Amino acid sequence constraints and co-evolution across a metabolic enzyme pair

Thuy N. Nguyen[1], Saumel M. Thompson[2], and Kimberly A. Reynolds[1,3]

***Short Abstract* — Several studies which have investigated the pattern of amino acid interactions between proteins in physical complexes show that a subset of contacting interfacial residues are thermodynamically coupled and co-evolve. However, it is less clear how the amino acid sequence of one protein is constrained by another protein to which it is functionally coupled, but does not directly bind. Here I describe a strategy and initial data for experimentally quantifying the sequence constraints imposed by functional coupling between metabolic enzymes. These measurements are then compared to sequence variation across homologs.**

***Keywords* — epistasis, co-evolution, statistical coupling analysis, fitness, saturation mutagenesis, next-generation sequencing**

I. Epistasis beyond physical interfaces

The amino acid sequence of an enzyme encodes information necessary for it to fold, catalyze biochemical reactions, and interact with partners in the cell. Maintaining these properties over evolutionary time constrains sequence variation, and can drive co-evolution between functionally coupled (or epistatic) positions. For example, positions at the interface of two-component signal transduction systems are thermodynamically coupled and co-evolve [1,2]. However, epistasis between proteins is not limited to physical complexes. In the general case of functionally coupled, but not necessarily physically interacting proteins, the pattern of epistasis between amino acids is poorly characterized. I am using: (1) experimental deep mutational scanning and (2) statistical analysis of sequence variation to map the constraints introduced by coupling in metabolism. Together, these data will provide a comprehensive and quantitative illustration of how epistasis between proteins shapes amino acid sequence variation.

II. The model system

To study the structural basis of epistasis, I am using the model system of dihydrofolate reductase (DHFR) and thymidylate synthase (TYMS), two enzymes in the folate metabolic pathway. Statistical analysis of gene presence/absence and synteny across 1445 bacterial genomes show that they are highly co-evolving with each other but are independent from the rest of the genome [3]. Though the two enzymes do not physically interact, recent experimental measurements of metabolomics and epistasis indicate that the relative biochemical activities of DHFR and TYMS are constrained by a need to limit accumulation of the metabolic intermediate dihydrofolate (DHF) [3]. In particular, mutations in DHFR that reduce catalytic activity are buffered by loss-of-function mutations in TYMS.

III. Quantitative measurements of epistasis

To determine how sequence variation in DHFR is modulated by the catalytic activity of TYMS, I am conducting fitness measurements for a saturation mutagenesis library of DHFR in the context of several TYMS mutants. The library contains all possible single point mutations at every position of DHFR (of 5,088 total). To assay fitness, I transform the library into *E. coli* and grow this mixed culture in continuous culture. I take samples of this culture during the experiment and use next-generation sequencing to track allelic frequencies of each mutant. The fitness of each mutant in the library is then calculated from a linear fit of the allelic frequency over time [3]. These data will reveal a fitness landscape across the entire structure of DHFR. I am currently repeating the assay in the background of five TYMS that range in catalytic activity from WT to catalytically dead. For each TYMS mutant, I will calculate epistasis from these fitness measurements and map these epistatic couplings onto the structure of DHFR. At the end, these data will reveal how the fitness landscape of DHFR is modulated by TYMS activity.

IV. Comparing Epistasis to co-evolution

I am comparing the pattern of epistasis from the experiments described above to the pattern of co-evolution between DHFR and TYMS homologs. To compute co-evolution, I am using the Statistical Coupling Analysis [4], which has recently been shown to reasonably estimate pair-wise couplings between positions in a single protein [5]. This comparison of experimental and statistical couplings will test the hypothesis that co-evolution can be used to infer sequence constraints linking non-binding but functionally coupled proteins.

Acknowledgements: Funding by Gordon and Betty Moore Foundation Data-Driven Discovery grant: GBMF4557.

[1]Green Center for Systems Biology, University of Texas Southwestern Medical Center, Dallas, TX, E-mail: thuy.nguyen@utsouthwestern.edu
[2]Graduate Program in Biophysics, University of California, San Francisco, E-mail: samuel.thompson@ucsf.edu
[3]Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, E-mail: kimberly.reynolds@utsouthwestern.edu

References

[1] Skerker, J.M., et al. (2008) "Rewiring the specificity of two-component signal transduction systems." *Cell* 133.6: 1043-1054.
[2] Podgornaia, A. I., & Laub, M. T. (2015). Pervasive degeneracy and epistasis in a protein-protein interface. *Science*, 347(6222), 673-677.
[3] Schober, Andrew F., et al. (2017) "An evolutionary module in central metabolism." *bioRxiv*: 120006.
[4] Rivoire, O., Reynolds, K.A., & Ranganathan,R. (2016). Evolution-based functional decomposition of proteins. *PLoS computational biology*, 12(6), e1004817.
[5] Salinas, V. H., & Ranganathan, R. (2017). Inferring amino acid interactions underlying protein function. *bioRxiv*, 215368.