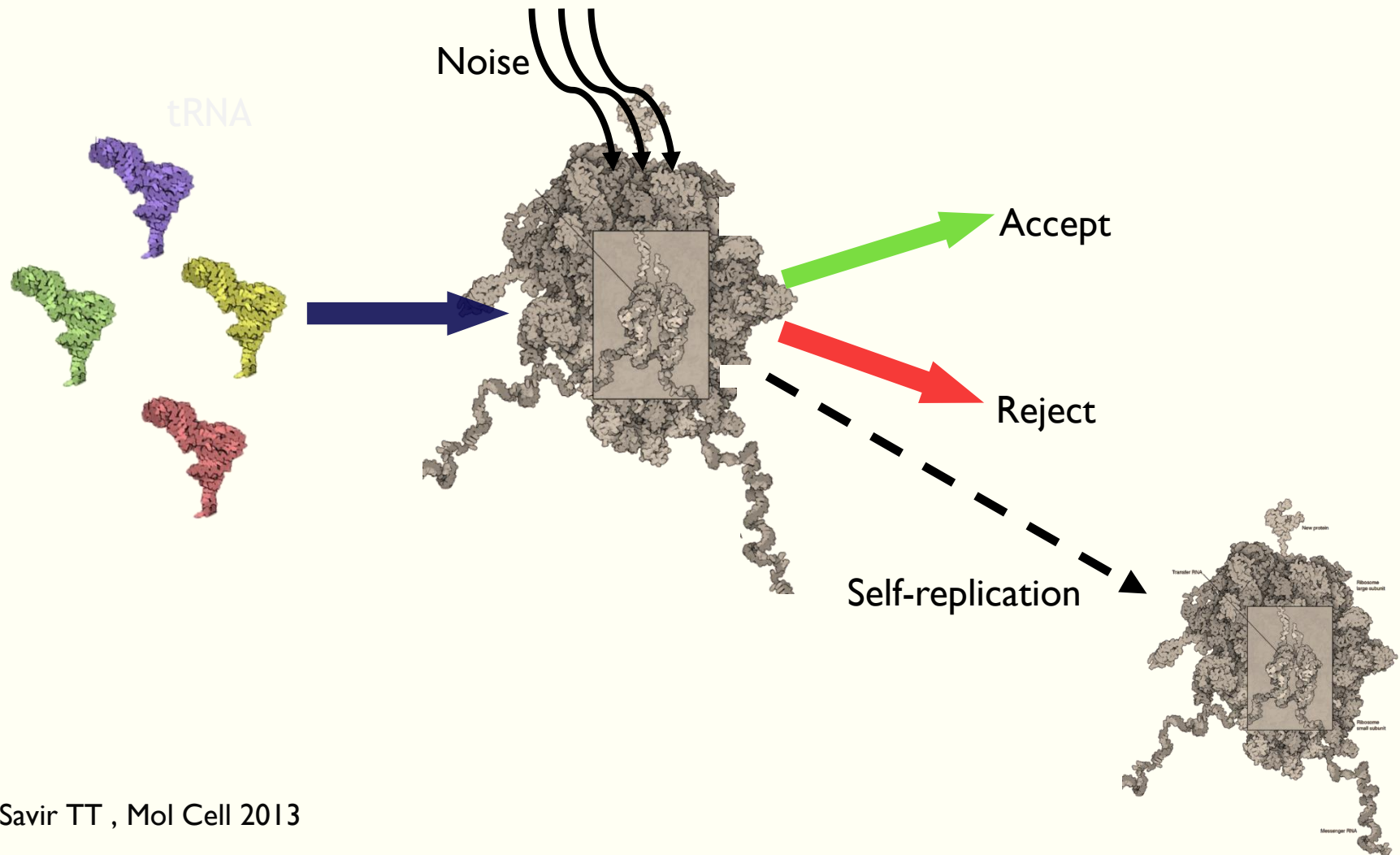
The genetic code is a noisy information channel emerging at a critical point on the rate-distortion curve

- The geometry of the **representation** problem.

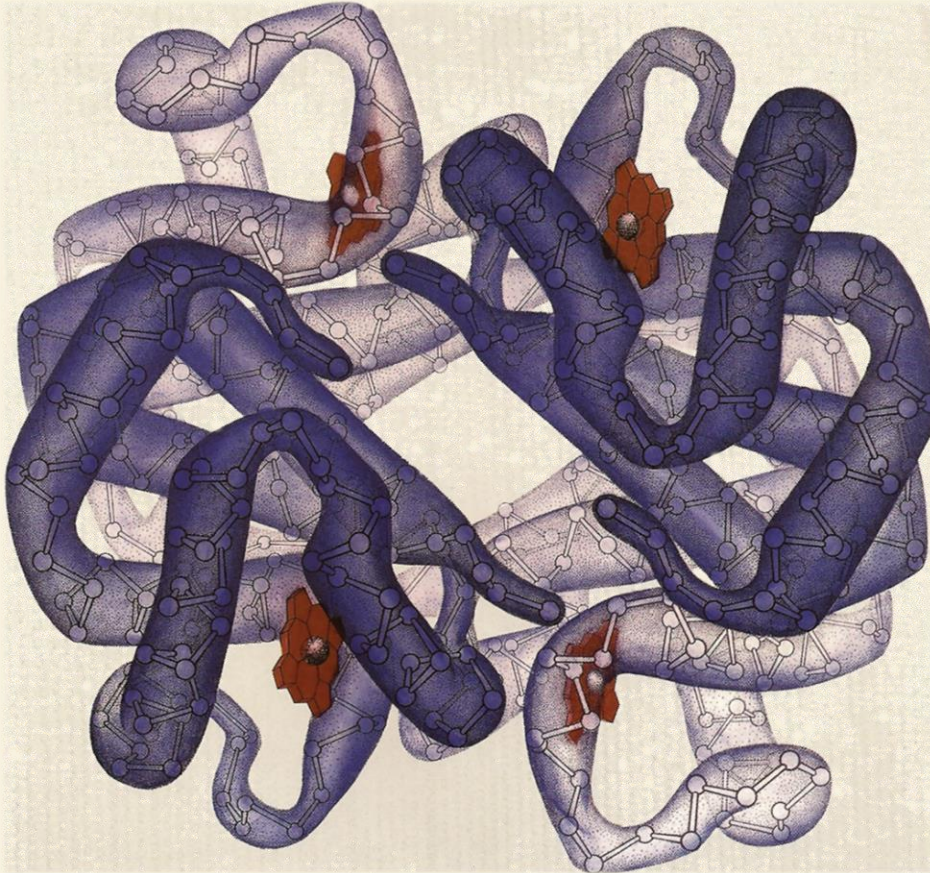


Ribosome, the engine of *self-replication*, is tuned to solve the tRNA decision problem

- Shaping the recognition energy landscape optimized ribosome performance:



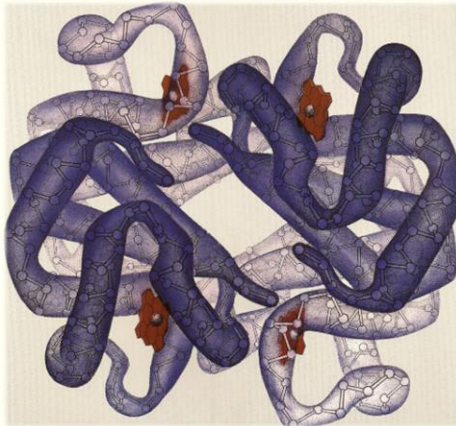
Proteins are complex systems made of hundreds to thousands amino acids arranged in a 3D structure



- » Non-random, information-rich.
- » Small, heterogeneous.
- » Collective interactions
- » Evolutionary memory.

Hemoglobin,
Irving Geiss, 1978

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



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

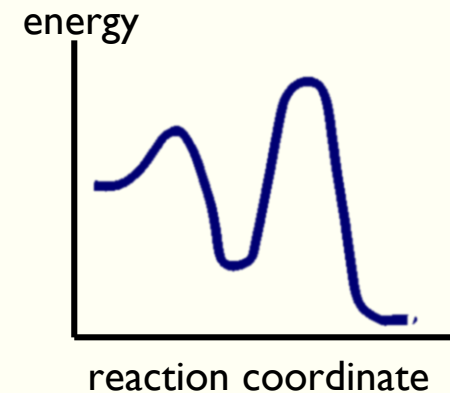
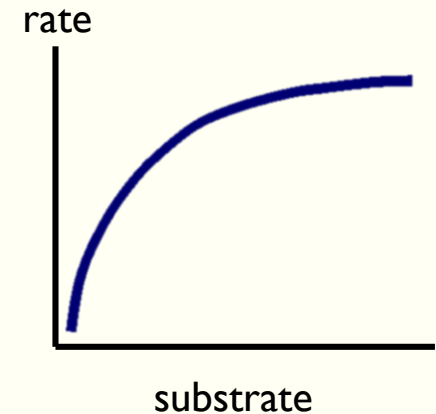
TGG GAG CGG TTC CGC TGC AAC GTG
Trp Glu Arg Phe Arg Cys Asn Val

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

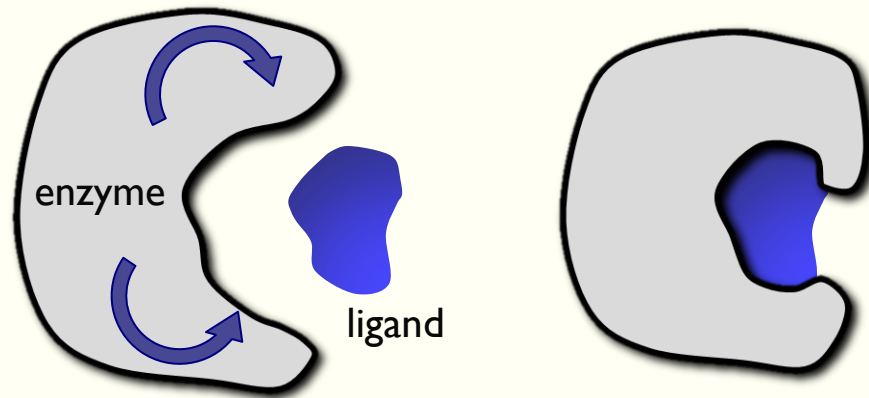
dimensional
reduction



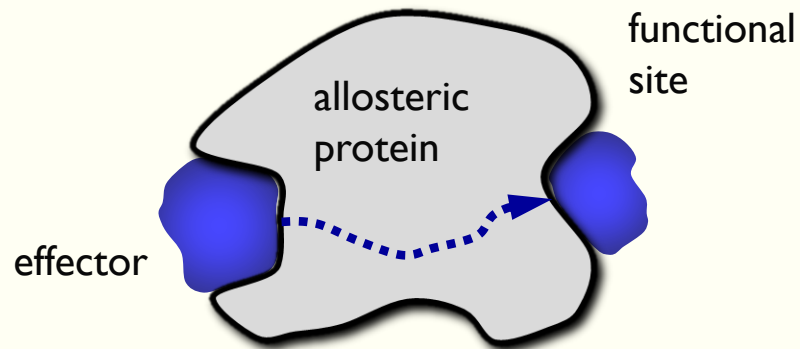
Information ~ gene length $\xrightarrow{\text{compression}}$ Information ~ a few bits

Many proteins use *large-scale collective motions* to process and convey information

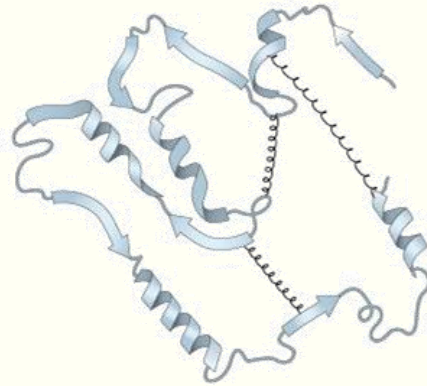
Induced fit:



Allostery:

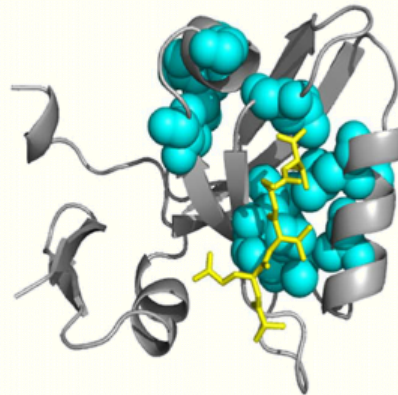
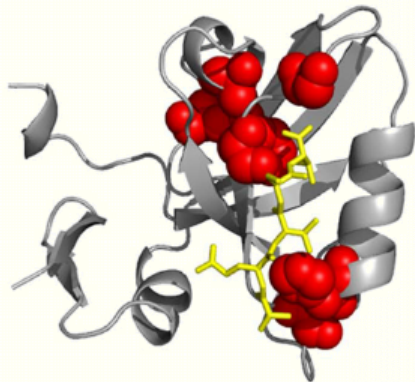


Protein sequences are also remarkably collective, with correlations indicating structure and function



Correlations predict 3D contacts.

Sander Marks Colwell Hopf et al.



Sectors: coevolving protein domains

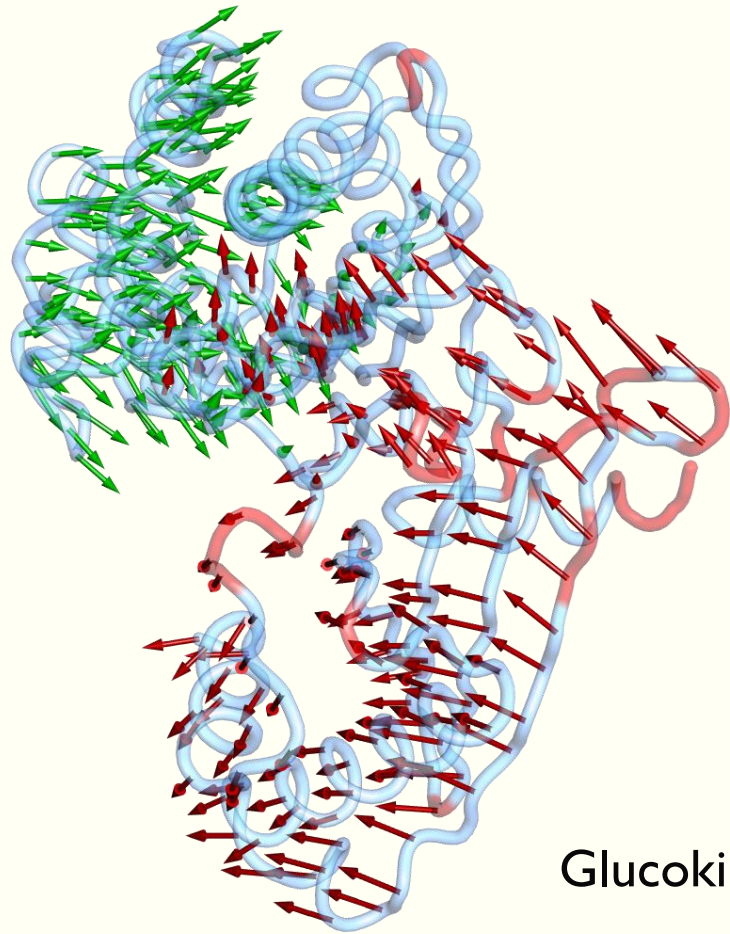
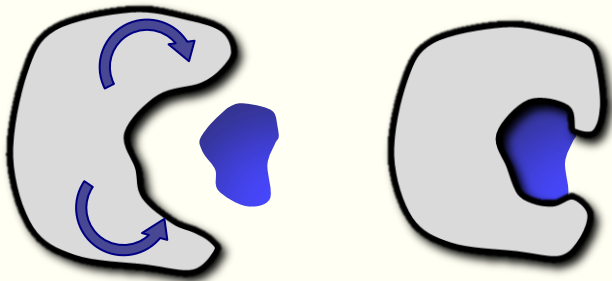
Halabi Rivoire Leibler Ranganathan et al.

Epistasis = interaction among mutations

Looking at dynamics suggests a mechanical basis for mapping genetic information to collective function

Motion (displacement):

unbound → bound

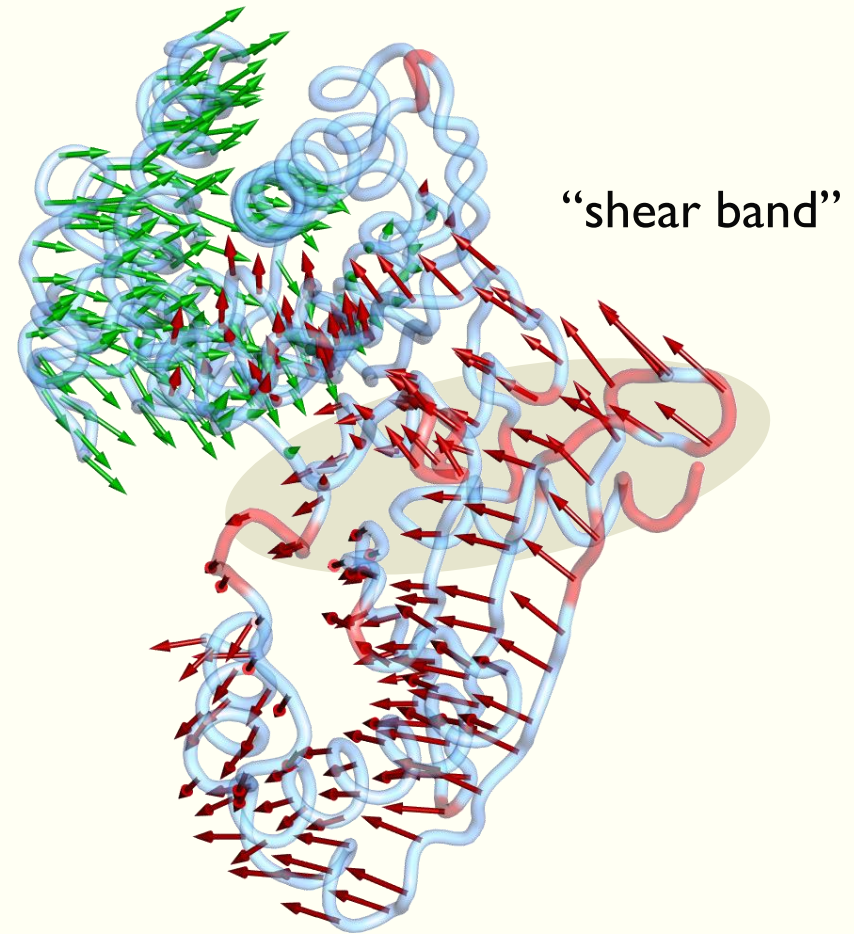


Glucokinase

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

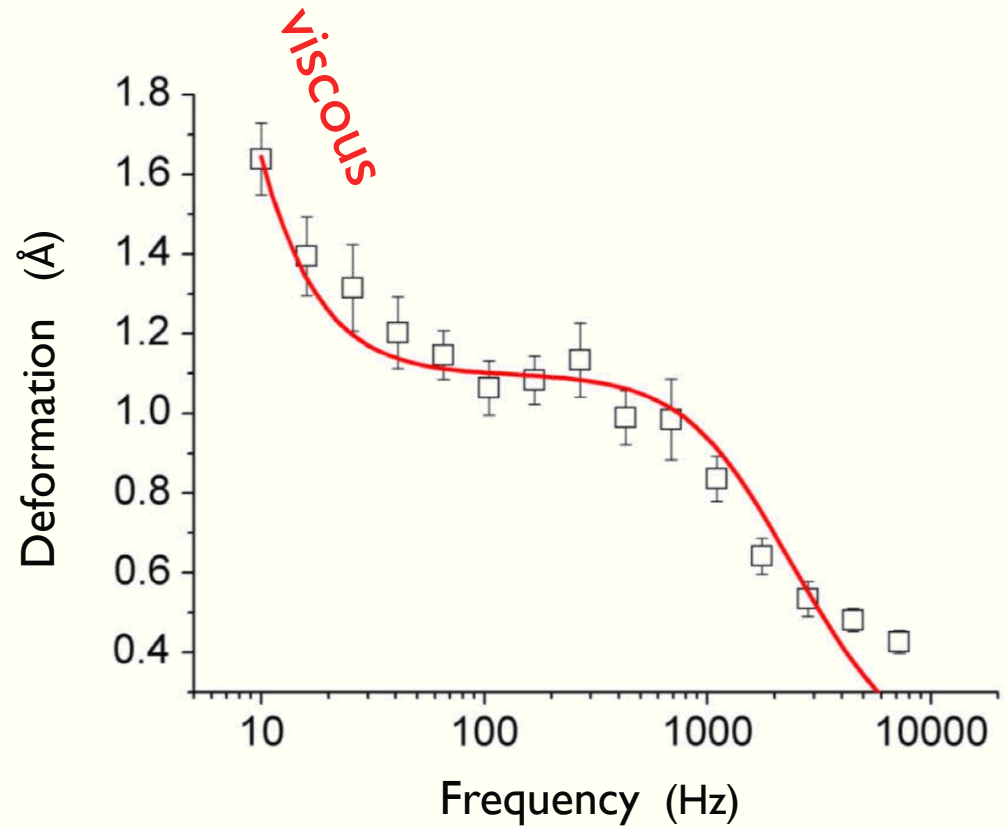
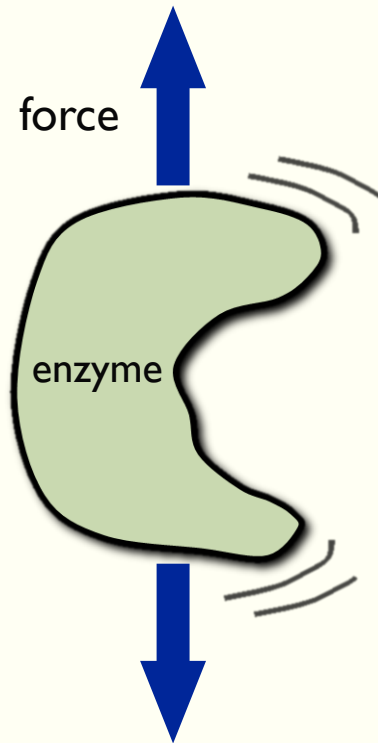
Shear = gradient (motion)

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

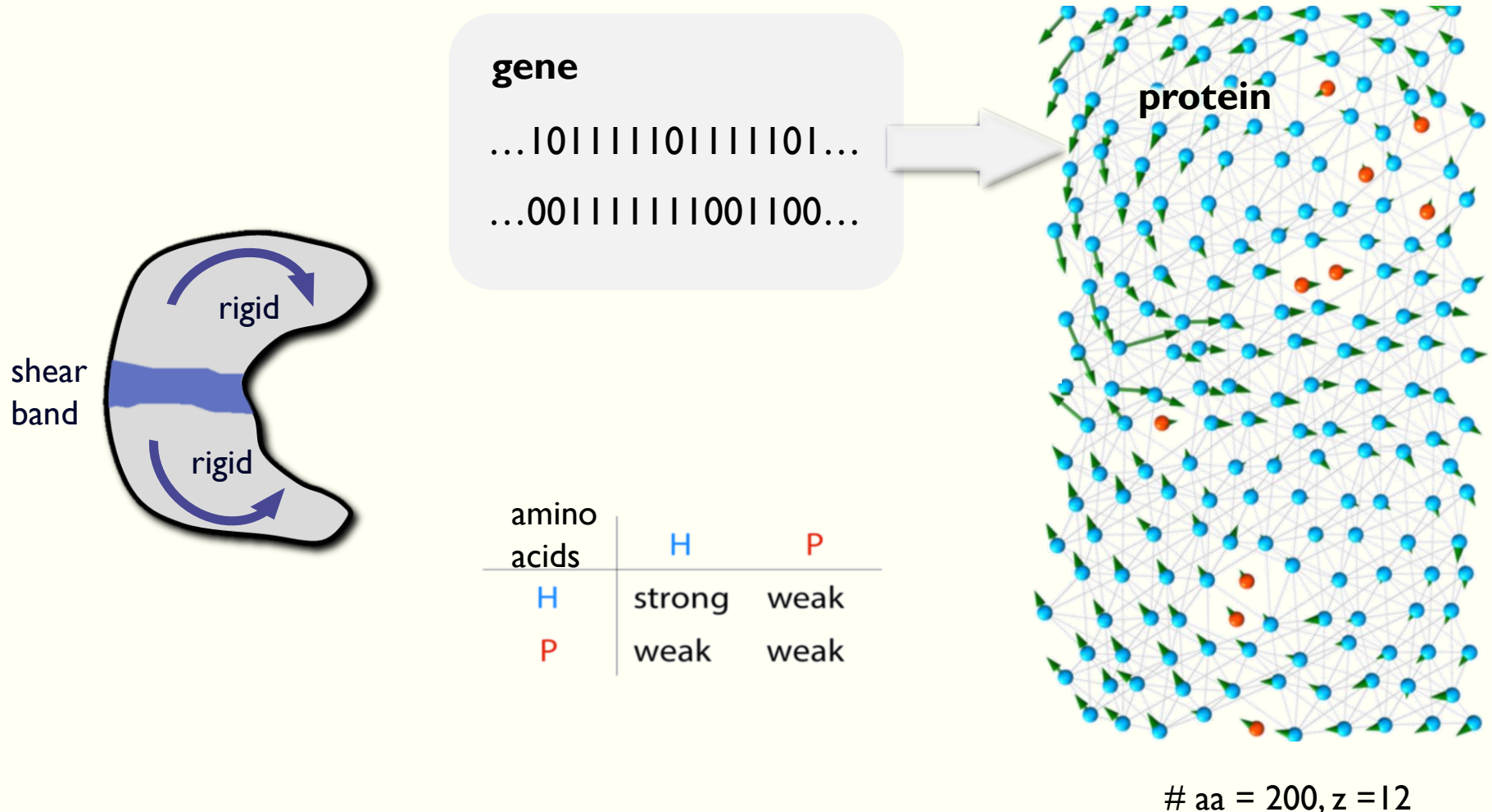


Glucokinase (PDB: 1v4s → 1v4t)

Proteins are soft and *viscous* at low frequency

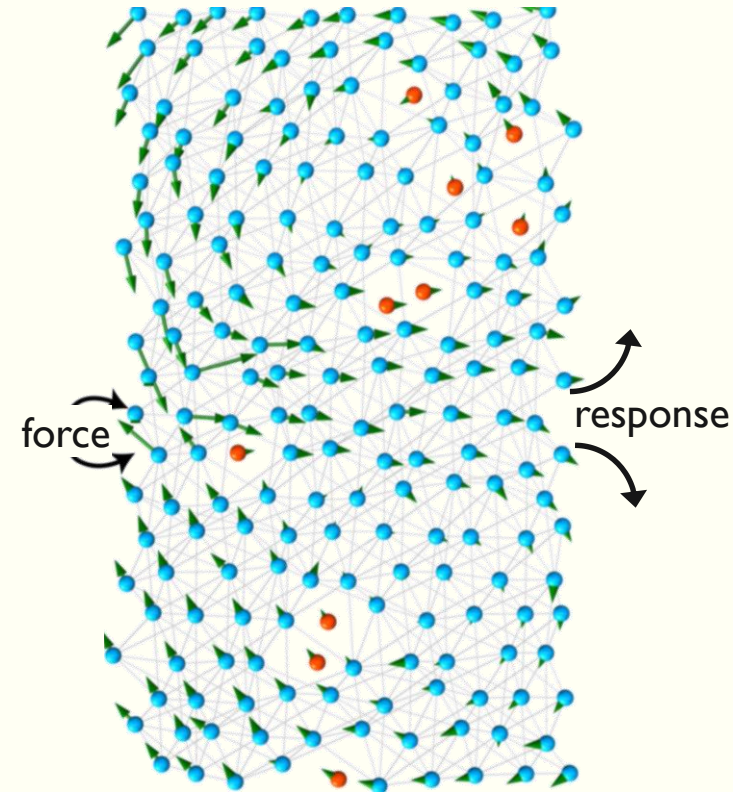
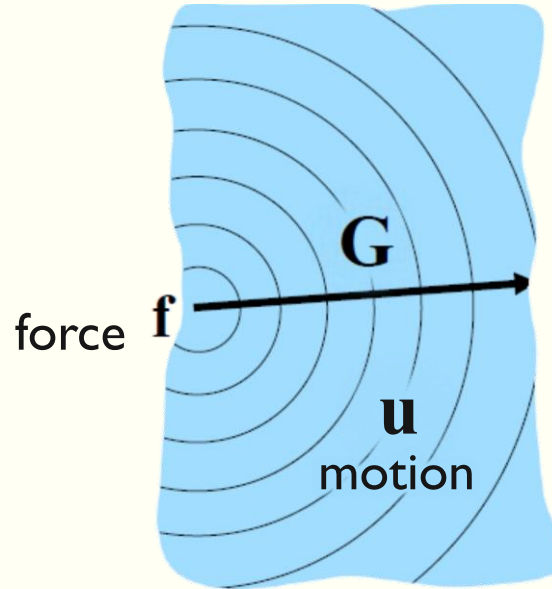


Minimal model of protein as learning matter to give evolution a mechanical interpretation



The **Green** function of the protein measures the transmission of force induced by a local impulse, such as ligand binding

$$\mathbf{G}(\text{gene}) = \mathbf{G}(\text{aa net})$$

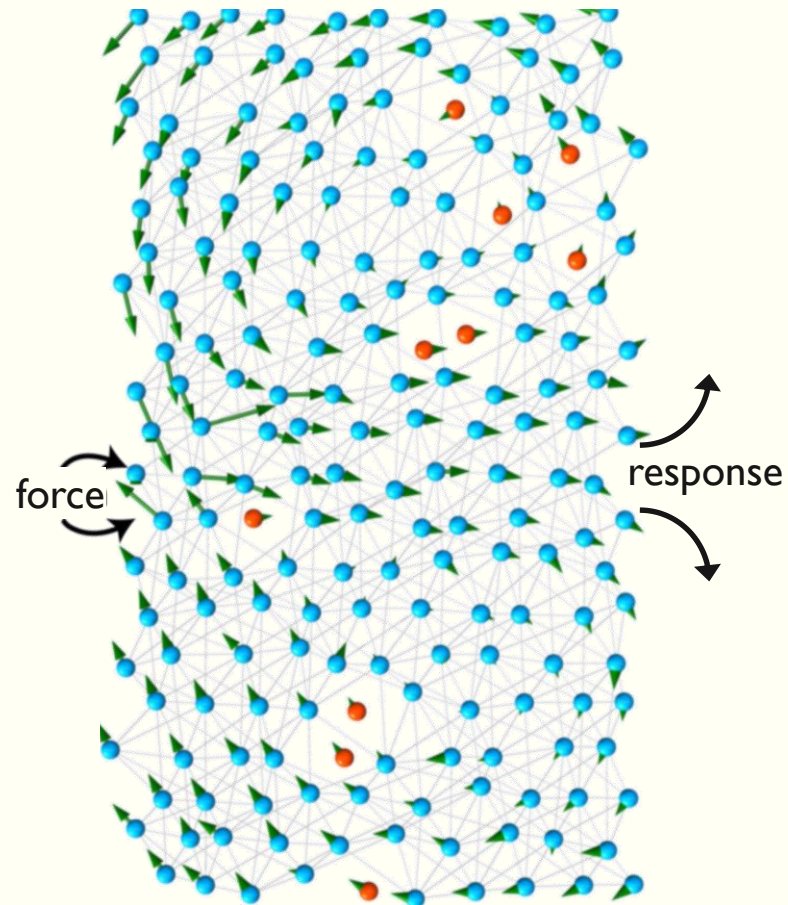


- Green's function solves the inverse problem

$$\left. \begin{array}{l} |\text{force}\rangle = \mathbf{H} |\text{motion}\rangle \\ |\text{motion}\rangle = \mathbf{G} |\text{force}\rangle \end{array} \right\} \Rightarrow \mathbf{G} = \mathbf{H}^{-1} \quad (\text{or } \mathbf{H}^+)$$

- Green (1828) An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism

The Green function defines the genotype-to-phenotype map and the fitness landscape



Genotype-to-phenotype (gene $\rightarrow$ motion)

$$|\text{motion}\rangle = \mathbf{G}(\text{gene})|\text{force}\rangle$$

Fitness landscape (gene $\rightarrow$ fitness)

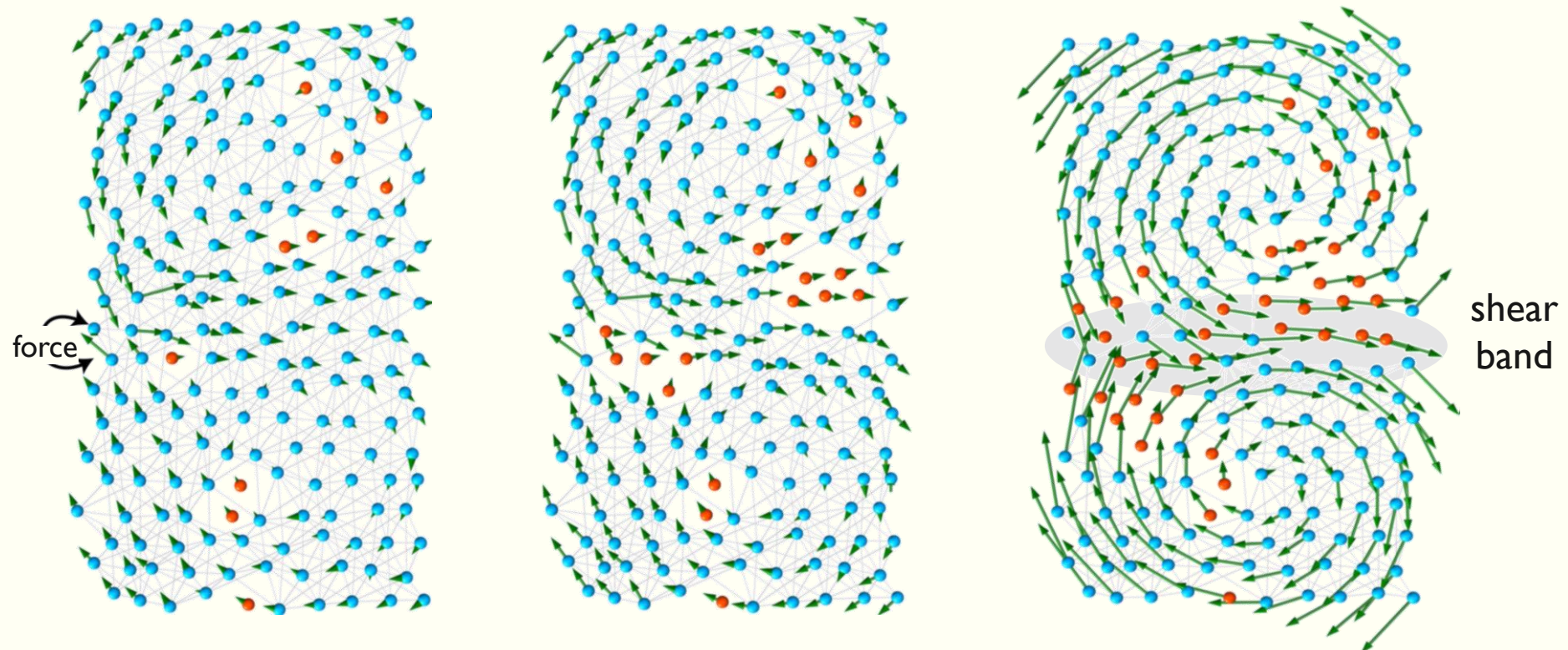
$$\begin{aligned}\text{Fitness}(\text{gene}) &= \langle \text{response} | \text{motion} \rangle \\ &= \langle \text{response} | \mathbf{G}(\text{gene}) | \text{force} \rangle\end{aligned}$$

Evolution searches in the mechanical fitness landscape for a solution with shear band and soft mode

initial

intermediate

solution

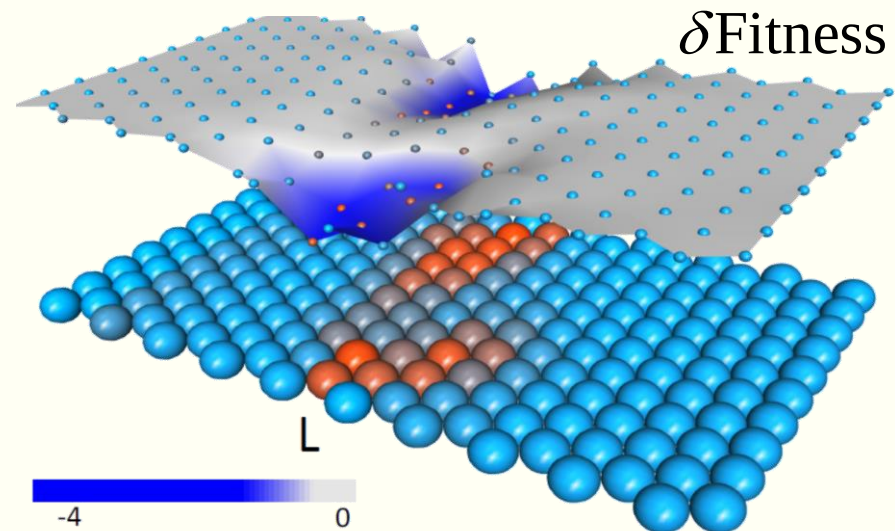
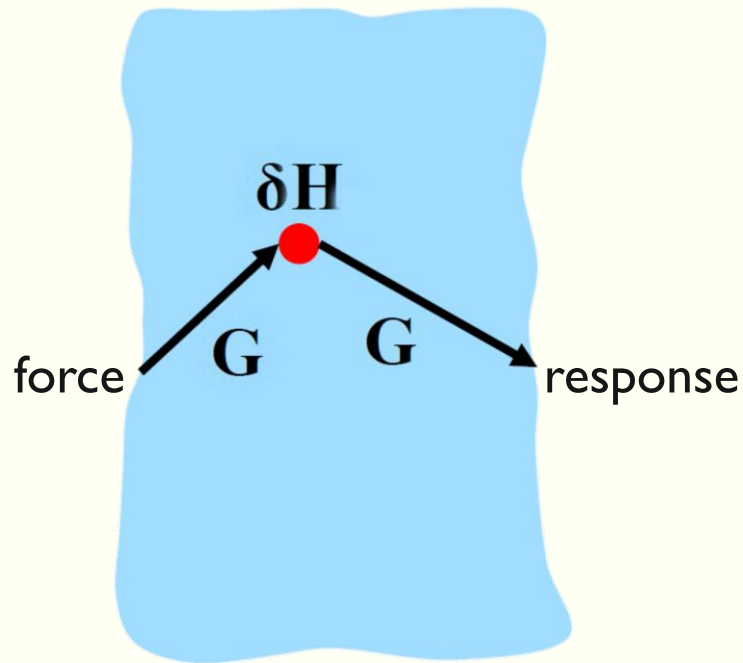


$$\text{Fitness} = \langle \text{response} | \mathbf{G}(\text{gene}) | \text{force} \rangle$$

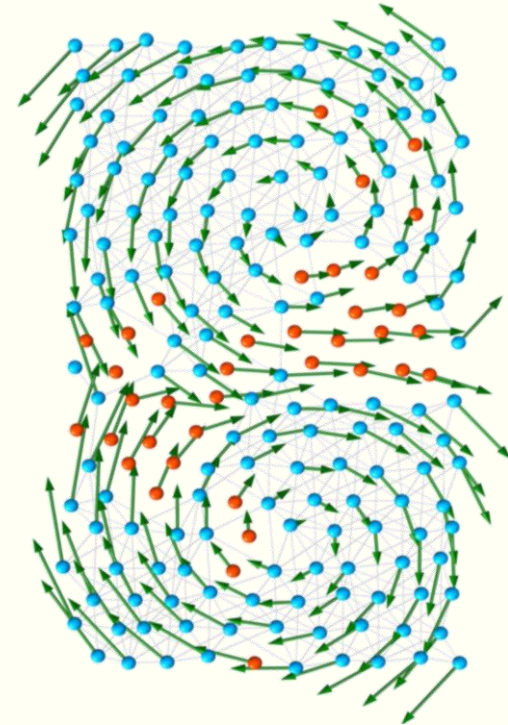
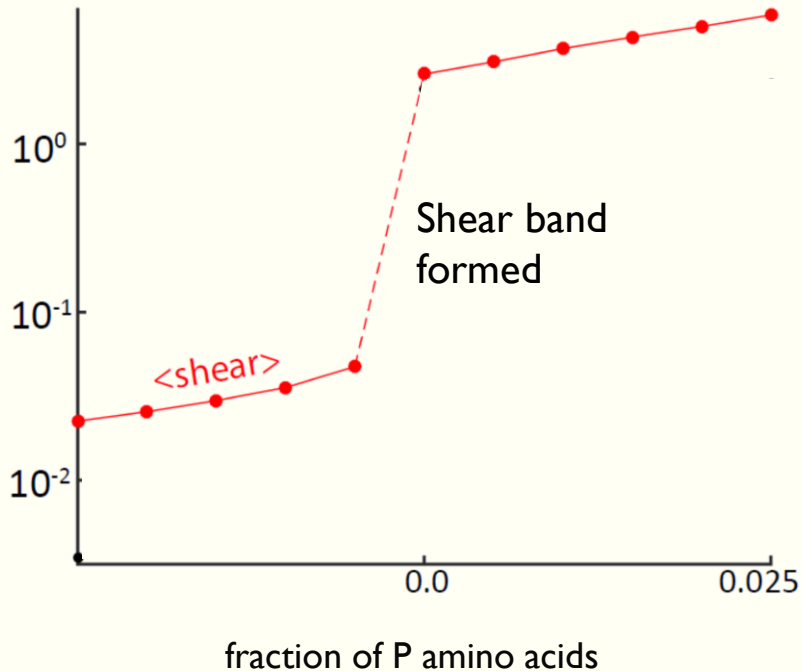
Mutations locally perturb the amino acid network, scatter the force, and thereby change the fitness

$$\delta \mathbf{G} = -\mathbf{G} \delta \mathbf{H} \mathbf{G} + \mathbf{G} \delta \mathbf{H} \mathbf{G} \delta \mathbf{H} \mathbf{G} - \dots$$

$$\delta \text{Fitness} = \langle \text{response} | \delta \mathbf{G} | \text{force} \rangle$$



Protein function emerges at a *topological transition* when a shear band of mutations divides the protein and a *soft mode* appears



$$\text{Fitness} = \langle \text{response} | \mathbf{G}(\text{gene}) | \text{force} \rangle \sim \frac{1}{\text{energy}(\text{soft mode})}$$

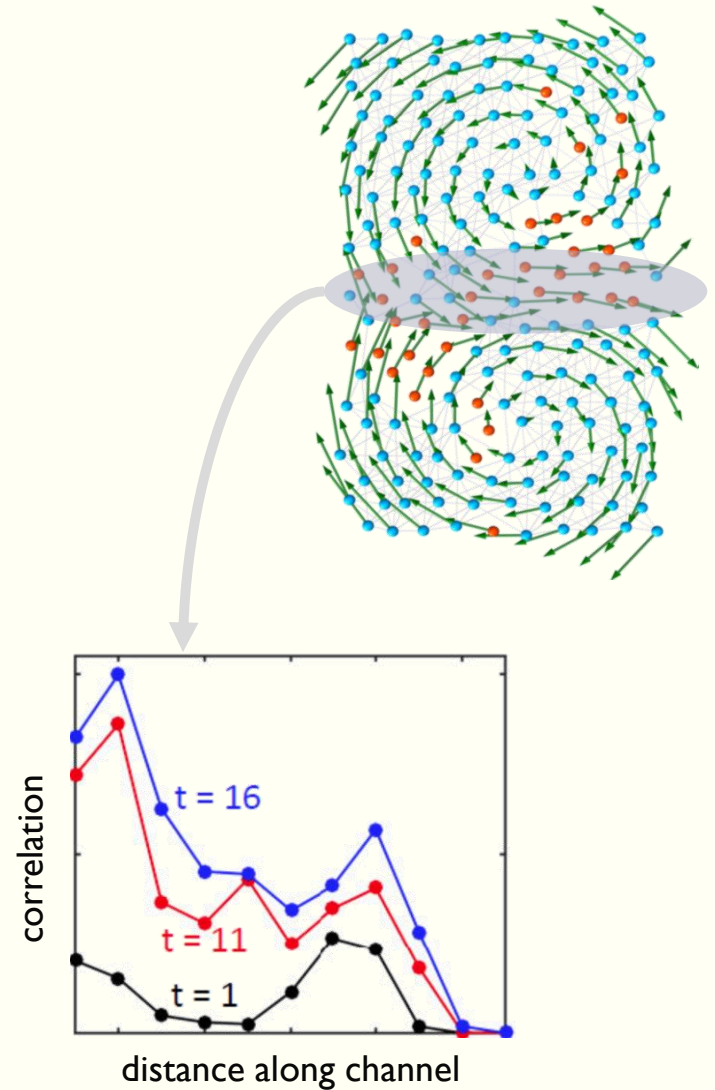
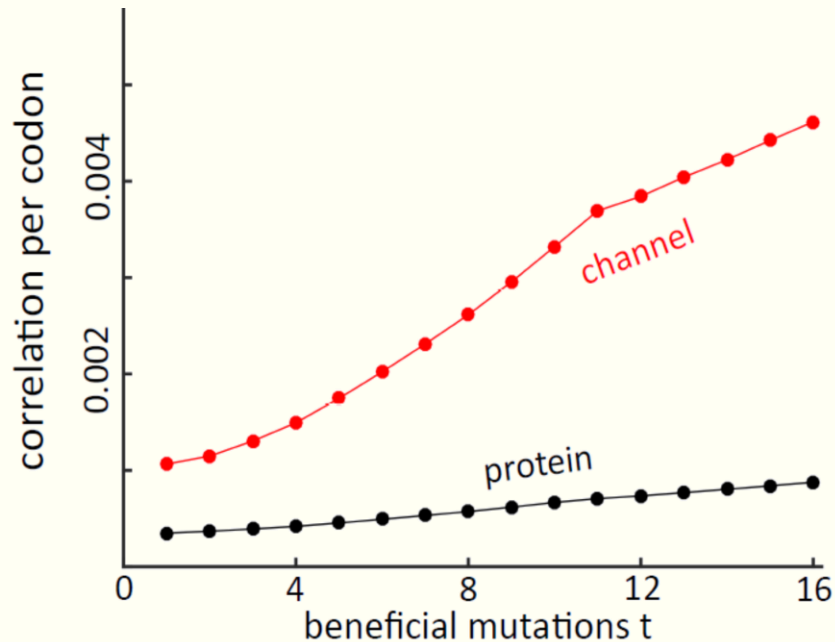
The genetic correlation among solutions increases towards the emergence of mechanical function

10^6 solutions {

```

11100101011111000010101...
10010001000101011101110...
001111110111110100111010...
10011111110011001111000...
...
    
```

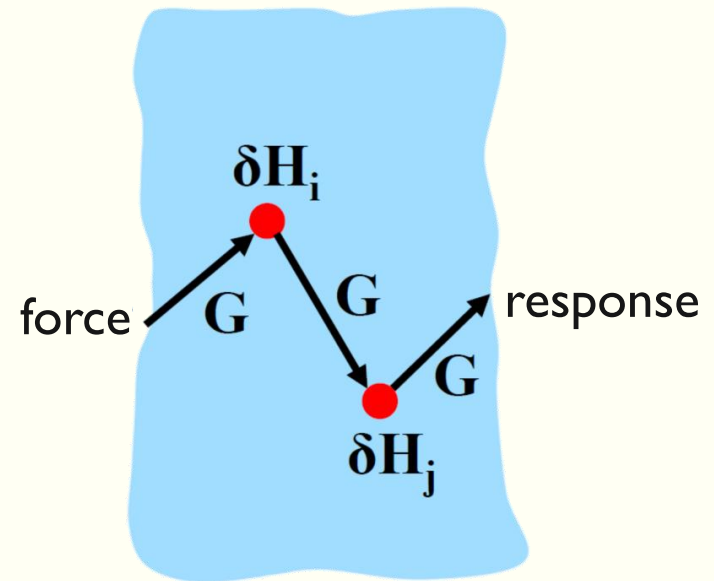
$$\text{correlation}_{ij} = \langle \text{gene}_i \text{gene}_j \rangle - \langle \text{gene}_i \rangle \langle \text{gene}_j \rangle$$



Epistasis is interaction of mutations resulting in nonlinearity of the mechanical fitness

$$\text{epistasis}_{i,j} = \delta\text{Fitness}_{i,j} - (\delta\text{Fitness}_i + \delta\text{Fitness}_j)$$

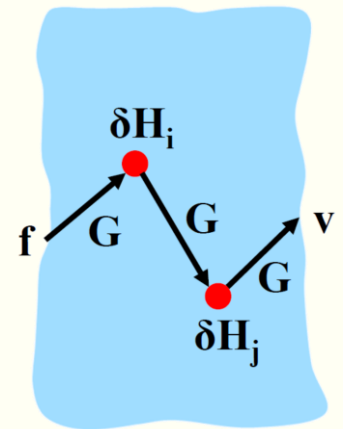
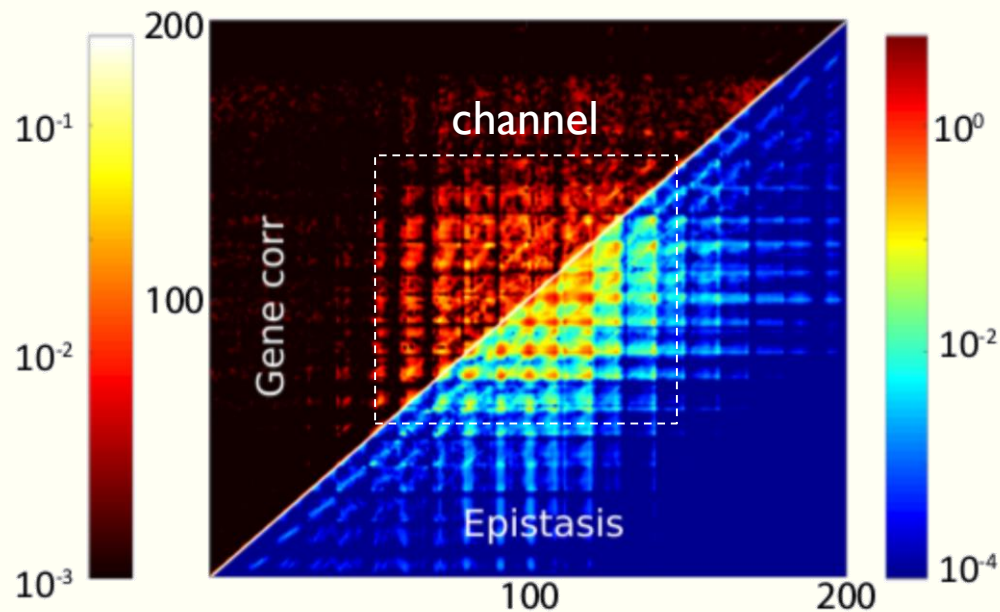
- Epistasis $\sim$ “curvature” of G
 $\sim$ sum over scattering paths:



$$\text{epistasis}_{i,j} = \langle \text{force} | \mathbf{G}\delta\mathbf{H}_i\mathbf{G}\delta\mathbf{H}_j\mathbf{G} + \mathbf{G}\delta\mathbf{H}_j\mathbf{G}\delta\mathbf{H}_i\mathbf{G}... | \text{response} \rangle$$

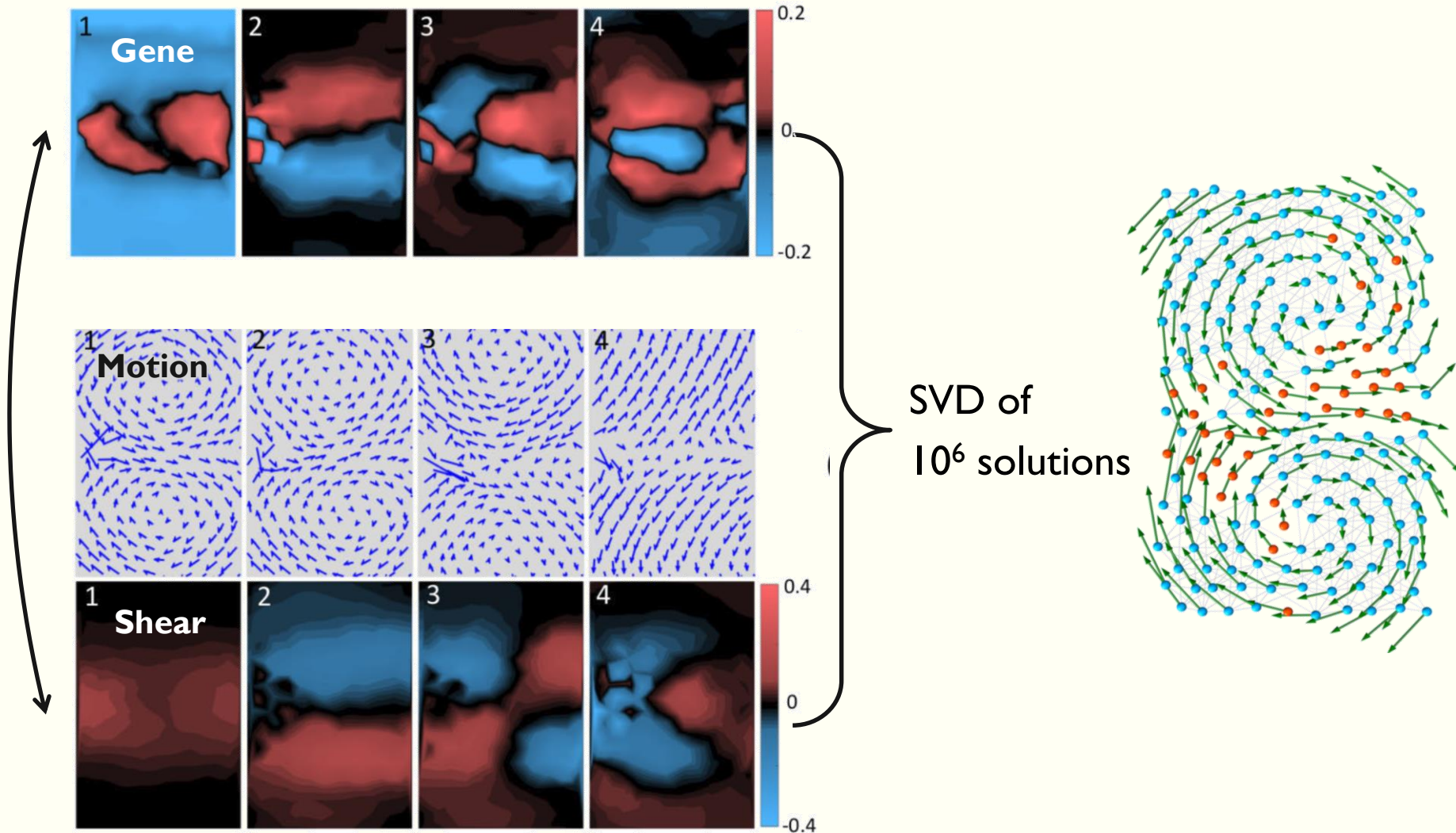
Mechanical forces link epistasis to genetic correlation, and both exhibit similar patterns

$$\text{correlation}_{ij} = \langle \text{gene}_i \text{gene}_j \rangle - \langle \text{gene}_i \rangle \langle \text{gene}_j \rangle$$



$$\text{epistasis}_{i,j} = \delta \text{Fitness}_{i,j} - \delta \text{Fitness}_i - \delta \text{Fitness}_j$$

Phenotype and genotype modes capture the relevant degrees-of-freedom of mechanical function and evolution

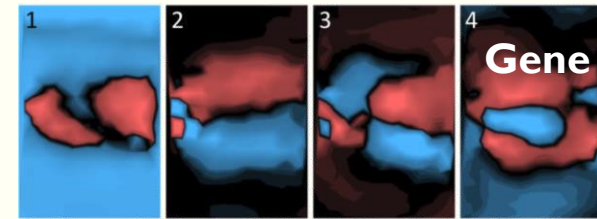
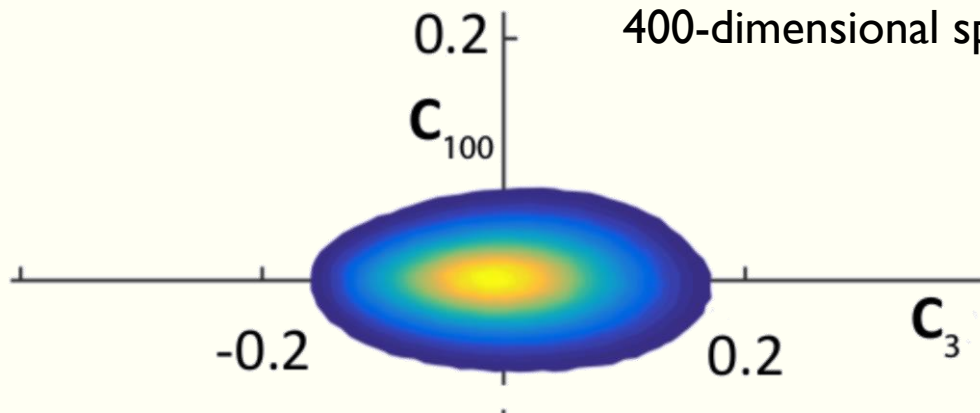


The genotype-to-phenotype map involves huge dimensional reduction

10^6 solutions

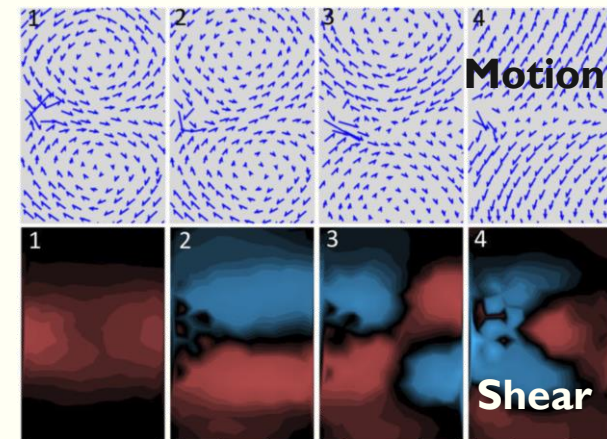
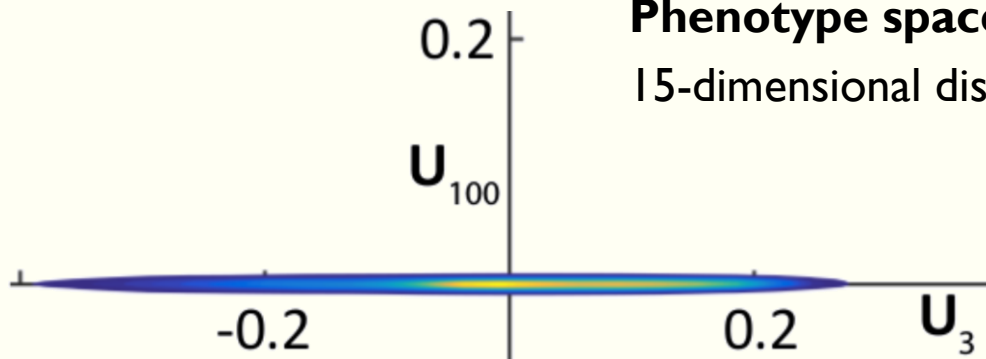
Genotype space

400-dimensional spheroid

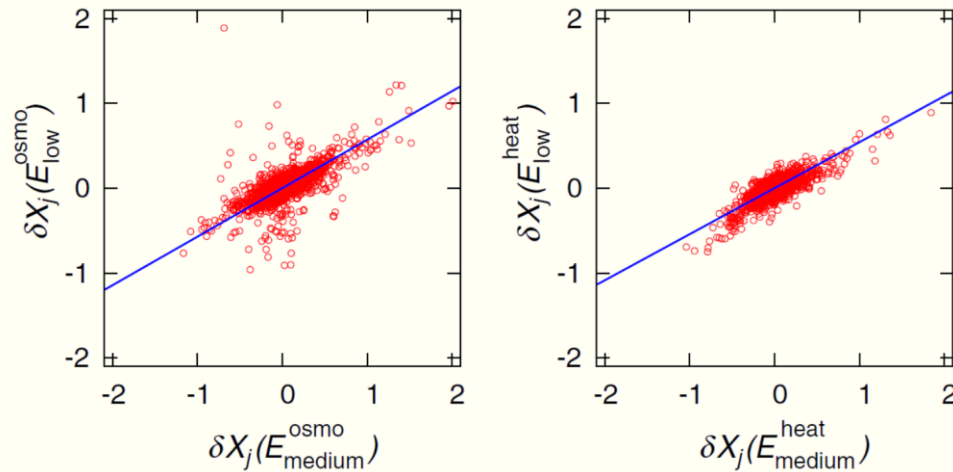


Phenotype space

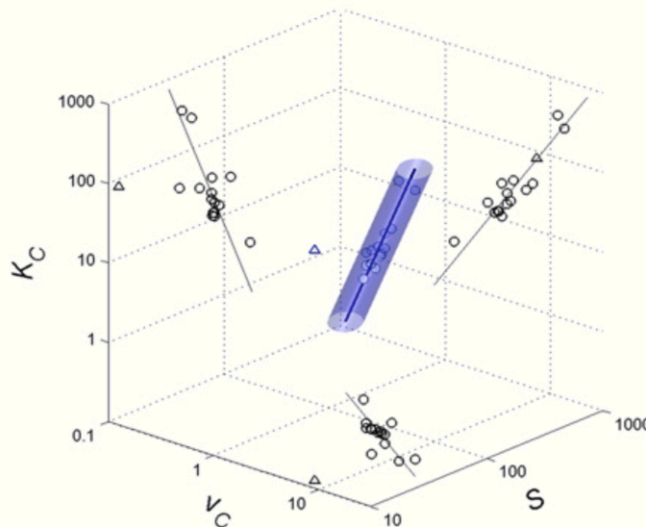
15-dimensional discoid



Dimensional reduction is a general phenomenon in living systems



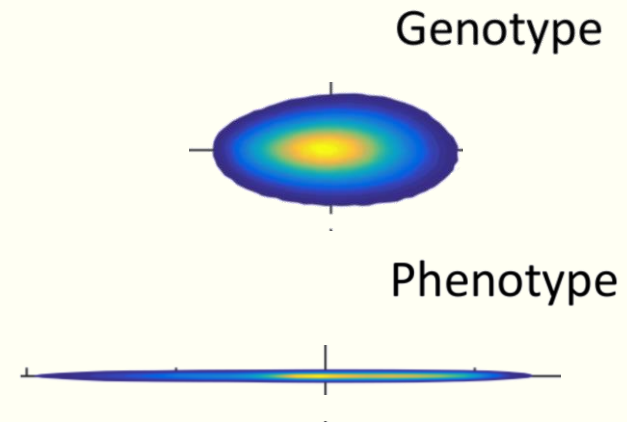
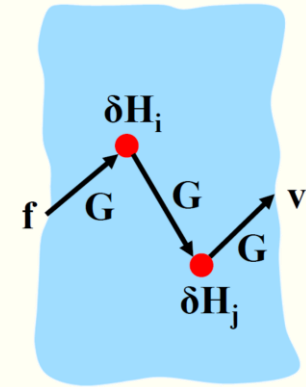
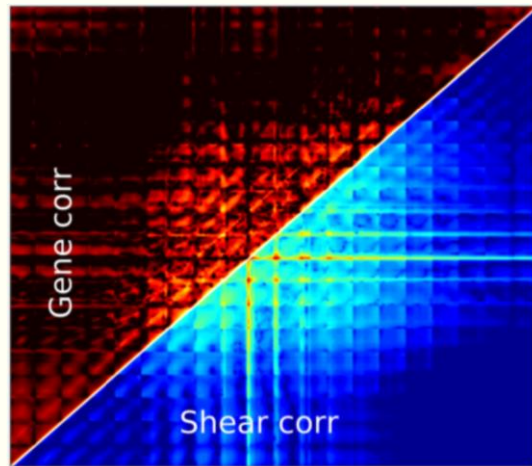
1D response of gene expression to stress
(Kaneko, Furusawa & Yomo, PRX 2015)



1D scaling of kinetic parameters in Rubisco
(Savir, Noor, Milo & TT, PNAS 2010)

Genetic information and mechanical function are linked by a minimal model of protein

- **Mutations:** mechanical perturbations
- **Epistasis:** multiple scattering
- **Genotype-to-phenotype map**
- **Fitness landscape**
- **Soft modes at topological transition.**
- **Dimensional reduction**
- **Correspondence:** mechanical modes – coevolution modes



Dutta Eckmann Libchaber & TT, in review
TT Eckmann & Libchaber, Phys Rev X 2017

- **Applicable to other strongly-coupled living systems?**

Geometric mechanical view on the evolution of living matter

- 12molecules (within 12cells)
- Geometry of protein evolution:
 - Geometry of the amino acid itself (Green function = inverse of the network metric)
 - Phenotype space: geometry of all possible AA network geometries.
- Mechanical logic (a simple one): develop specific mechanical response.
- What is the right language to describe this “logic matter?”
- What about self-reference?

Thanks:

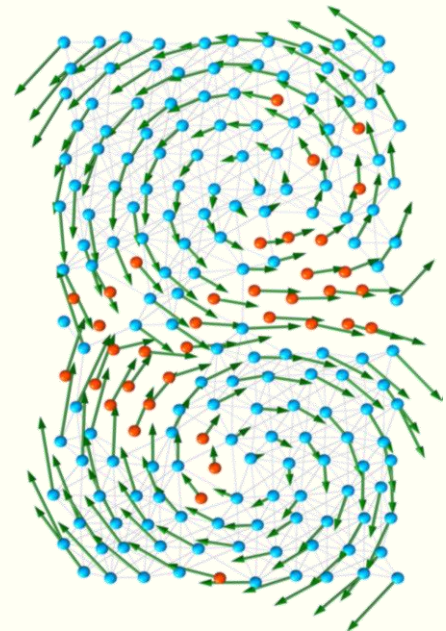
S. Dutta (IBS)

J.-P. Eckmann (Geneva)

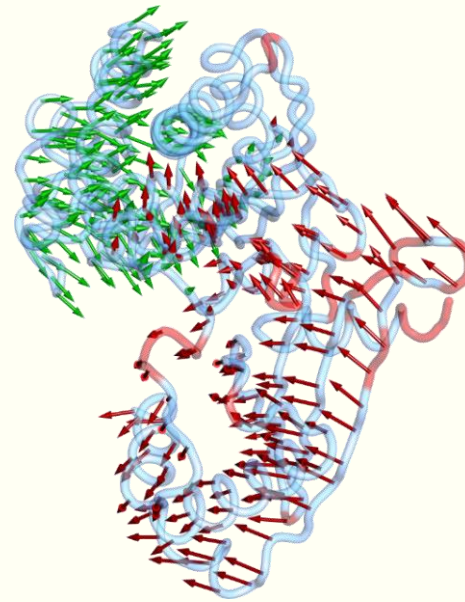
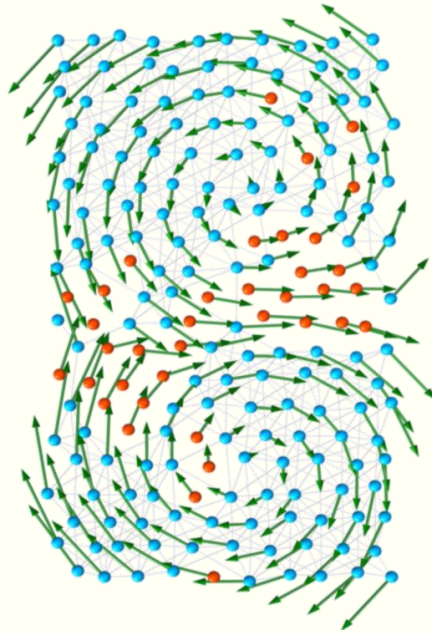
A. Libchaber (Rockefeller)



**Center for Soft
& Living matter**



Looking at protein dynamics suggests simple mechanical understanding of protein as evolving amorphous matter



Thanks:

S. Dutta (IBS)

J.-P. Eckmann (Geneva)

A. Libchaber (Rockefeller)

S. Leibler & M. Mitchell (IAS)



**Center for Soft
& Living matter**