

Harvard - May 2018

PROTEINS AS LEARNING MACHINES

Green Functions of Correlated Genes and The Mechanical Evolution of Protein

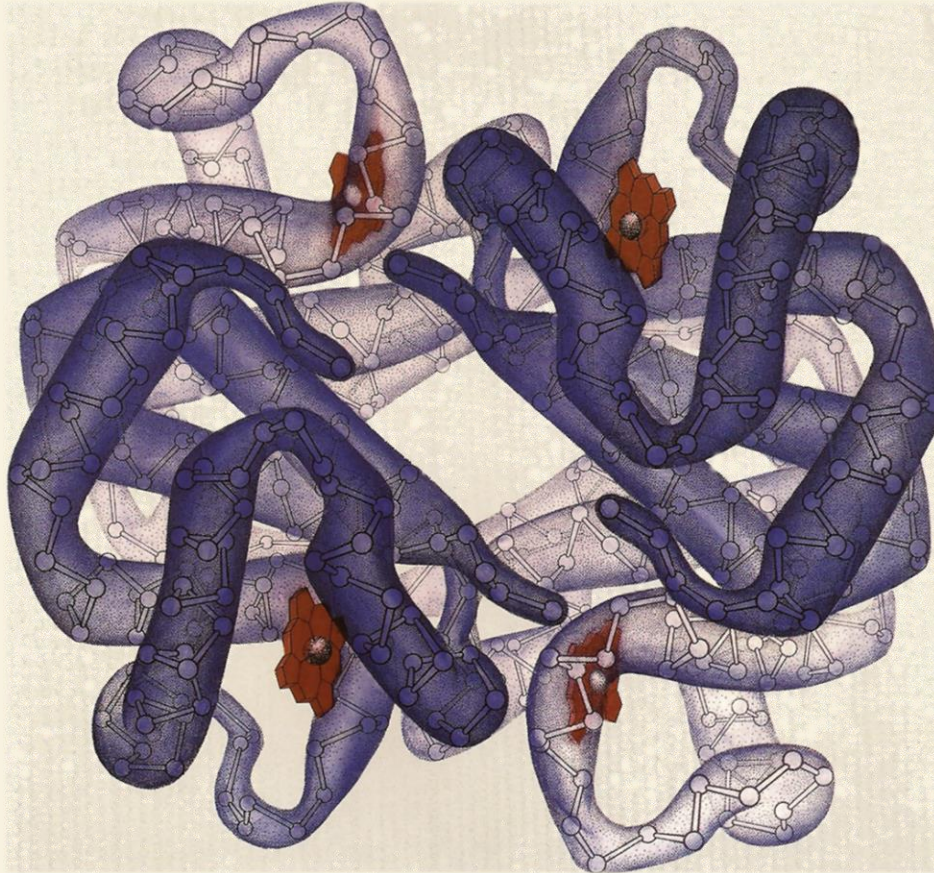
Sandipan Dutta & TT (IBS)

Albert Libchaber (Rockefeller)

Jean-Pierre Eckmann (Geneva)

Michael Mitchell and Stan Leibler (IAS & Rockefeller)

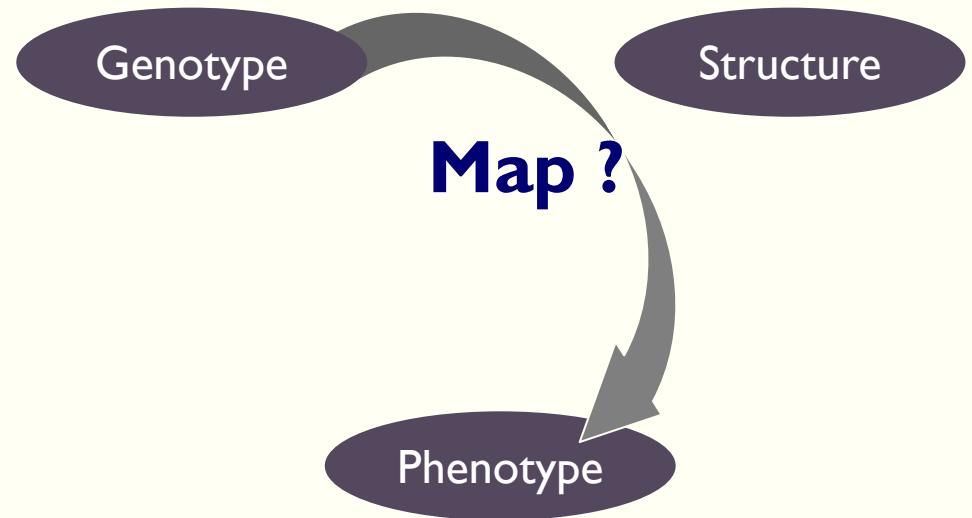
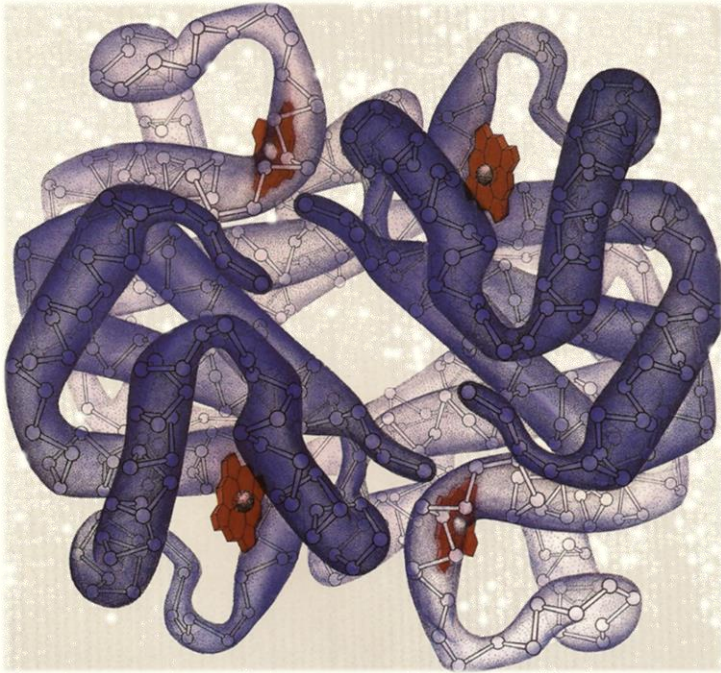
Proteins are complex systems made of hundreds to thousands amino acids arranged in a 3D structure



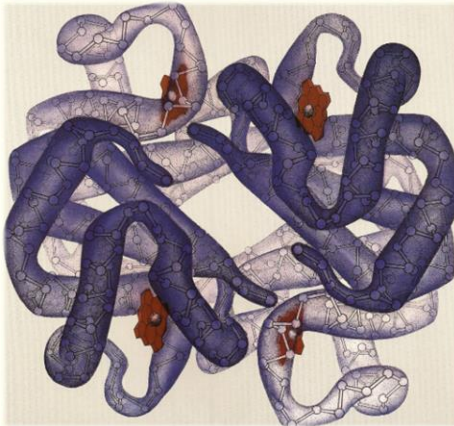
- » Non-random, information-rich.
- » Long-term evolutionary memory.
- » Small, heterogeneous.
- » Collective interactions.

Hemoglobin,
Irving Geiss, 1978

The physics of the genotype-to-phenotype map is not fully understood



Genes and protein structures have *many* parameters but protein function is described by only a *few*



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

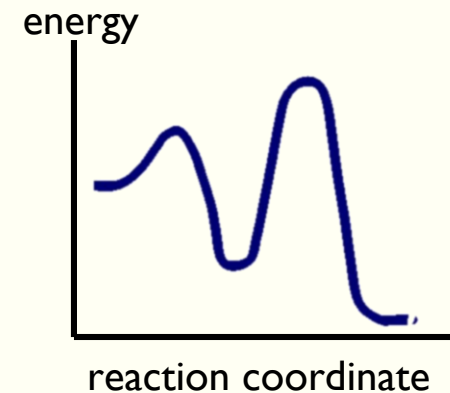
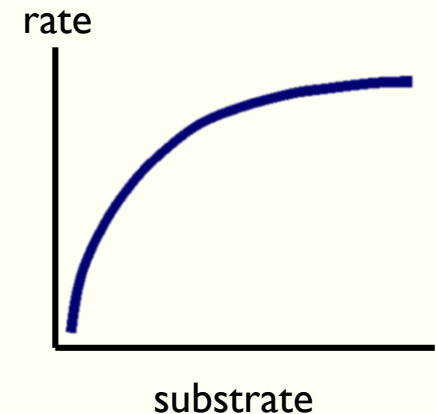
TGG GAG CGG TTC CGC TGC AAC GTG
Trp Glu Arg Phe Arg Cys Asn Val

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

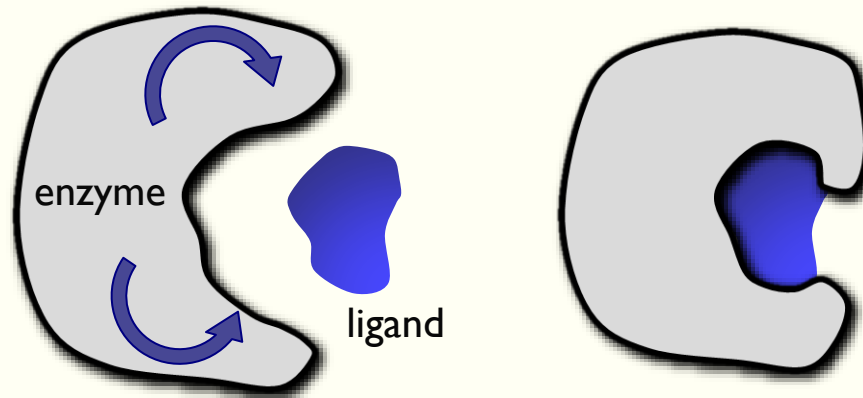


compression

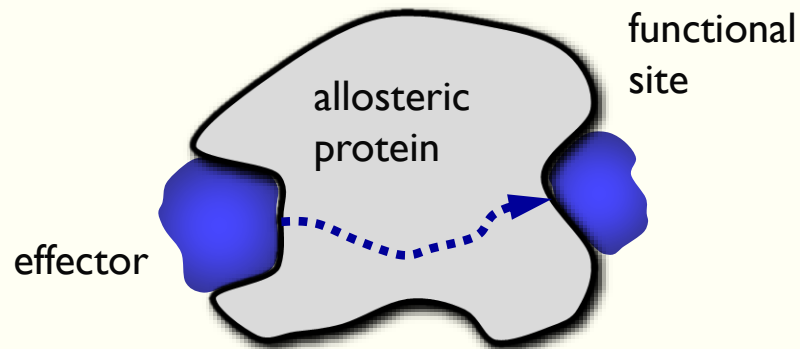
Information ~ gene length → Information ~ a few bits

Many proteins use *large-scale collective* motions to function

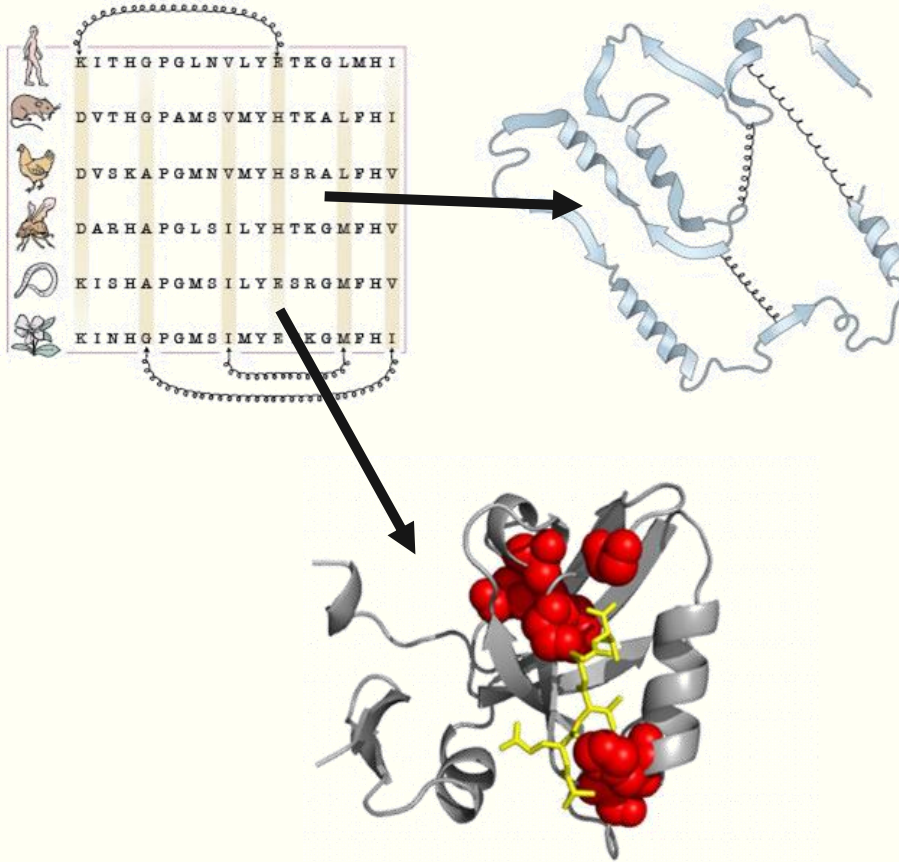
Induced fit:



Allostery:



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



Correlations predict 3D contacts.

Sander Marks Colwell Hopf et al.

coevolving protein domains (“sectors”)

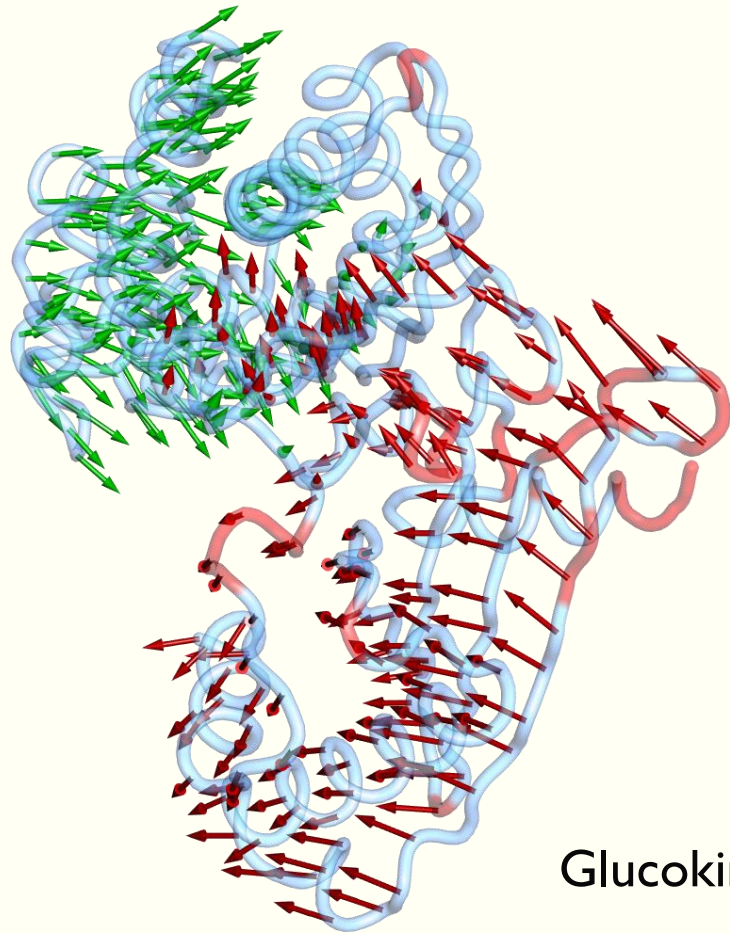
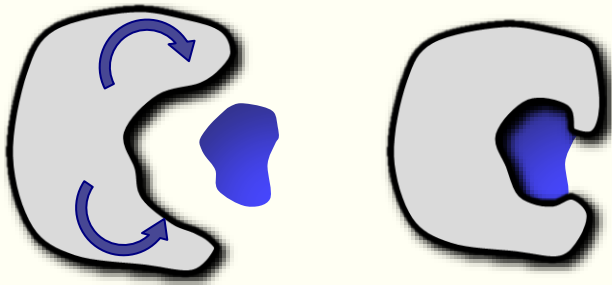
Halabi Rivoire Leibler Ranganathan et al.

- **Epistasis** = interaction among mutations.
- *How is **epistasis** encoded in the gene?*

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

Motion = $\Delta(\text{Structure})$

unbound $\rightarrow$ bound



Glucokinase

Dutta Eckmann Libchaber & TT, PNAS 2018

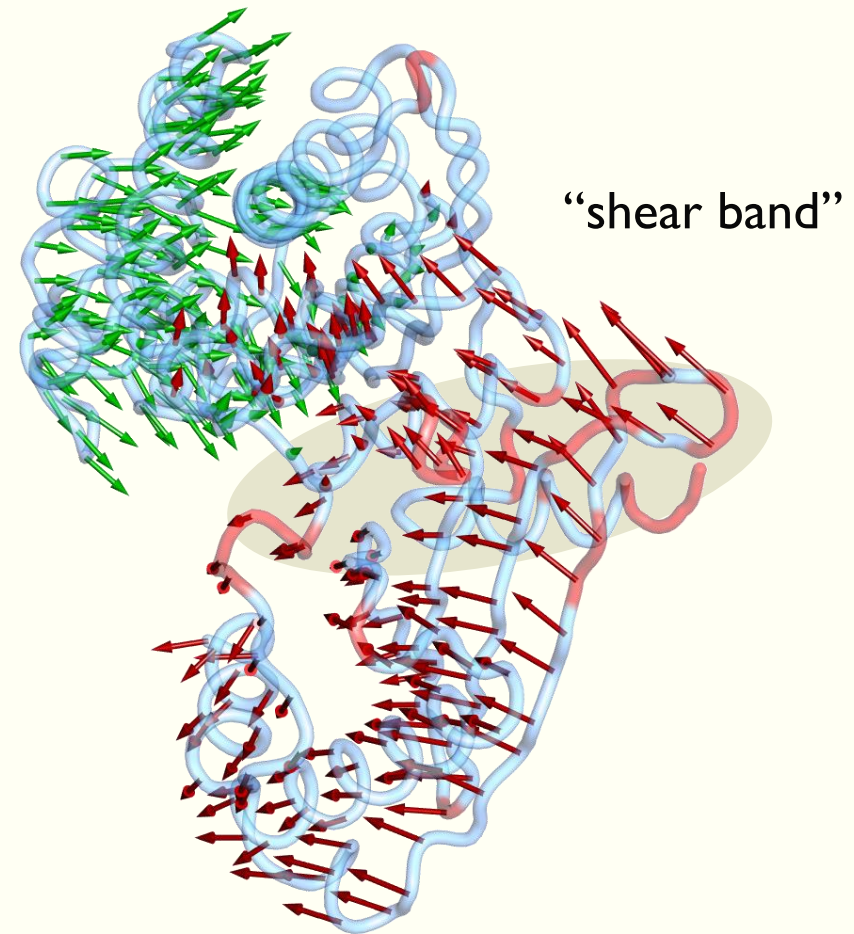
Mitchell TT Leibler, PNAS 2016

Eckmann Libchaber & TT, PRX 2017

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

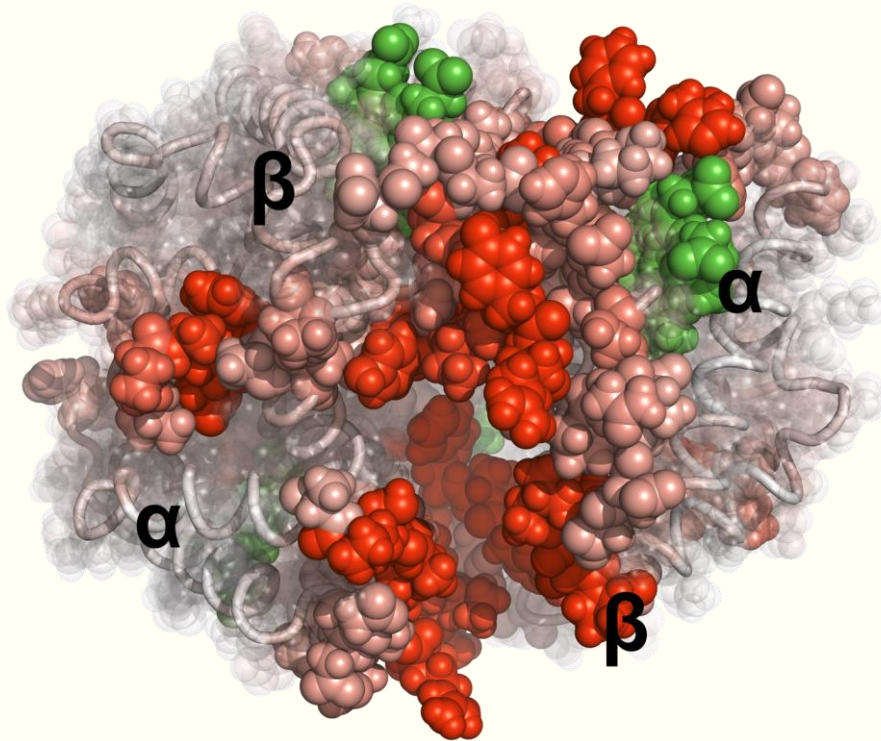
Shear = Δ Motion = $\Delta\Delta$ Structure

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



Glucokinase (PDB: 1v4s → 1v4t)

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

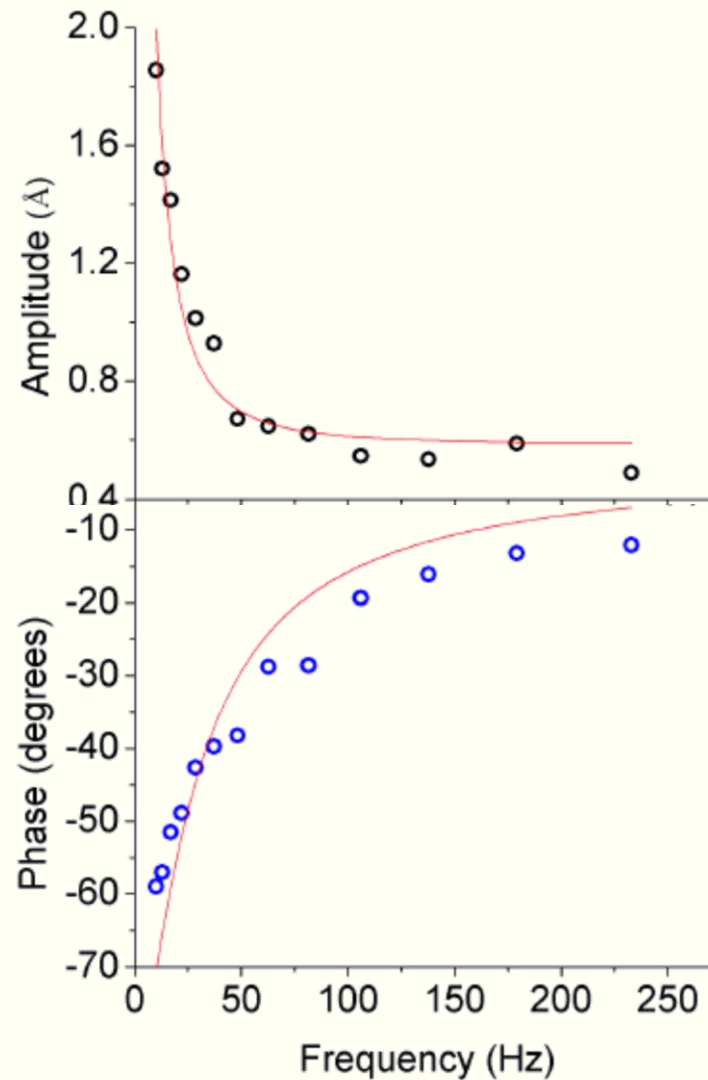
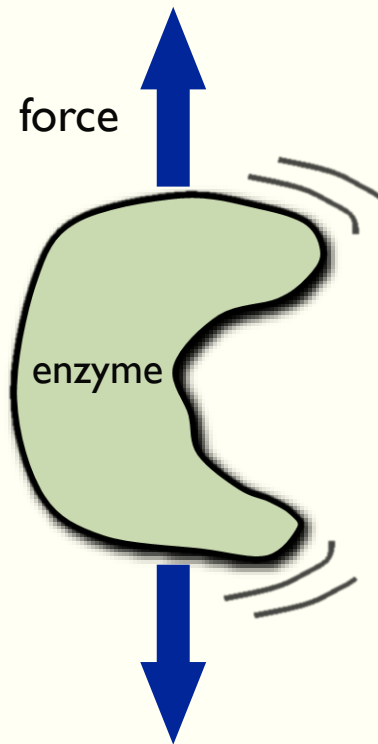


Hemoglobin

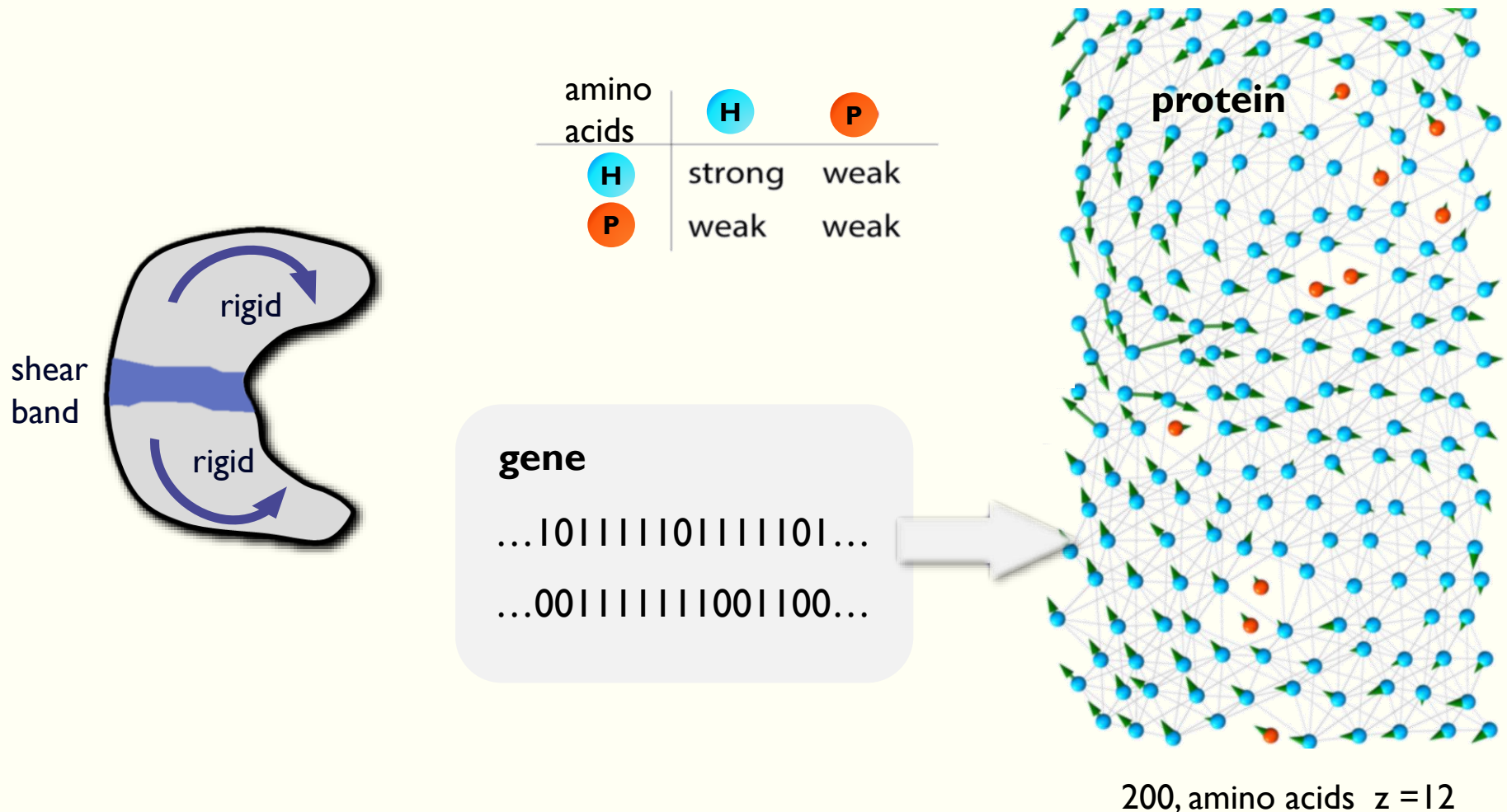
Shear band fractal dimension

- Hemoglobin: 2.3
- Glucokinase: 2.0
- ATCase : 2.3
- Albumin : 1.9

Proteins are soft and *viscous* at low frequency

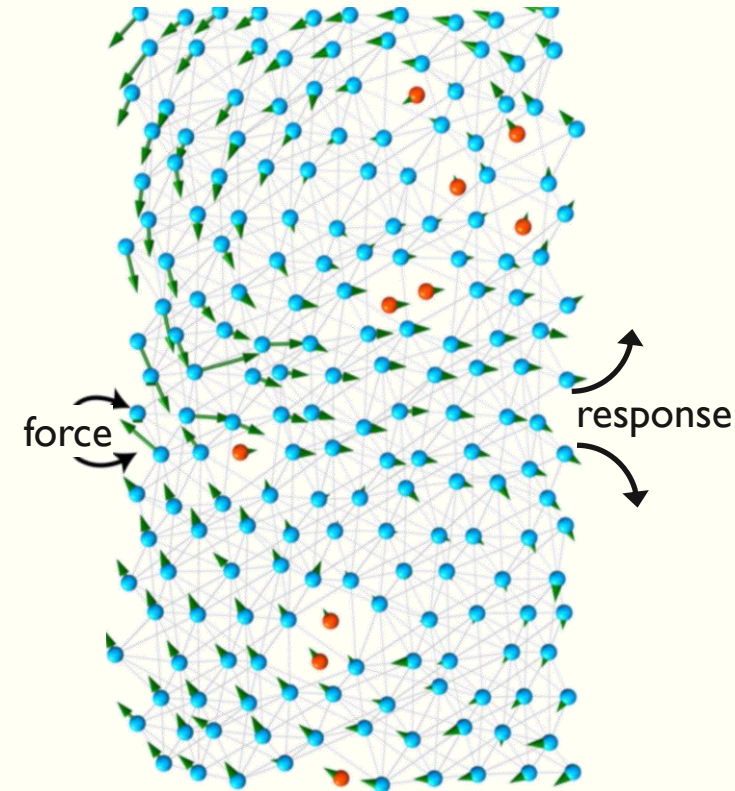
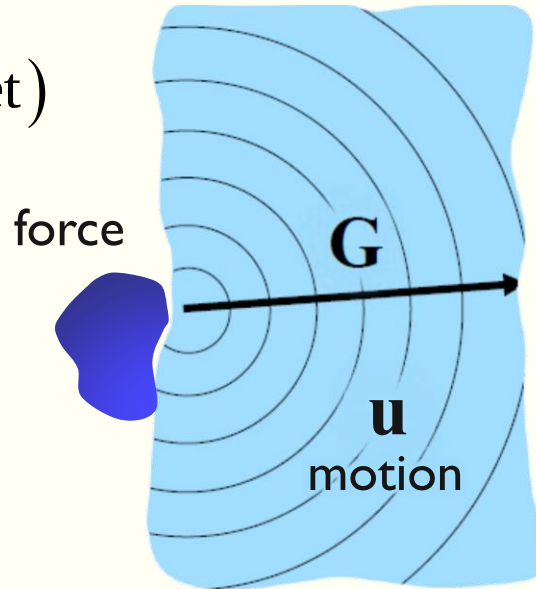


Minimal model of protein as learning matter to give evolution a mechanical interpretation



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

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

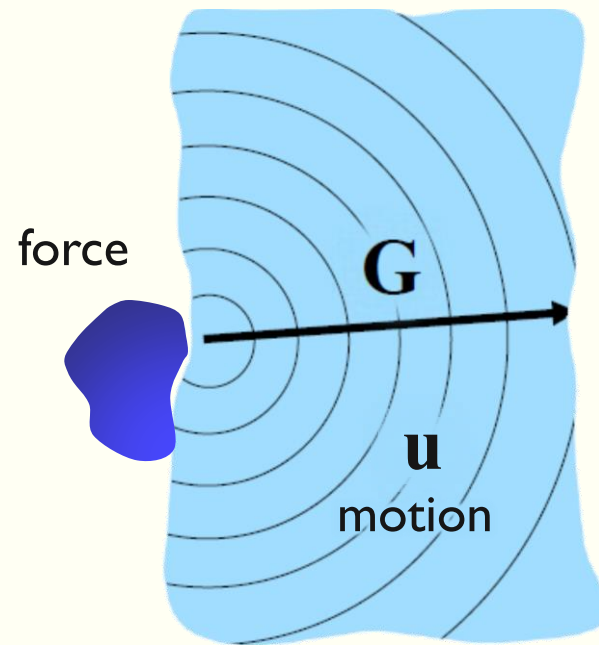


- Green's function solves the inverse problem

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

- **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 map** is the induced motion

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

- **Fitness** is overlap with functional motion, $|\text{response}\rangle$

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

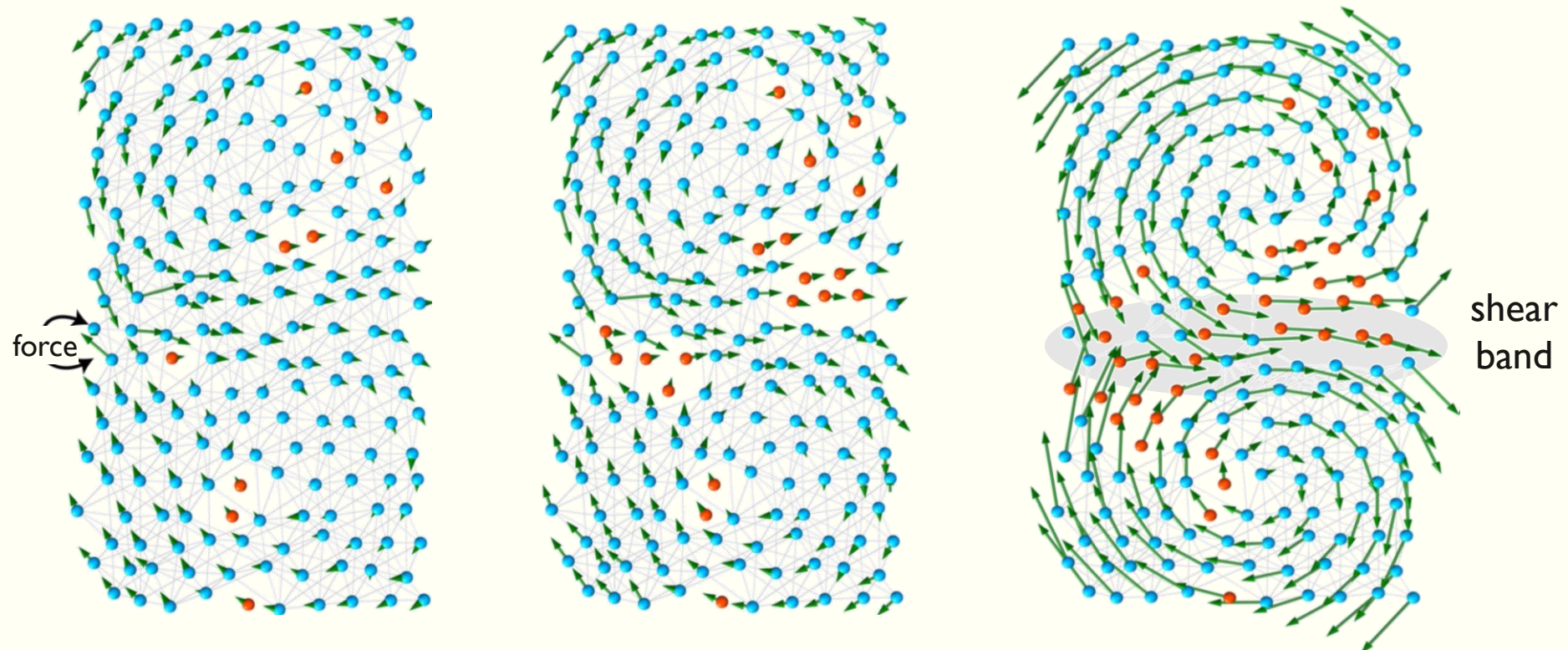
$\mathbf{G}$: gene $\rightarrow$ amino acid network $\rightarrow$ protein dynamics $\rightarrow$ function

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

initial

intermediate

solution



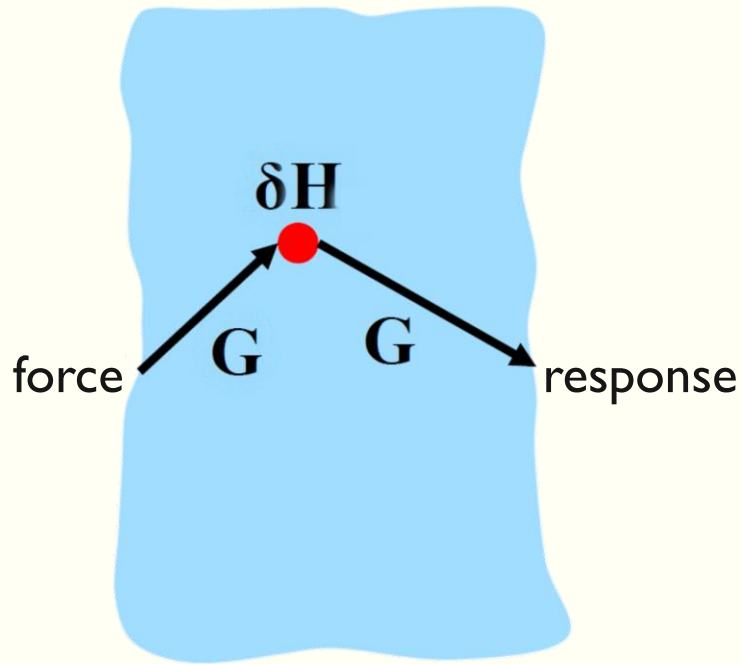
$\text{Fitness} = \langle \text{response} | \mathbf{G}(\text{gene}) | \text{force} \rangle$

* One can look at *more general* patterns of forces and response.

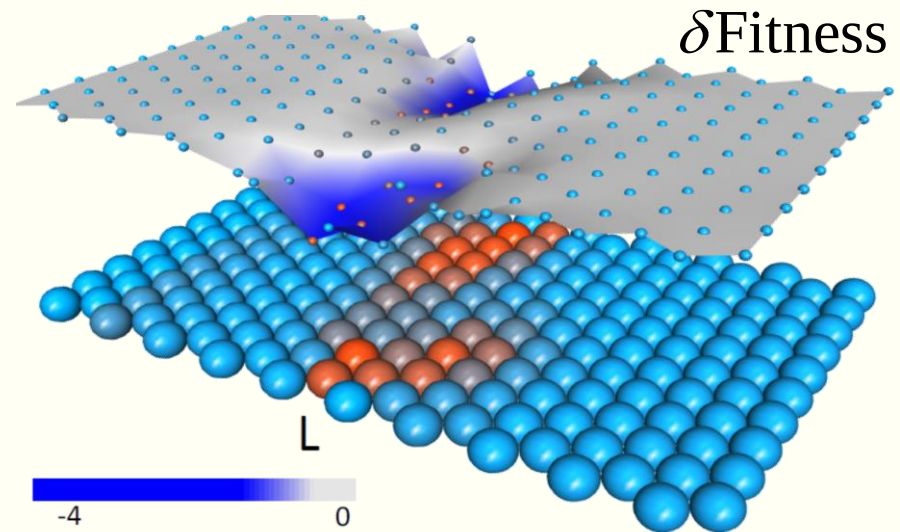
Mutations locally perturb the amino acid network, scatter the force, and thereby change the fitness

$\delta\mathbf{H}$ = mutation's effect on aa network

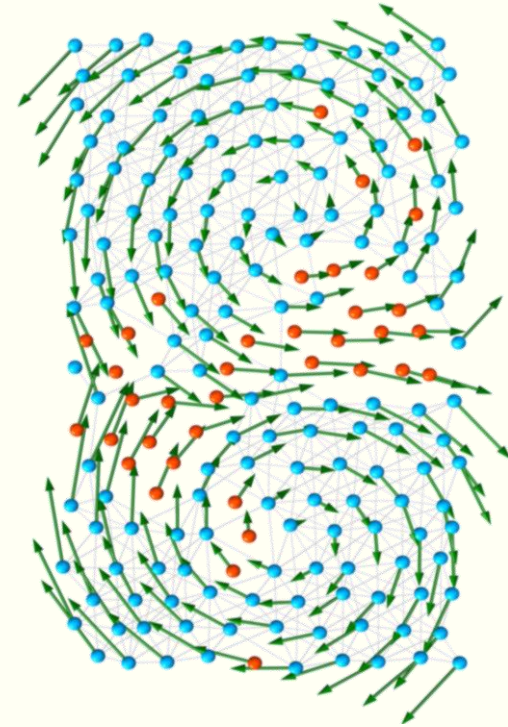
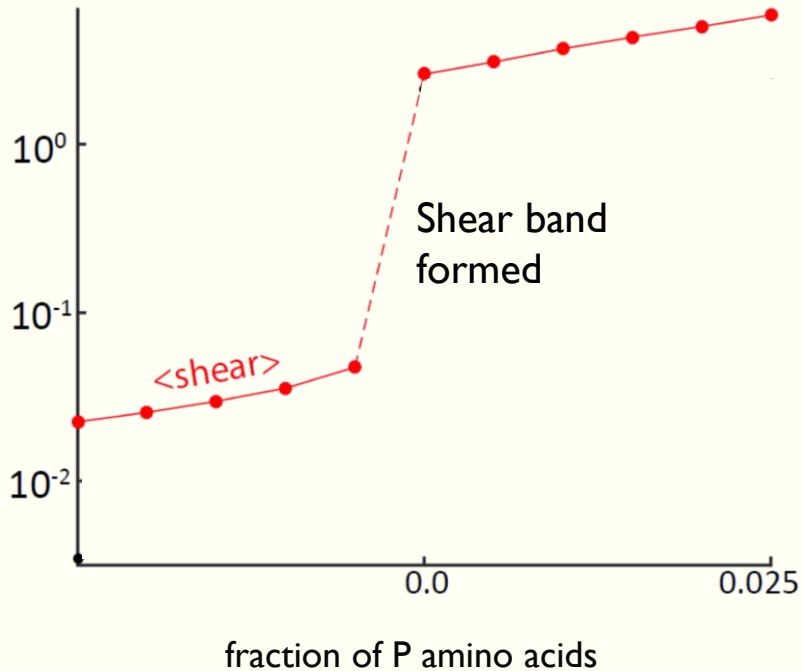
$$\delta\mathbf{G} = -\mathbf{G}\delta\mathbf{H}\mathbf{G} + \mathbf{G}\delta\mathbf{H}\mathbf{G}\delta\mathbf{H}\mathbf{G} - \dots$$



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



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



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

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

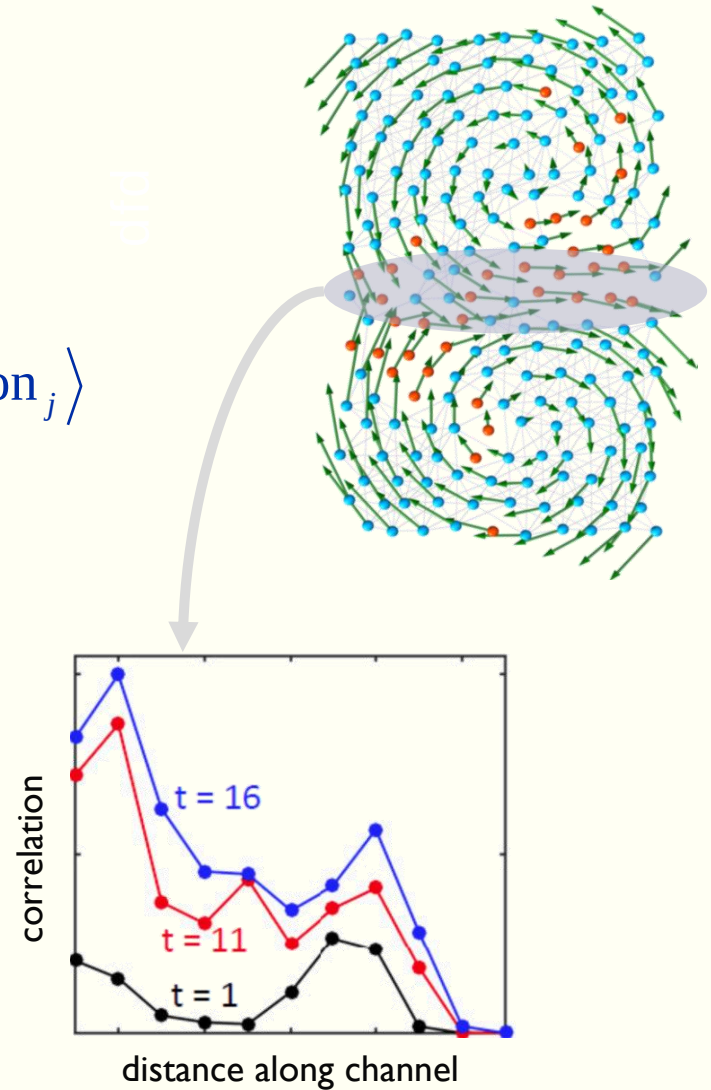
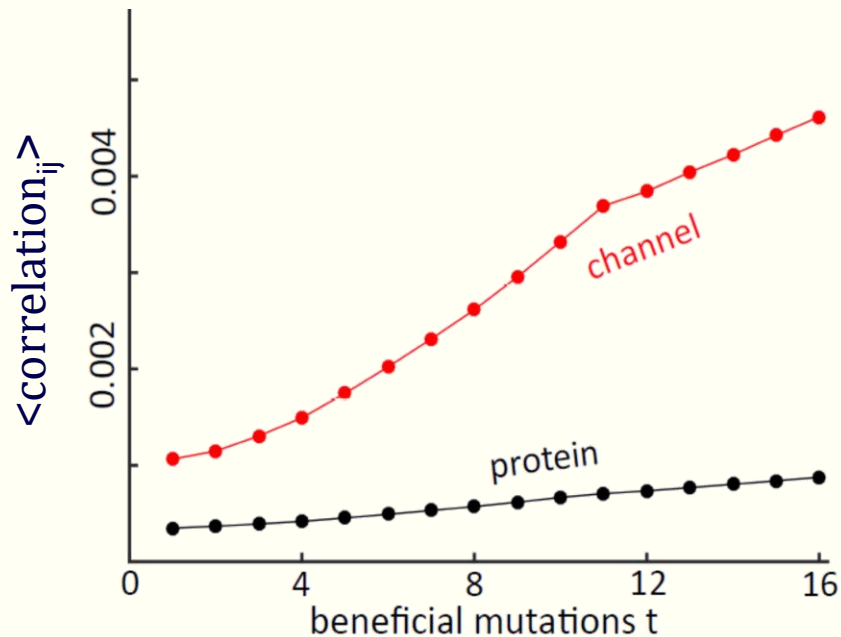
10^6 solutions

```

{
  11100101011111000010101...
  10010001000101011101110...
  00111110111110100111010...
  10011111110011001111000...
  ...

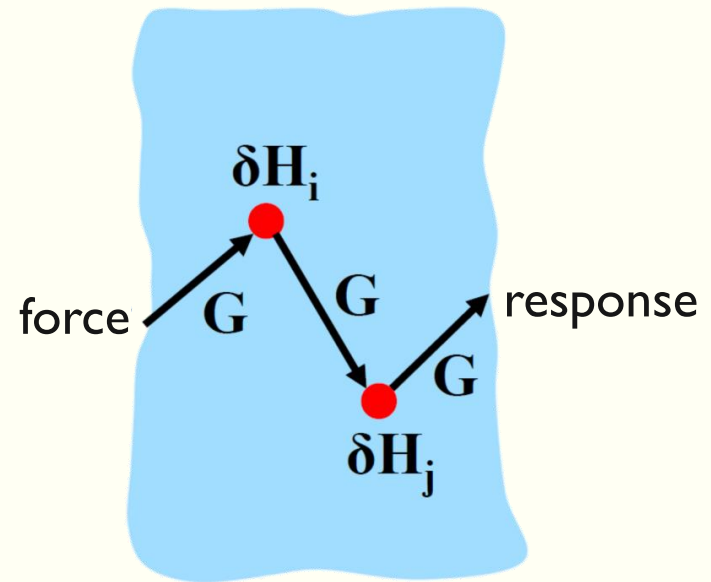
```

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



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

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

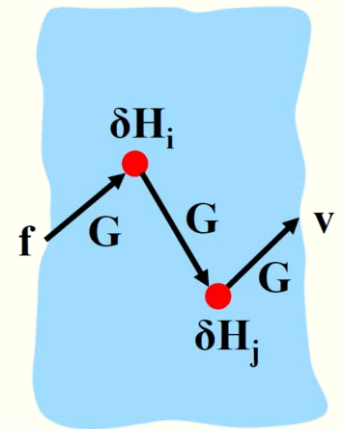
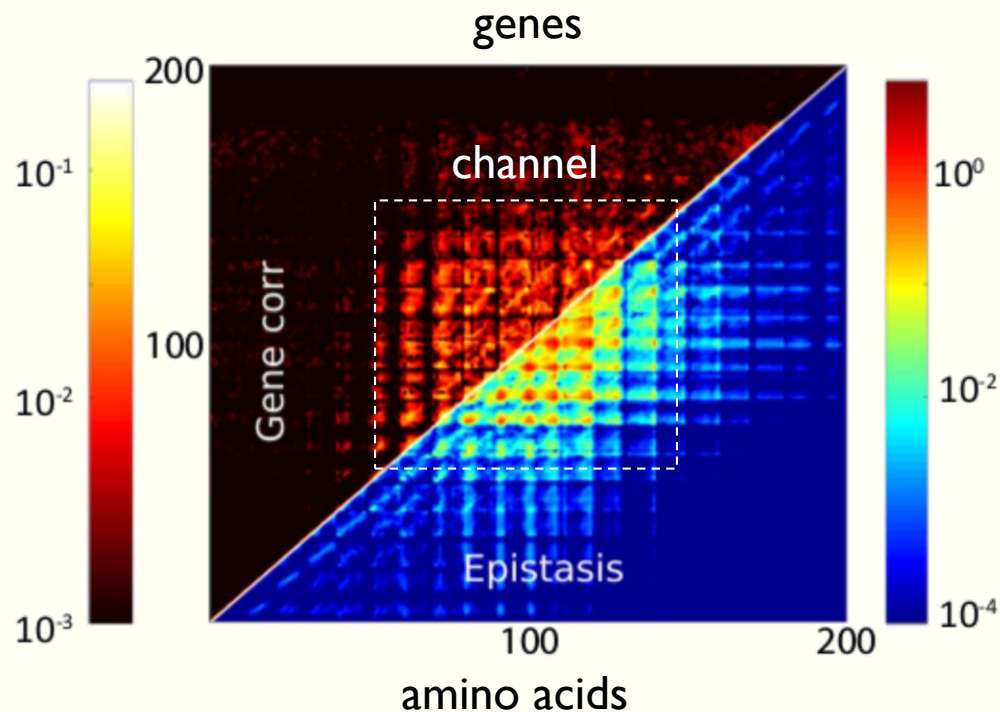


- Epistasis = sum over all scattering paths passing through i and j

$$\text{epistasis}_{i,j} = \langle \text{force} | \mathbf{G}\delta H_i \mathbf{G}\delta H_j \mathbf{G} + \mathbf{G}\delta H_j \mathbf{G}\delta H_i \mathbf{G} \dots | \text{response} \rangle$$

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

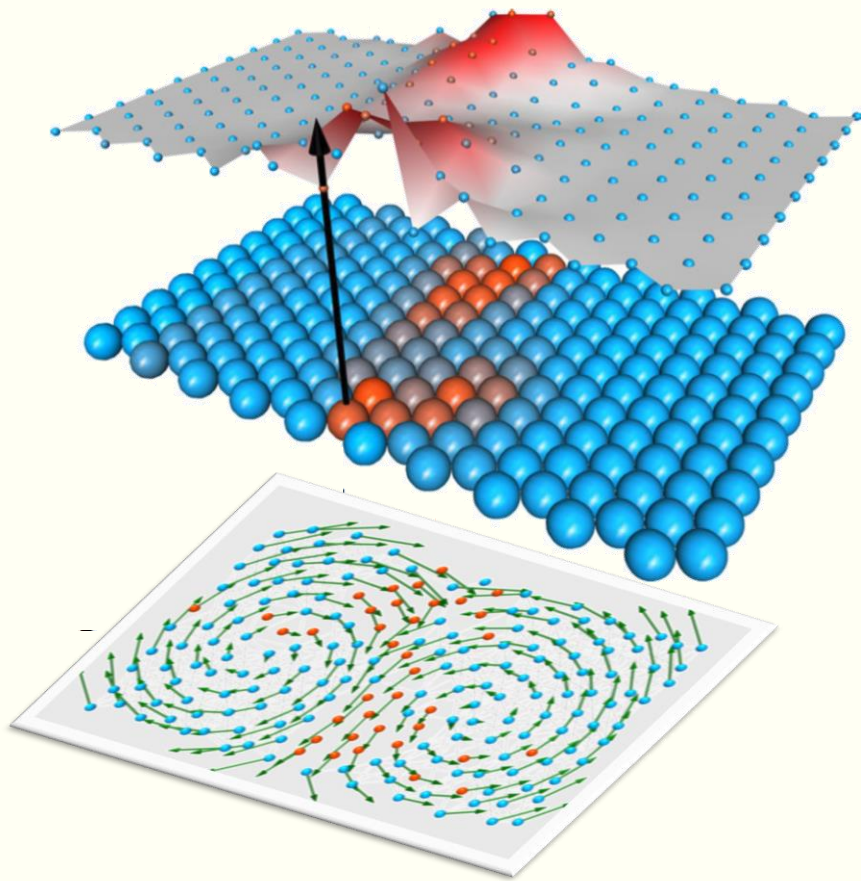
$$\text{correlation}_{ij} = \langle \text{codon}_i \text{codon}_j \rangle - \langle \text{codon}_i \rangle \langle \text{codon}_j \rangle$$



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

Epistasis and sequence correlation are strongest in the high shear region separating the rigid subdomains

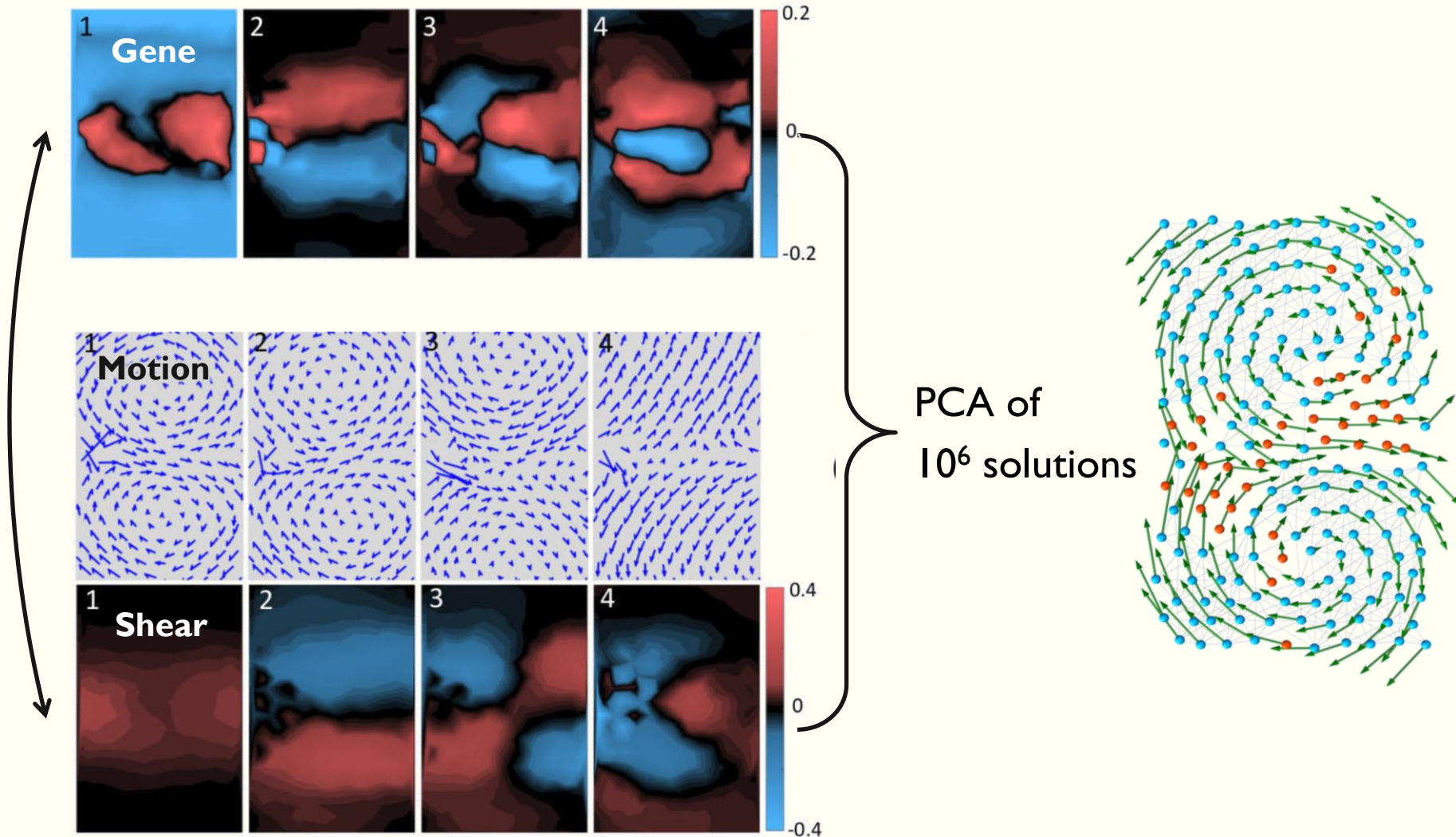
$$\text{epistasis}_{i,j} = \langle \text{force} | \mathbf{G} \delta \mathbf{H}_i \mathbf{G} \delta \mathbf{H}_j \mathbf{G} + \mathbf{G} \delta \mathbf{H}_j \mathbf{G} \delta \mathbf{H}_i \mathbf{G} \dots | \text{response} \rangle$$



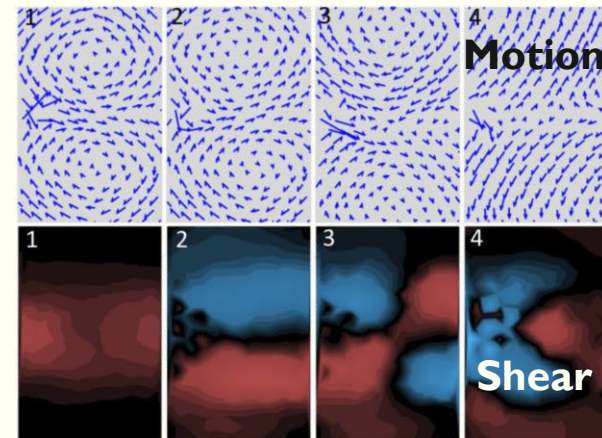
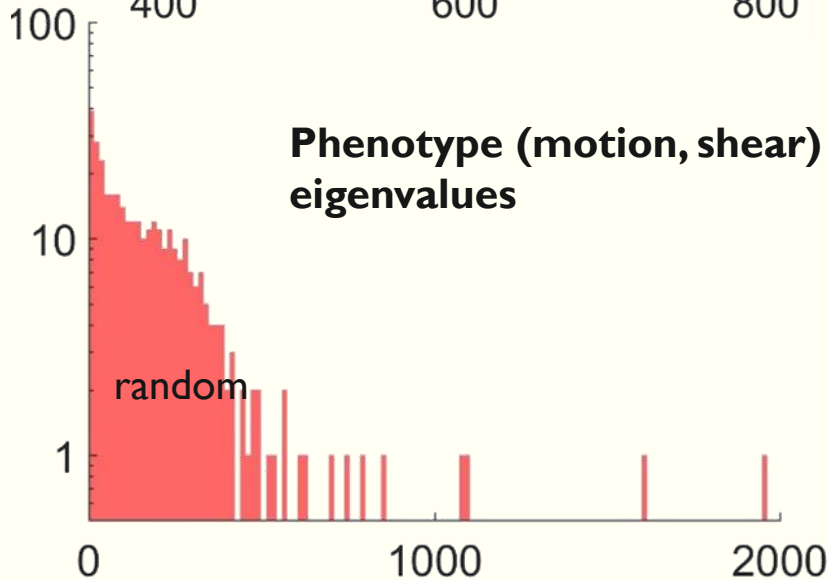
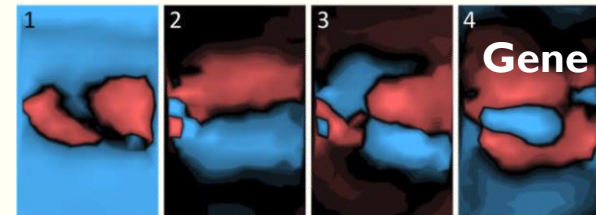
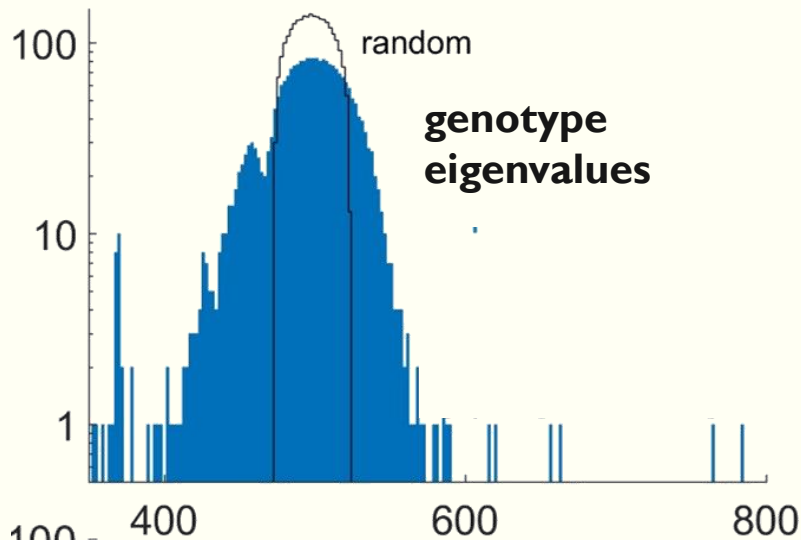
$$\frac{\text{epistasis}_{ij}}{\text{fitness}} \simeq \frac{h_i}{1+h_i} + \frac{h_j}{1+h_j} - \frac{h_i+h_j}{1+h_i+h_j}$$

$$h_i = \langle \text{motion} | \delta \mathbf{H}_i | \text{motion} \rangle = \Delta \text{energy}$$

PCA of phenotype and genotype capture the relevant modes of mechanical function and evolution



A few non-random eigenvectors capture the functionally relevant bit of information in the gene

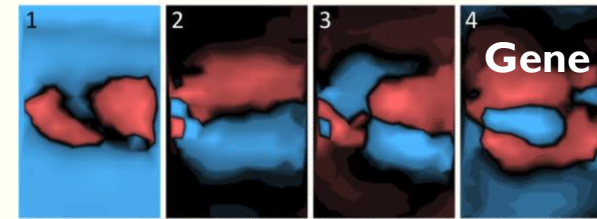
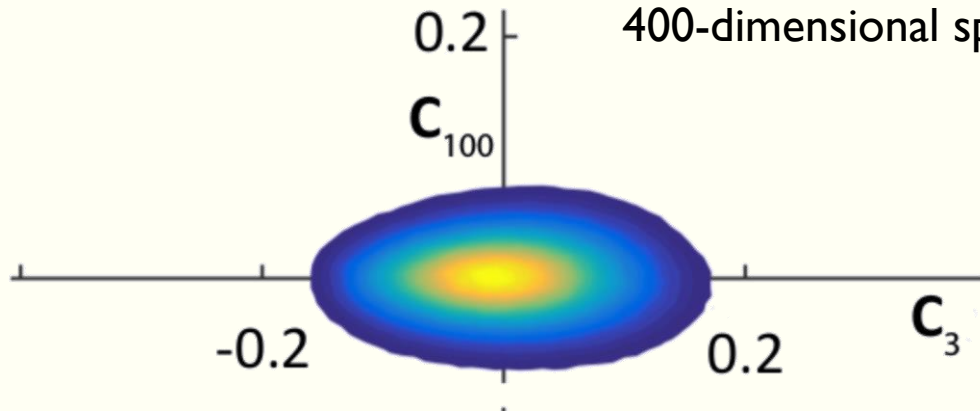


The genotype-to-phenotype map involves huge dimensional reduction

10^6 solutions

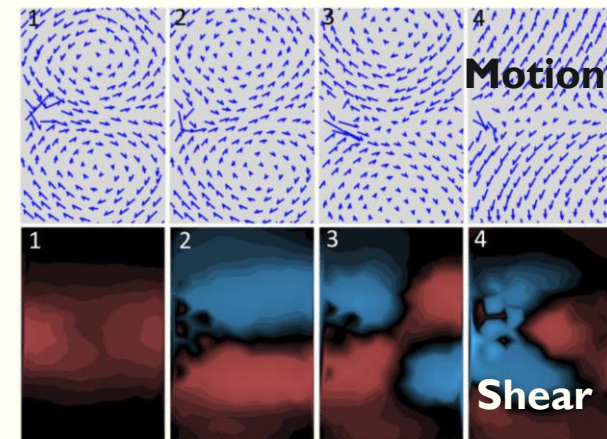
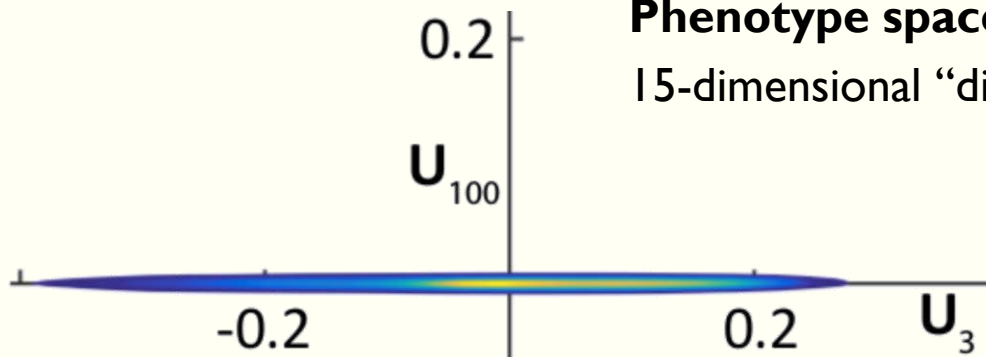
Genotype space

400-dimensional sphere

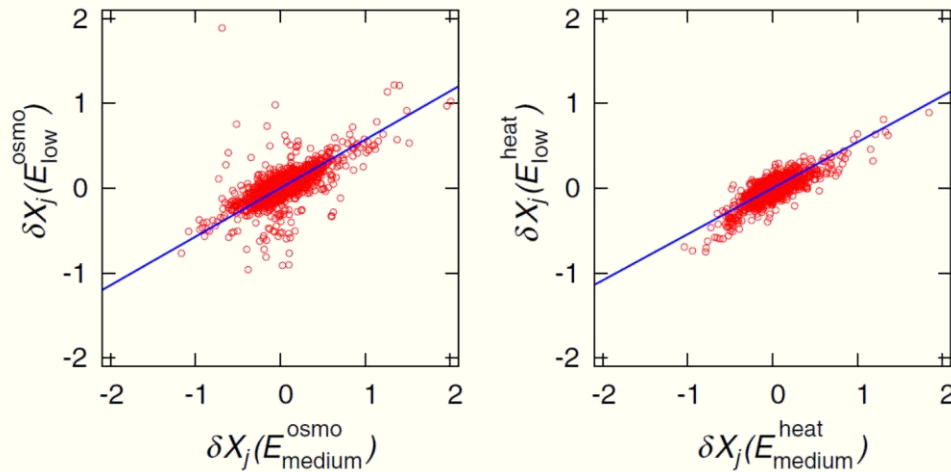


Phenotype space

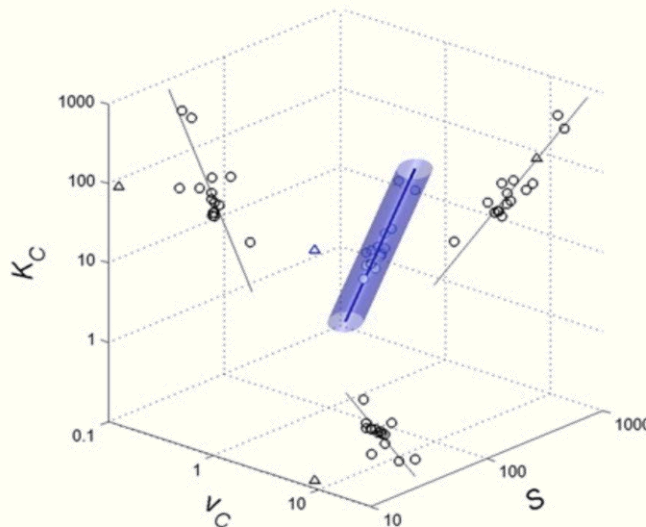
15-dimensional “disc”



Dimensional reduction is a general phenomenon in living systems



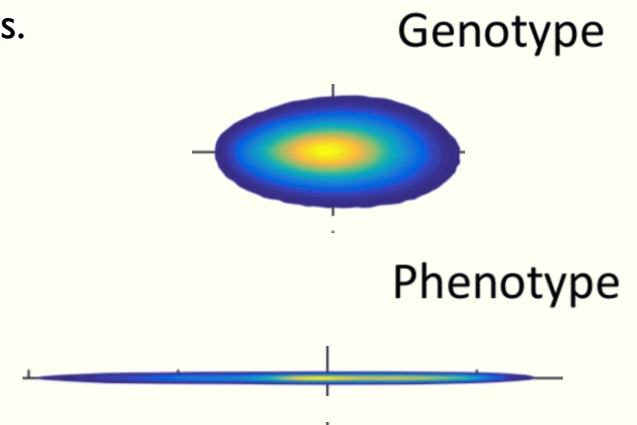
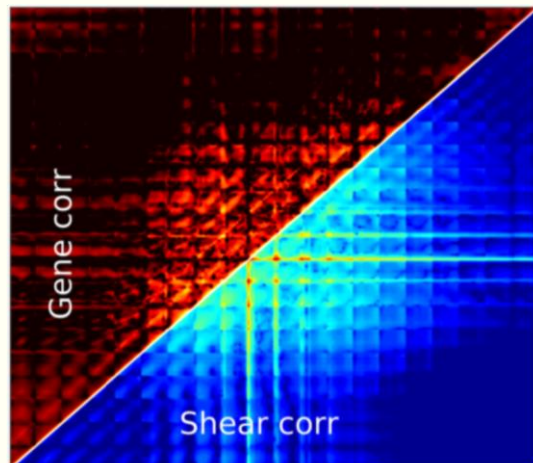
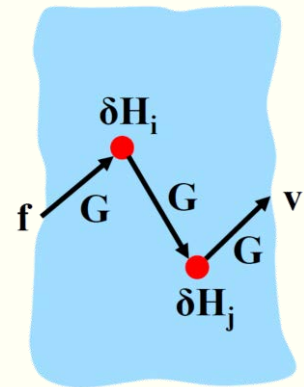
1D response of gene expression to stress
(Kaneko, Furusawa & Yomo, PRX 2015)



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

Genetics and function are linked by a minimal mechanical theory of protein

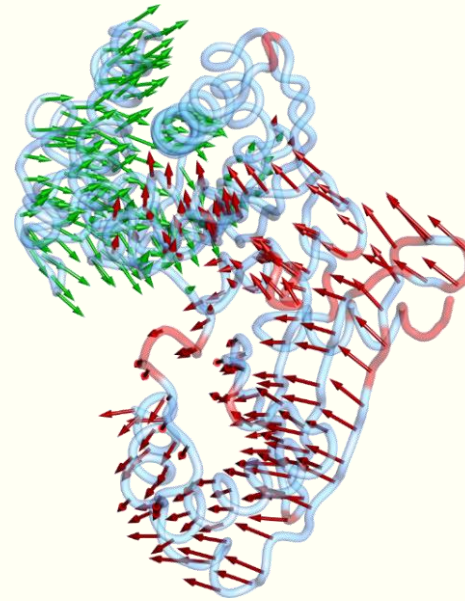
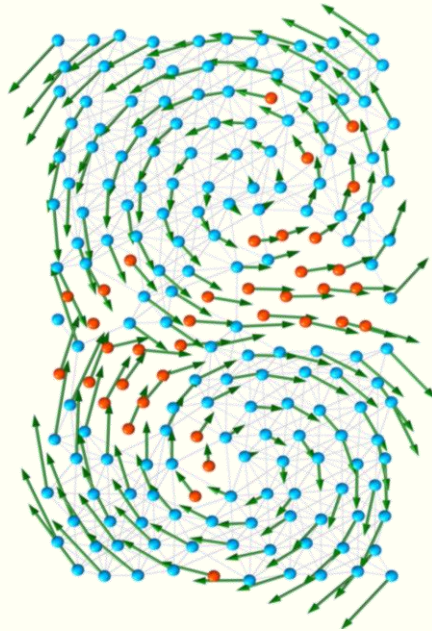
- **Mutations:** mechanical perturbations (“defects”).
- **Epistasis:** multiple scattering among mutations, which shapes:
- **Genotype-to-phenotype map** and **fitness landscape**
- **Soft modes** emerge at a **topological transition**.
- Mechanics (Green function) explain **dimensional reduction** and
- **Correspondence:** mechanical modes – coevolution modes.



Dutta Eckmann Libchaber & TT, PNAS 2018
TT Eckmann & Libchaber, Phys Rev X 2017

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

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



Thanks:

S. Dutta (IBS)

J.-P. Eckmann (Geneva)

A. Libchaber (Rockefeller)

S. Leibler & M. Mitchell (IAS)



**Center for Soft
& Living matter**