Memory Formation: From Condensed Matter to Biological Matter and Beyond

KITP 2018

MECHANICS AND INFORMATION
IN A MINIMAL MODEL OF PROTEIN EVOLUTION

Sandipan Dutta & TT (IBS)
Albert Libchaber (Rockefeller)
Jean-Pierre Eckmann (Geneva)
Michael Mitchell & Stanislas Leibler (IAS & Rockefeller)
Proteins are complex systems made of hundreds to thousands amino acids arranged in a 3D structure

» Non-random, information-rich.

» Small, heterogeneous.

» Collective interactions

» Evolutionary memory.

Hemoglobin, Irving Geiss, 1978
Genes and proteins have many degrees-of-freedom but protein function is described by a few parameters.

ATG CTG GAG AAC GGG CTG GCC 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 GCC TGG AGG GAG CTG GCC 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

Information ~ gene length  Dimensional reduction

rate

substrate

energy

reaction coordinate

Information ~ a few bits
Many proteins use *large-scale collective motions* to function.

**Induced fit:**
- Enzyme
- Ligand

**Allostery:**
- Allosteric protein
- Functional site
- Effector
Protein sequences are also remarkably collective, with correlations indicating structure and function.

Correlations predict 3D contacts.
Sander Marks Colwell Hopf et al.

Sectors: coevolving protein domains
Halabi Rivoire Leibler Ranganathan et al.

**Epistasis** = interaction among mutations
Looking at dynamics suggests a mechanical basis for mapping genetic information to collective function.

Motion (displacement):

unbound $\rightarrow$ bound

Glucokinase

Dutta Eckmann Libchaber & TT, arXiv
Mitchell TT Leibler, PNAS 2016
Eckmann Libchaber & TT, PRX 2017
The shear highlights the relative motion of amino acids within the protein

Shear = gradient (motion)

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

Glucokinase (PDB: 1v4s → 1v4t)
A two-dimensional “shear band” communicates mechanical signals between distant sites

Shear band dimension
- Hemoglobin: 2.3
- Glucokinase: 2.0
- ATCase: 2.3
- Albumin: 1.9

Hemoglobin

Mitchell TT Leibler, PNAS 2016
Proteins are soft and viscous at low frequency

Zocchi et al., 2012
Minimal model of protein as evolving matter to give evolution a mechanical interpretation

Dutta Eckmann Libchaber & TT, in review, arxiv
TT Eckmann & Libchaber, Phys Rev X 2017
Coarse-grained network models of protein examine large-scale dynamics

**Floppy modes in amorphous matter**
- “fracton” spectrum in glasses
- Floppy modes in amorphous solids
- Constraint theory in glasses

**Elastic networks and normal modes in proteins**
- Protein normal-modes.
- Large amplitude elastic modes in proteins.
- Elastic network models
  - Bahar et al. (many works since 1990s).

**Allostery in elastic network models**
- Allostery in coarse-grained models of proteins.
- Low-frequency modes describe allostery.
- Large-scale allosteric deformations

**Weakly-connected regions and viscoelasticity in proteins and biologically-inspired allosteric matter**
- “Protein-quakes”.
- Viscoelastic properties of enzymes.
- Mechanical stress in functional enzymes.
- Strain analysis of protein structures and low-dim.
  - Mitchell, TT & Leibler. PNAS (2016)
- Allostery-inspired response in mechanical networks.
- Architecture and coevolution of allosteric materials.
  - Yan, Ravasio, Brito & Wyart. PNAS (2017)
- Elastic networks models of allosteric proteins.
- Optimal cooperativity in allosteric materials
  - Yan, Ravasio, Brito & Wyart. arxiv (2017)

**Mechanical genotype-to-phenotype map of proteins**
- Self-referring DNA and protein.
  - TT, Phil. Trans A (2016).
- Physical model of genotype-to-phenotype map.
The Green function of the protein measures the transmission of force induced by a local impulse, such as ligand binding.

\[ G(gene) = G(aa \text{ net}) \]

- Green’s function solves the inverse problem

\[ |\text{force}\rangle = H |\text{motion}\rangle \]
\[ |\text{motion}\rangle = G |\text{force}\rangle \]

\[ \Rightarrow \quad G = H^{-1} \quad (\text{or} \quad H^+) \]
The Green function defines the genotype-to-phenotype map and the fitness landscape

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

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

**Fitness landscape** (gene $\rightarrow$ fitness)

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

$$= \langle\text{response}|G(\text{gene})|\text{force}\rangle$$
Evolution searches in the mechanical fitness landscape for a shear band and soft mode

Fitness = \( \langle \text{response} | G(\text{gene}) | \text{force} \rangle \)
Mutations locally perturb the amino acid network, scatter the force, and thereby change the fitness.

\[ \delta G = -G \delta H G + G \delta H G \delta H G - \ldots \]

\[ \delta \text{Fitness} = \langle \text{response} | \delta G | \text{force} \rangle \]
Protein function emerges at a topological transition when a shear band of mutations divides the protein and a soft mode appears.

Fitness = \langle \text{response} \mid G(\text{gene}) \mid \text{force} \rangle \sim \frac{1}{\text{energy(soft mode)}}
The genetic correlation among solutions increases towards the emergence of mechanical function.
Epistasis is interaction of mutations resulting in nonlinearity of the mechanical fitness

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

- Epistasis ~ “curvature” of G
  ~ sum over scattering paths:

\[ \text{epistasis}_{i,j} = \langle \text{force} | G\delta H_i G\delta H_j G + G\delta H_j G\delta H_i G \ldots | \text{response} \rangle \]
Mechanical forces link epistasis to genetic correlation, and both exhibit similar patterns

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

\[
\text{epistasis}_{i,j} = \delta\text{Fitness}_{i,j} - \delta\text{Fitness}_i - \delta\text{Fitness}_j
\]
Epistasis and sequence correlation are strongest in the high shear region separating the rigid subdomains.

\[ \text{epistasis}_{i,j} = \langle \text{force} | G\delta H_i G\delta H_j G + G\delta H_j G\delta H_i G \ldots | \text{response} \rangle \]
Phenotype and genotype modes capture the relevant degrees-of-freedom of mechanical function and evolution.

SVD of $10^6$ solutions
The genotype-to-phenotype map involves huge dimensional reduction

10^6 solutions

**Genotype space**
400-dimensional spheroid

**Phenotype space**
15-dimensional discoid
Dimensional reduction is a general phenomenon in life

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

1D scaling of kinetic parameters in Rubsico
(Savir, Noor, Milo & TT, PNAS 2010)
Genetics and mechanical function are linked by a minimal theory

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

Dutta Eckmann Libchaber & TT, in review

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

TT Eckmann & Libchaber, Phys Rev X 2017
Looking at protein dynamics suggests simple mechanical understanding of protein as evolving amorphous matter.

Thanks:
S. Dutta (IBS)
J.-P. Eckmann (Geneva)
A. Libchaber (Rockefeller)
S. Leibler & M. Mitchell (IAS)