Center for Soft
& Living matter

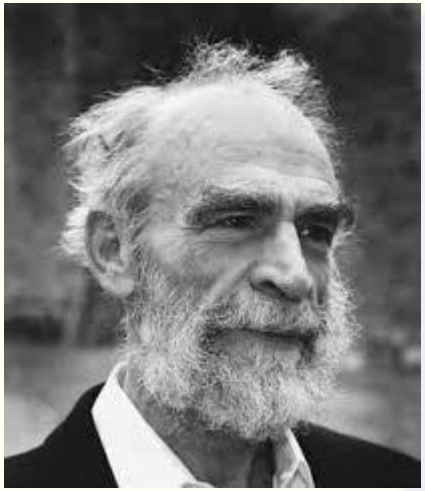
2019 Soft Matter Summer School:
Soft Matter and Slow Relaxation Dynamics
IBS CSLM, Ulsan, June 2019

MECHANICS OF EVOLUTION

Sandipan Dutta (IBS)
Jean-Pierre Eckmann and Jacques Rougemont (Geneva)
Albert Libchaber (Rockefeller)
Michael Mitchell and Stanislas Leibler (Rockefeller & IAS)

General motivation:

Is there a mathematical physical language of living matter?

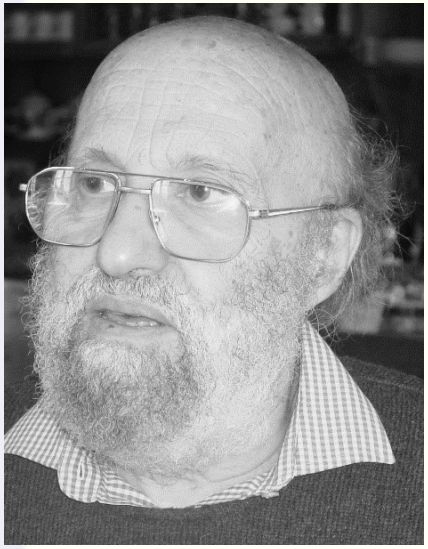


"…You feel there must be a new world of mathematical structures shadowing what we see in Life, a ***new language*** we do not know yet, something in the spirit the 'language' of calculus we use when describing physical systems."

Misha Gromov, *In a search for a structure*, 2013

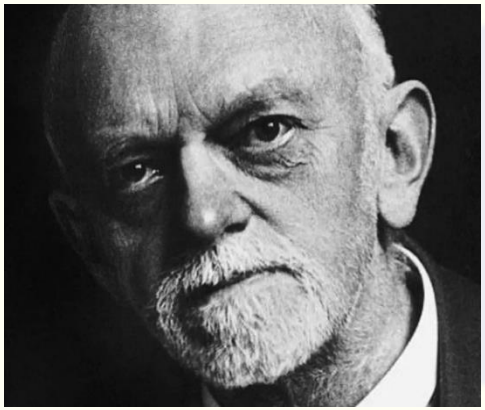
"***Theoretical physics*** was recognized as an independent field of research only at the end of the 19th century… the time has come for a more precise characterization of the research field of ***theoretical biology***, and for an assessment of its scope."

Giovanni Jona-Lasinio, *Mathematical models and languages in the study of biological problems*, 2012



General motivation:

Is there a mathematical language of living matter?



"Every kind of science, if it has only reached a certain degree of maturity, automatically becomes a part of mathematics. "

David Hilbert *Axiomatic Thought* (1918)

So before we start, four things:

- Only time will tell whether there is
  such a **"*general language*"** of theoretical biology.

- The aim of his lecture is to demonstrate a trial to think about a basic biological problem in terms of a ***physical theory***.

- ***Everything I say is wrong ****

- ***Much of this talk is about what we don't know.***

This is more or less our plan

I. Introduction

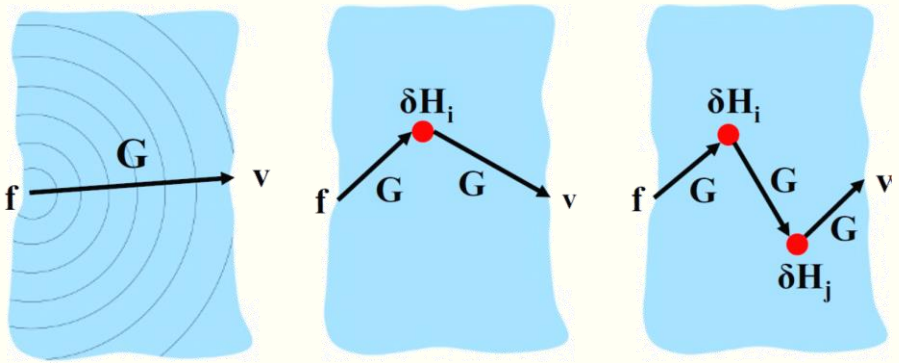
II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

} *informal introduction:* *basic ideas and questions*

IV. Mechanical evolution of protein

- *Condensed matter theory of protein using Green's functions.*
- *Linking to biological phenomena*



I. Introduction

- Why bother?
- ***What is life?***
- Stuff living matter is made from.
- Proteins.

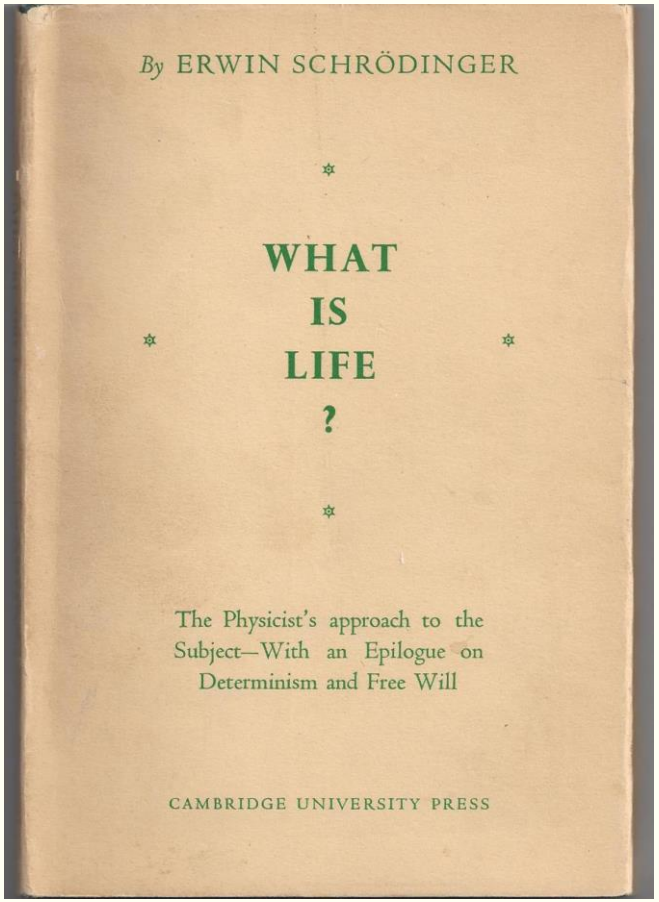
II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

What is Life?

- We know it when we see it,
  or at least that's what we think.



Schrodinger 1944: *"The large and important and very much discussed question is: How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?*

The preliminary answer can be summarized as follows:

The obvious inability of present-day physics and chemistry to account for such events is no reason at all for doubting that they can be accounted for by those sciences."

Can life be defined?

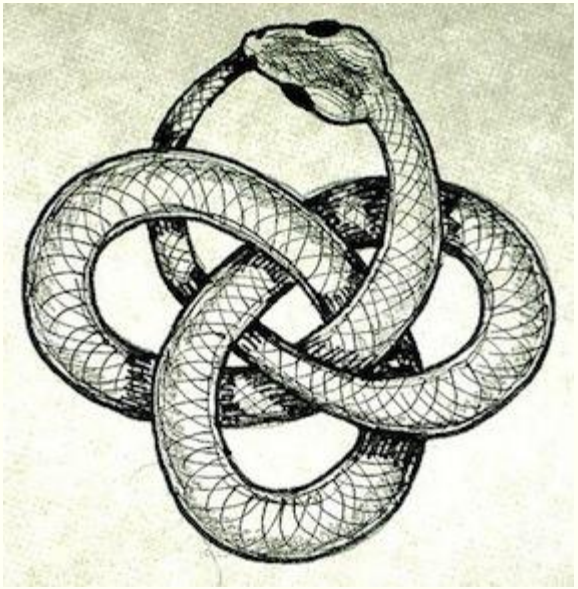
Webster dictionary

LIFE ***1a*** **:** the quality that distinguishes a vital and functional being from a dead body.

b **:** a principle or force that is considered to underlie the distinctive quality of animate beings

c **:** an organismic state characterized by capacity for metabolism, growth, reaction to stimuli, and reproduction

2a **:** the sequence of physical and mental experiences that make up the existence of an individual



The ouroboros symbolizes self-reflexivity, introspection, or cyclicality.

Wikipedia

LIFE is a characteristic distinguishing physical entities having biological processes, such as signaling and self-sustaining processes, from those that do not.

More on definitional loops in *Levari et al.* PRX 2012.

Plan B: define life by its *symptoms*

Descriptive "definition": A living thing has ***all or most*** of the following traits:

- **Homeostasis**: regulation (pH, temperature)
- **Organization**: composed of one or more cells.
- **Metabolism**: chemicals and energy into cellular components (anabolism) and decomposing organic matter (catabolism).
  Living things require energy to live.
- **Growth and development**.
- **Adaptation**: response to environment. This ability is fundamental to the process of evolution.
- **Response to stimuli**: senses of multicellular organisms. Motions toward the sun (phototropism), or nutrient (chemotaxis).
- **Reproduction**: the ability to produce new individual organisms.

I. Introduction

- Why bother?
- What is life?
- ***The stuff living matter is made from.***
- Proteins.

II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

Biological macromolecules are made of a small set of elements

carbon (C), oxygen (O), nitrogen (N), hydrogen (H), phosphorous (P)

***Four main classes*:**

(A) Nucleic acids (DNA):

store and process information…

(B) **Protein** (hemoglobin):

catalysis, structure, signaling…

(C) **Lipids** (phosphatidylcholine):

membrane, organelles…

(D) **Carbohydrates** (polysaccharide):

energy storage, surface, cell wall…

- Huge diversity in each class.

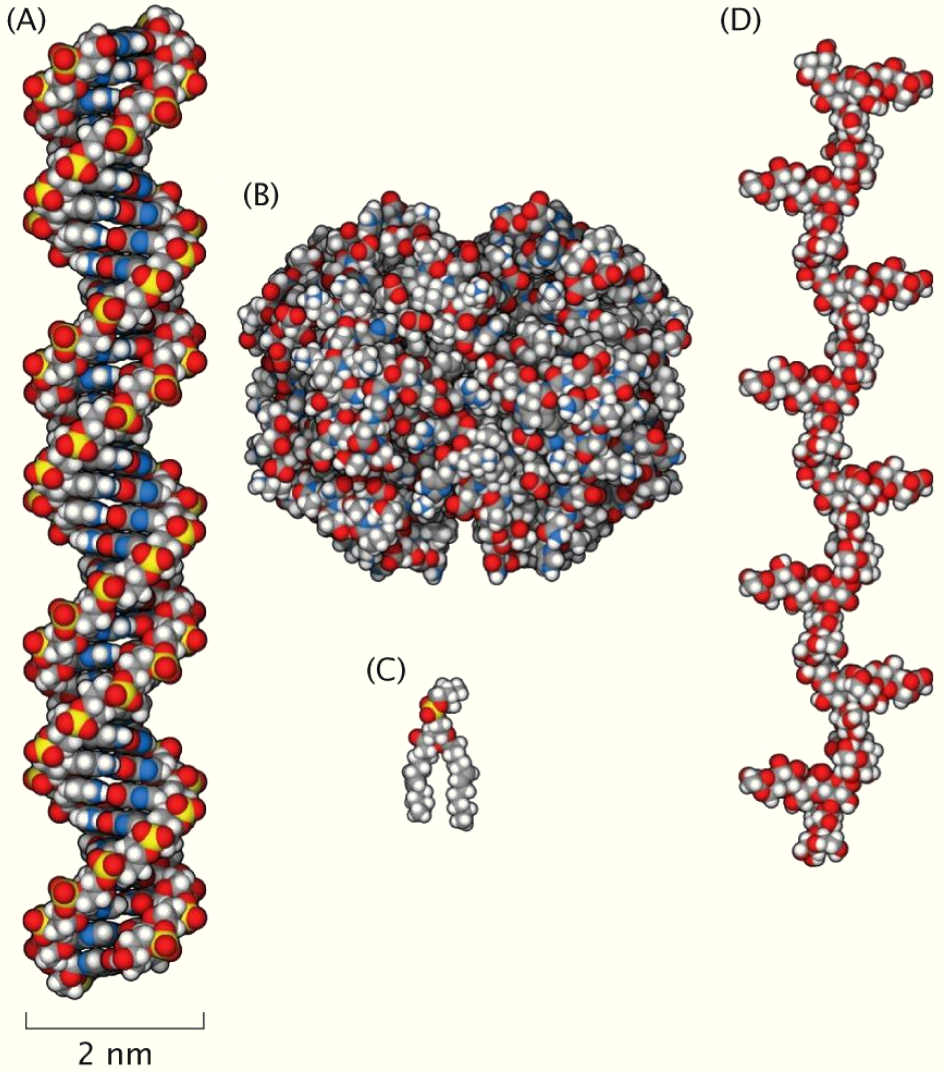


Figure 1.1 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

Why life molecules are so “simple”?

- Only four classes of molecules?
- Why these classes and no others?

- Origin of Life –
  many theories, none testable.

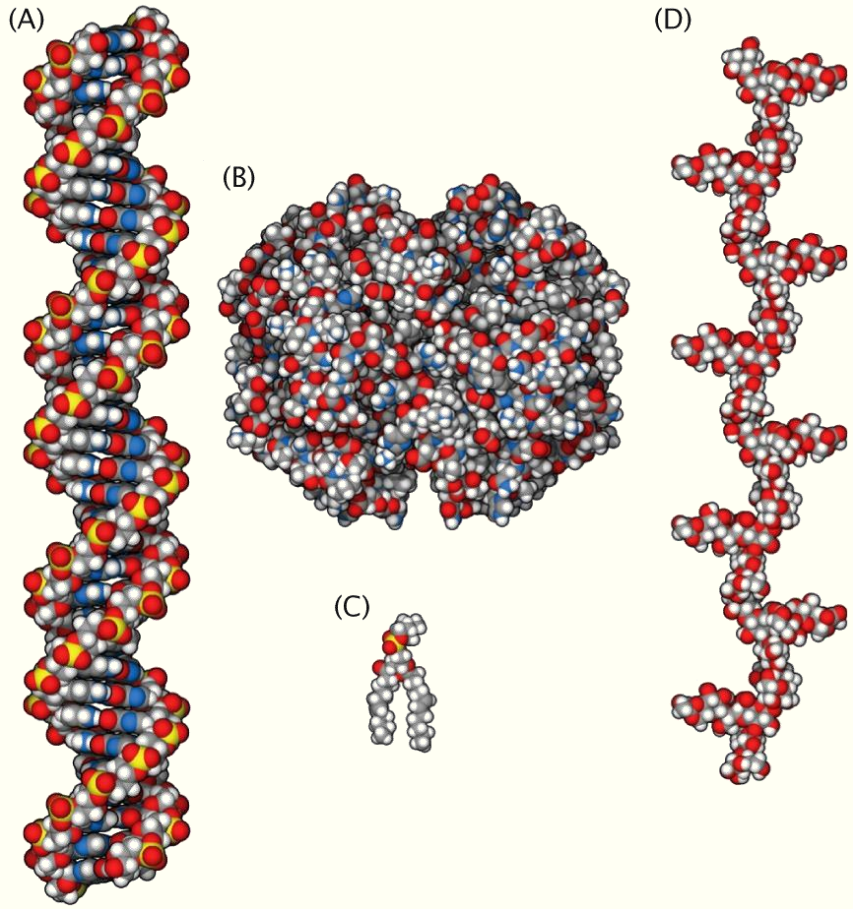
Advantages:

- Small number of basic elements and precursors.

- Modular combinatorial chemistry.

- Universal to all organisms that can eat\decompose each other.

- ***These are hallmarks of languages ****



I. Introduction

- Why bother?
- What is life?
- The stuff living matter is made from.
- ***Proteins.***

II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

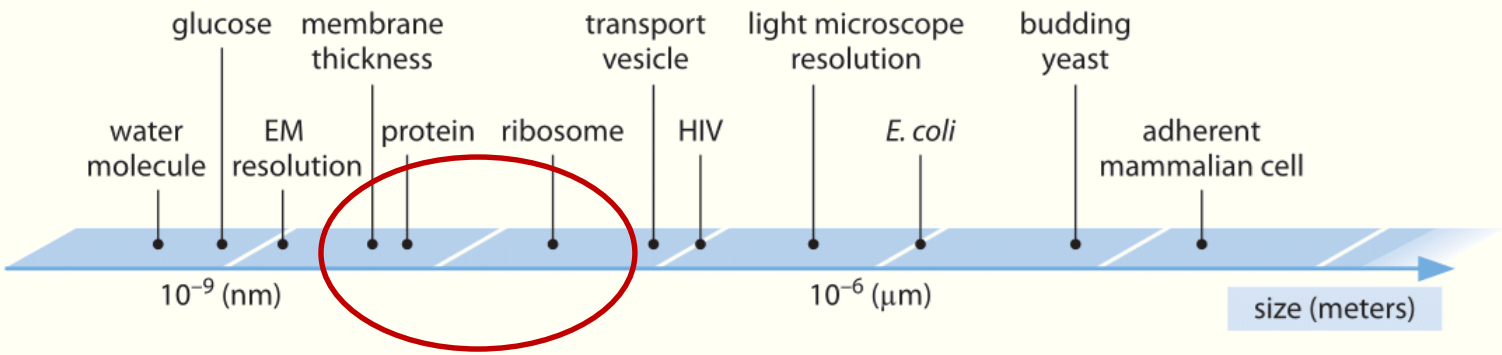
IV. Mechanical evolution of protein

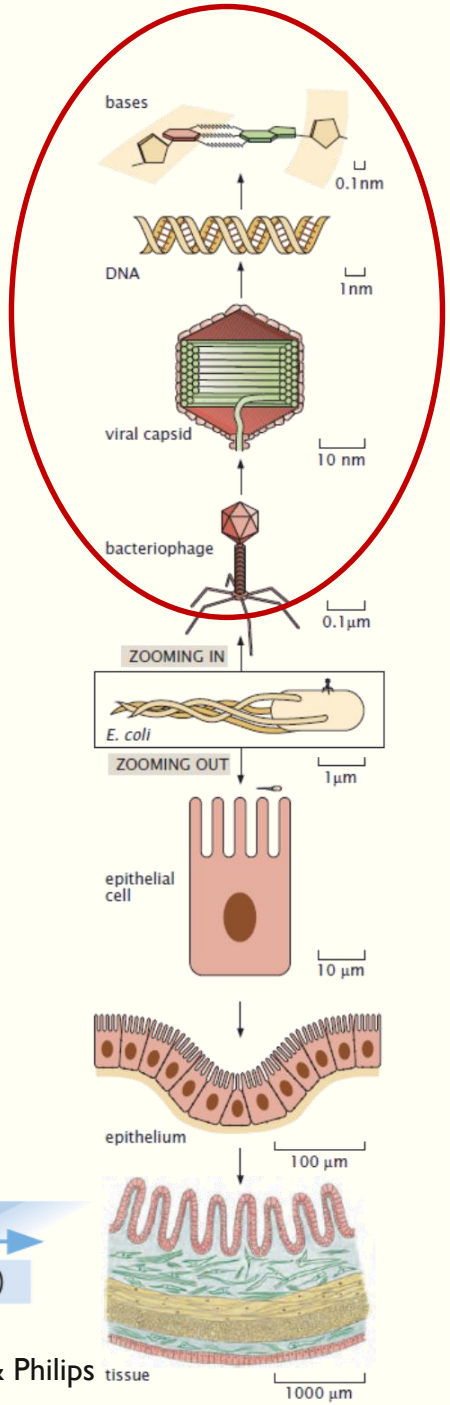
Living matter is assembled into multiscale hierarchal structures

We will stay the ***nanoscale*** (DNA, proteins):

- These objects are not alive by themselves.
- But since they are components of a living system:
- They are selected by evolution and therefore ***information-rich.***

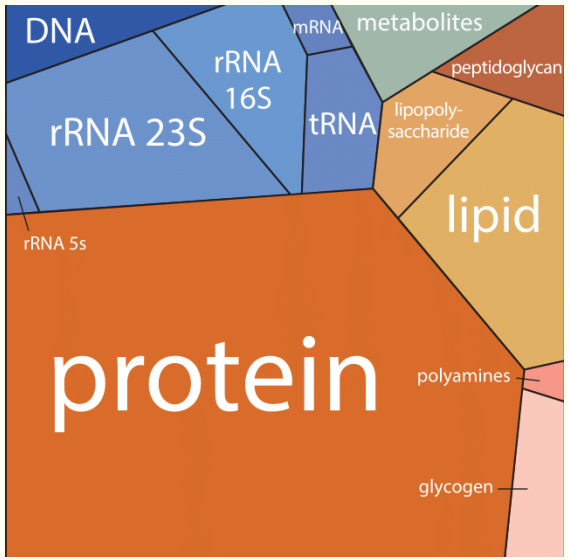
- So at the nanoscale the question is: ***how the information about life and evolution is mapped into physical objects.?***



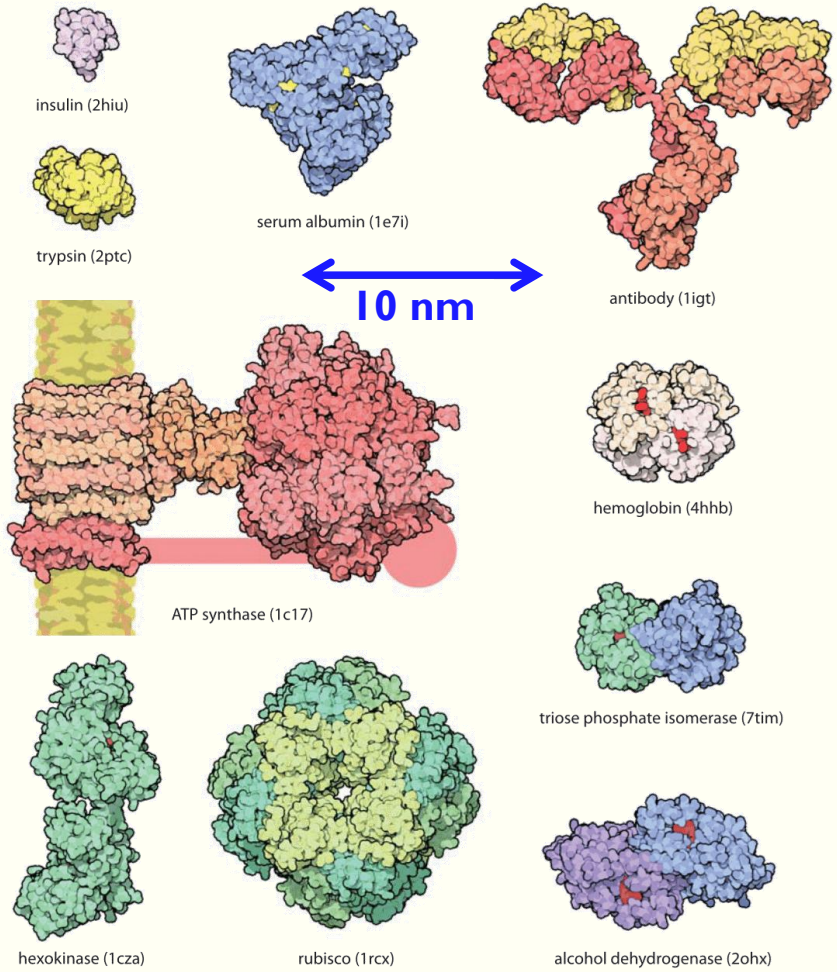


Specifically, we will examine proteins: molecular machines that perform numerous essential life functions

Structure, metabolism, defense, communication, storage, transport, information processing…



Dominant in cell biomass: ~ 10^6 proteins/μm^3



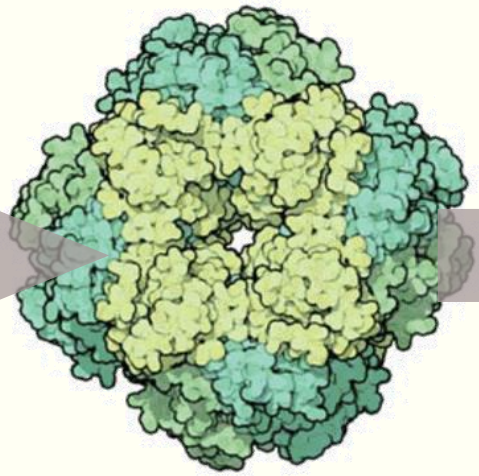
The macromolecules that make living matter,

lipids, hydrocarbons, nucleic acids, and especially **proteins**,

are among the most studied objects of Nature.

- Crystallography,
- Biochemical assays
- Mass spectrometry
- Fluorescence imaging
- Electron microscopy
- Directed evolution
- Deep sequencing
- Computational models

…



Physical Principles?

I. Introduction

II. Imperfect duality in living matter

- ***The DNA/protein schism (RNA...)***
- Representation-object duality.
- Genotype-to-phenotype map.
- The protein question.

III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

Two dual languages: *nucleic acids* and *proteins*

- Small alphabets:
  - 4 nucleic bases
  - 20 amino acids.
- Translation:
  - 3-letter words (codons)
  - Each encodes 1 AA.
- Hierarchical languages: modular subunits.

* *DNA encodes also many non-protein entities which we will not discuss here.*

**There are other "languages" which we will not discuss (cell states, epigenetics…)*

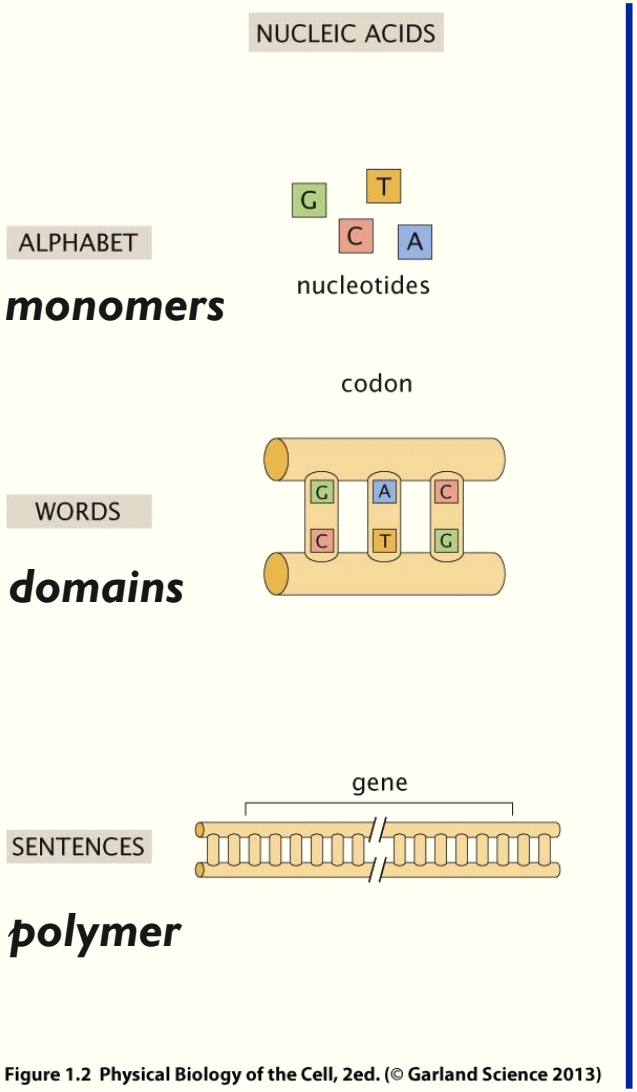


Figure 1.2 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

One may think of

genes and proteins as

***symbols* and their meanings or**

representations* and *objects

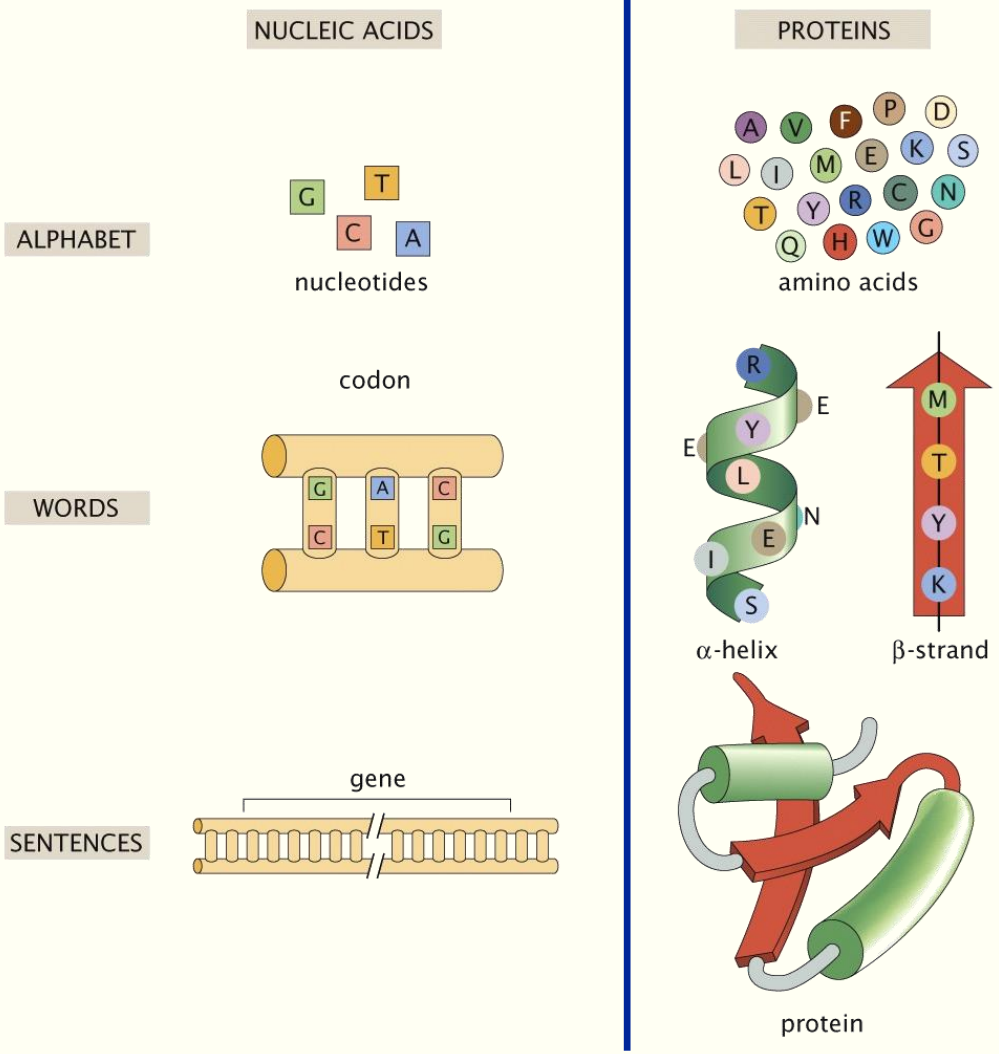


Figure 1.2 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

In terms of von Neumann's universal constructor:

DNA is the tape and proteins make the constructor *

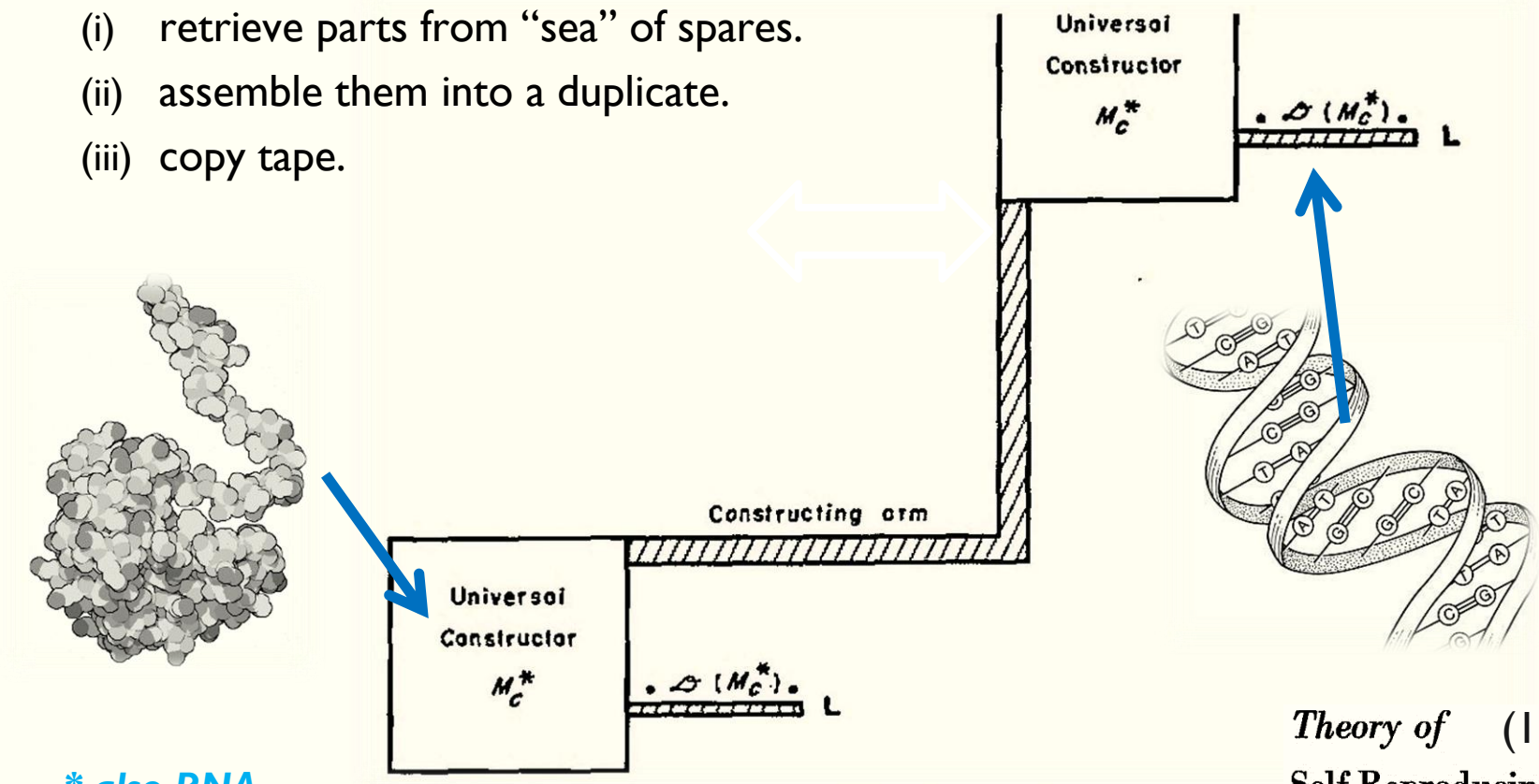
Self-reproducing machine: constructor + tape **(1948/9)**

- Program on tape:

(i) retrieve parts from "sea" of spares.

(ii) assemble them into a duplicate.

(iii) copy tape.



Theory of (1966)
Self-Reproducing Automata

* *also RNA…*

I. Introduction

II. Imperfect duality in living matter

- The DNA/protein schism (RNA...)
- ***Representation-object duality.***
- Genotype-to-phenotype map.
- The protein question.

III. Protein: amorphous evolving matter

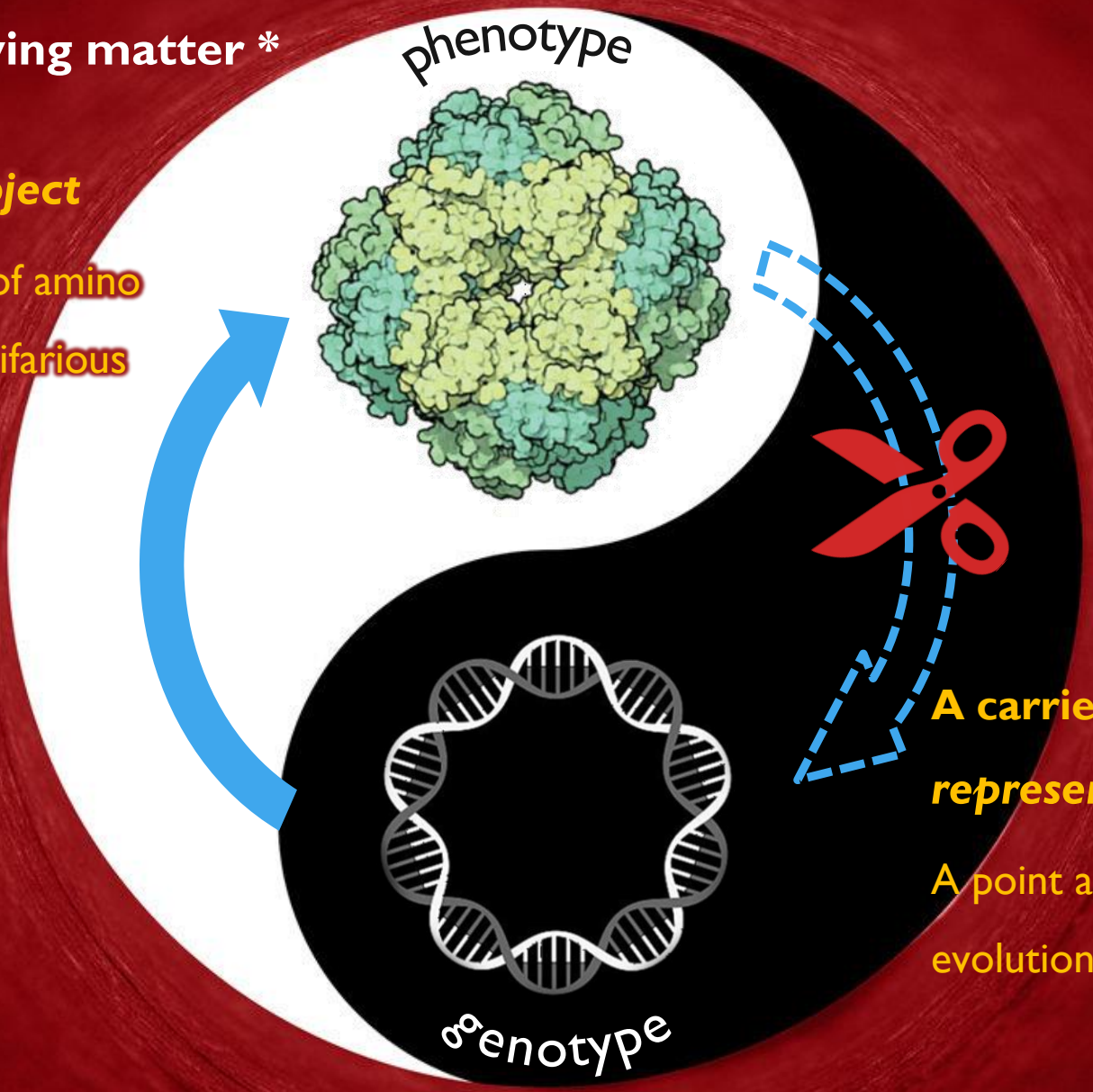
IV. Mechanical evolution of protein

Proteins are among the smallest entities that reflect the **object-representation duality of living matter** *

A physical *object*

a folded chain of amino acids with multifarious biochemistry.

•



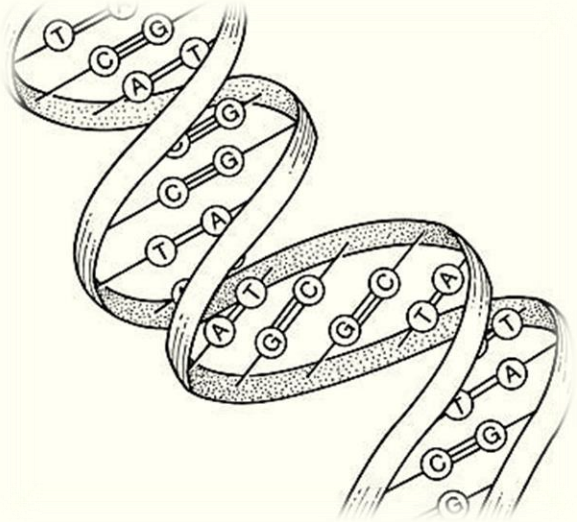
A carrier of genetic *representation:*

A point along an evolutionary trajectory.

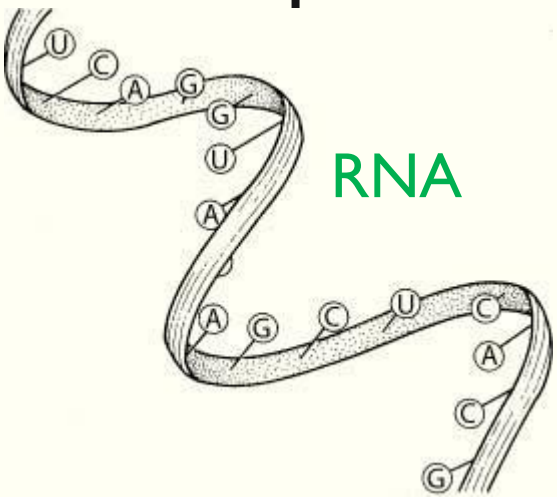
**** This is an imperfect duality***

RNA intermediates can be both objects and representation (tapes and machines)

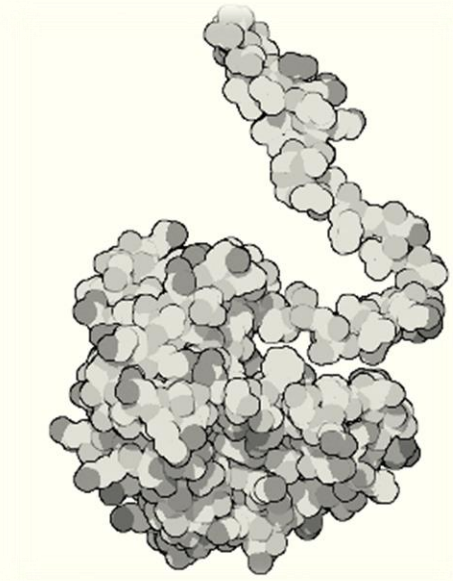
DNA



RNA



protein



Stable, slow, digital, almost perfect *

Noisy, dissipative, analog *

- Before the DNA-protein schism: Primordial "RNA world"
- RNA molecules are both information carriers (DNA) and executers (proteins).
- RNAs have several other functions in the cell*

I. Introduction

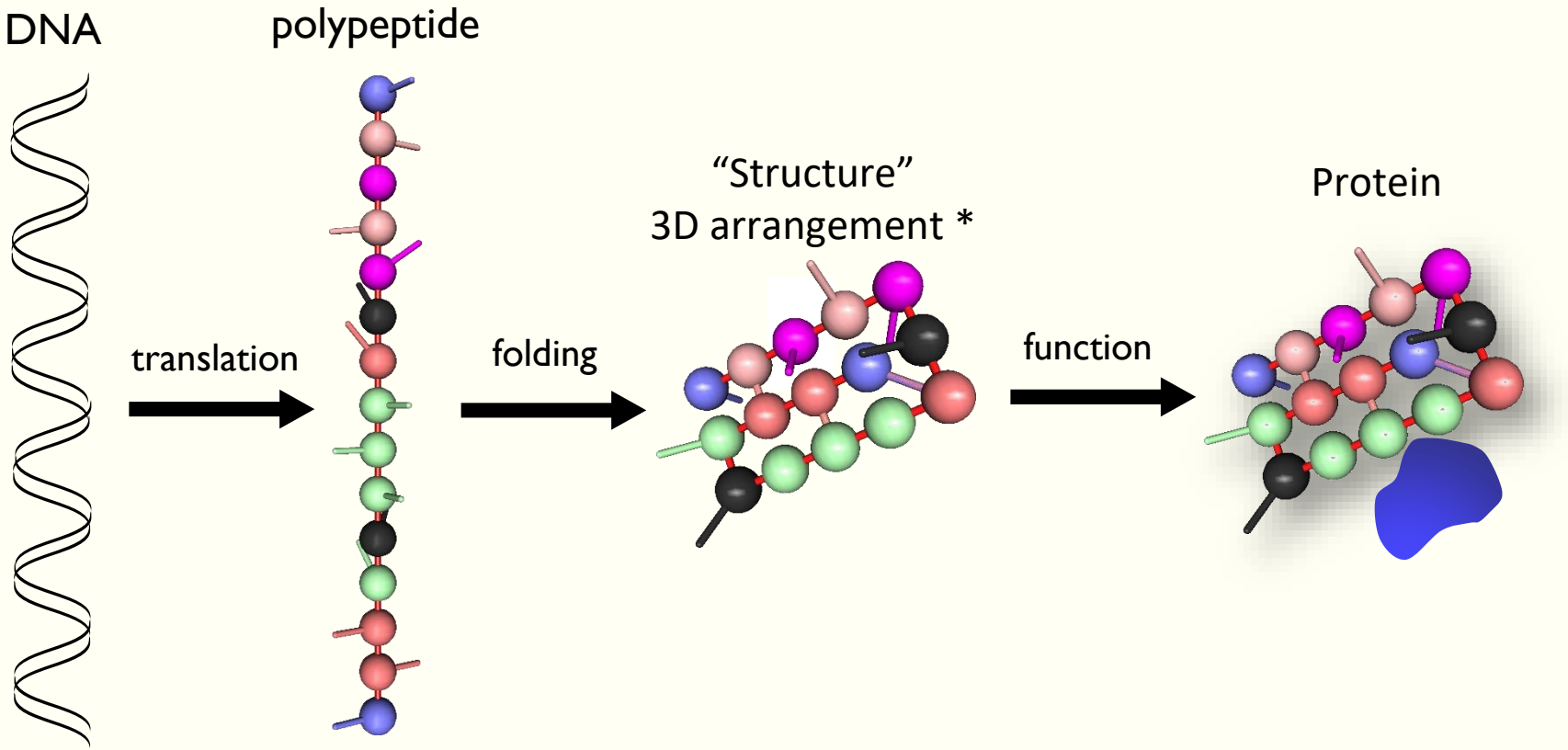
II. Imperfect duality in living matter

- The DNA/protein schism (RNA...)
- Representation-object duality.
- ***Genotype-to-phenotype map.***
- The protein question.

III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

Protein are mappings from sequence (genotype) to functions (phenotype)



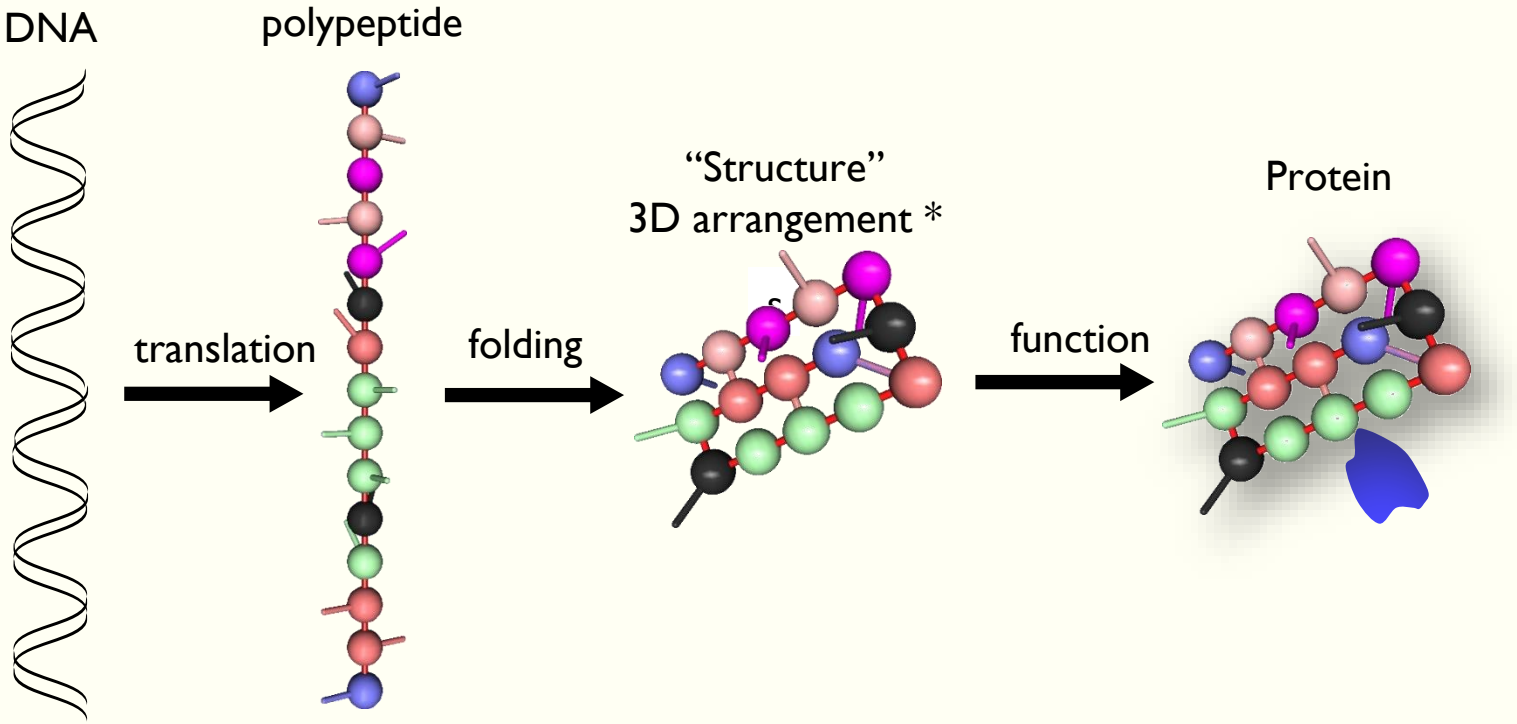
Formally, the map is

$$\textbf{protein} = F(\textbf{gene}) = \text{function} \circ \text{fold} \circ \text{tranlsate}(\textbf{gene})$$

Rev Mod Phys (in press)

This talk is mostly about a physical approach

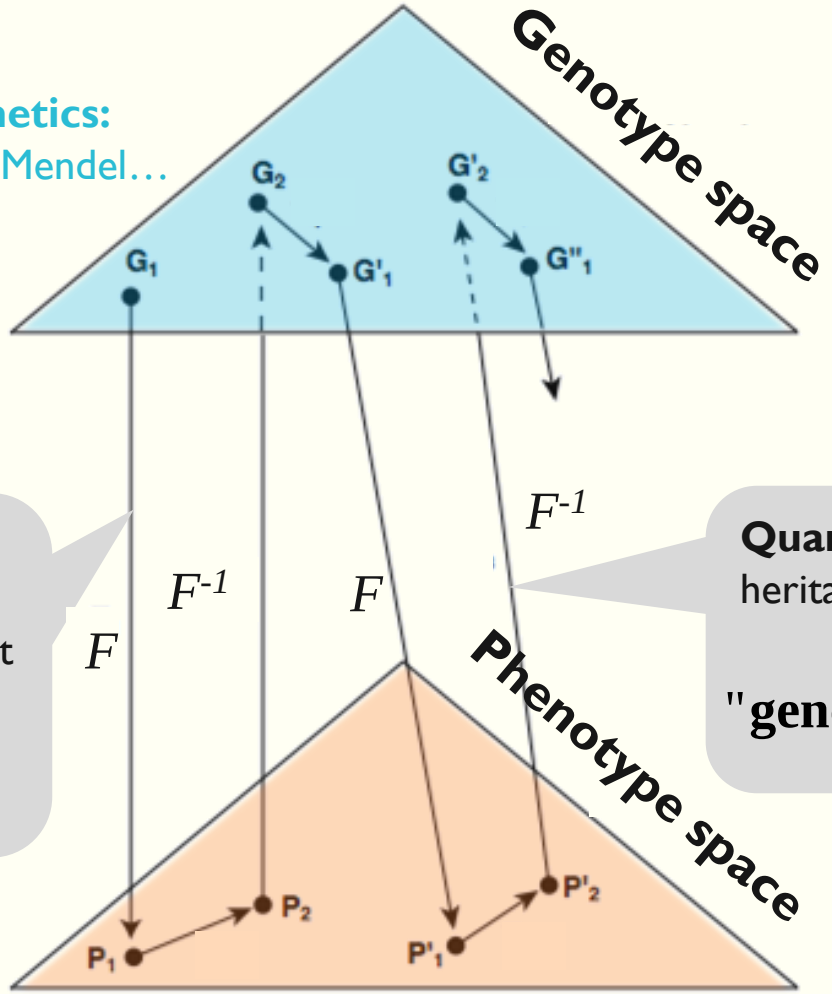
trying to explain certain basic features of the map



$$\textbf{protein} = F(\textbf{gene}) = \text{function} \circ \text{fold} \circ \text{tranlsate}(\textbf{gene})$$

Duality allows efficient search in a high-dimensional genotype space and selection in phenotype space

Stochastic population genetics:
mutation, horizontal transfer, Mendel…



Phenotype construction:
developmental and cellular biology, physiology, environment

$$\textbf{protein} = F(\textbf{gene})$$

Quantitative genetics:
heritability of phenotypes

$$\text{"gene"} = F^{-1}(\textbf{protein})^{*}$$

Richard Lewontin, 1974

Evolutionary biology of phenotypes:
evolutionary ecology, mating, migration, **selection**…

I. Introduction

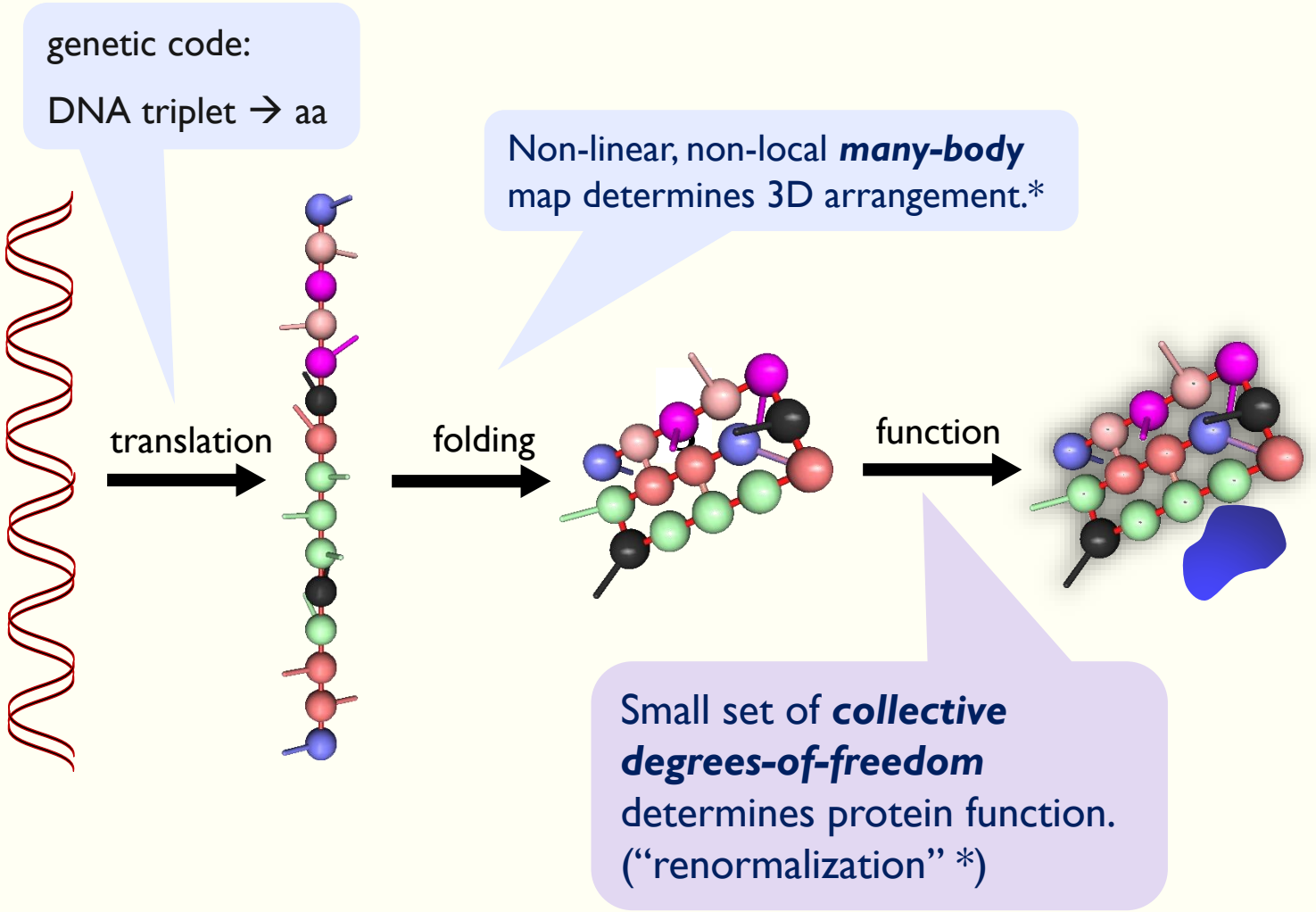
II. Imperfect duality in living matter

- The DNA/protein schism (RNA...)
- Representation-object duality.
- Genotype-to-phenotype map.
- ***The protein question.***

III. Protein: amorphous evolving matter

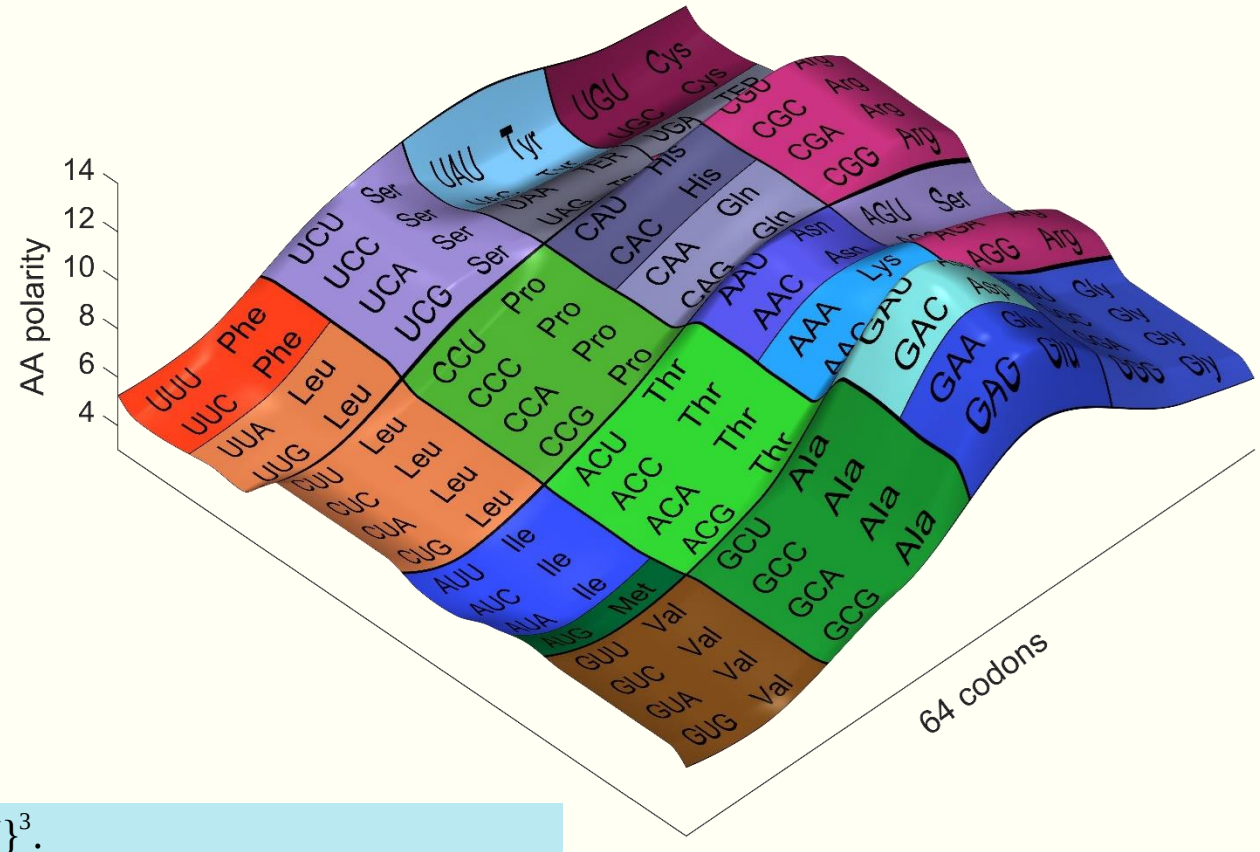
IV. Mechanical evolution of protein

The main question about protein is: ***how are the collective interactions that determine protein function encoded in the gene?***



* *We will not discuss folding*

The genetic code is a *smooth* one-body map from DNA codons to amino acids



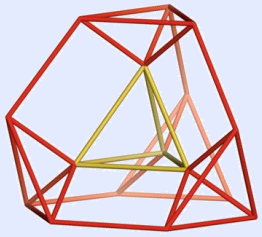
- 64 **codons** $\in \{A,G,C,U\}^3$.
- **Genetic code :** codon $\rightarrow$ amino acid.
- **Degeneracy :** 64 codons $\rightarrow$ 20 amino acids.
- **Smoothness**: close codons $\rightarrow$ similar/same amino acids.

Rev Mod Phys (in press)

In other words,

What is the geometry of the gene-to-function map?

Gene = sequence of *N* DNA bases from 4-letter alphabet {A,C,G,T}

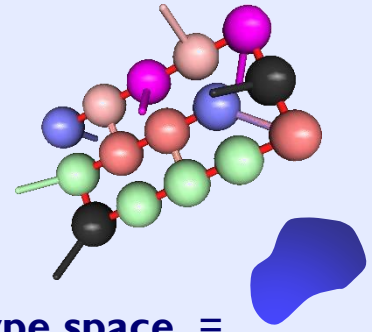


Genotype space =
N-dimensional hyper-tetrahedron

$$F : \{\mathrm{A,C,G,T}\}^N \rightarrow \mathbb{R}^d$$

$$F(\textbf{gene}) = \textbf{protein}$$

Function typically described by a few relevant numbers



Phenotype space =
low-dimensional $\mathbb{R}^d$

This talk is mostly about a physical approach

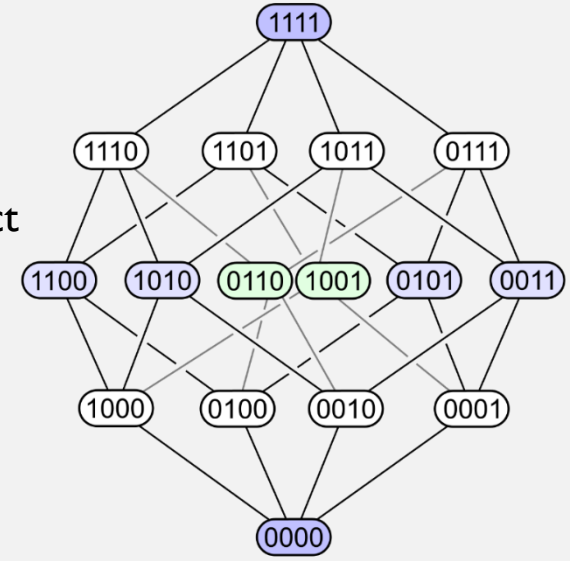
trying to explain certain basic features of the map

Exercise: draw a 6D hypercube.

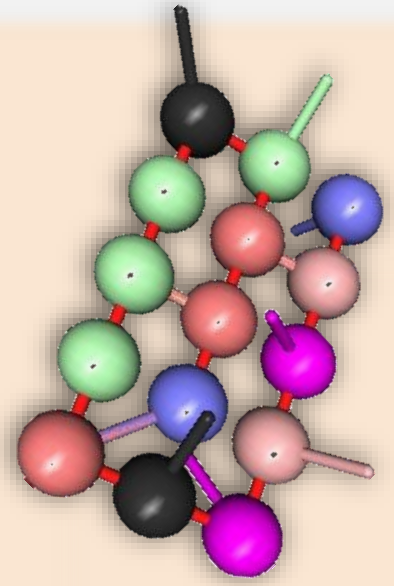
A general gene-to-function mapping is immensely complex but biophysical constraints make it much simpler

- Consider general Boolean functions (logical propositions) of *N* variables

$$F: \; x_1 x_2 \ldots x_N \in \{0,1\}^N \rightarrow \{0,1\}$$



- There are 2^{2^N} such functions because all variables x_i may interact
  (# of {0,1} combinations on N-cube)



But in protein

complexity is reduced

- ***Short-range*** many-body interactions are limited to neighboring amino acids .
- ***A few*** collective and local degrees-of-freedom determine the function.

  → ***much fewer*** physically consistent ***F*** functions.

I. Introduction

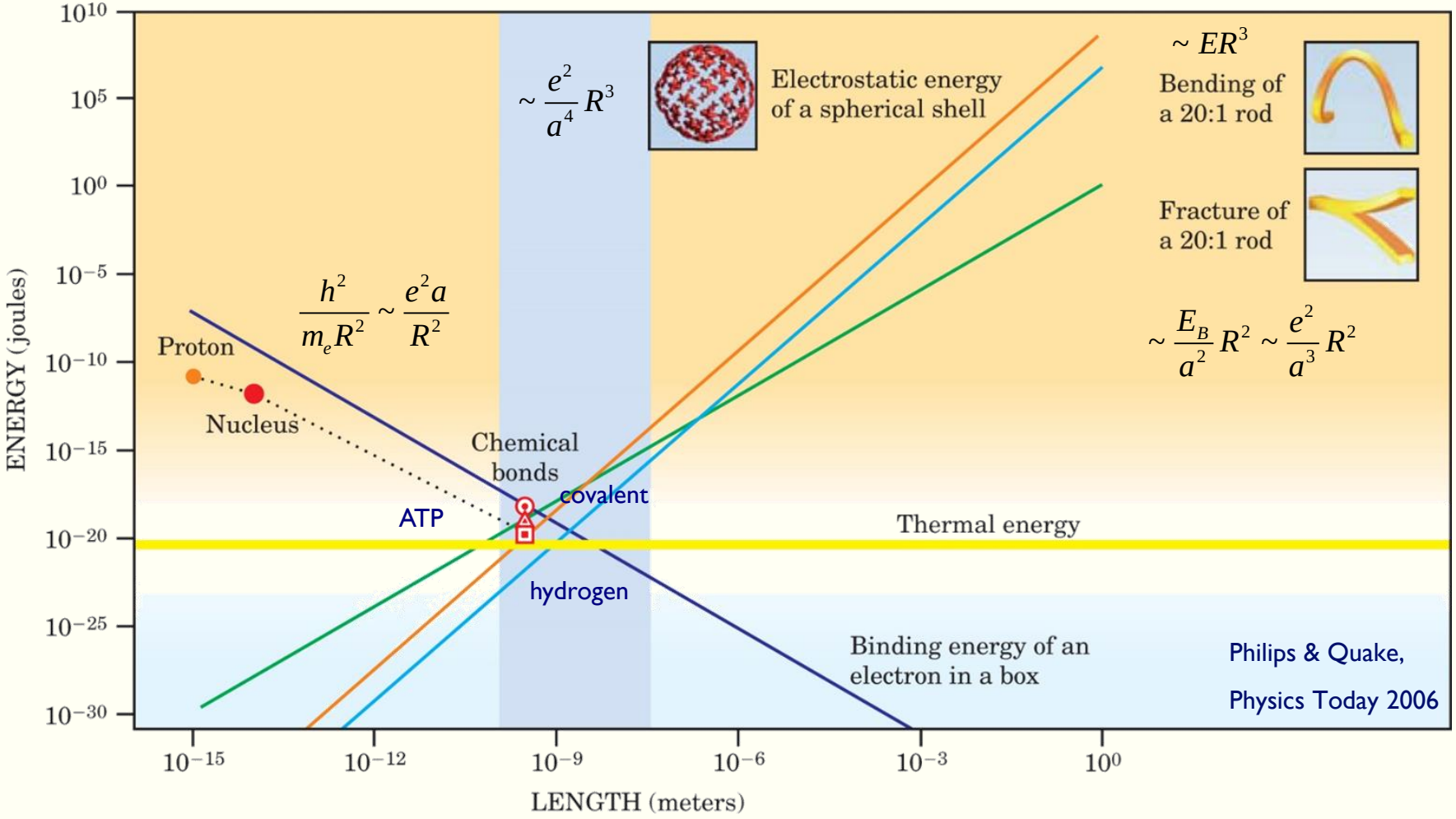
II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

- ***Physical scales and interactions.***
- Coarse-grained view of protein.
- Long-range deformation and viscoelasticity in proteins.
- Mechanical models of protein: spring nets.
- Amorphous matter: rigidity, spectrum.

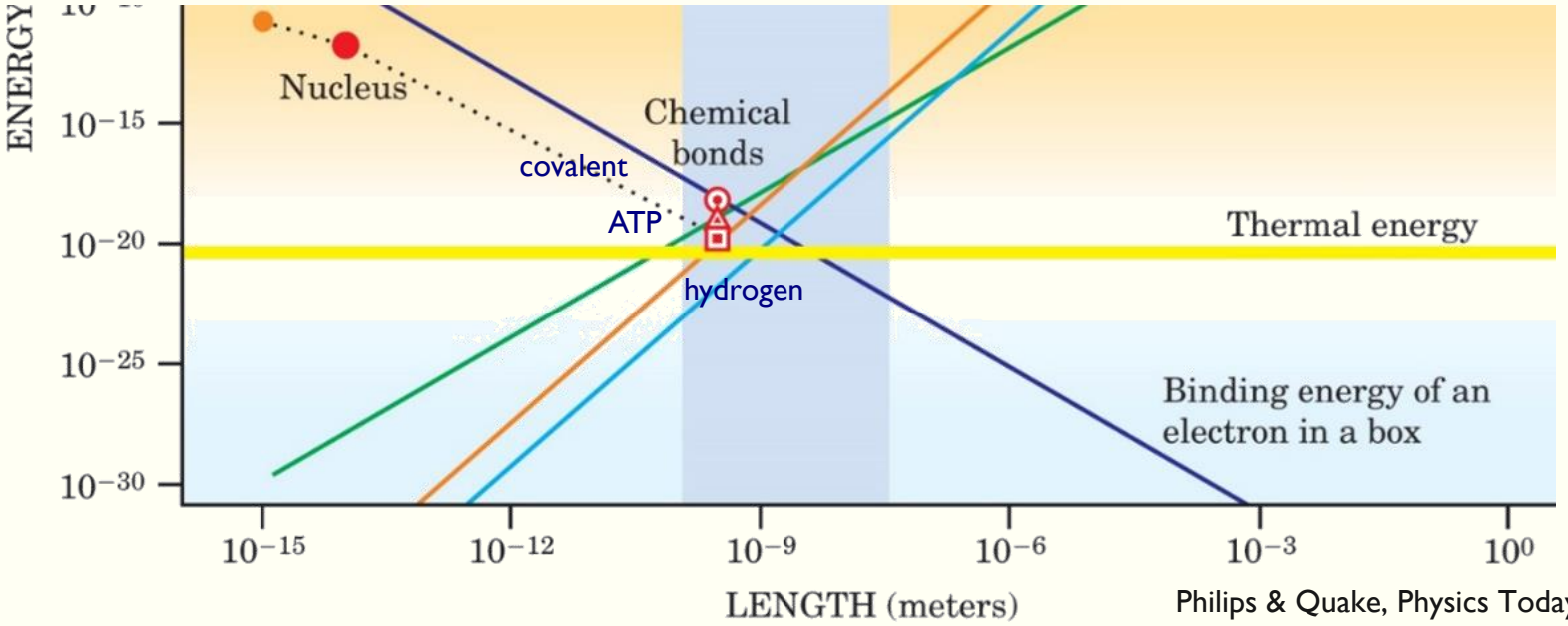
IV. Mechanical evolution of protein

At the nanoscale, energy scales converge to ~thermal energy



Exercise: derive these scalings, especially the bending energy.

Nanoscale regime, especially in living matter has many interplays



Philips & Quake, Physics Today 2006

- Zoom: convergence zone is 3-4 orders-of-magnitude in each direction.
- Typical elastic, electrostatic, chemical are not too many orders-of-magnitude apart.
- Interplays are expected.
- **Thermal fluctuations** are relevant and excite these degrees-of-freedom: ***soft matter*** *.

At the fundamental level, all nanoscale interactions are electromagnetic

Primary	Secondary
	Hydrogen bonding
	Hydrophobic
Electrodynamic (van der Waals):	Hydration
London, dispersion, Debye,	Osmotic
Induction, Keesom, orientation	Disjoining
	Structural
Electrostatic (Coulombic)	Steric
	Depletion
	Entropy-driven
	Enthalpy-driven
	Cross-binding
Polar (acid-base)	Specific
	Magnetic

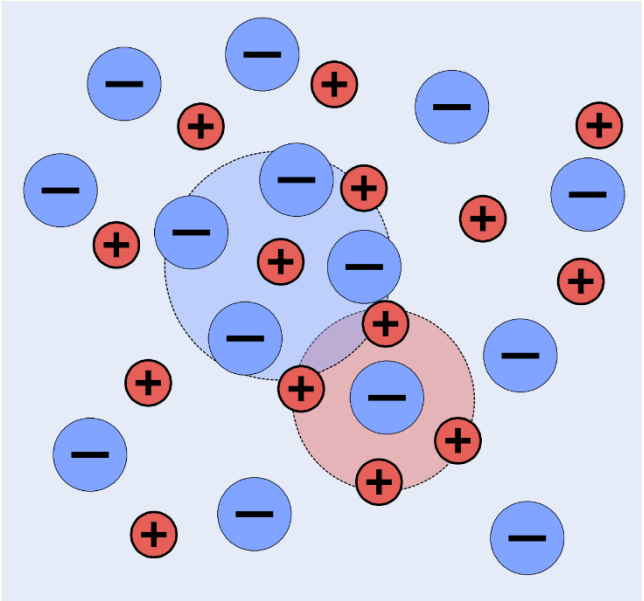
* "The distinction is quite arbitrary. Secondary interactions either are subsets of primary interactions, specific to a class of materials, or are a more macroscopic phenomenon"

Long range interactions in nanoscale science *Reviews of Modern Physics 2010*

French, Parsegian, Podgornik, Rajter, Jagota, Luo, Asthagiri, Chaudhury, Chiang, Granick, Kalinin, Kardar, Kjellander, Langreth, Lewis, Lustig, Wesolowski, Wettlaufer, Ching, Finnis, Houlihan, von Lilienfeld, van Oss, Zemb

Life happens in salty solutions where electrostatic forces are screened

- In principle, electrostatic interactions are long-range ~1/r.



However:

- Life began in the oceans in salt water (~500 mM of ions).
- Inside the cell: 150 mM.

→ Strongly screened interactions by charge correlations.

- The typical screening length scale, the **Debye length** *,

$$\lambda_D = \sqrt{\frac{\varepsilon k_B T}{4\pi n e^2}} \approx 1\ \text{nm}\,.$$

- This is the scale of the Poisson-Boltzmann equation.

$$\left.\begin{aligned} \nabla^2\phi &= -\frac{4\pi n e}{\varepsilon} \\ n &= n_0 \exp\left(\frac{-e\phi(r)}{k_B T}\right) \end{aligned}\right\} \Rightarrow \nabla^2\phi = -\frac{4\pi n_0}{\varepsilon}\exp\left(\frac{-e\phi(r)}{k_B T}\right)$$

* *There are other length scales (Bjerrum, Gouy-Chapman)*

Exercise:

- Show that λ~ 1 nm.
- How would Life look if λ ~ 1 mm?

At nanoscale, all interactions are zero range*

- From all this hen coarse-grained at a resolution of ~1 nm forces are **"zero-range":** interaction is only with neighboring molecules.
- We are in a regime of ***coarse-grained theories*** such an elasticity and hydrodynamics.

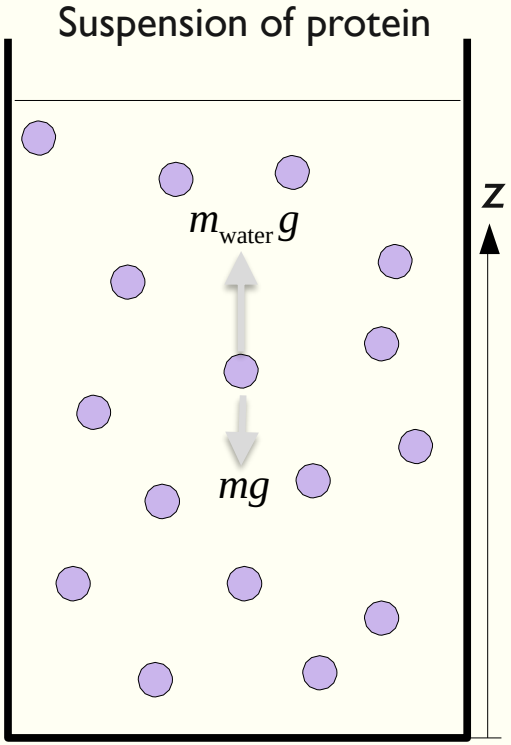
> *"An important fact in the theory of elasticity is that the molecular forces have a* ***very short range*** *of action. Their effect extends only to the* ***neighborhood*** *of the molecule exerting them, over a distance of the same order as that between the molecules...* ***The range of action of the molecular forces should therefore be taken as zero*** *in the theory of elasticity. We can say that the forces which cause the internal stresses are, as regards the theory of elasticity,* ***"near-action"*** *forces, which act from any point only to neighboring points".*
>
> (Landau & Lifshitz, *Theory of Elasticity*, p. 4)

- Typically*, the dynamics take the form of local conservation laws, e.g. Cauchy's equation:

$$\frac{D\mathbf{u}}{Dt} = \frac{1}{\rho}\nabla\cdot\sigma + \mathbf{f}$$

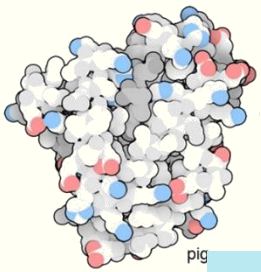
* *Short or long-range is relative to a scale … in "Long range interactions in nanoscale science" (Rev Mod Phys 2010) long-range means relative to bond length*

Gravity remains long-range but is irrelevant at the nanoscale



Net force: $m_{\text{net}}\mathrm{g} = \Delta\rho \mathrm{V g}$ (Archimedes)

- ***What happens after a long time?***



$$P(z) \propto \exp\left[-\frac{m_{\text{net}} g z}{k_B T}\right]$$

protein mass: $m = 10^4$ Da $= 2\cdot 10^{-23}$ kg

net mass: $m_{\text{net}} \approx m/4 = 0.5\cdot 10^{-20}$ kg

length scale $z_* = \dfrac{k_B T}{m_{\text{net}} g} = \dfrac{4\cdot 10^{-23}}{0.5\cdot 10^{-20}\times 10} = 100$ m

- Density in test-tube is ~ constant.

Exercise: what is the sedimentation velocity? Is it relevant?

I. Introduction

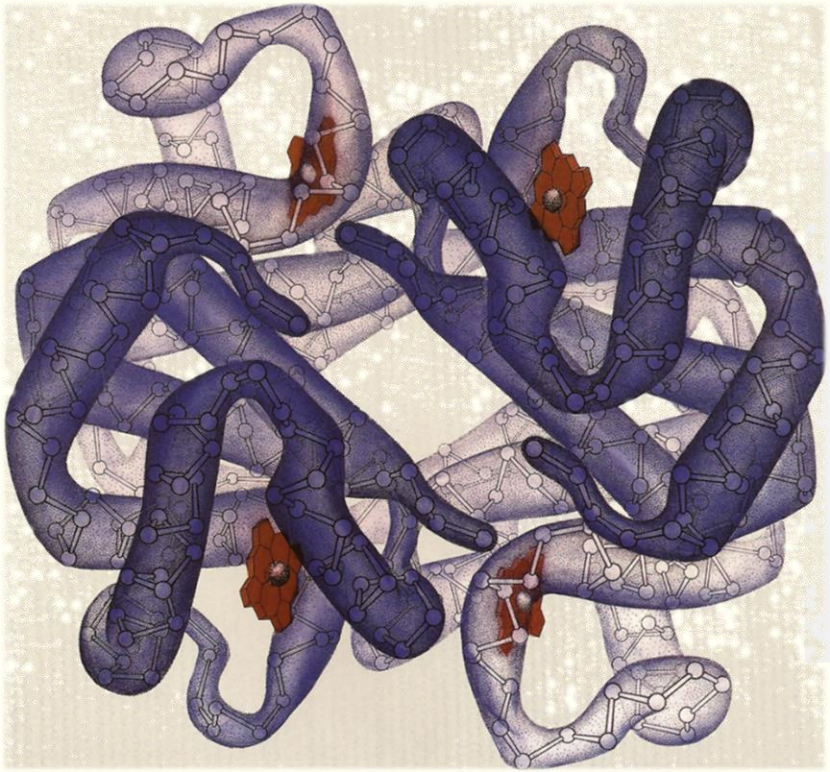
II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

- Physical scales and interactions.
- ***Coarse-grained view of protein.***
- Long-range deformation and viscoelasticity in proteins.
- Mechanical models of protein: spring nets.
- Amorphous matter: rigidity, spectrum.

IV. Mechanical evolution of protein

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

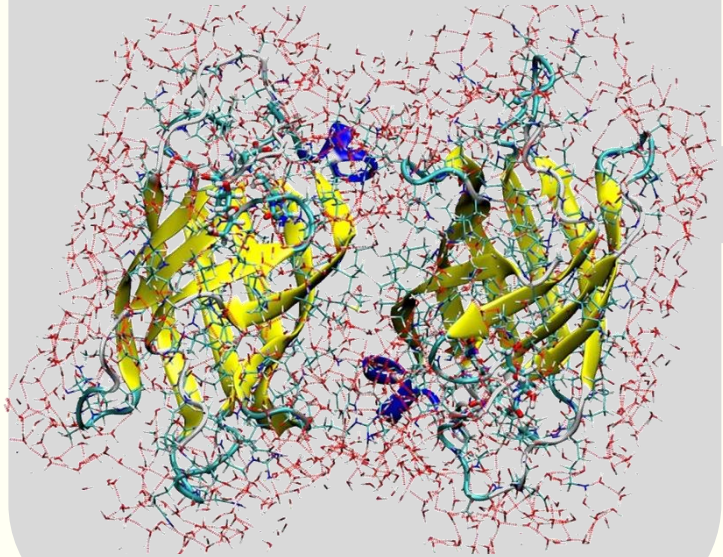


- » ***Non-random information-rich matter***
- » ***Long-term evolutionary memory.***
- » ***Sparsity of data*** in high-D genotype space.
- » ***Collective interactions:*** many-body physics.
- » ***Small, heterogeneous*** – no TD limit.

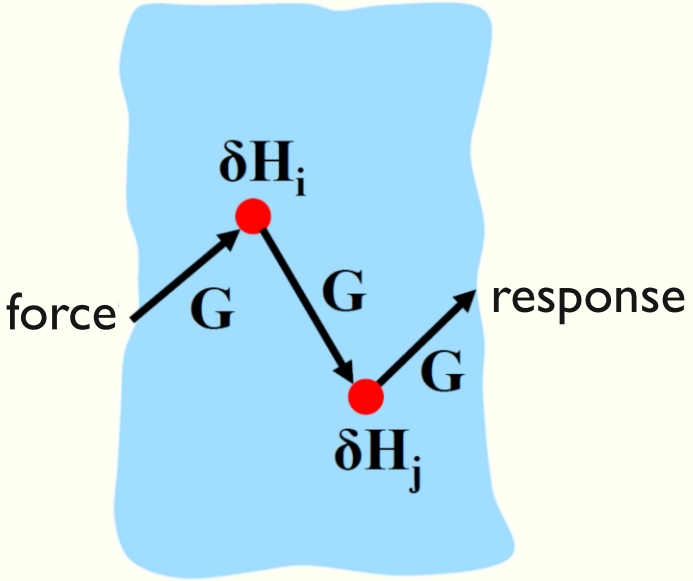
Hemoglobin,
Irving Geiss, 1978

Complementary theoretical approach puts the protein problem in the realm of condensed matter physics

Realistic simulations predict dynamics and function of concrete proteins



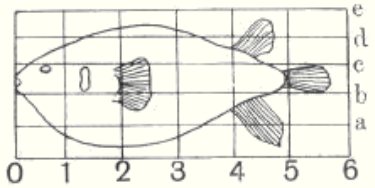
Minimal, coarse-grained models to examine questions of protein evolution

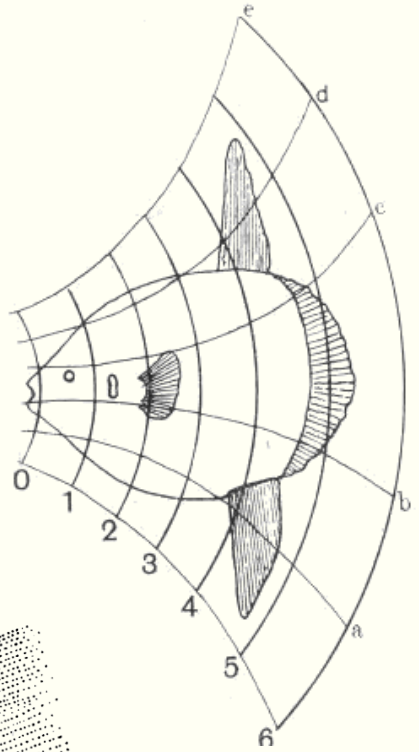


There is a long history of abstraction and simplification in modeling living systems*

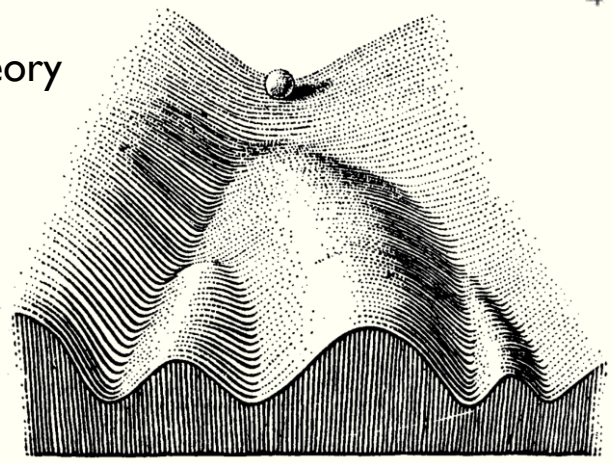
- **D'Arcy Thompson** (1917):

  shape transformations as morphogenesis.





- **René Thom** (1970): Catastrophe theory

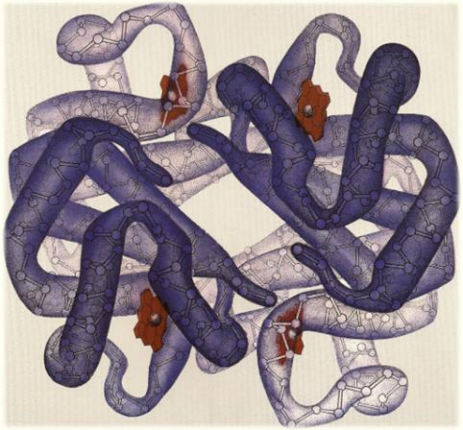


Waddington: A Catastrophe Theory of Evolution 3

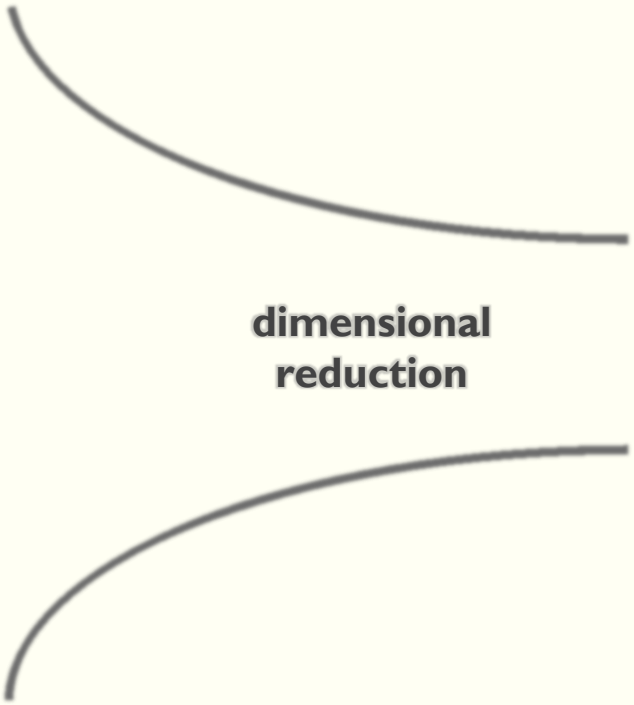
** Many aspects these models are "wrong" but they still conceptually useful*

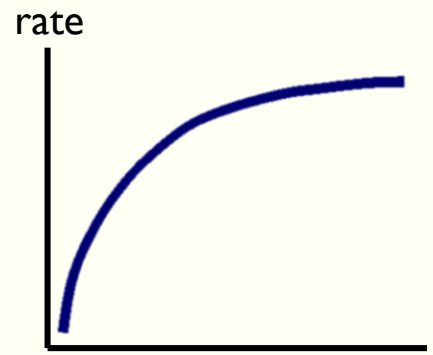
Genes and protein structures have *many* parameters

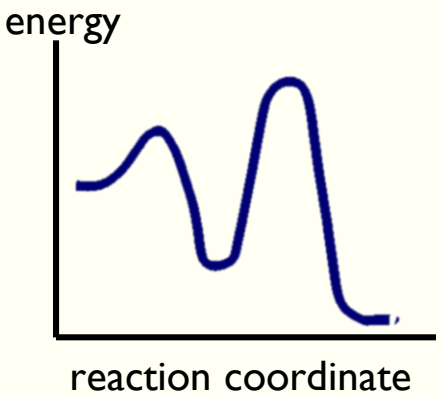
but protein function is described by only a *few*



```
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
```







Genotype
Information ~ gene length

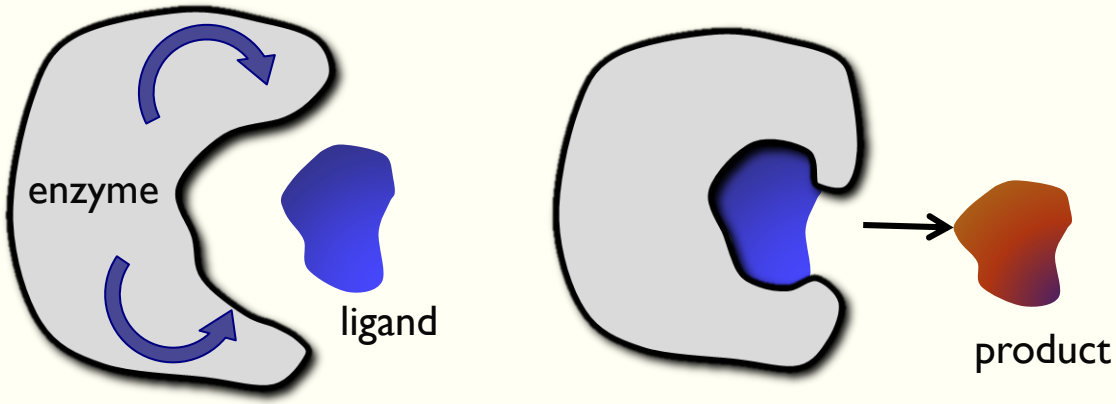
compression

Phenotype
Information ~ a few bits

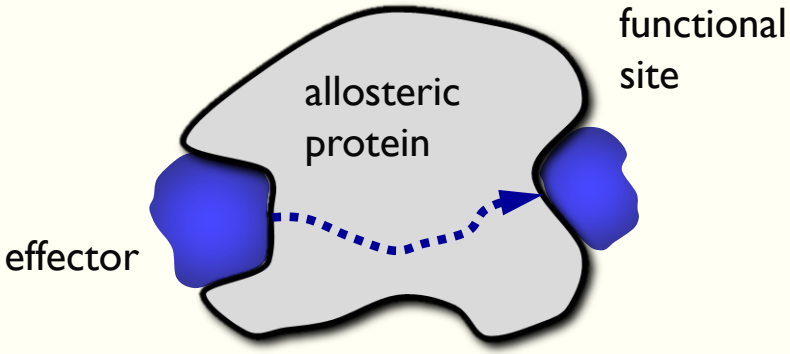
Hint about degrees-of-freedom:

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

Induced fit:

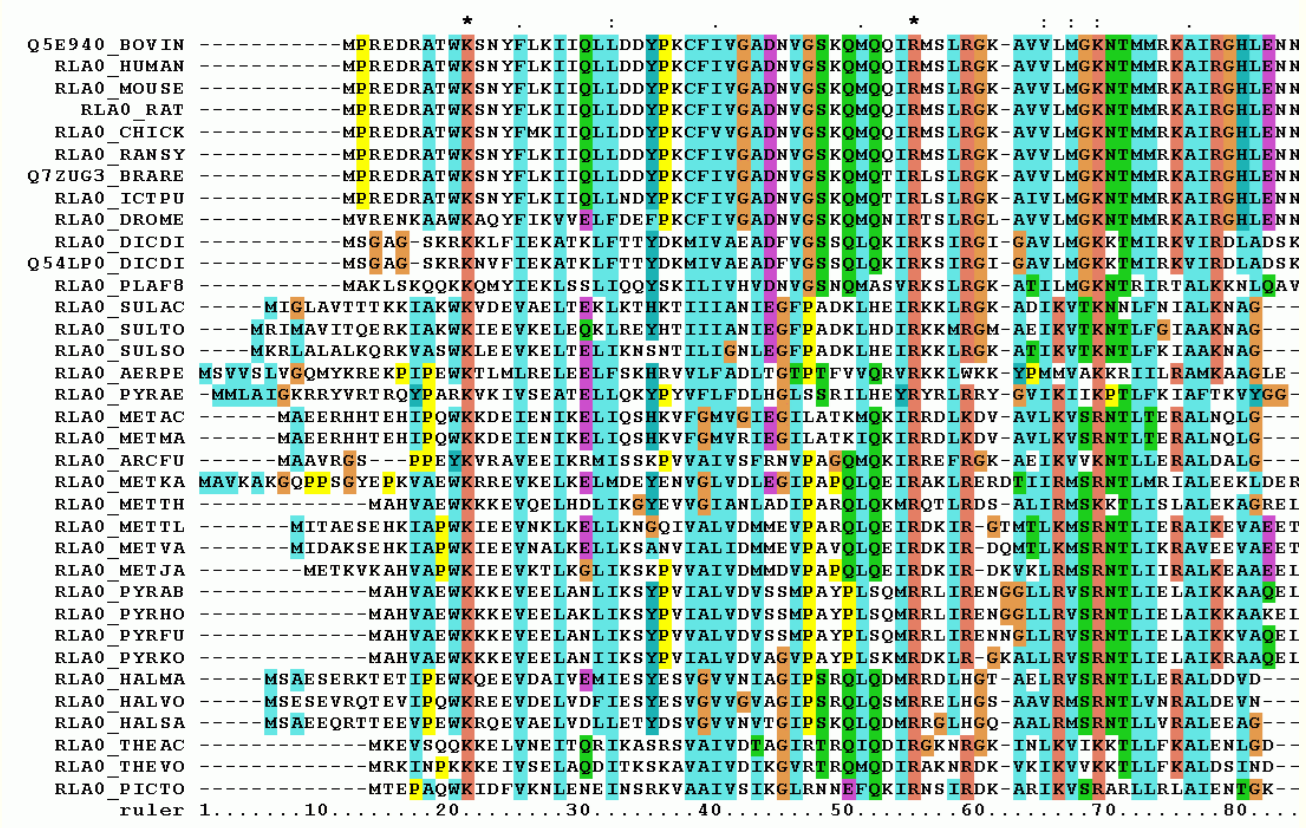


Allostery:



- **Slow global** changes in enzymes, ion-channels, ribosome…

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

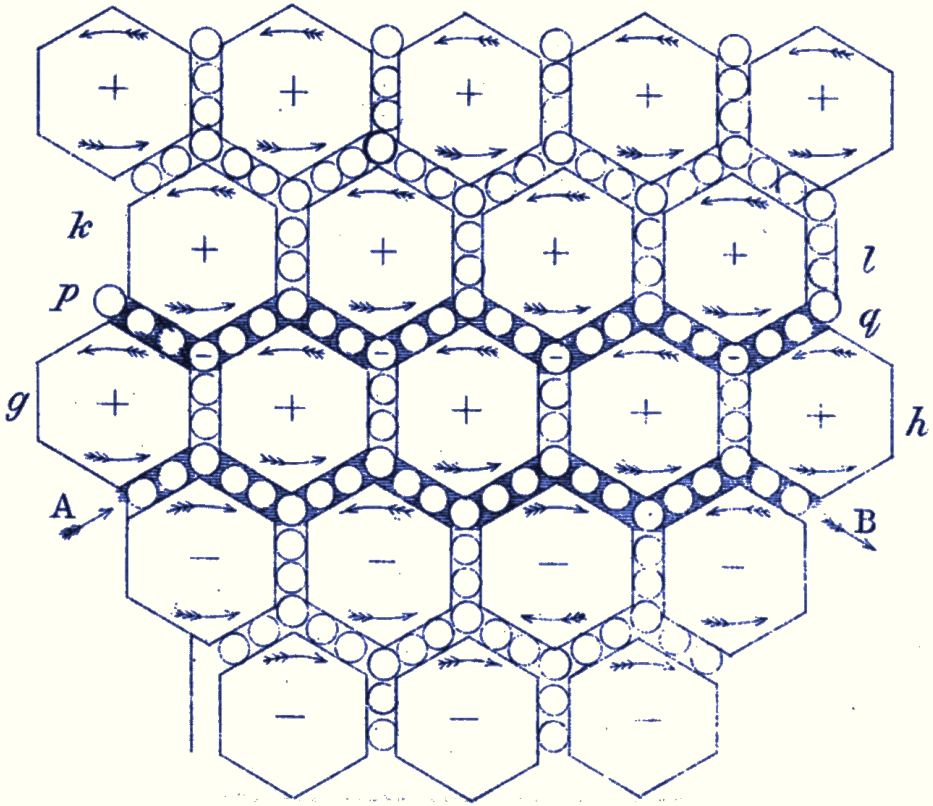


- **Epistasis** = interaction among mutations.
- *How is **epistasis** encoded in the gene?*

How to get long-range interaction from zero-range forces?

- **Maxwell's elastic model of electromagnetism (*Phil Mag* 1861):**

- Space is filled with vortex tubes (***aether***)
- Charges are frictionless 'ball–bearings'.
- In conductors: charges flow between the vortices and rotate them.
- In insulators and free space: charges are displaced and deform the vortices .
- Maxwell calls this: '***electric elasticity***'.



More earthly example: waves in a pond.

I. Introduction

II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

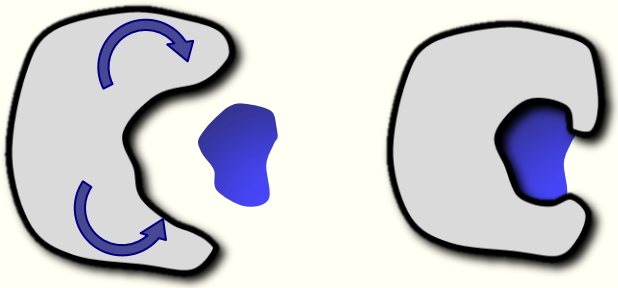
- Physical scales and interactions.
- Coarse-grained view of protein.
- ***Long-range deformation and viscoelasticity in proteins.***
- Mechanical models of protein: spring nets.
- Amorphous matter: rigidity, spectrum.

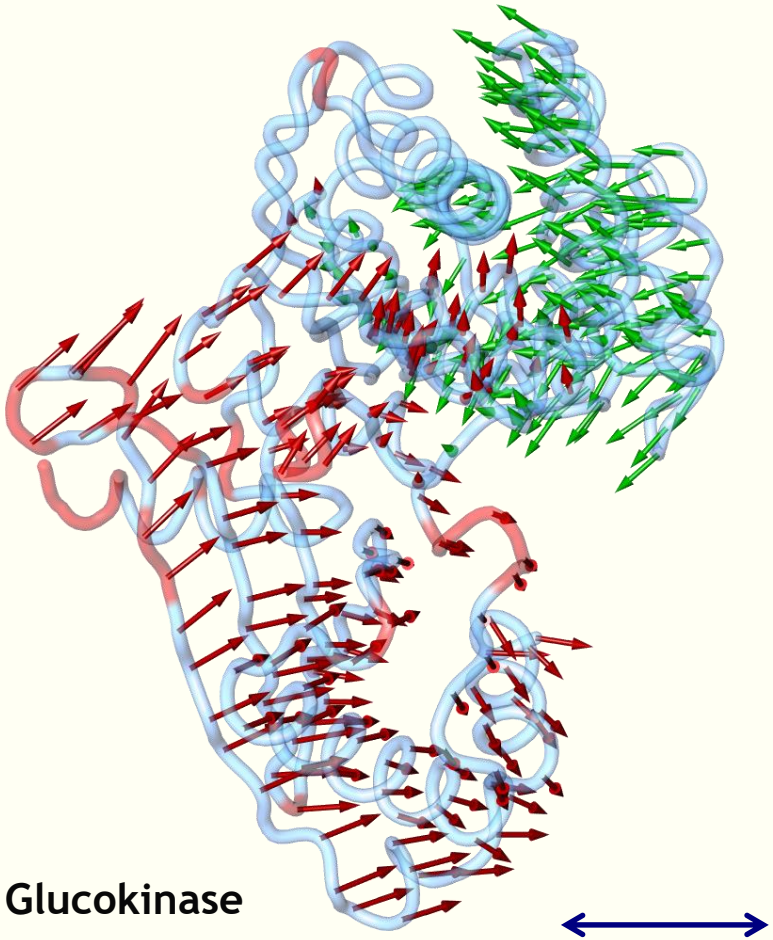
IV. Mechanical evolution of protein

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

Motion = Δ(Structure)

unbound → bound





Glucokinase

(PDB: 1v4s → 1v4t)

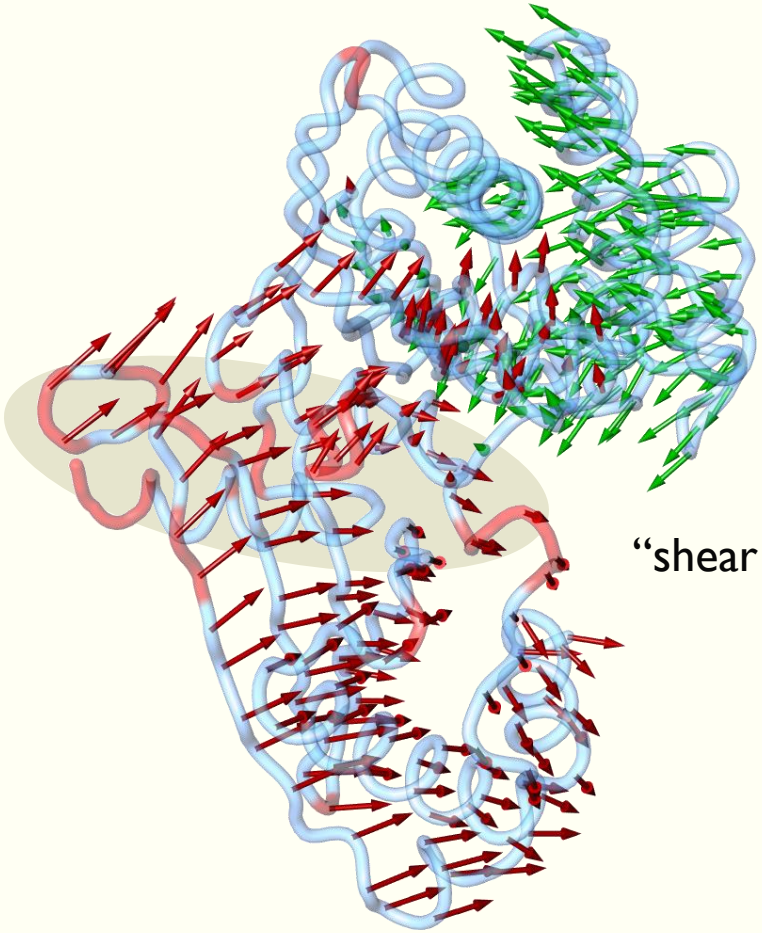
Mitchell Leibler PNAS 2016
Dutta Eckmann Libchaber PRX 2018
Eckmann Libchaber PNAS 2017

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

Shear = Δ Motion =ΔΔ Structure

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

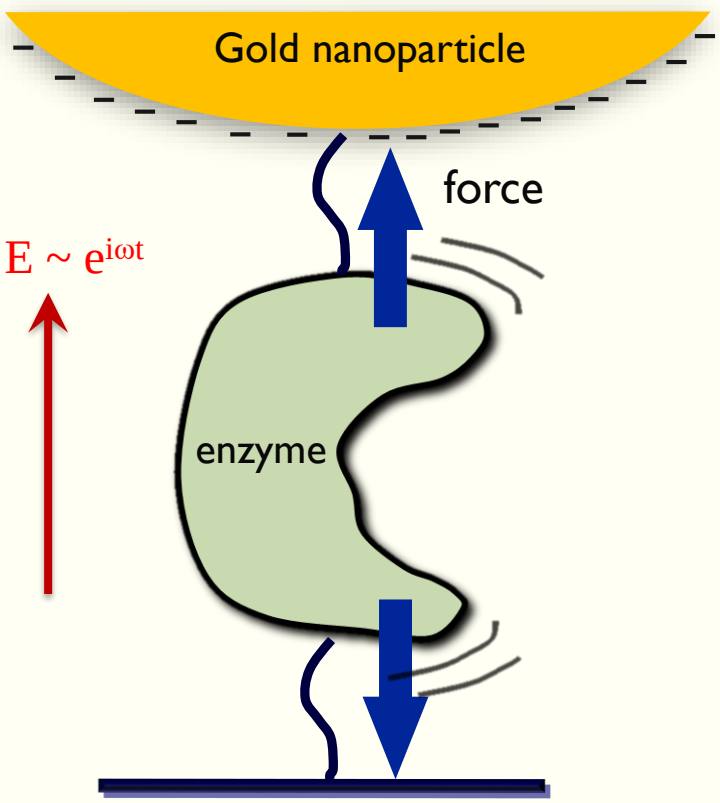
Q: why do we look at non-linear terms?

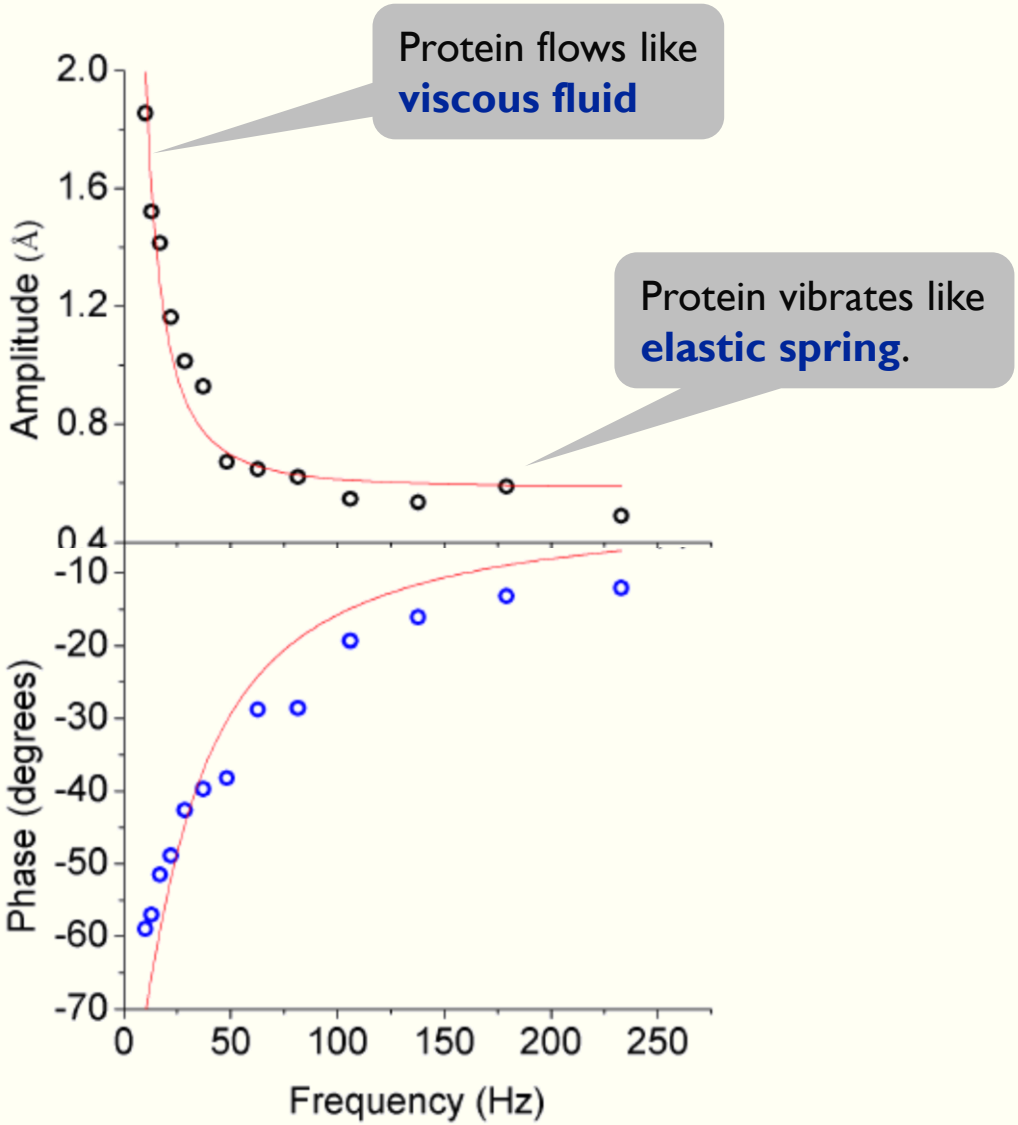


"shear band"

Glucokinase (PDB: 1v4s → 1v4t)

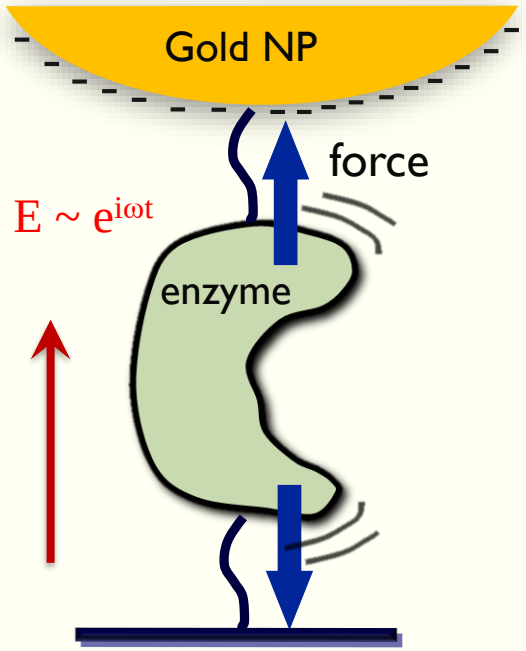
Motivation: Proteins are soft and *viscous* at low frequency

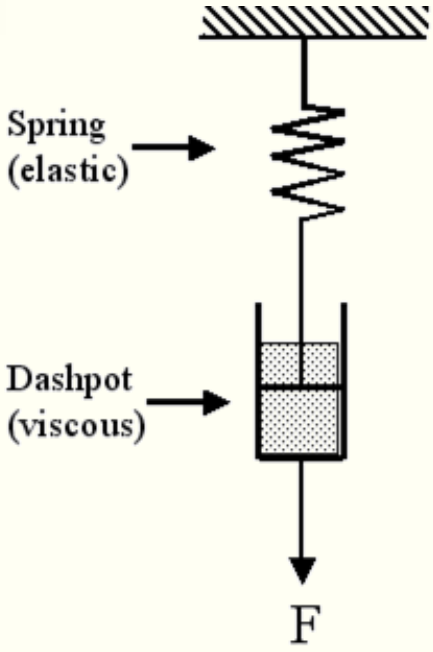


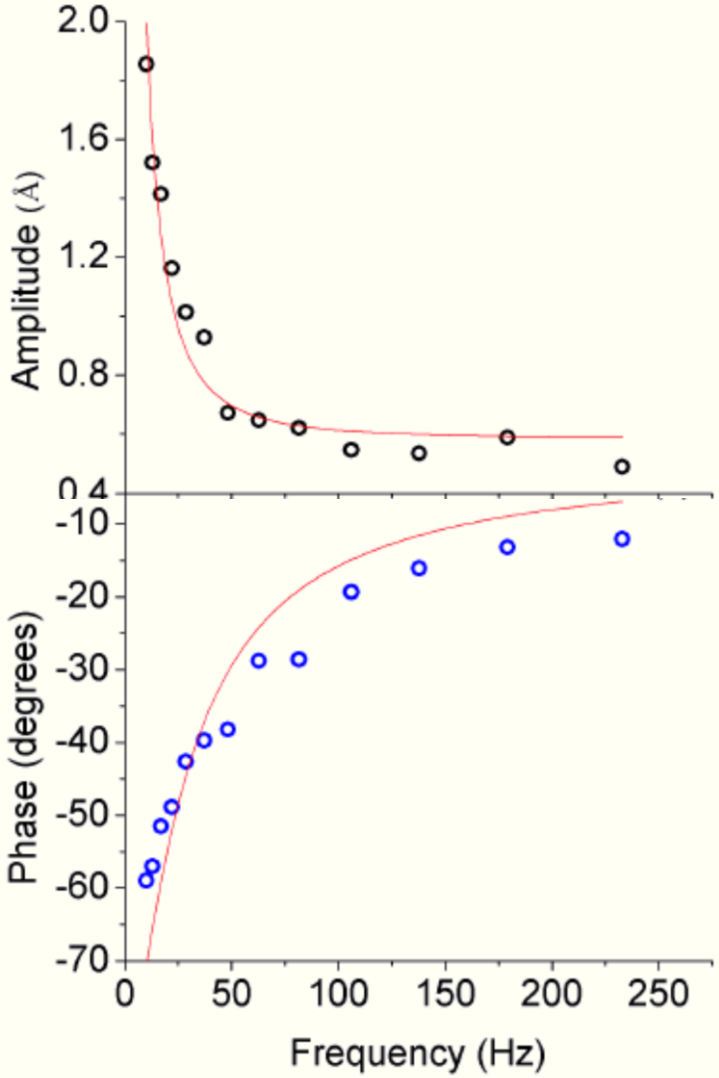


Zocchi et al., 2012-2018

Viscoelastic dynamics in protein are similar to Maxwell's dashpot model







Exercise: derive the response curves of the dashpot.

Zocchi et al., 2012-2018

I. Introduction

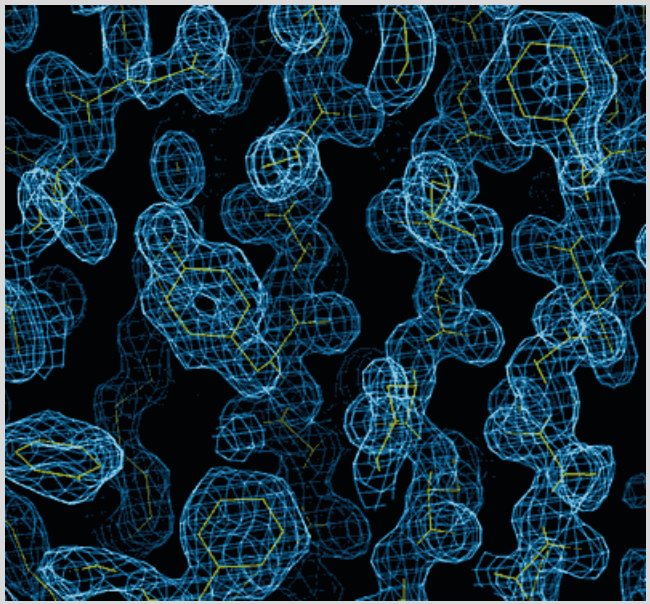
II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

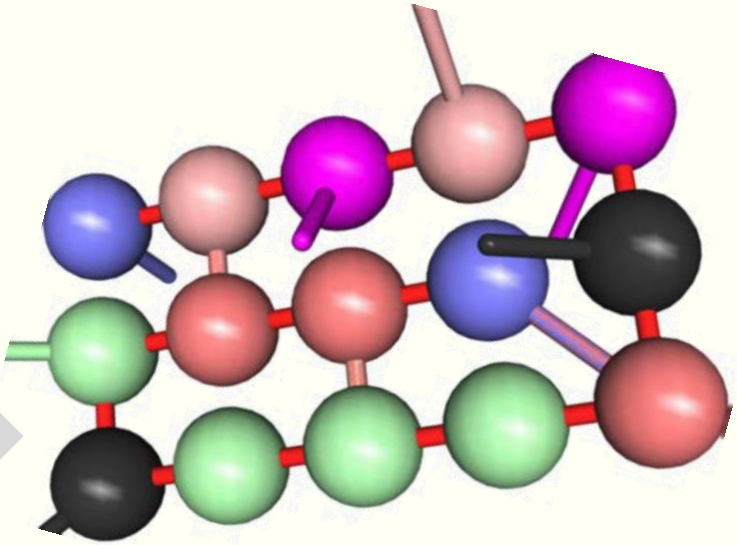
- Physical scales and interactions.
- Coarse-grained view of protein.
- Long-range deformation and viscoelasticity in proteins.
- ***Mechanical models of protein: spring nets.***
- Amorphous matter: rigidity, spectrum.

IV. Mechanical evolution of protein

Simplifications translate questions of biology to analogous questions in the physics of amorphous networks



- X-ray and cryo-EM see proteins structure at angstrom resolution.
- Detailed inter-atomic force fields.



- **Coarse grained:** amino acid are points.
- Simplified forces and energies.
- Fewer types of amino acids.

Elastic networks and normal mode analysis capture large-scale motions in proteins

- Several variants:: Gaussian, anisotropic...

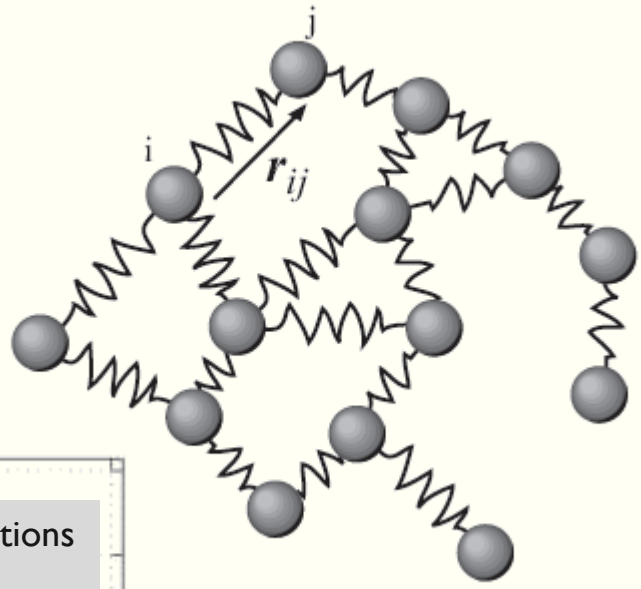
» ***Protein normal-modes:*** *Levitt, Sander, Stern.(1985).*
» ***Large amplitude modes in proteins:*** *Tirion (1996).*
» ***Elastic network models:*** *Bahar et al. (since 1990s).*

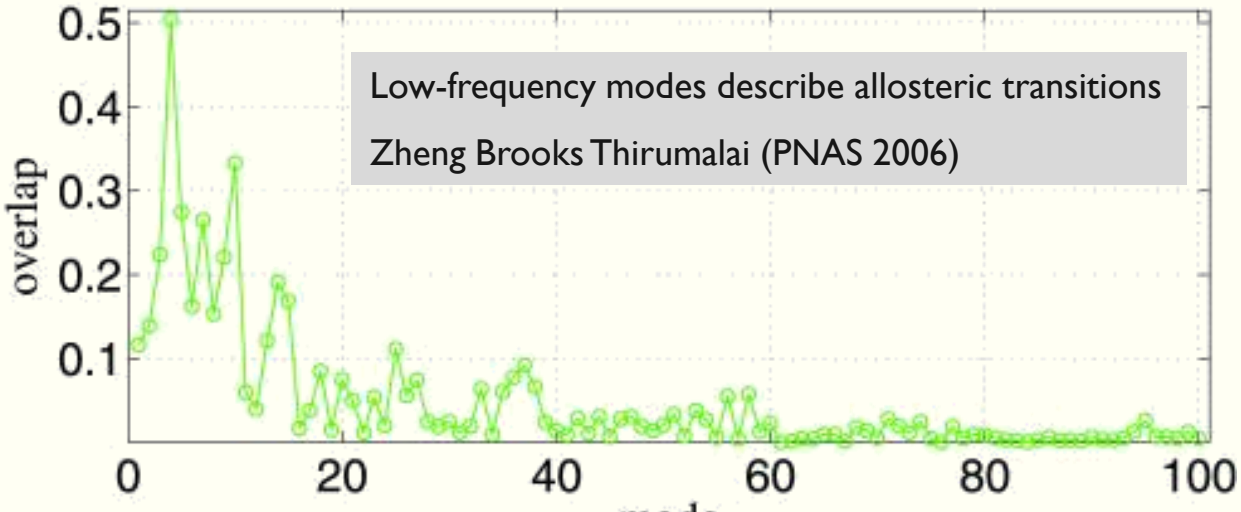
- Basic principle is the same:

(A) Amino acids closer than cutoff (7-10 angstrom) are connected by harmonic springs.

(B) Calculate normal modes, esp. at low energy.

(C) Successful in analyzing large-scale motion.





Low-frequency modes describe allosteric transitions
Zheng Brooks Thirumalai (PNAS 2006)

Recent Review (Chem Rev, 2019)
Thirumalai, Hyeon, Zhuravlev, Lorimer

I. Introduction

II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

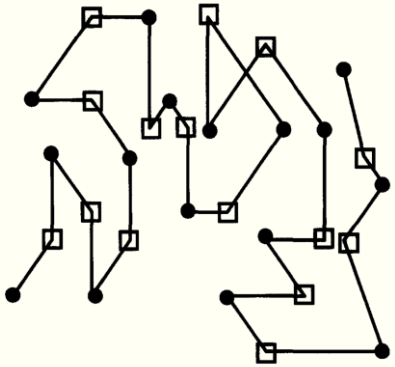
- Physical scales and interactions.
- Coarse-grained view of protein.
- Long-range deformation and viscoelasticity in proteins.
- Mechanical models of protein: spring nets.
- ***Amorphous matter: rigidity, spectrum.***

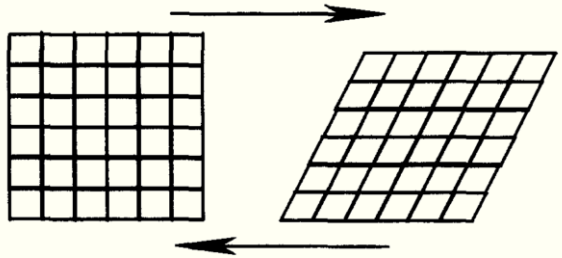
IV. Mechanical evolution of protein

The coarse-grained approach we examine is based on ideas from amorphous solid theory (glasses)

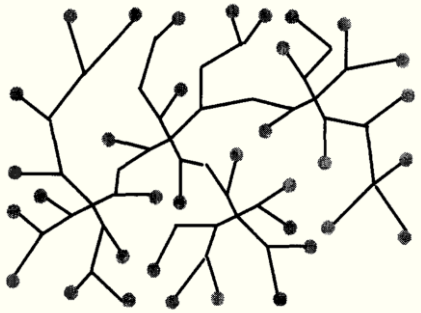
Especially from an inspiring paper by Shlomo Alexander: ***Amorphous solids: their structure, lattice dynamics and elasticity.*** *Physics reports* (1998)

- Macroscopic and microscopic ***floppy modes***





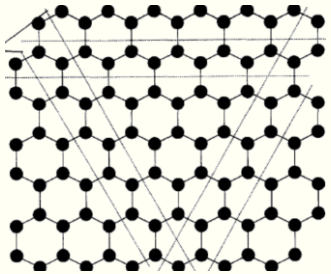
Nearest-neighbor square and freely-linked chain are ***macroscopically floppy***



Leaves of tree are ***microscopically floppy***

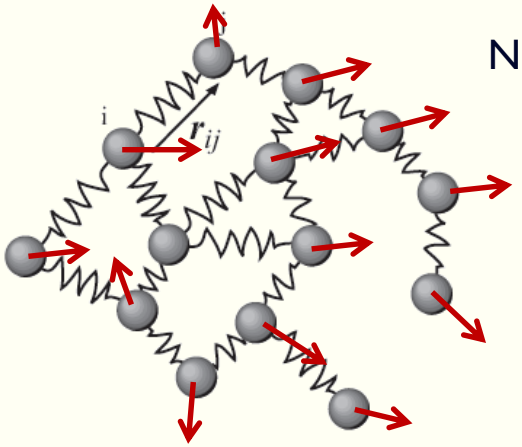
Exercise: (a) is this network floppy?

(b) How many bonds are needed to stabilize the square lattice?



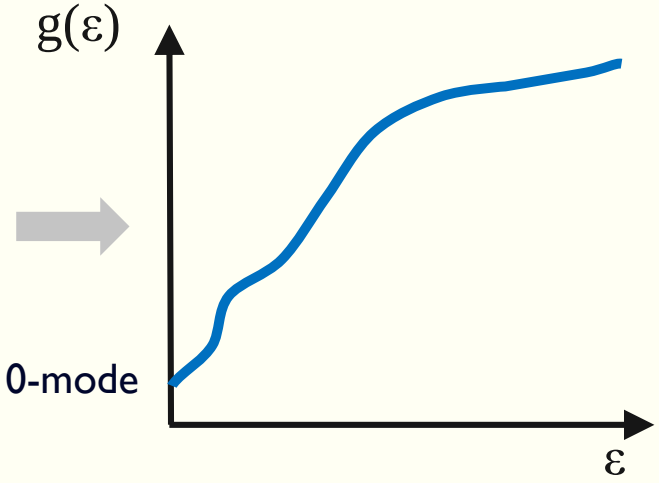
Floppy modes are zero modes of spring networks

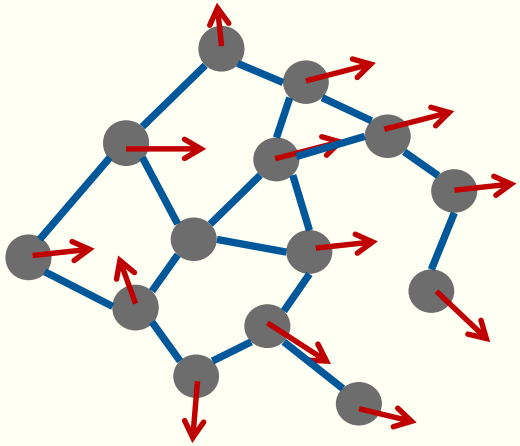
or free modes of rigid networks



Normal mode **u**

$$E \sim \langle \mathbf{u} | H | \mathbf{u} \rangle$$





Rigidity calculation → 0-mode

Rigidity can be estimated by counting degrees-of-freedom

Maxwell's rigidity criterion

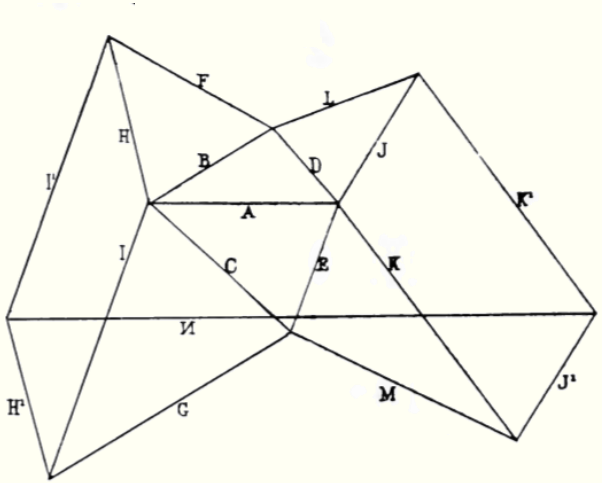
Point d.o.f. : $d \cdot N$

Constraints: $-c \cdot N$

Galilean: $-\frac{1}{2} d(d+1)$

$$F = d \cdot N - c \cdot N - \frac{1}{2} d(d+1) = 0$$

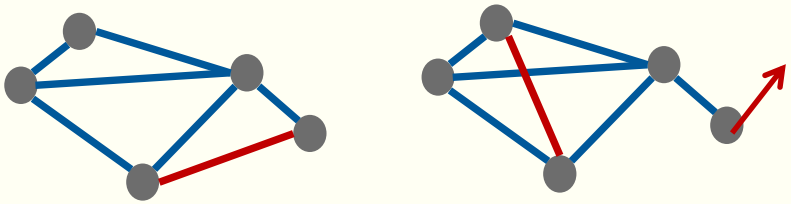
$$\text{for } N \gg 1: \quad f = \frac{F}{N} = d - c = 0$$



Maxwell, *On reciprocal figures and diagrams of forces* (1864)

This is a bound:

one has to consider topology



Constraint theory in glasses
Thorpe & Philips. Solid State Comm, (1985)

Center for Soft
& Living matter

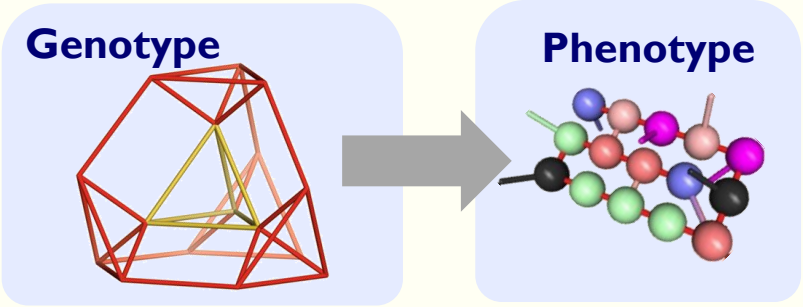
2019 Soft Matter Summer School:
Soft Matter and Slow Relaxation Dynamics
IBS CSLM, Ulsan, June 2019

MECHANICS OF EVOLUTION

PART B

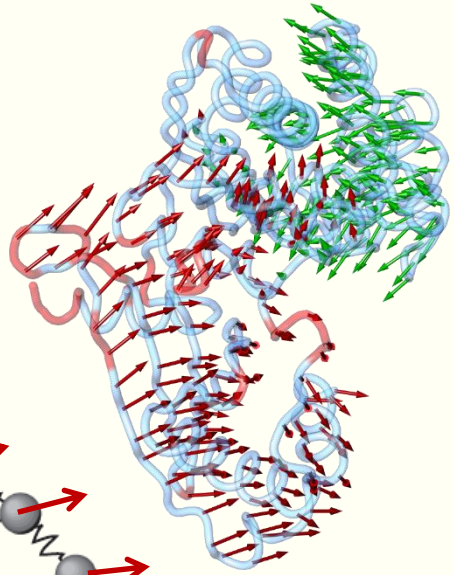
Sandipan Dutta (IBS)
Jean-Pierre Eckmann and Jacques Rougemont (Geneva)
Albert Libchaber (Rockefeller)
Michael Mitchell and Stanislas Leibler (Rockefeller & IAS)

Reminder

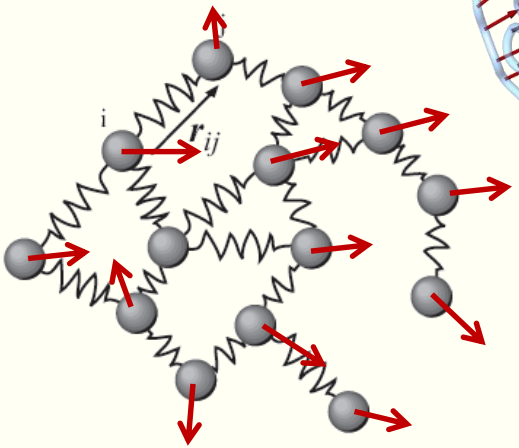


- Genotype-to-phenotype map of proteins involves dimensional reduction.

- The function of many proteins involves large-scale deformations.



- At the nanoscale, one can coarse-grain protein into elastic (or viscoelastic) networks.



- Amorphous matter: floppy zero-energy modes from energy spectrum or constraint theory.

I. Introduction

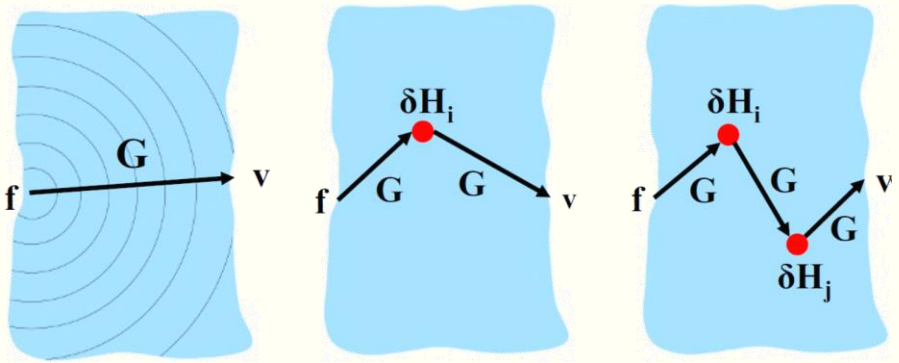
II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

} *informal introduction:* *basic ideas and questions*

IV. Mechanical evolution of protein

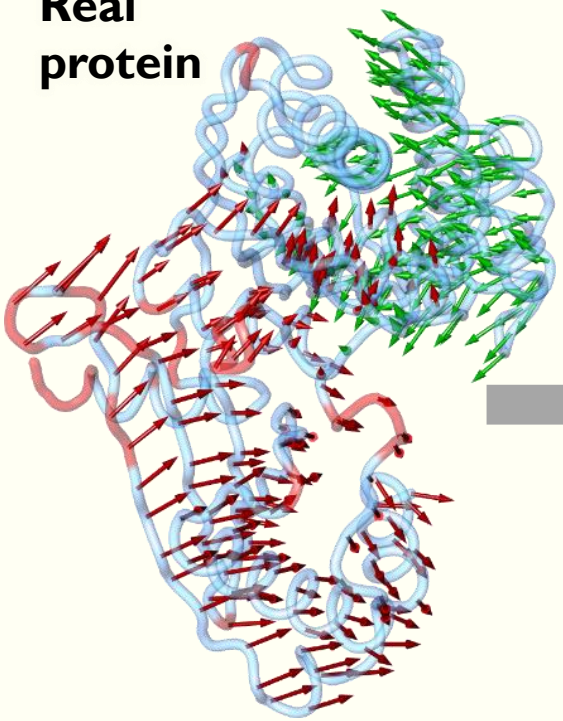
- *Condensed matter theory of protein using Green's functions.*
- *Linking to biological phenomena*



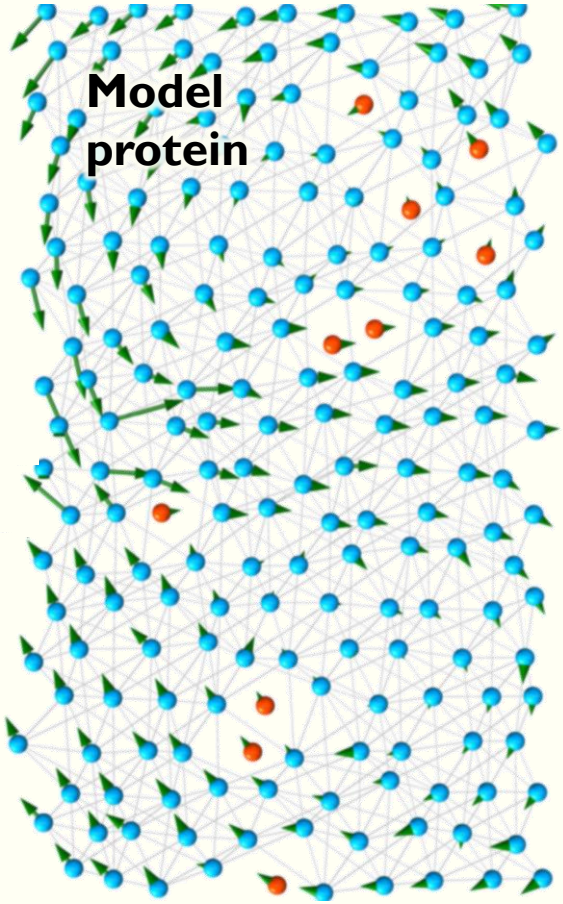
Minimal coarse-grained model of protein as adaptive matter

gives evolution a physical interpretation

Real protein



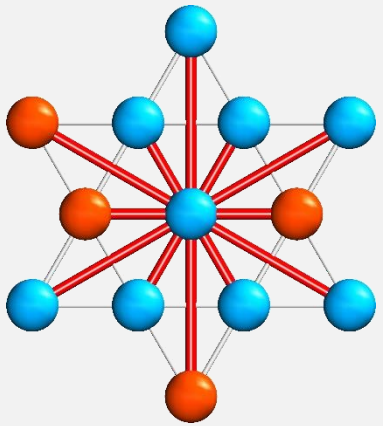
Model protein



Only 2 types of amino acids connected by harmonic springs

amino acids	H	P
H	strong	weak
P	weak	weak

z =12 neighbors



200 beads = amino acids
at randomized 2D positions.

Dutta Eckmann Libchaber et al, PNAS 2018
Eckmann Libchaber et al., PRX 2017

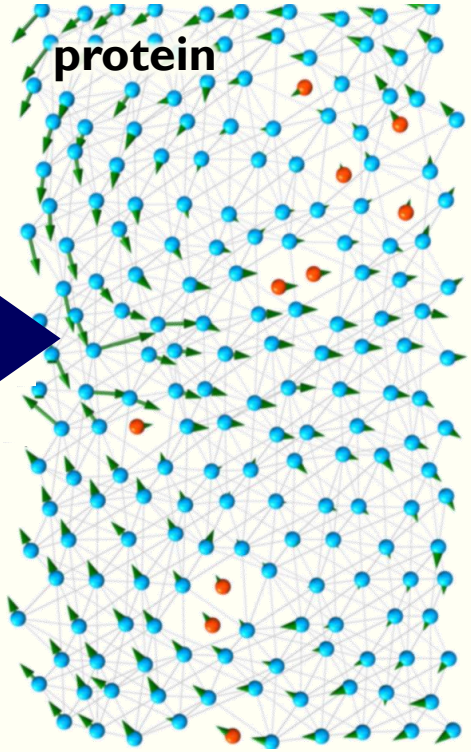
Minimal coarse-grained model of protein as adaptive matter gives evolution a physical interpretation

gene

…10111110111110I…

…00111111100II00…

map



amino acids	H	P
H	strong	weak
P	weak	weak

- There are $2^{200} = 10^{60}$ genes.
- ***How many of them are functional?***
- ***Where are they on the 200-cube?***

I. Introduction

II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

- Minimal Coarse-grained model
- ***Hooke's law of networks: Laplacian etc.***
- Green's function
- Mechanical evolution and scattering
- Epistasis and genetic correlation
- The dimension question
- Examples and conclusion

Mechanics of harmonic spring networks

- Two-stage construction: ***connect*** then ***embed*** in space.

- Network is a ***graph***:

  AA are vertices $i, j, k...$; Bonds are springs $\alpha, \beta, \gamma...$

  $n_a = 10 \times 20 = 200$ AA and n_b bonds.

- ***Connectivity*** recorded in adjcancy matrix $\mathbf{A}$ $(n_a \times n_a)$:

$A_{ij} = 1$ if i, j connected by a bond, 0 otherwise.

- ***Gradient*** operator connects spaces of vertices and bonds:

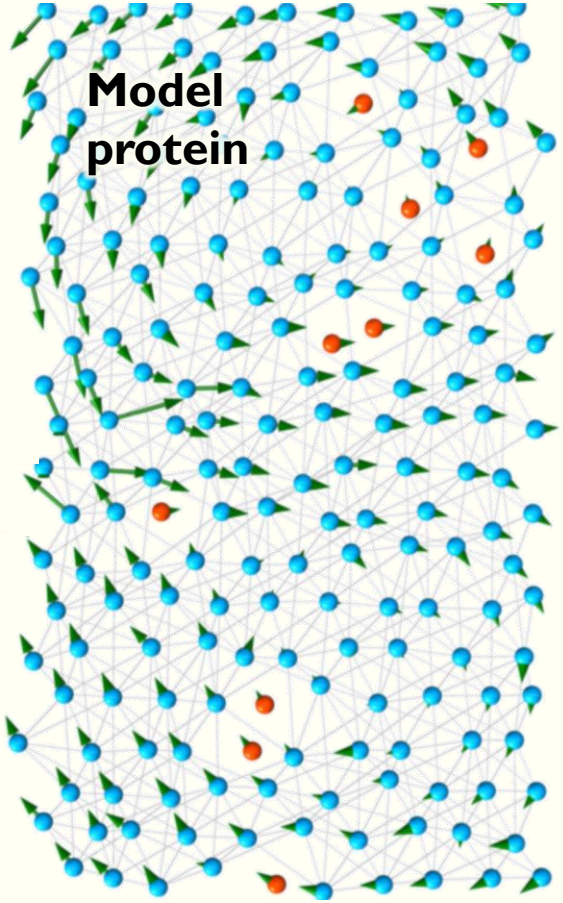
  if i, j connected by bond α, $\nabla_{\alpha i} = 1$ and $\nabla_{\alpha j} = -1$.

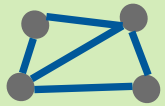
- As in continuum, the ***Laplacian*** is $\Delta = \nabla^{\mathsf{T}} \nabla$:

  $\Delta_{ij} = -1$ if i, j connected, 0 otherwise.

  $\Delta_{ii} = z_i$ degree of vertex i.

  so $\Delta = Z - A$ (Z diagonal matrix z_i)



Exercise: what is the Laplacian and spectrum of  ?

The Laplacian is the averaging or curvature operator

- $\mathbf{f}_i$: function on vertices $\mathbf{f}_i$ (n_a-vector)

$$(\Delta \mathbf{f})_i = \sum_{\alpha=(i,j)} \left(\mathbf{f}_i - \mathbf{f}_j\right) = \sum_{\alpha=(i,j)} \nabla \mathbf{f}$$

(analogue of $\Delta = \nabla \cdot \nabla$)

- As in continuum, ones examines the ***spectrum*** :

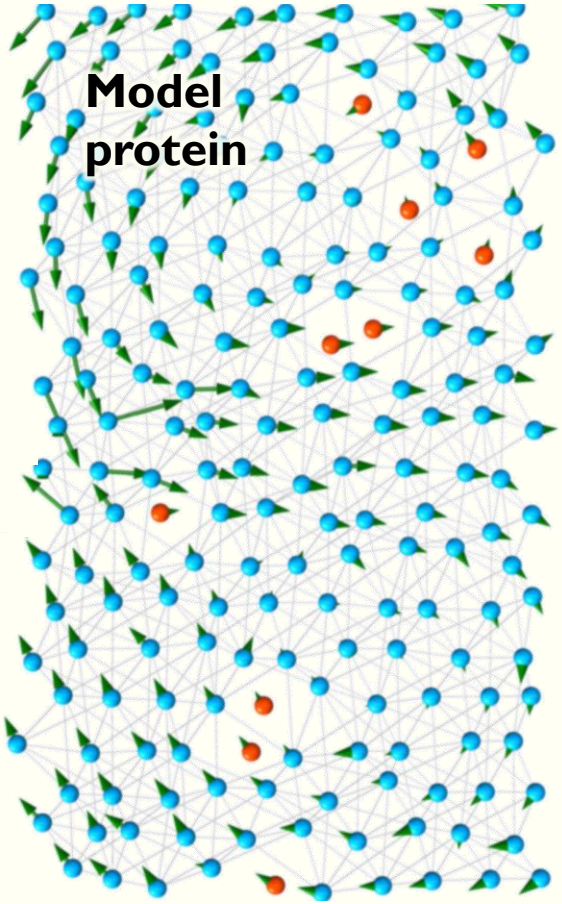
$$\Delta \mathbf{f} = \lambda \mathbf{f}$$

(i) λ are real (Δ is self-adjoint).

(ii) $\lambda \geq 0$ ($\mathbf{f}^{\mathsf{T}} \Delta \mathbf{f} = \sum_{\alpha=(i,j)} (\mathbf{f}_i - \mathbf{f}_j)^2$).

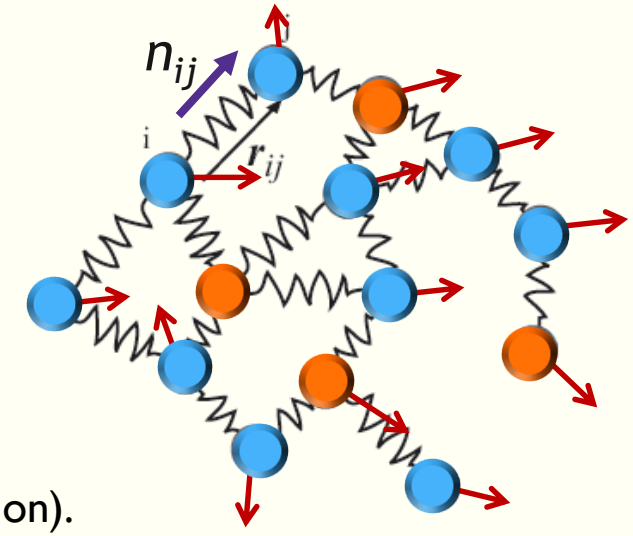
(iii) $\min \lambda = 0$.

0 eigenvactors = # connected components.



Textbook: Fan Chung, ***spectra of graphs***, AMS

Embedding the graph in space



- Embedding: $i \to \mathbf{r}_i \in \mathbb{R}^d$ ($d = 2$, $\mathbf{r}$ is $n_a \cdot d$ vector)

- Spring matrix $k_{\alpha\alpha} = k_w + (k_S - k_w) \cdot c_i \cdot c_j$

  if bond $\alpha = (i, j)$ is H-H: $k_\alpha = k_S$ otherwhise $k_\alpha = k_w$.

- Embedded ***gradient*** $\mathbf{D}$: $\mathbf{D}_{\alpha i} = \nabla_{\alpha i} \mathbf{n}_{ij}$ ($\mathbf{n}_{ij}$: $i \to j$ direction).

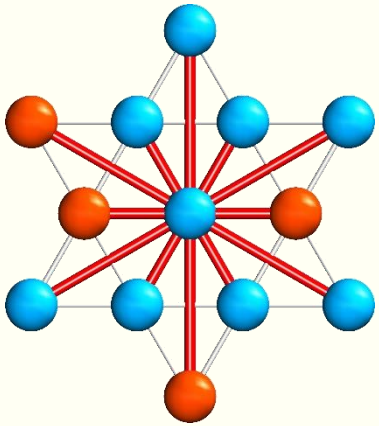
  (each $D_{(\alpha)}$ row correpsonds to bond α)

- For small deviations $\mathbf{r}_i \to \mathbf{r}_i + \mathbf{u}_i$, the elastic energy is

$$\mathcal{E} = \frac{1}{2}\mathbf{u}^\mathsf{T}\mathbf{H}\mathbf{u} \quad \text{with the Hessian } \mathbf{H}_{ij} = \frac{\partial^2 \mathcal{E}}{\partial \mathbf{u}_i \partial \mathbf{u}_j}$$

Elasticity tensor $\mathbf{H}=\mathbf{D}^\mathsf{T}\mathbf{K}\mathbf{D}$ has the structure of the ***Laplacian***

The mechanical response of the protein is described by its elasticity tensor H



- Gene **c** = sequence of 200 0's and 1's (H and P aa).
- Elasticity tensor (Hamiltonian) is a function of the gene **H**(**c**):

$$\mathbf{H}_{ij}(c_i, c_j) = \Delta_{ij} \cdot k_{ij}(c_i, c_j) \cdot \mathbf{n}_{ij} \times \mathbf{n}_{ij}$$

Laplacian matrix

(=-1 iff *i* and *j* are neighbors)

2D embedding

$\mathbf{n}_{ij}$= direction from *i* to *j*

H-P interaction: Spring constant strength k_{ij}

$$k_{ij}(c_i, c_j) = c_i \text{ AND } c_j = \begin{cases} \text{strong} & : c_i = c_j = 1 \\ \text{weak} & : \text{otherwise} \end{cases}$$

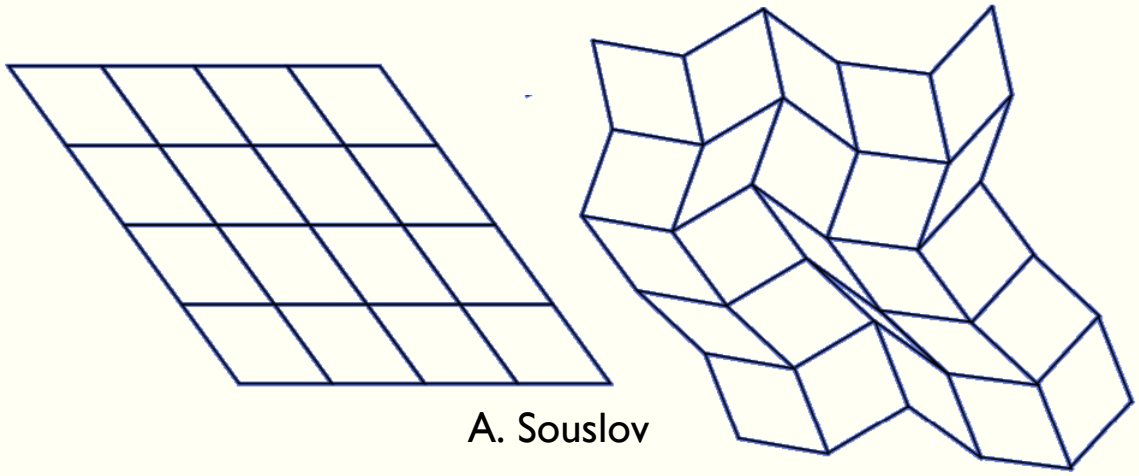
- **H** has d(d+1)/2 Galilean symmetries (rotations and translation)

  that conserve relative positions and are therefore zero-modes.

Zero elasticity modes may originate also from embedding, not only connectivity

$$\mathbf{H}_{ij}(c_i, c_j) = \Delta_{ij} \cdot k_{ij}(c_i, c_j) \cdot \mathbf{n}_{ij} \times \mathbf{n}_{ij}$$

- **H** has d(d+1)/2 Galilean symmetries (rotations and translation) that conserve relative positions and are therefore zero-modes.



A. Souslov

Exercise: find the floppy modes by (a) identifying the modes (b) counting constrains. How many are there?

- Some soft modes are outcome of the tensorial part of the elasticity (embedding).

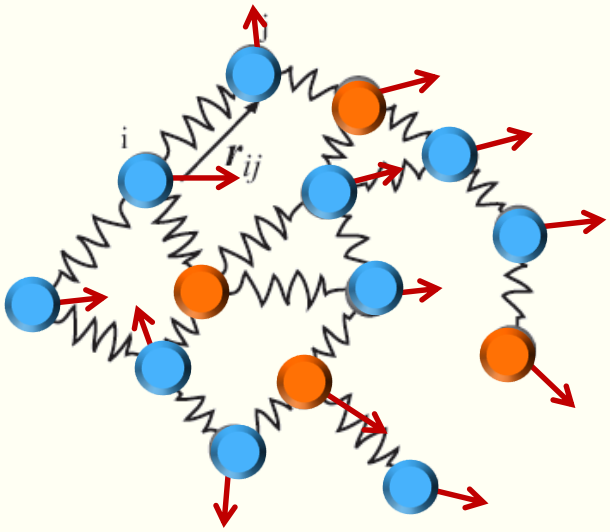
This is just the network version of Hooke's law

For point:

$$f = ku, \qquad \mathcal{E} = \frac{1}{2}ku^2$$

For network:

$$\mathbf{f}=\mathbf{H}\mathbf{u}, \qquad \mathcal{E} = \frac{1}{2}\mathbf{u}^{\mathsf{T}}\mathbf{H}\mathbf{u}$$

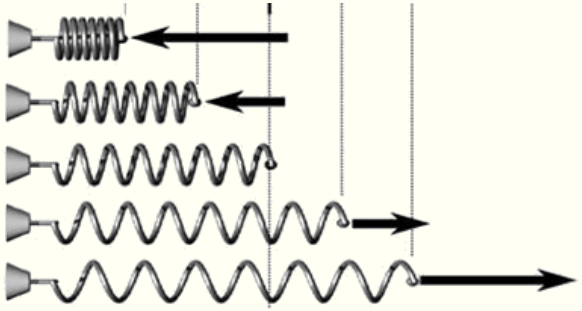


1676

3. The true Theory of Elafticity or Springineſs, and a particular Explication thereof in ſeveral Subjects in which it is to be found: And the way of computing the velocity of Bodies moved by them. ceiiinosssttuu. ut tenſio ſic vis

4. A very plain and practical way of counterpoiſing Liquors, of great uſe in Hydraulicks. Diſcovered.

5. A new ſort of Object-Glaſſes for Teleſcopes and Microſcopes, much exceeding any yet uſed. Diſcovered.

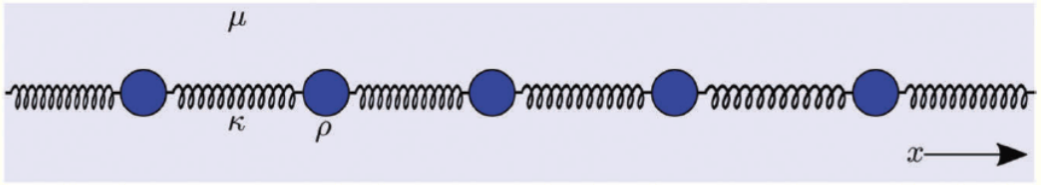


About two years ſince I printed this Theory in an Anagram at the end of my Book of the Deſcriptions of Helioſcopes, *viz.* c e i i i n o s s s t t u u, *id eſt*, *Ut tenſio ſic vis*; That is, The Power of any Spring is in the ſame proportion with the Tenſion thereof: That is, if one power ſtretch or bend it one ſpace, two will bend it two, and three will bend it three, and ſo forward. Now as the Theory is very ſhort, ſo the way of trying it is very eaſie.

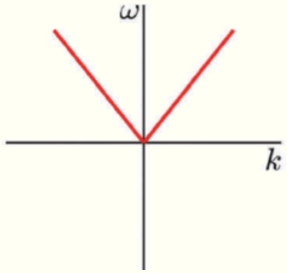
1678

Digression: what if there is no Hooke's law?

$$\rho \frac{\partial^2 A}{\partial t^2} + \mu \frac{\partial A}{\partial t} = \kappa \frac{\partial^2 A}{\partial x^2}$$



Inertial ($\mu = 0$): $\rho \frac{\partial^2 A}{\partial t^2} = \kappa \frac{\partial^2 A}{\partial x^2}$

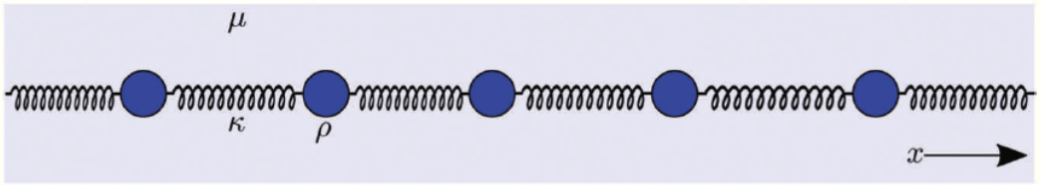


$\omega \sim \sqrt{\frac{\kappa}{\rho}}|k|$

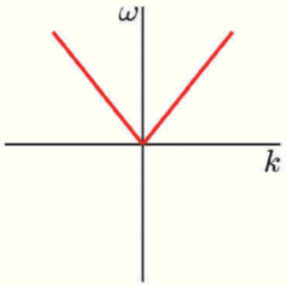
Hooke's law

What if there is no Hooke's law?

$$\rho\frac{\partial^2 A}{\partial t^2}+\mu\frac{\partial A}{\partial t}=\kappa\frac{\partial^2 A}{\partial x^2}$$



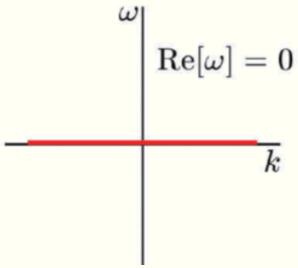
Inertial ($\mu = 0$): $\rho\frac{\partial^2 A}{\partial t^2}=\kappa\frac{\partial^2 A}{\partial x^2}$



$\omega \sim \sqrt{\frac{\kappa}{\rho}}|k|$

Hooke's law

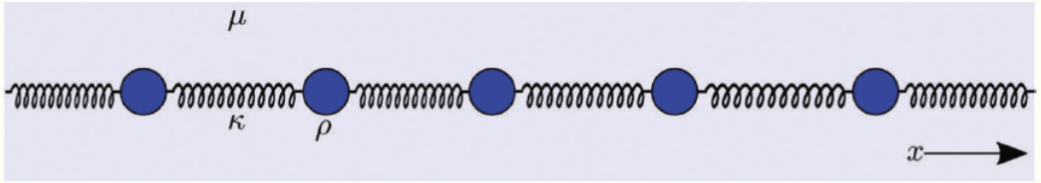
Damped ($\rho = 0$): $\mu\frac{\partial A}{\partial t}=\kappa\frac{\partial^2 A}{\partial x^2}$



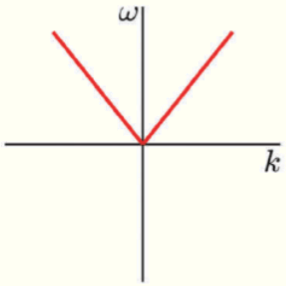
$\omega \sim i\left(\frac{\kappa}{\mu}\right)k^2$

What if there is no Hooke's law?

$$\rho\frac{\partial^2 A}{\partial t^2}+\mu\frac{\partial A}{\partial t}=\kappa\frac{\partial^2 A}{\partial x^2}$$



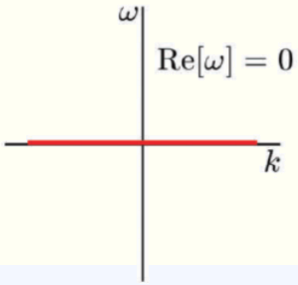
Inertial ($\mu = 0$): $\rho\frac{\partial^2 A}{\partial t^2}=\kappa\frac{\partial^2 A}{\partial x^2}$



$\omega \sim \sqrt{\frac{\kappa}{\rho}}|k|$

Hooke's law

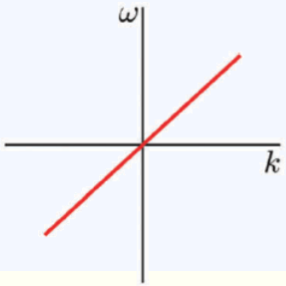
Damped ($\rho = 0$): $\mu\frac{\partial A}{\partial t}=\kappa\frac{\partial^2 A}{\partial x^2}$



$\omega \sim i\left(\frac{\kappa}{\mu}\right)k^2$

Sedimentation, driven μ-fluidics

Driven: $\mu\frac{\partial A}{\partial t}=\xi\frac{\partial A}{\partial x}$



$\omega \sim \left(\frac{\xi}{\mu}\right)k$

Sound velocity

Exercise: (a) why the simplest wave equation is not more common?

*(b) Derive the 1st-order wave equation in microfluidics (Beatus, Bar-Ziv et al.)?

I. Introduction

II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

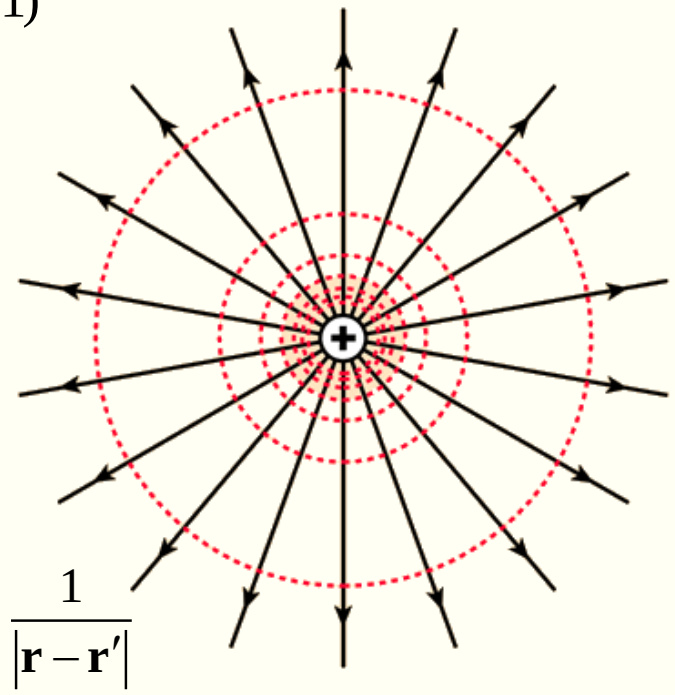
- Minimal Coarse-grained model
- Hooke's law of networks: Laplacian etc.
- ***Green's function***
- Mechanical evolution and scattering
- Epistasis and genetic correlation
- The dimension question
- Examples and conclusion

Green function is the response of a physical system to a local perturbation

- **Poisson' equation**: $\Delta\phi(\mathbf{r}) = \rho(\mathbf{r})$ $\quad (4\pi \to 1)$
- What charge density $\rho(\mathbf{r})$ induces a potential $\phi(\mathbf{r})$?
- **Green function** G answers the inverse question:
- What $\phi(\mathbf{r})$ is induced by $\rho(\mathbf{r})$?

$$\phi(\mathbf{r}) = \int d\mathbf{r}' G(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')$$

- With $G(\mathbf{r},\mathbf{r}')$ given by Coulomb's law: $G(\mathbf{r},\mathbf{r}') = \dfrac{1}{|\mathbf{r}-\mathbf{r}'|}$
- Green function is the inverse of the Laplacian



$$\Delta G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') \quad \Rightarrow \quad G = \Delta^{-1}$$

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

AN ESSAY

ON THE

APPLICATION

OF

MATHEMATICAL ANALYSIS TO THE THEORIES OF ELECTRICITY AND MAGNETISM.

BY

GEORGE GREEN.

Nottingham:

PRINTED FOR THE AUTHOR, BY T. WHEELHOUSE.

SOLD BY HAMILTON, ADAMS & Co. 33, PATERNOSTER ROW; LONGMAN & Co.; AND W. JOY, LONDON; J. DEIGHTON, CAMBRIDGE;

AND S. BENNETT, H. BARNETT, AND W. DEARDEN, NOTTINGHAM.

1828.

FACSIMILE-DRUCK IN 100 EXEMPLAREN. BERLIN. MAYER & MÜLLER. 1889.

Read the story of George Green:

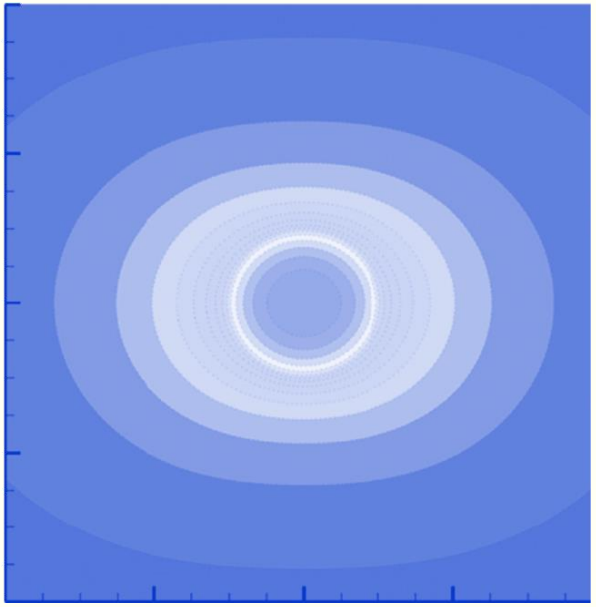
- *Dyson*, ***George Green and Physics***, World of Physics 1993
- *Challis & Sheard*, ***The Green of Green's Functions***, Physics Today 2007.

Elastic and hydrodynamic Green function are long-range

- Like the electrostatic one, but tensorial:

$$\mathbf{u}_i(\mathbf{r}) = \mathbf{G}_{ik}(\mathbf{r})\mathbf{f}_k \quad \text{with a point force } \mathbf{f}_k\delta(0)$$

$$\mathbf{G}(\mathbf{r}) \sim \frac{1}{E}\left[(3-4\sigma)\delta_{ik} + \mathbf{n}_i\mathbf{n}_k\right]\cdot\frac{1}{r}$$

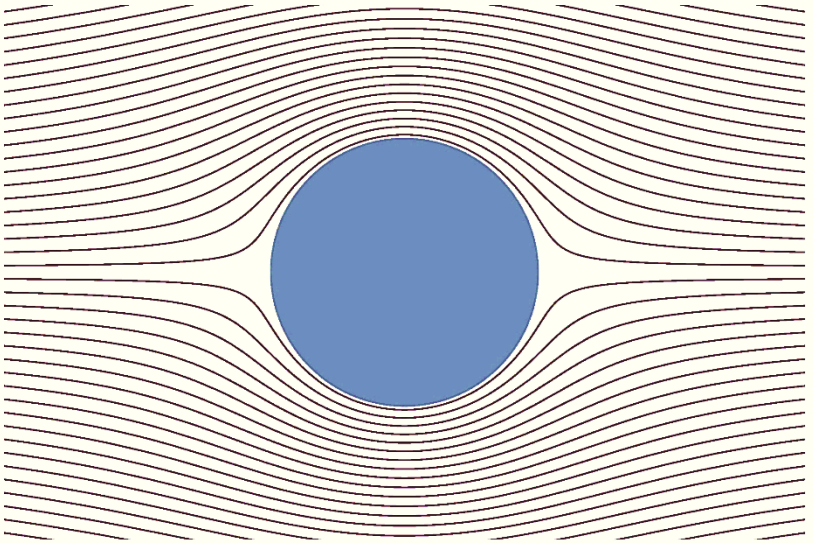


- Similarly in viscous flow:

$$\mathbf{v}_i(\mathbf{r}) = \mathbf{G}_{ik}(\mathbf{r})\mathbf{f}_k \quad \text{with a velocity } \mathbf{v}(\mathbf{r})$$

$$\mathbf{G}(\mathbf{r}) \sim \frac{1}{\eta}\left[\delta_{ik} + \mathbf{n}_i\mathbf{n}_k\right]\cdot\frac{1}{r}$$

Solves Stokes equation: $\nabla p = \eta\Delta\mathbf{v}$



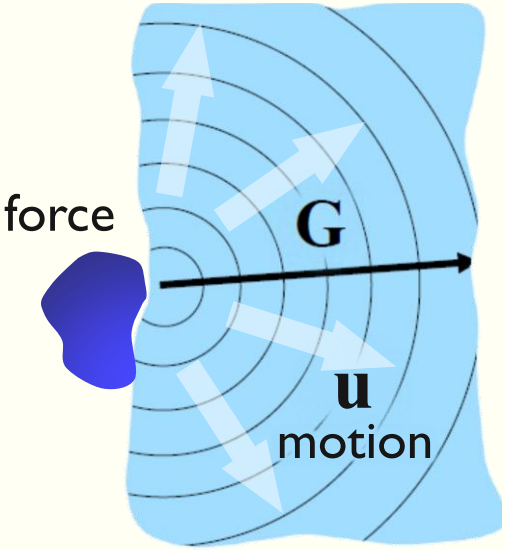
Exercise: get 1/r scaling from momentum conservation.

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

- Green's function solves the inverse problem

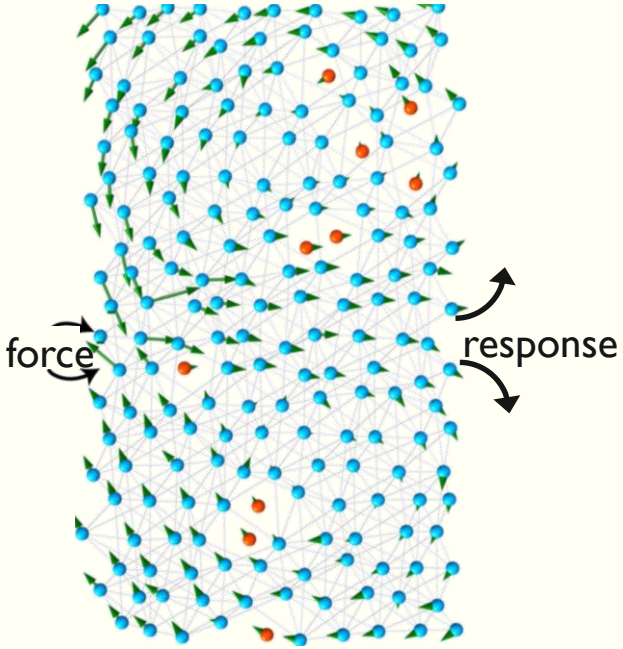
$$\left.\begin{aligned} |\text{force}\rangle &= \mathbf{H}\,|\text{motion}\rangle \\ |\text{motion}\rangle &= \mathbf{G}\,|\text{force}\rangle \end{aligned}\right\} \Rightarrow \quad \mathbf{G} = \mathbf{H}^{-1}$$

(the pesudoinverse $\mathbf{H}^{+}$)



$$|\text{motion}\rangle = (u_1..u_{200})$$
$$|\text{force}\rangle = (f_1...f_{200})$$

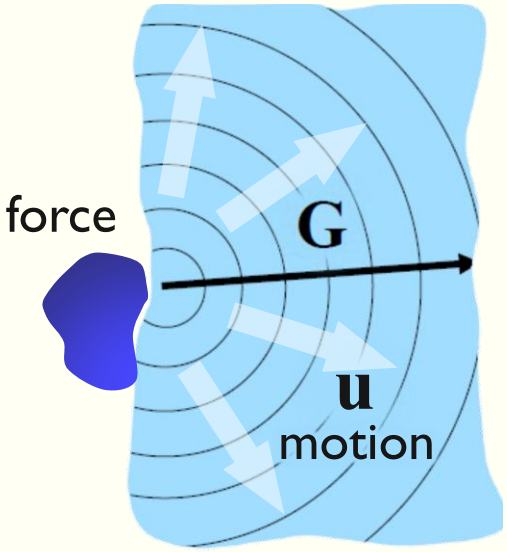
$\mathbf{H}$ = Δ of spring network



$$\mathbf{G}(\text{gene}) = \mathbf{G}(\text{network})$$

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

The Green function defines the gene-to-function map and the fitness landscape



$$|\text{motion}\rangle = (u_1, u_2, u_3 ..., u_{200})$$
$$|\text{force}\rangle = (f_1, f_2, f_3 ..., f_{200})$$

- **Genotype-to-phenotype map** is the induced motion

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

- **Fitness** measures performance of the protein
  = overlap with functional motion, $|\text{response}\rangle$

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

G: ***gene → amino acid network → protein dynamics → function → fitness***

I. Introduction

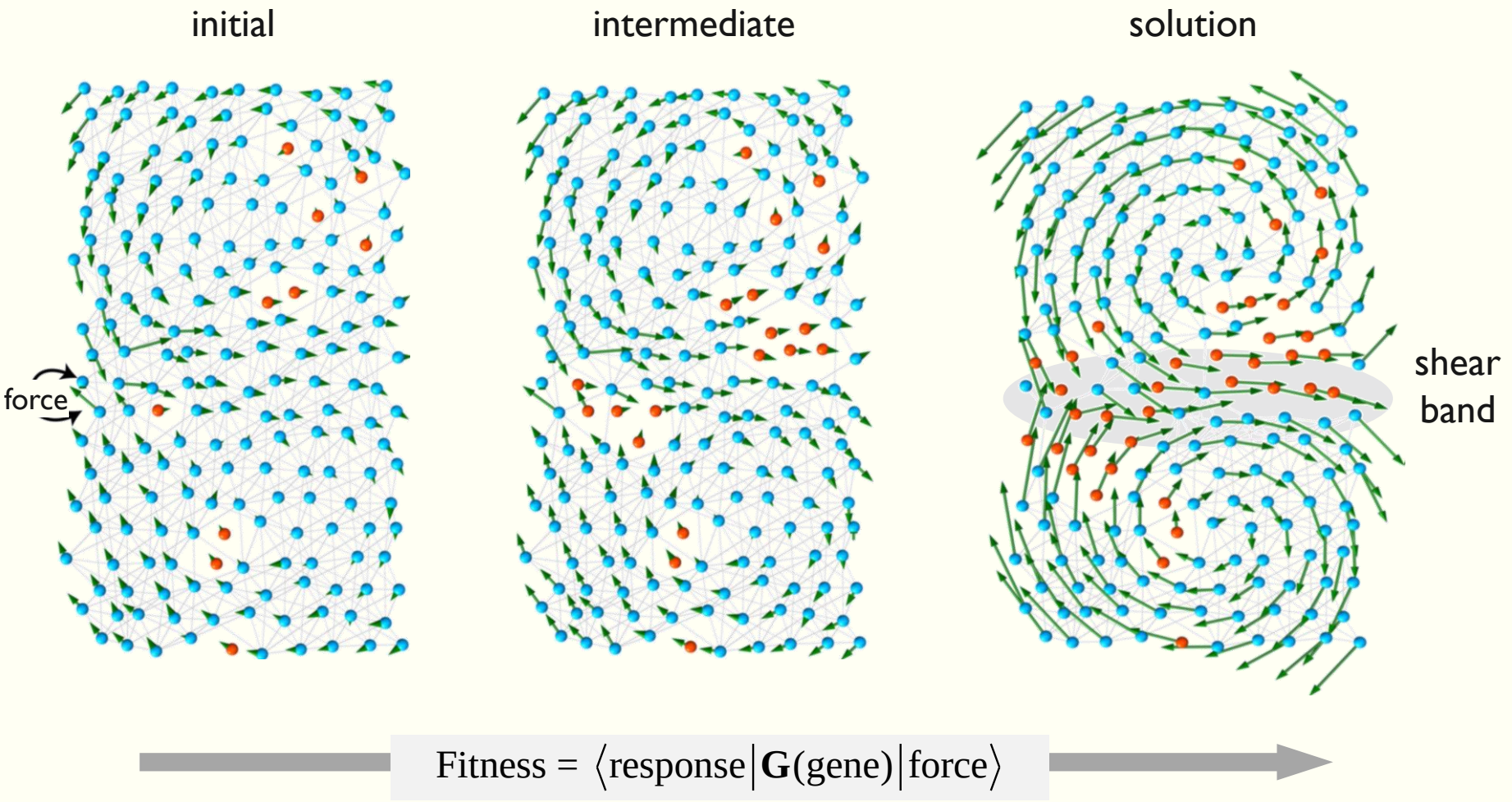
II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

- Minimal Coarse-grained model
- Hooke's law of networks: Laplacian etc.
- Green's function
- ***Mechanical evolution and scattering***
- Epistasis and genetic correlation
- The dimension question
- Examples and conclusion

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



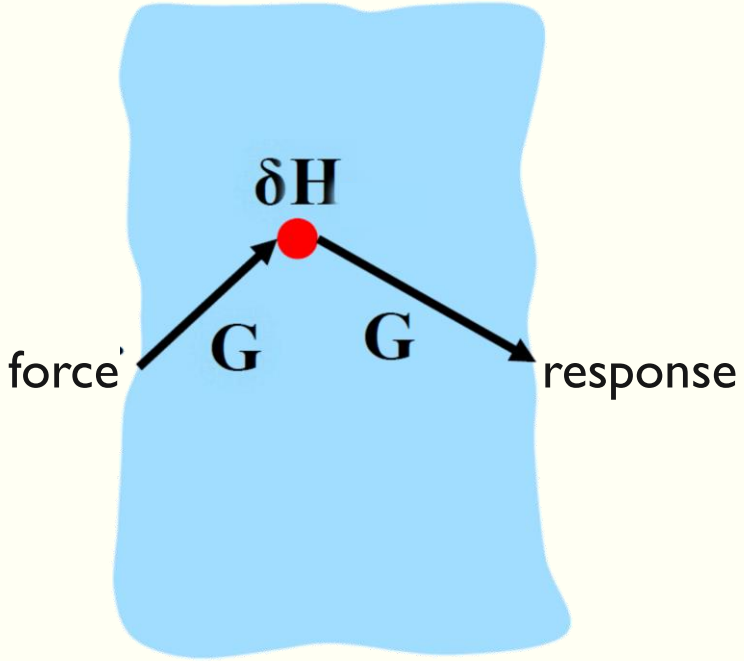
* One can look at *more general* patterns of forces and response.

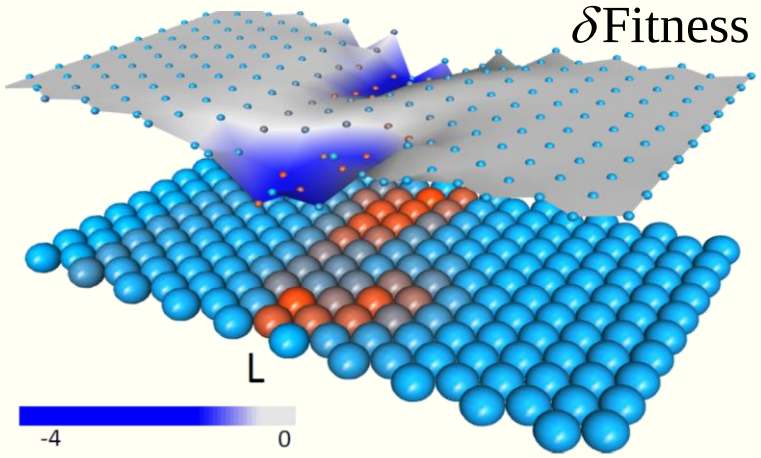
Mutations locally perturb the amino acid network, scatter the force, and thereby change the fitness

$$\delta\mathbf{H} = \mathbf{H}(c') - \mathbf{H}(c) = \text{mutation effect}$$

Dyson's series

$$\delta\mathbf{G} = \mathbf{G}(c') - \mathbf{G}(c) =$$
$$-\mathbf{G}\delta H\,\mathbf{G} + \mathbf{G}\delta H\,\mathbf{G}\delta H\,\mathbf{G} - \ldots$$





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

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

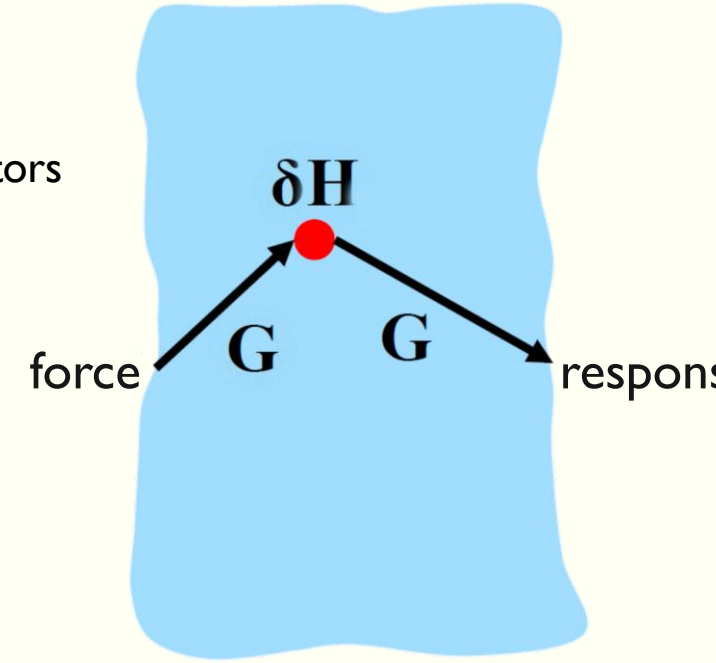
Woodbury's formula leverages the local nature of mutation to calculate the Green function fast

$\delta\mathbf{H} = \mathbf{M}\mathbf{B}\mathbf{M}^{\mathrm{T}}$

with $\mathbf{B}$ diagonal $r \times r$ and (r = # mutated bonds)

and $\mathbf{M}$ $n_a \cdot d \times r$ matrix whose columns are $D_{(\alpha)}$ bond vectors

$r = \mathrm{rank}(\delta\mathbf{H}) = \#$ bonds changed $< z = 12$.



- No need to invert the whole Hamiltonian:

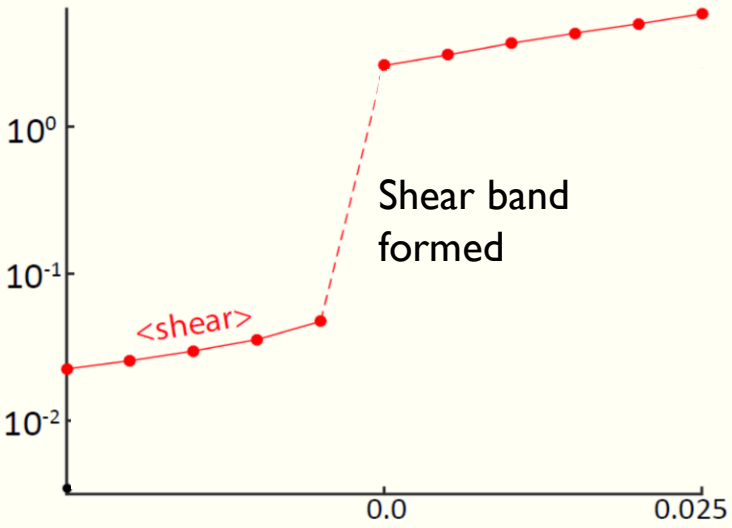
$$\delta\mathbf{G} = -\mathbf{G}\mathbf{M}(\mathbf{B}^{-1} + \mathbf{M}^{\mathrm{T}}\mathbf{G}\mathbf{M})^{+}\mathbf{M}^{\mathrm{T}}\mathbf{G}$$

- Accelerates the calculation by ~$(z/N)^3$~10,000

Exercise: show the scalar version –the resolvent formula

$$\frac{1}{x+\delta x} - \frac{1}{x} = -\frac{1}{x}\left(\frac{1}{\frac{1}{x}+\frac{1}{\delta x}}\right)\frac{1}{x}$$

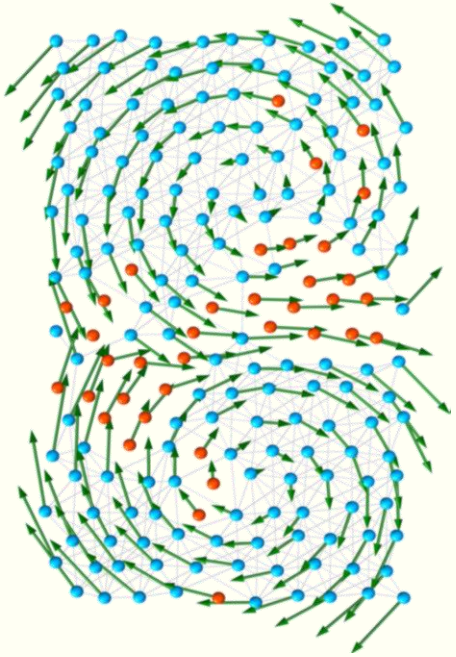
Protein function emerges at a *topological transition* when a shear band of mutations divides the protein and a *soft mode* appears



$$\mathbf{H} = \sum_i \lambda_i |\mathbf{v}_i\rangle\langle\mathbf{v}_i| \quad \Rightarrow \quad \mathbf{G} = \sum_i \frac{|\mathbf{v}_i\rangle\langle\mathbf{v}_i|}{\lambda_i}$$

$$\lambda_* = \text{soft mode} \to 0$$

$$\text{Fitness} = \langle \text{response} | \mathbf{G}(\text{gene}) | \text{force} \rangle \sim \frac{1}{\lambda_*} \to \infty$$

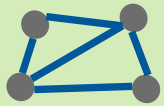


- The matrix-tree theorem (Kirchhoff, 1847)

$$\#\ \text{spanning trees} = \frac{1}{n}\lambda_*\lambda_2 \ldots \lambda_{n-1} \quad \text{(non-zero e.v. of } \Delta)$$

Dutta Eckmann Libchaber (2018)

Exercise: demonstrate the matrix-tree theorem for:

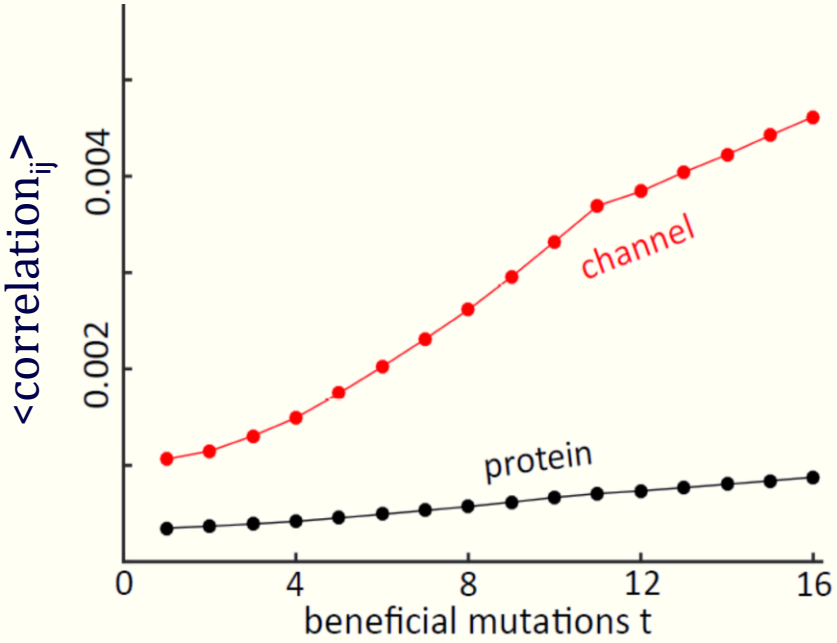


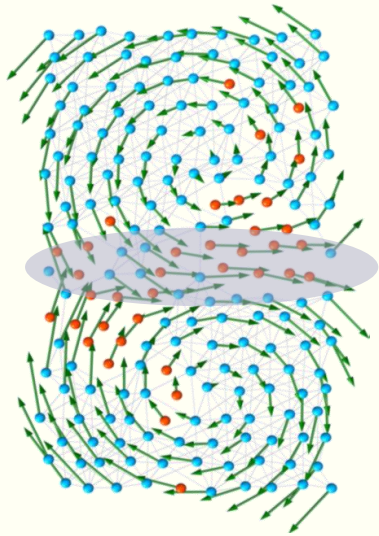
The genetic correlation among solutions increases towards the emergence of mechanical function

10^6 solutions

1 1 1 0 0 1 0 1 0 1 1 1 1 1 0 0 0 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 1 1 1 0 …
0 0 1 1 1 1 1 0 1 1 1 1 1 0 1 0 0 1 1 1 0 1 0 …
1 0 0 1 1 1 1 1 1 1 1 0 0 1 1 0 0 1 1 1 1 0 0 0 …
…

$$\text{correlation}_{ij} = \langle \text{codon}_i \, \text{codon}_j \rangle_{\text{genes}} - \langle \text{codon}_i \rangle_{\text{genes}} \langle \text{codon}_j \rangle_{\text{genes}}$$





I. Introduction

II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

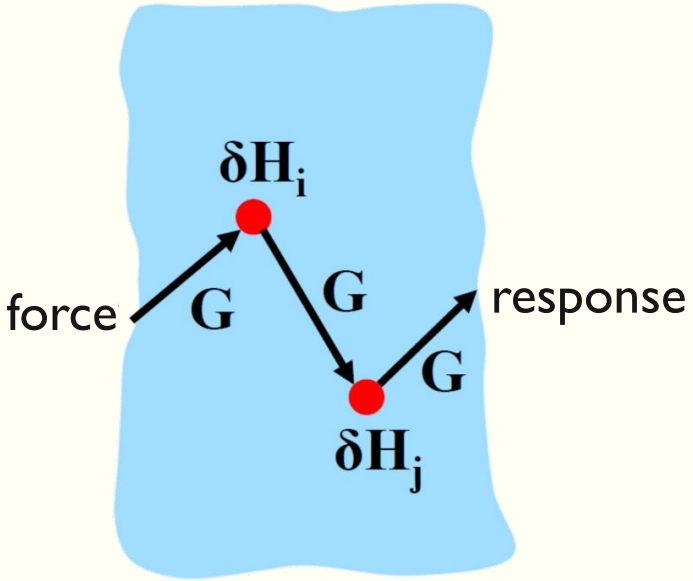
IV. Mechanical evolution of protein

- Minimal Coarse-grained model
- Hooke's law of networks: Laplacian etc.
- Green's function
- Mechanical evolution and scattering
- ***Epistasis and genetic correlation***
- The dimension question
- Examples and conclusion

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

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

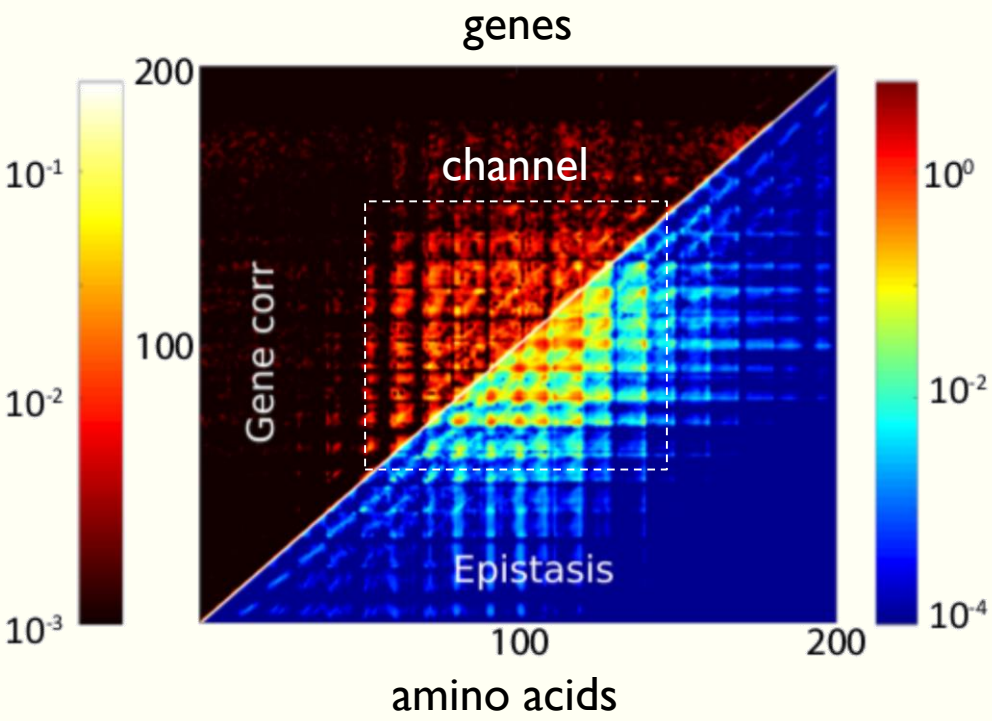
- Mechanical interpretation:
- Epistasis = sum over all scattering paths passing through *i* and *j*

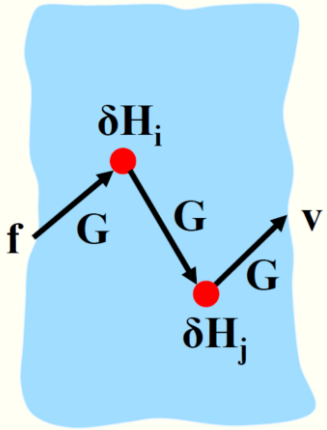


$$\text{epistasis}_{i,j} = \left\langle \text{force} \,|\, \mathbf{G\delta H_i G\delta H_j G + G\delta H_j G\delta H_i G}\ldots \,|\, \text{response} \right\rangle$$

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

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

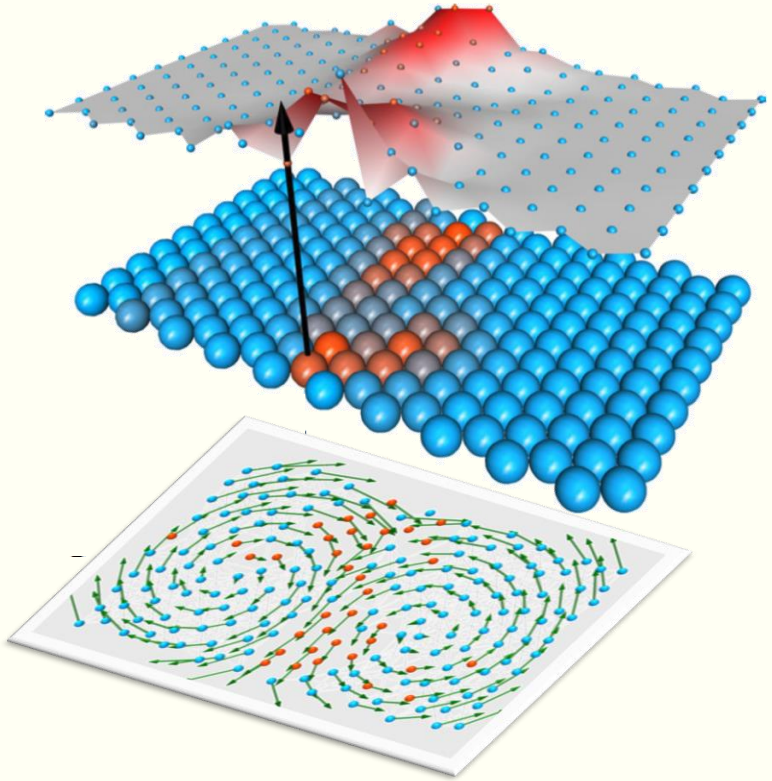




$$\text{epistasis}_{i,j} = \delta\text{Fitness}_{i,j} - \delta\text{Fitness}_i - \delta\text{Fitness}_j$$

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

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



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

$$h_i = \left\langle \text{motion} \left| \mathbf{\delta H_i} \right| \text{motion} \right\rangle = \Delta E_{\text{shear}}$$

Exercise: show that epistasis is approximately a rank-1 tensor

$$\text{epistasis}_{ij} \simeq \min(1, h_i) \cdot \min(1, h_j)$$

I. Introduction

II. Imperfect duality in living matter

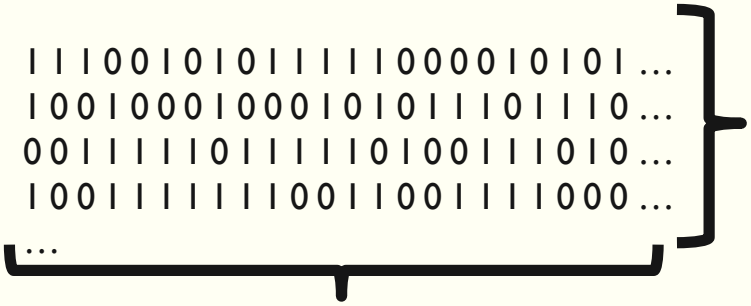
III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

- Minimal Coarse-grained model
- Hooke's law of networks: Laplacian etc.
- Green's function
- Mechanical evolution and scattering
- Epistasis and genetic correlation
- ***The dimension question***
- Examples and conclusion

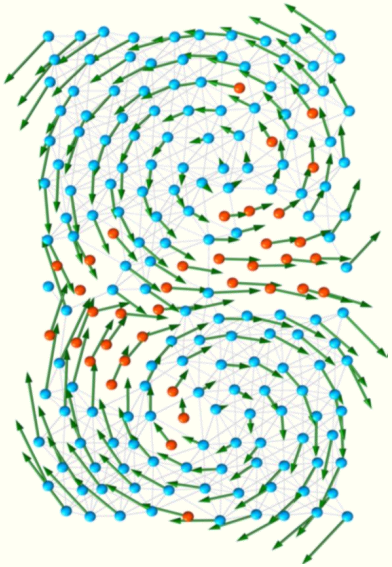
Spectra of phenotypes and genotypes capture the relevant modes of mechanical function and evolution

- Gene matrix M =



10^6 solutions

Gene length = 200



- Singular Value Decomposition (SVD) yields the solution spectrum

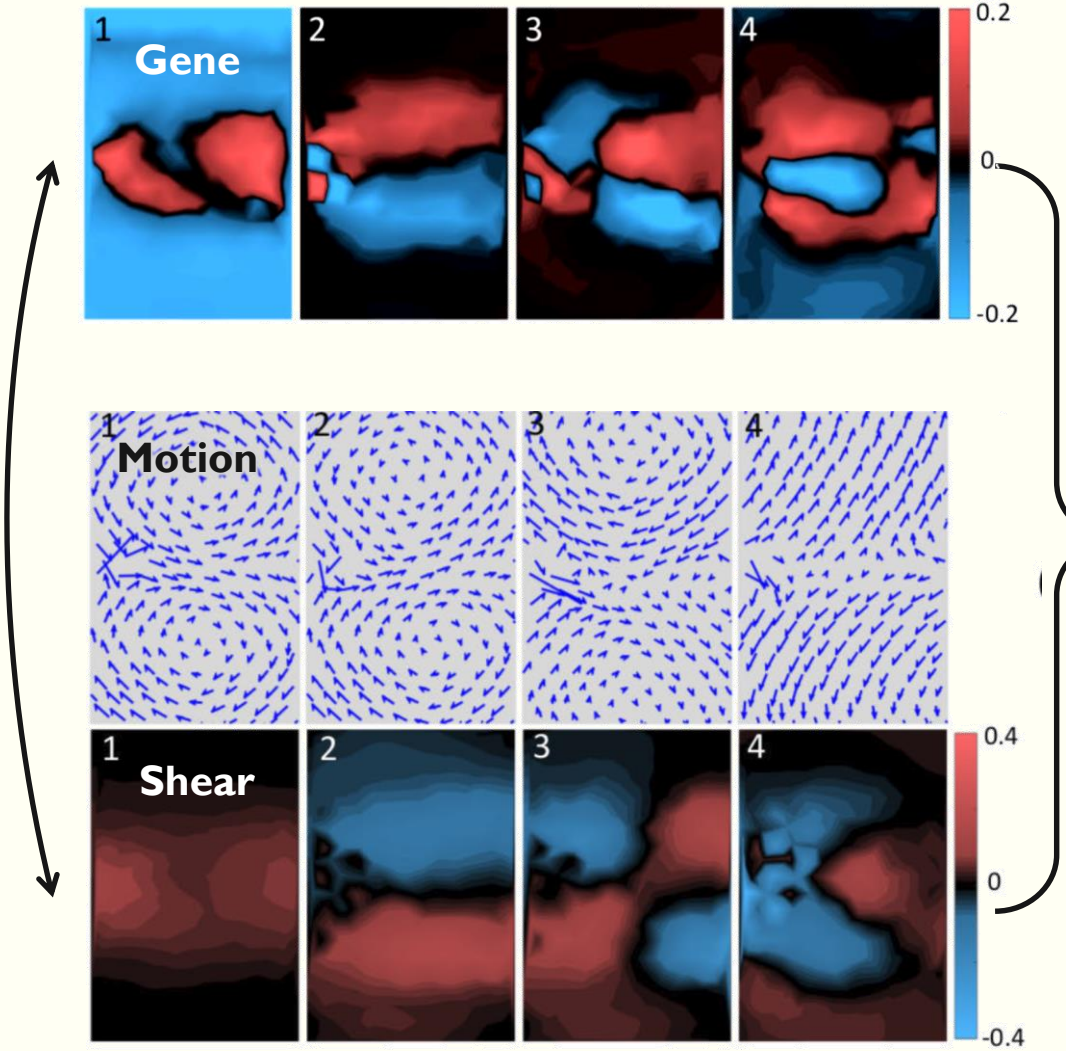
eigenvalues

$$\mathbf{M} = \mathbf{U} \begin{pmatrix} \lambda_1 & 0 & & \\ 0 & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_{200} \\ 0 & 0 & \cdots & 0 \\ & & \cdots & \end{pmatrix} \mathbf{V}^{\mathrm{T}}$$

eigenvectors

$10^6 \times 200$ $10^6 \times 10^6$ $10^6 \times 200$ 200×200

Spectra of phenotypes and genotypes capture the relevant modes of mechanical function and evolution



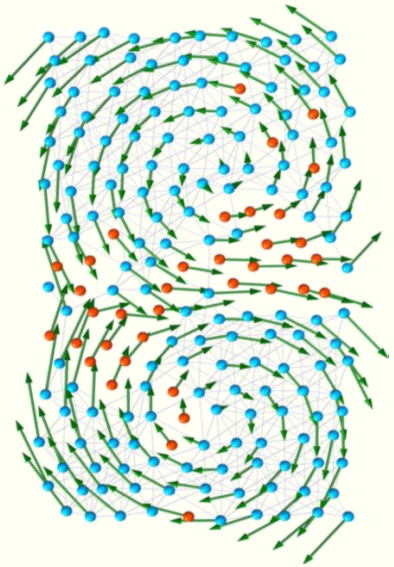
gene

…101111101111101…

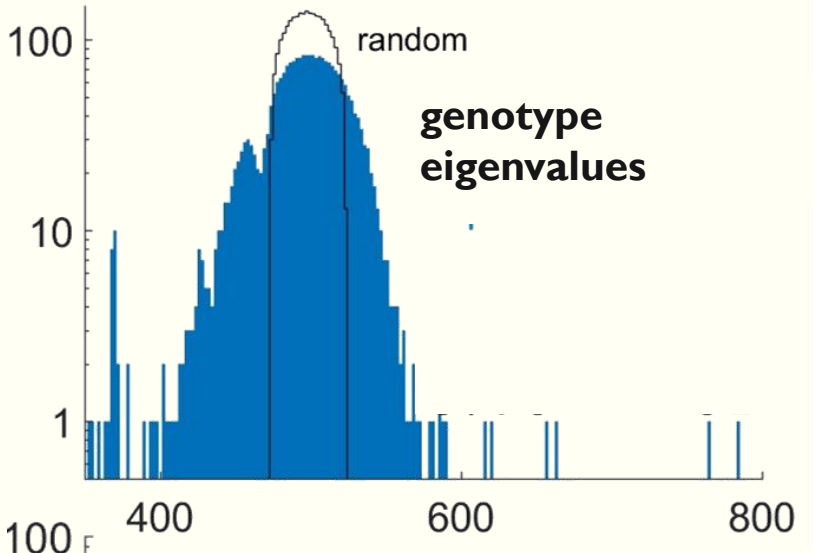
…001111111001100…

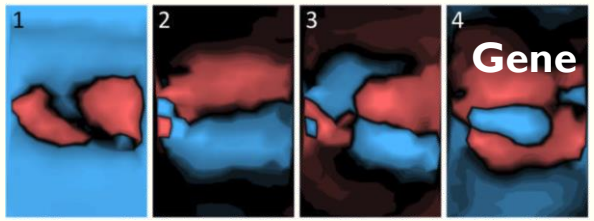
SVD of

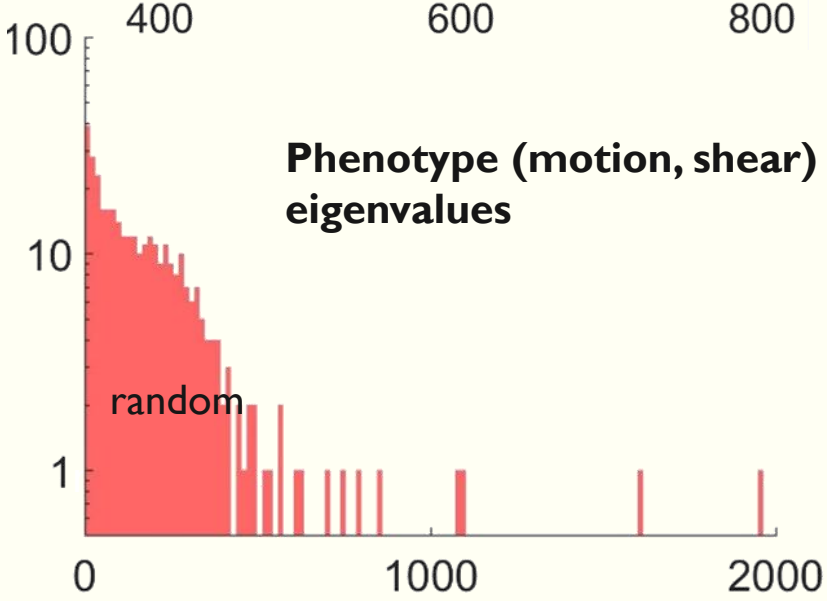
10^6 solutions

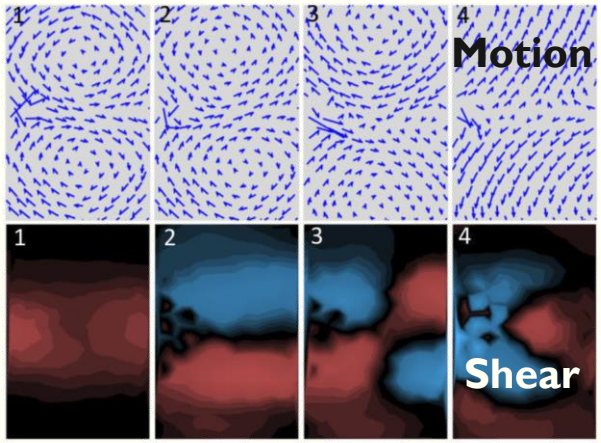


A few non-random eigenvectors capture the functionally relevant bits of information in the gene







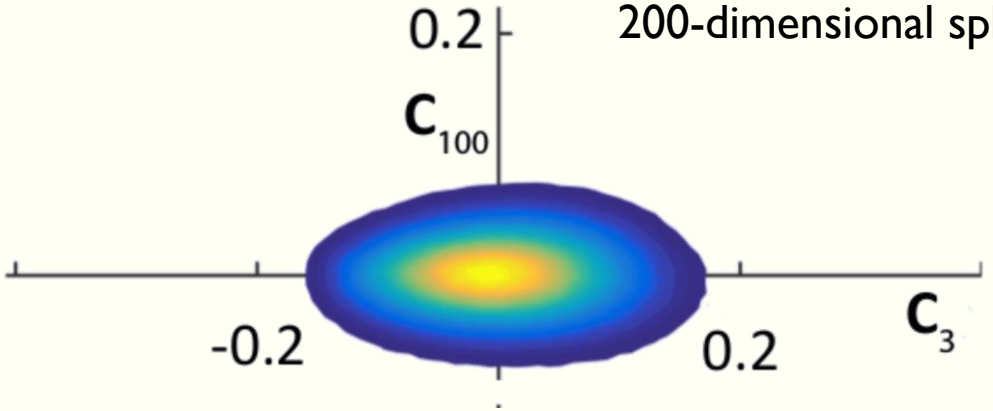


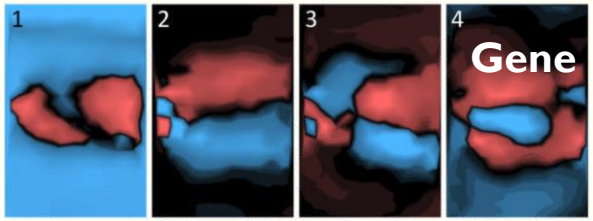
The genotype-to-phenotype map involves huge dimensional reduction whose origin is the mechanical interaction

10^6 solutions

Genotype space

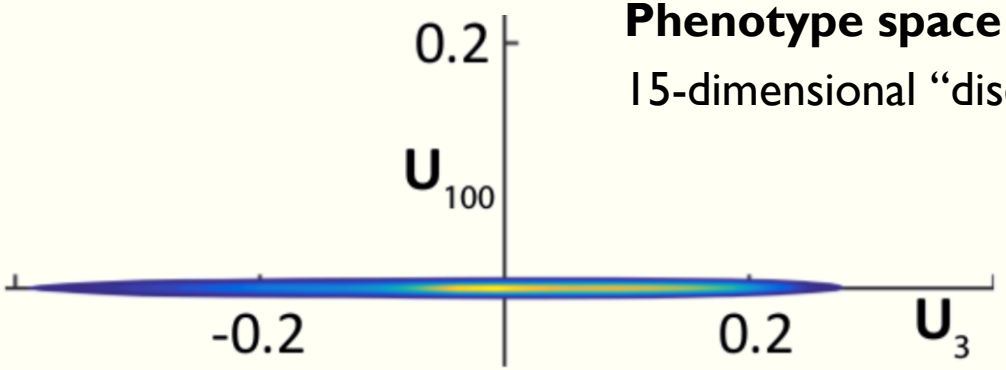
200-dimensional sphere

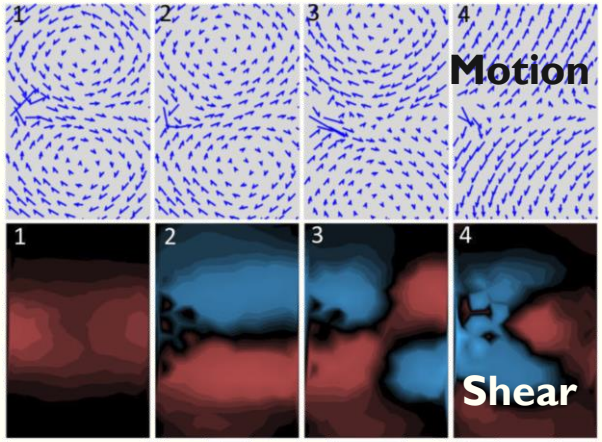




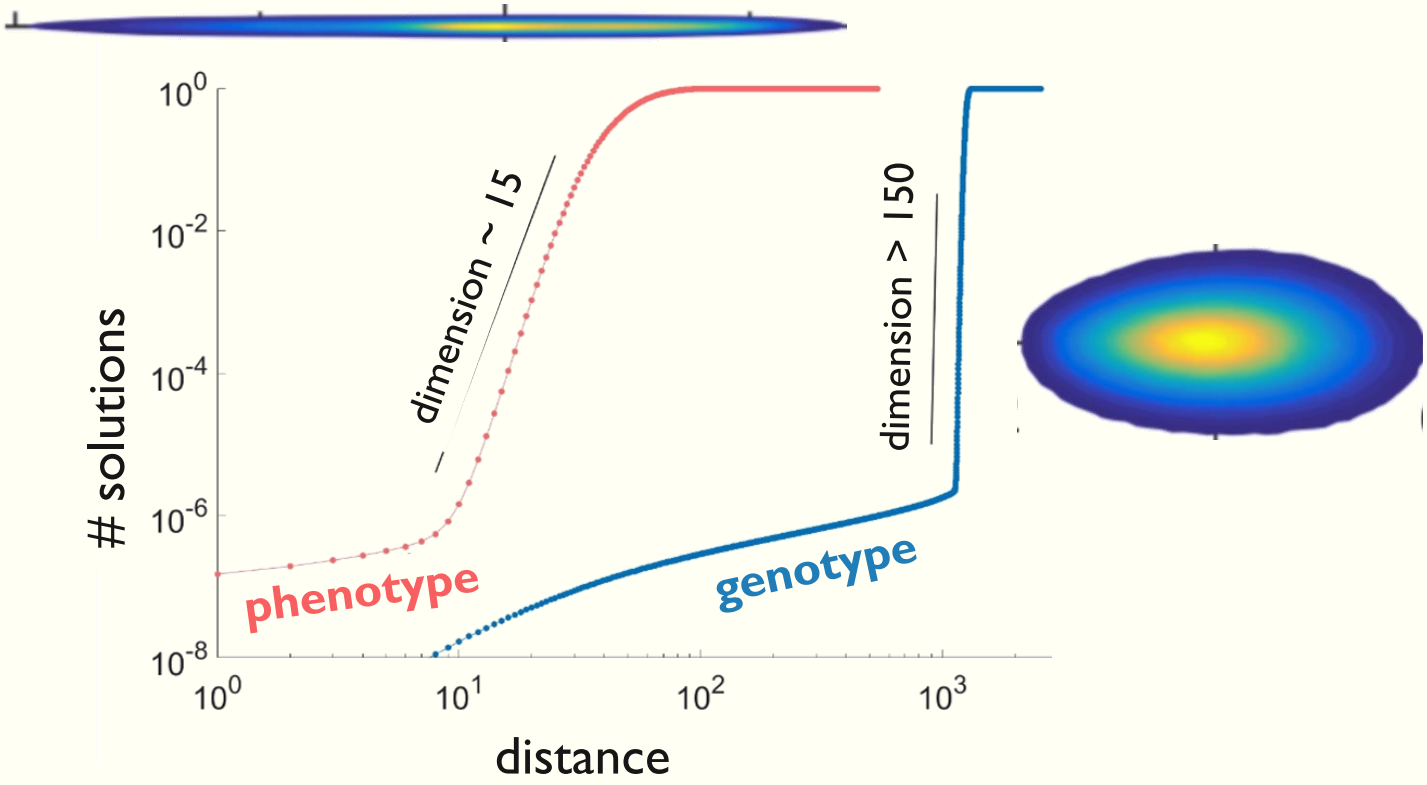
Phenotype space

15-dimensional "disc"





Dimension of genotype and phenotype can be found by box counting



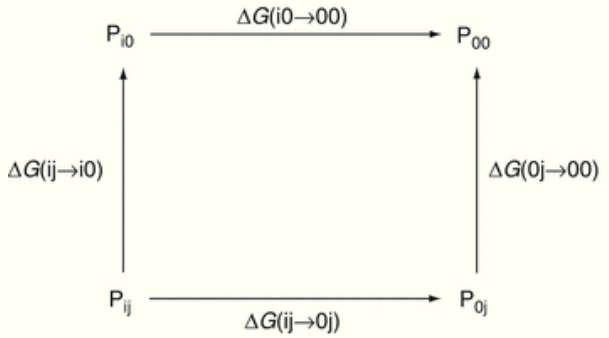
$$\#\text{ solutions} \sim (\text{distance})^{\textbf{dimension}}$$

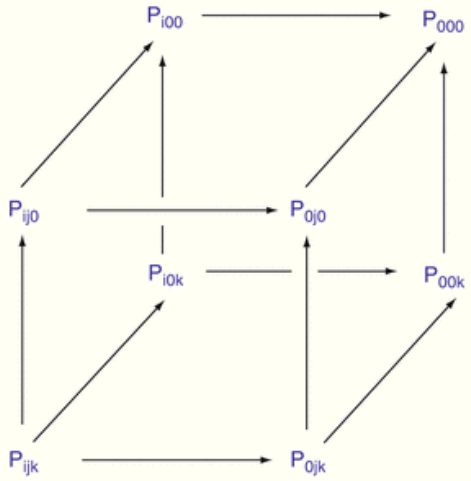
Higher-order epistatic interactions are higher-order curvatures on the hypercube

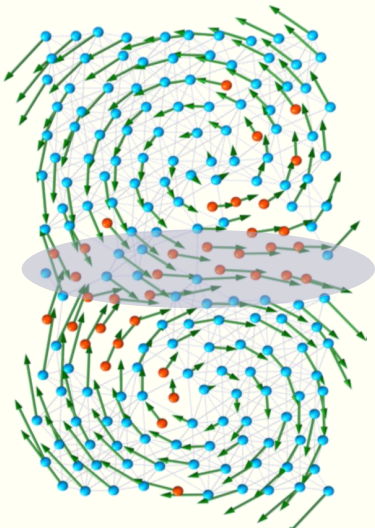
2^{nd}-order: $\mathrm{e}_{i,j} = \delta\mathrm{F}_{i,j} - \left(\delta\mathrm{F}_i + \delta\mathrm{F}_j\right)$

3^{rd}-order: $\mathrm{e}_{i,j,k} = \delta\mathrm{F}_{i,j,k} - \left(\delta\mathrm{F}_{i,j} + \delta\mathrm{F}_{j,k} + \delta\mathrm{F}_{i,k}\right) + \delta\mathrm{F}_i + \delta\mathrm{F}_j + \delta\mathrm{F}_k$

N^{th}-order: $\mathrm{e}_{i_1,i_2,\ldots,i_N} = \sum_{q=1}^{N} (-1)^{N-q} \sum_{i_1<\ldots<i_q} \delta\mathrm{F}_{i_1\ldots i_q}$,



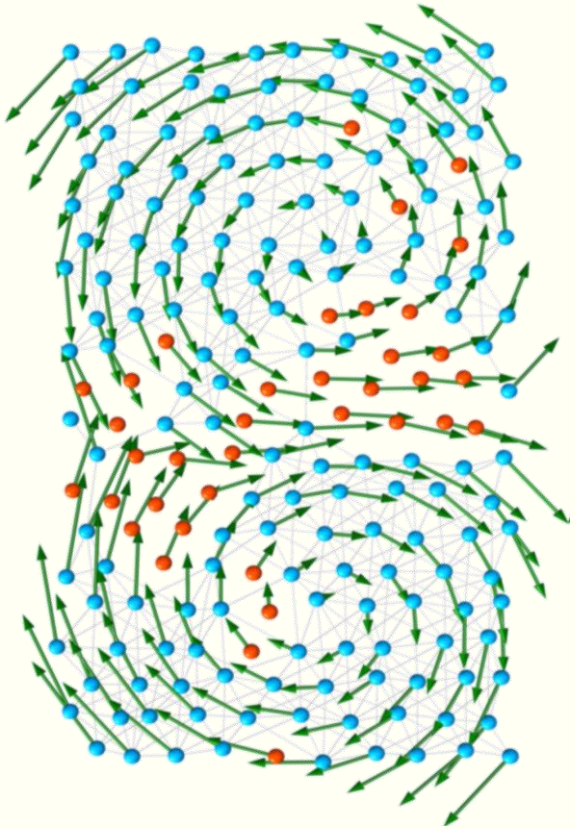




Mechanically critical regions are strongly coupled with ***many-body epistatic interactions*** among the mutations.

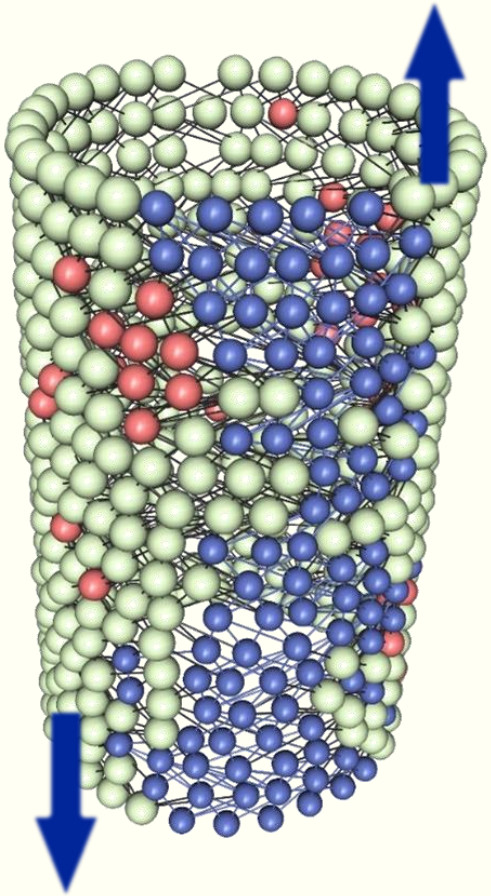
Exercise: derive 4-th order epistasis

The same phenomena are observed in a rigidity percolation model of protein



Green functions:

Dutta Eckmann Libchaber et al. PNAS 2018



Rigidity propagation:

Eckmann Libchaber et al., PRX 2017

I. Introduction

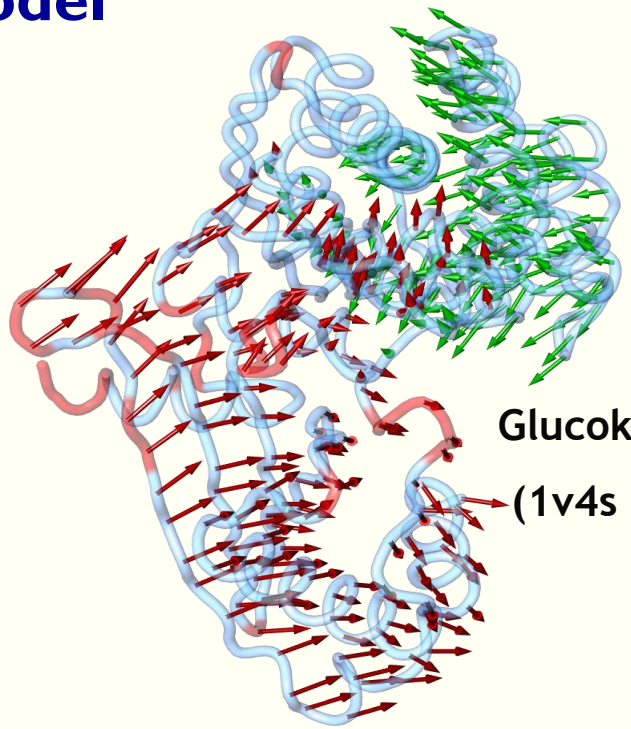
II. Imperfect duality in living matter

III. Protein: amorphous evolving matter

IV. Mechanical evolution of protein

- Minimal Coarse-grained model
- Hooke's law of networks: Laplacian etc.
- Green's function
- Mechanical evolution and scattering
- Epistasis and genetic correlation
- The dimension question
- ***Examples and conclusion***

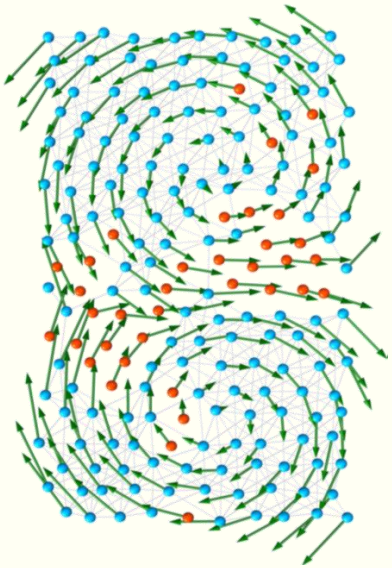
The simple predictions of the model can be tested in real proteins



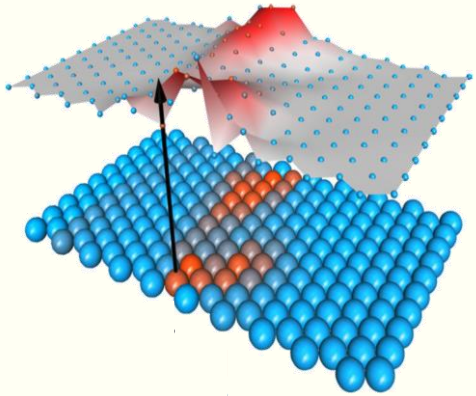
Glucokinase

(1v4s → 1v4t)

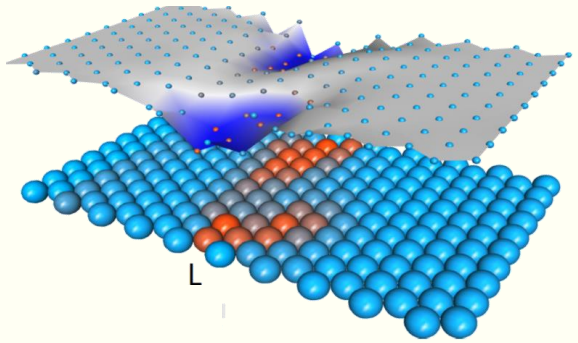
Shear

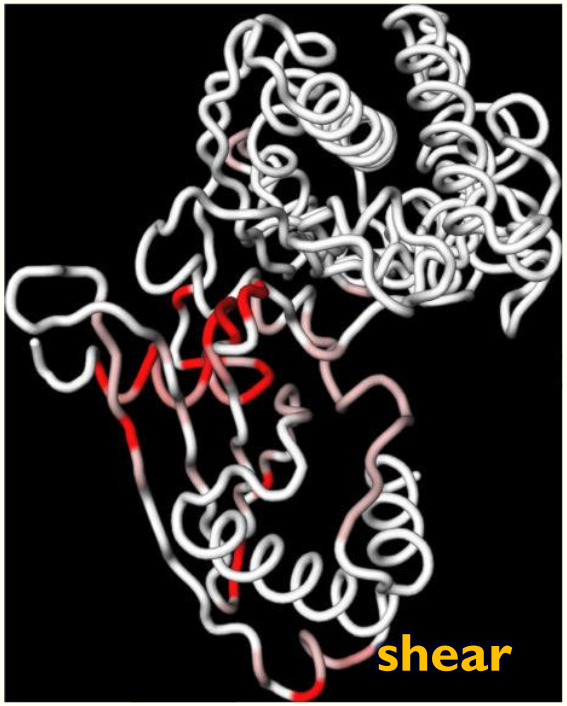


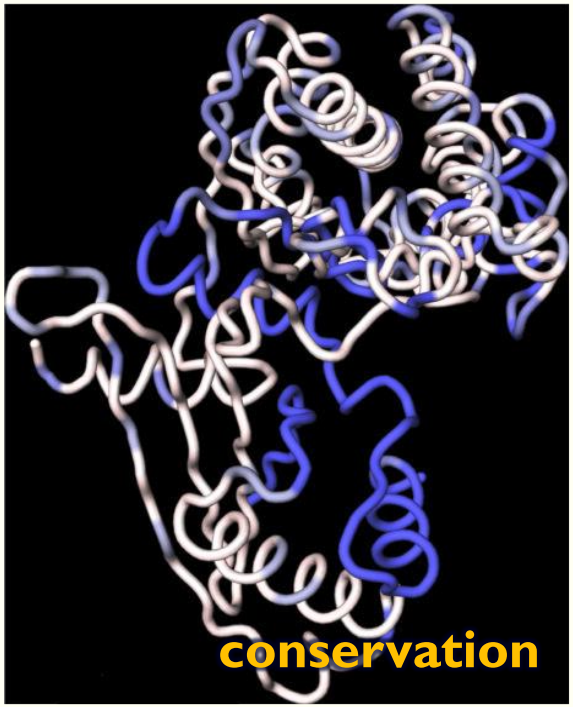
epistasis



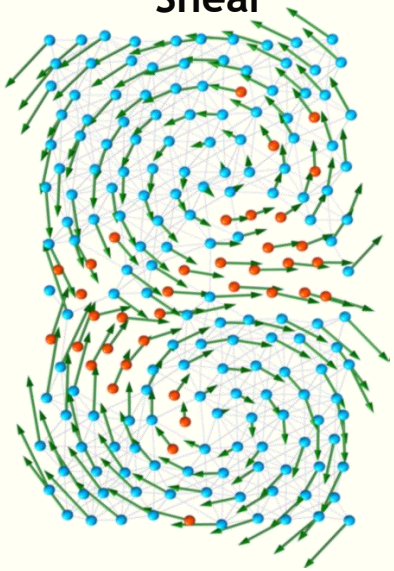
Fitness sensitivity



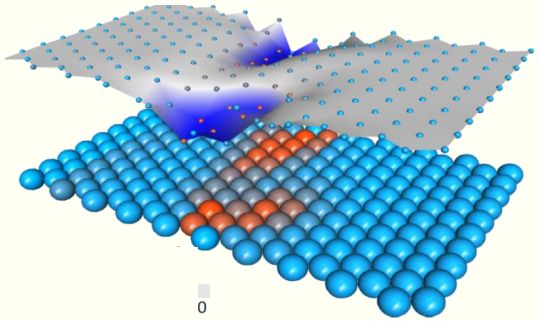




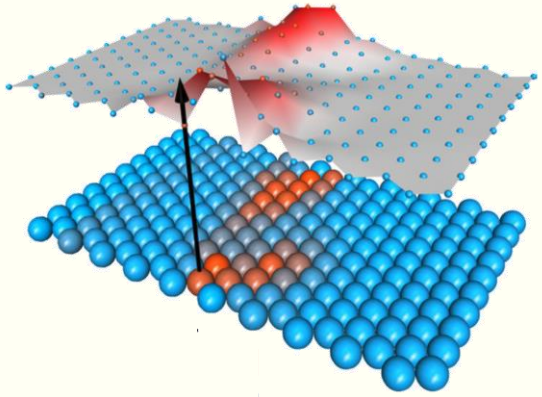
Shear



fitness sensitivity



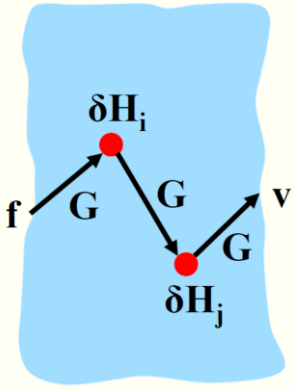
epistasis

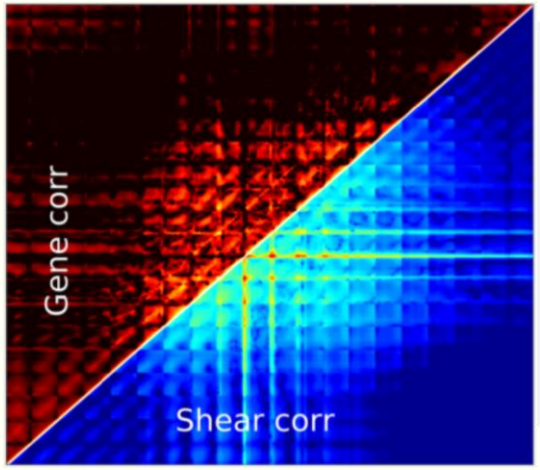


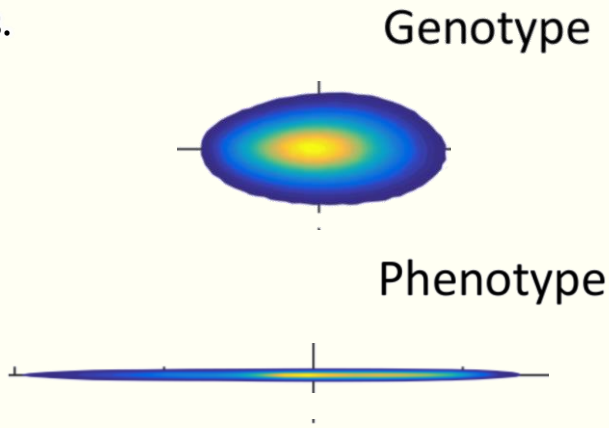
Rev Mod Phys (in press)

A minimal theory of protein links genotype and phenotype by mechanical propagators

- **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.





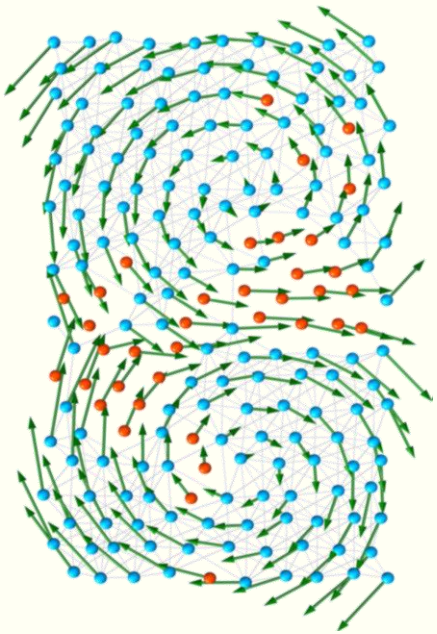


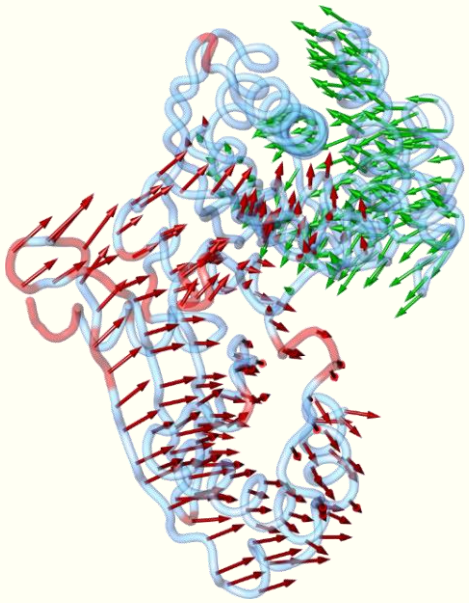
Dutta Eckmann Libchaber et al. PNAS 2018
Eckmann Libchaber et al., PRX 2017

Eckmann et al. Rev Mod Phys (in press)

- ***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 & J. Rougemont (Geneva)
- A. Libchaber (Rockefeller)
- S. Leibler & M. Mitchell (IAS & Rockefeller)

Center for Soft & Living matter