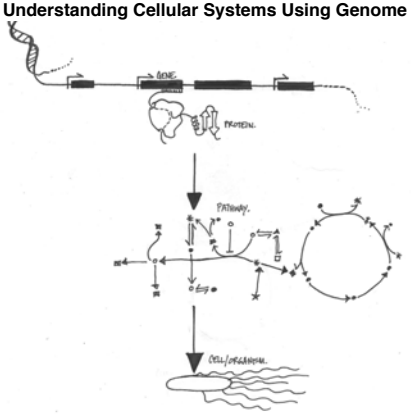
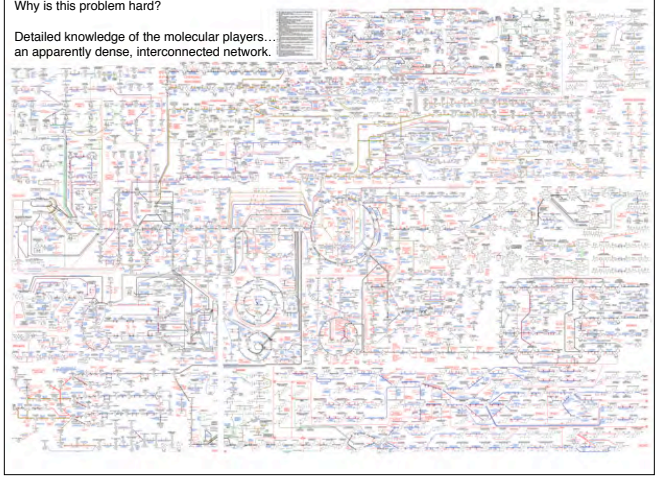
Understanding Cellular Systems Using Genome Data



Kim Reynolds, UT Southwestern, Sept. 2014

Why is this problem hard?

Detailed knowledge of the molecular players…
an apparently dense, interconnected network.



Why is this problem hard?

Detailed knowledge of the molecular players…
an apparently dense, interconnected network.

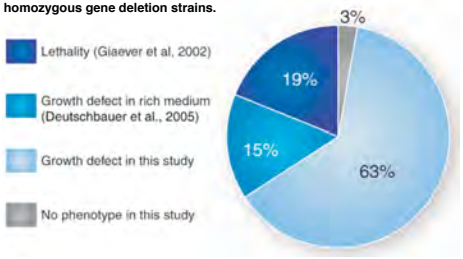
A **complex** biological system.

↓

what do we mean by complex?

First of all, not all genes are equally important.

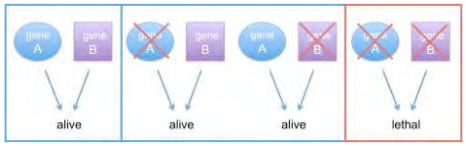
gene essentiality in yeast for ~5000 homozygous gene deletion strains.



Hillenmeyer, ME et al. (2008) Science 320, p 5874

Secondly, the effect of multiple genes can't be predicted from the effect of genes taken individually (things don't add)

An example - synthetic lethality:



Why is this problem hard?

Detailed knowledge of the molecular players…

an apparently dense, interconnected network.

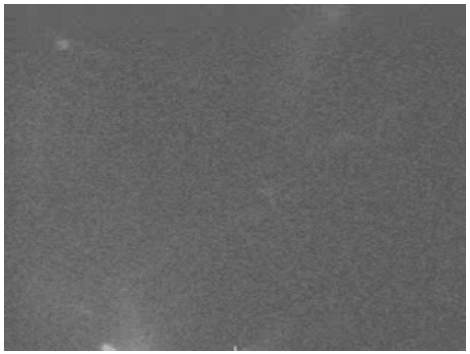
A **complex** biological system.

These two features:
(1) heterogeneity in functional importance (some things more important than others)
(2) non-additive (cooperative) interactions between components

Make it difficult to predict systems-level behavior from the individual parts.

We need a way to systematically and quantitatively measure these two properties!

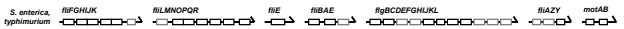
As an example, let's consider the bacterial flagellum.



Chilcott et al. (2000) Micro and Mol Bio Reviews 64 p.694
Fujii, T et al. (2009) Structure 17, p. 1425
Berg, HC (2003) Ann Rev Biochem 72, p 19

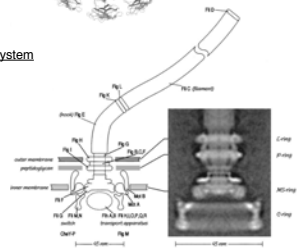
As an example, let's consider the bacterial flagellum.

Genes



Proteins

Cellular System



A multi-protein complex often described by analogy to man-made machines with words like "motor", "stator", "rotor", etc...

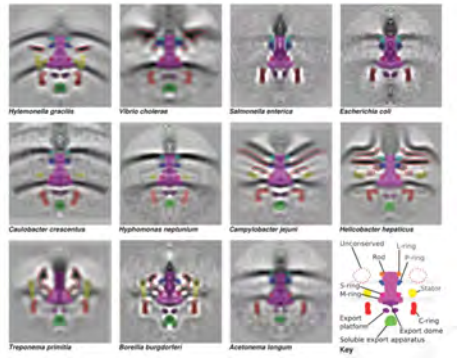
this suggests a precise arrangement of components necessary for function.

A challenge to understand how such systems are encoded by the genome, work and evolve.

Chilcott et al. (2000) Micro and Mol Bio Reviews 64 p.694
Fujii, T et al. (2009) Structure 17, p. 1425
Berg, HC (2003) Ann Rev Biochem 72, p 19

Evolution has arrived at a degeneracy of solutions to the problem of bacterial motility!

Visually comparing these, we see that a “core” motor element is conserved, and the peripheral elements are more variable. This hints that a simpler representation of such systems may be possible.



Chen et al. (2011) EMBO J 30 p. 2972

The central idea:
Comparison of genomes across many species can be used to make a ***statistical model*** for the design of biological systems.

1) Invariance over genomes as a measure of relevance - ***conservation***

2) Correlation over genomes as a measure of cooperative function - ***coevolution***

What do I mean here? Let’s look at an example….

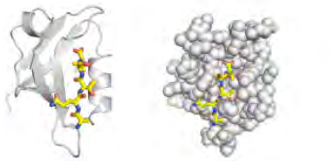
Here is one protein:

Amino acid sequence (primary structure):

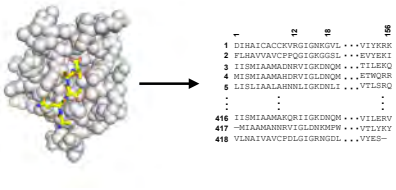
N-GEEDIPREPPRRIVIHRGSTGLGFNIVGGEDGEGIFISFILAG-GPADLSGELRKGDQILSVNGVDLRNASHEQAAIALKNAGQTVTIIAQYKPEE-C

Secondary structure:

Three dimensional structure:



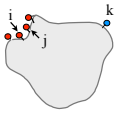
We can collect many amino acid sequences that encode this protein in different species, and align them to each other - a ***multiple sequence alignment***

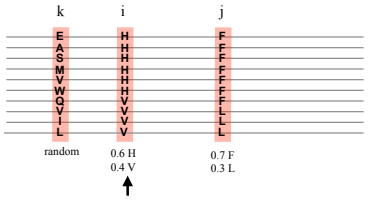


```
  1          12      18          166
1 DIHAICACCKVRGIGNKGVL···VIYKRK
2 FLHAVVAVCPFQGIGKGGSL···EVYEKI
3 IISMIAAMADNRVIGKDNQM···TILEKQ
4 MISMIAAMAHDRVIGLDNQM···ETWQRR
5 LISLIAALAHNNLIGKDNLI···VTLSRQ
  :        :                 :
416 IISMIAAMAKQRIIGKDNQM···VILERV
417 -MIAAMANNRVIGLDNKMPH···VTLYKY
418 VLNAIVAVCPDLGIGRNGDL···VYES-
```

Now we'd like to analyze this alignment to measure two things:
(1) which amino acid positions are most important
(2) and which interact/are cooperatively coupled

Conservation as a measure of functional importance



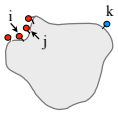


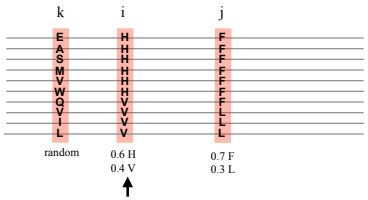
What we want to measure:

Are the amino acid frequencies at a particular position in the alignment more conserved than random?

Lockless and Ranganathan, Science 286, p.295
O. Rivoire, S. Leibler, and Ranganathan, in preparation.
N. Halabi, O.Rivoire, S. Leibler, and R. Ranganathan, Cell (2009) 138: 774-86.

Conservation as a measure of functional importance





What we want to measure:

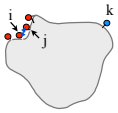
Are the amino acid frequencies at a particular position in the alignment more conserved than random?

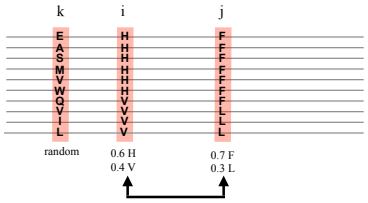
How we calculate this:

$$\tilde{C}_i^a = D(f_i^a \,\|\, q^a)$$
$$= f_i^a \ln \frac{f_i^a}{q^a} + (1 - f_i^a) \ln \frac{1 - f_i^a}{1 - q^a}.$$

Lockless and Ranganathan, Science 286, p.295
O. Rivoire, S. Leibler, and Ranganathan, in preparation.
N. Halabi, O.Rivoire, S. Leibler, and R. Ranganathan, Cell (2009) 138: 774-86.

Coevolution as a measure of interaction between two positions.

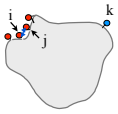


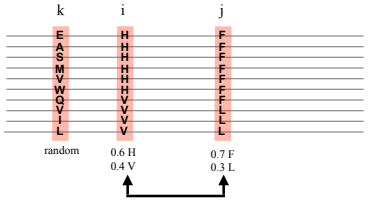


The basic premise: Functional Coupling of two amino acid positions will result in co-evolution... provided that the interaction contributes to the fitness of the protein.

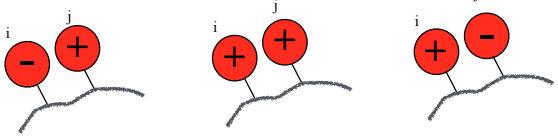
Lockless and Ranganathan, Science 286, p.295
O. Rivoire, S. Leibler, and Ranganathan, in preparation.
N. Halabi, O.Rivoire, S. Leibler, and R. Ranganathan, Cell (2009) 138: 774-86.

Coevolution as a measure of interaction between two positions.

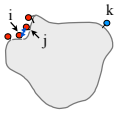


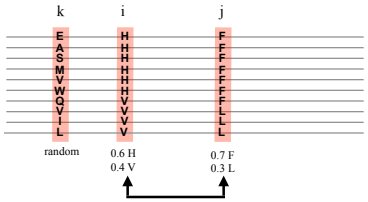


An example:



Coevolution as a measure of interaction between two positions.

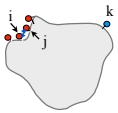


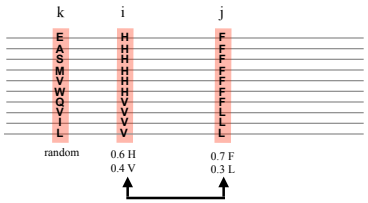


What we want to measure:

How independent are the amino acid frequencies at sites **i** and **j**?

Coevolution as a measure of interaction between two positions.





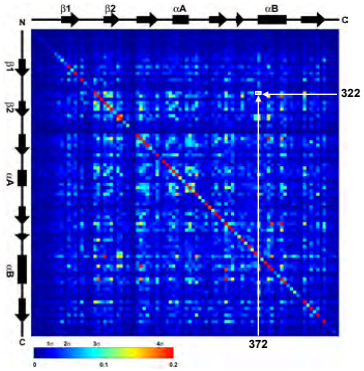
What we want to measure:

How independent are the amino acid frequencies at sites **i** and **j**?

How we calculate this:

$$\tilde{C}_{ij}^{(ab)} = \left|\frac{\partial D_i^{(a)}}{\partial f_i^{(a)}}\right| \left|\frac{\partial D_j^{(b)}}{\partial f_j^{(b)}}\right| \left[f_{ij}^{(ab)} - f_i^{(a)} f_j^{(b)} \right]$$

The result of these two calculations (conservation and co-evolution): Statistical Coupling Matrix

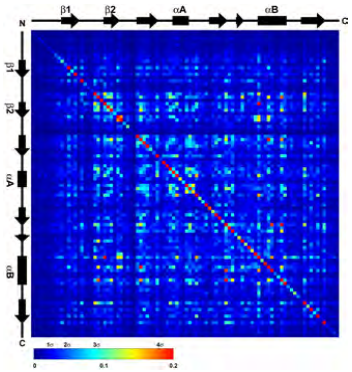


A global view of coupling between amino acids, based on evolution.

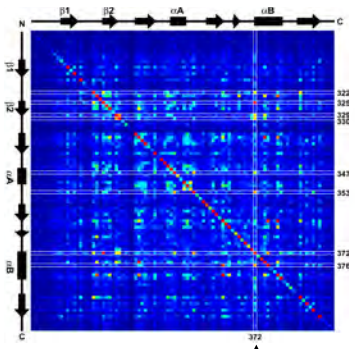
Each pixel is the coupling between a pair of amino acids.

The matrix is symmetric.

The result of these two calculations (conservation and co-evolution): Statistical Coupling Matrix



From initial inspection we can see that the matrix is:
(1) sparse
(2) shows no obvious arrangement in primary structure.

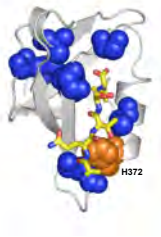


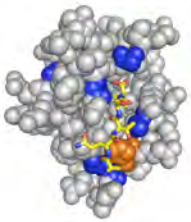
What does the pattern of couplings look like on the structure?
Let's take one position to start...

Lockless and Ranganathan (1999) Science 286, p.295

The global architecture of correlations...

For a single position, we can examine all other coupled residues...

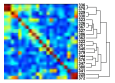


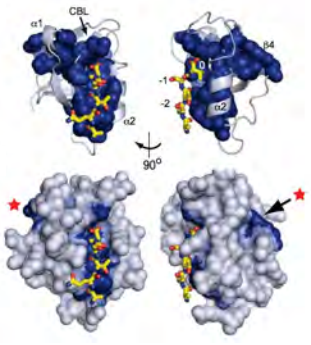


A map of interactions between H372 and the remainder of the PDZ domain.

Lockless and Ranganathan (1999) Science 286, p.295

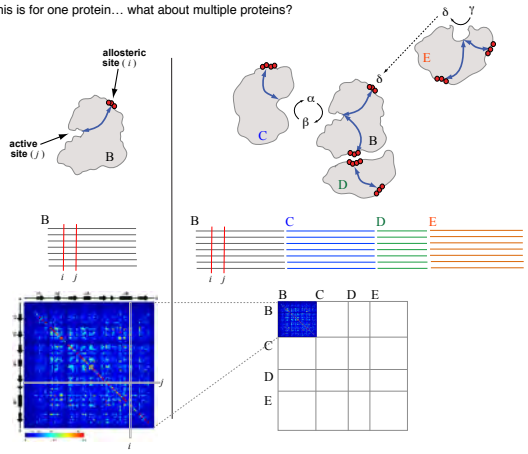
Mapping all the coupled positions to the structure, we see that they form a physically contiguous network… the ***protein sector.***





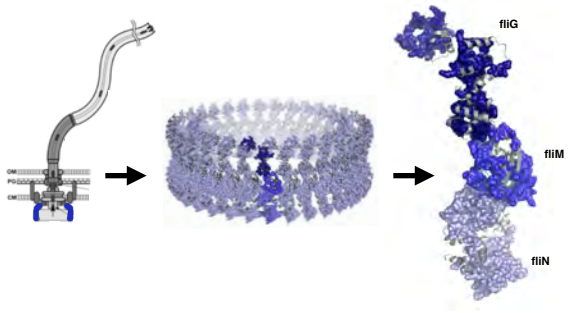
N. Halabi, O.Rivoire, S. Leibler, and R. Ranganathan, Cell (2009) 138: 774-86.

So this is for one protein… what about multiple proteins?



These co-evolving units seem to be “wired up” in larger cellular systems:

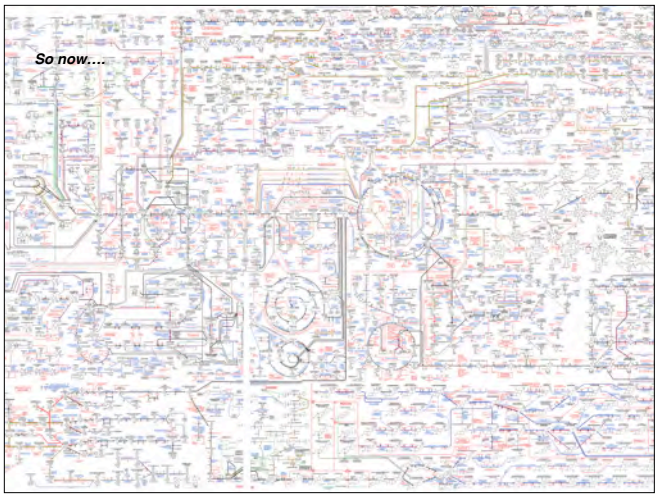
Flagellar Motor: **Torque-generation ring:** **Sector mapping:**



Minamino et al. (2008). Curr. Op. Struct Biol., 18, 693-701
Vartanian et al. (2012). JBC, 287, 35779-35783

Neal Sharma

So now....



Prior results at the level of single proteins and the current availability of complete genome data now motivate the application of this strategy genome wide.

The outcome (*if successful*):

A global decomposition of the genome into new cooperative units

A basis for:

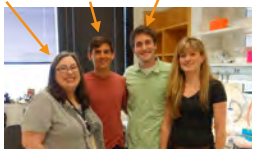
Rational strategy for control and design of cellular behavior

Context for interpreting disease-causing mutations and drug interactions

Better understanding of the design of cellular systems and how they might evolve.

Acknowledgements

Chris Ingle Neal Sharma Andrew Schober



http://systems.swmed.edu/krlab/Reynolds_Lab.html

Collaborators

Olivier Rivoire (CNRS, Grenoble)
Ivan Junier (CRG, Barcelona)

Funds

UT Southwestern Green
Center for Systems Biology