Protein Folding and Structure Prediction

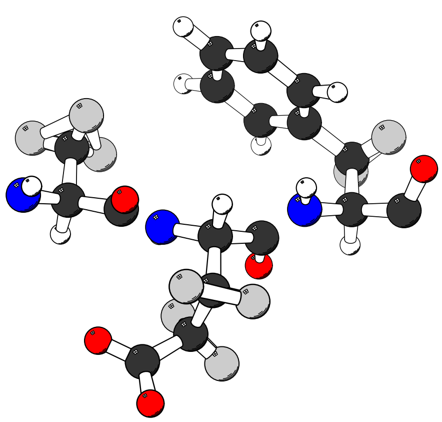
A Statistician's View

Ingo Ruczinski

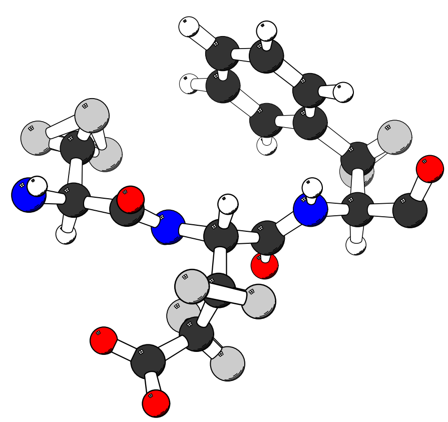
Department of Biostatistics, Johns Hopkins University

Proteins

Amino acids without peptide bonds.

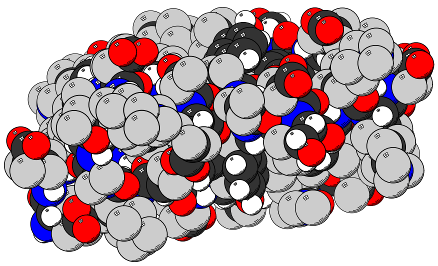


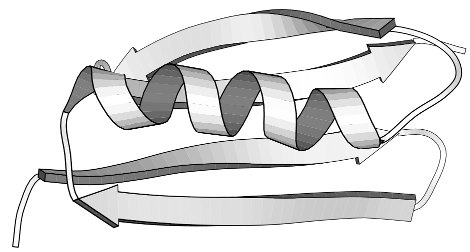
Amino acids with peptide bonds.



⟶ Amino acids are the building blocks of proteins.

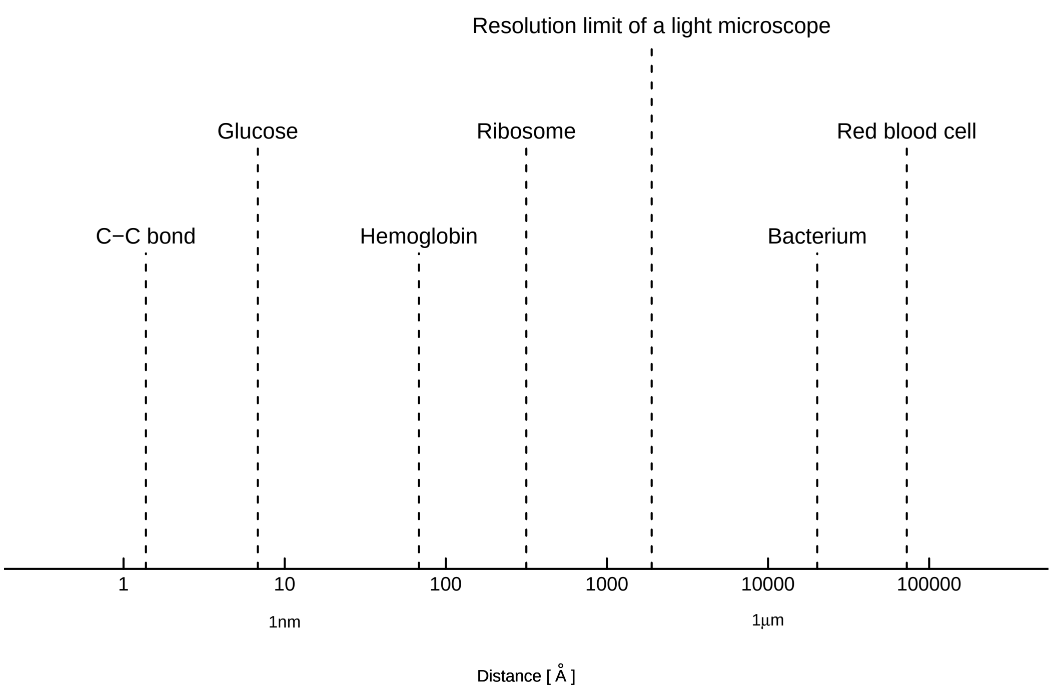
Proteins



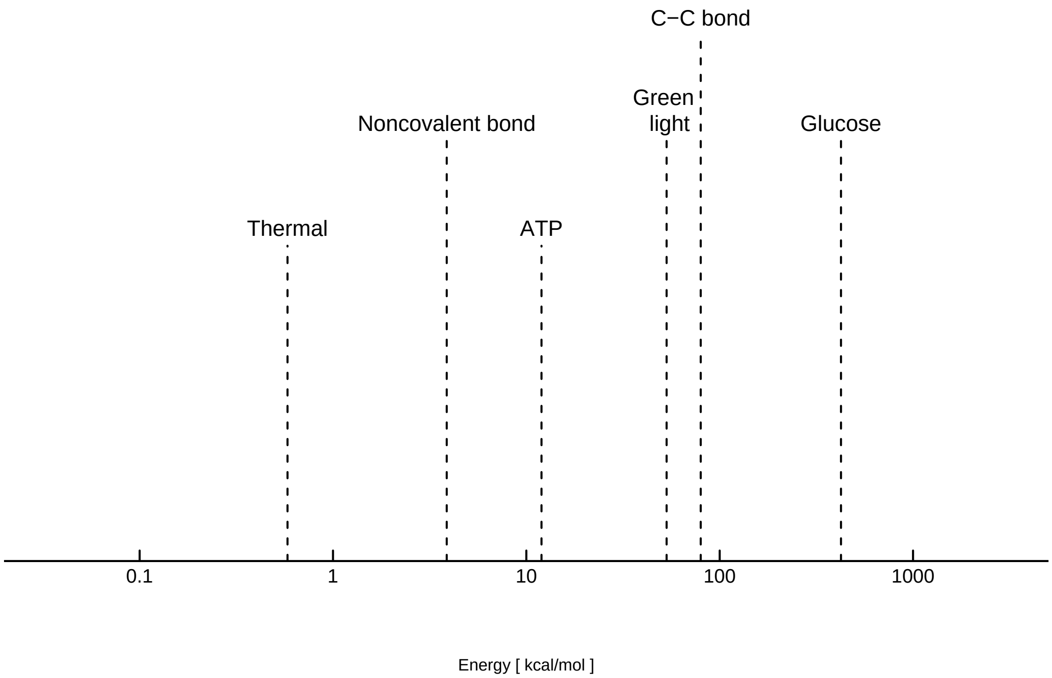


Both figures show the same protein (the bacterial protein L). The right figure also highlights the secondary structure elements.

Space



Energy



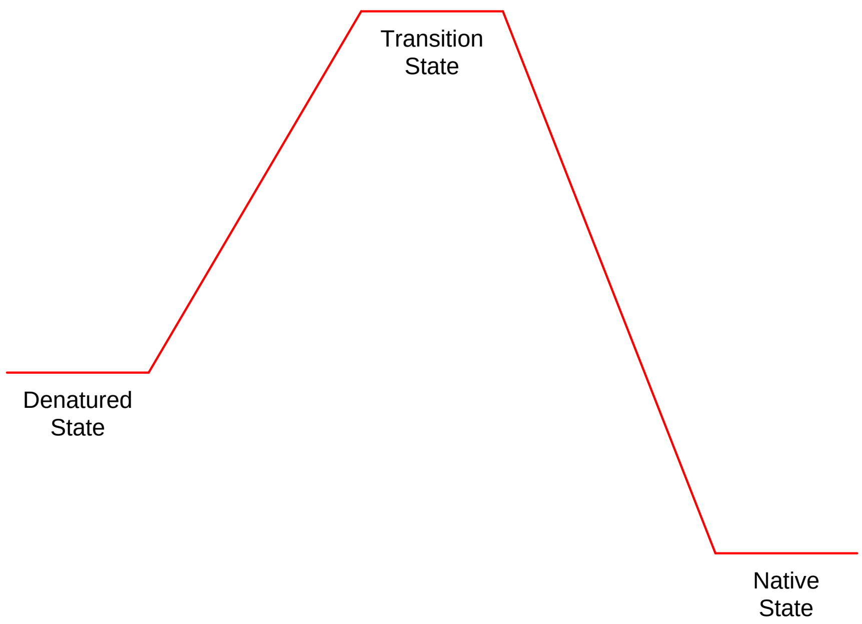
Non-Bonding Interactions

Amino acids of a protein are joined by covalent bonding interactions. The polypeptide is folded in three dimension by non-bonding interactions. These interactions, which can easily be disrupted by extreme pH, temperature, pressure, and denaturants, are:

- Electrostatic Interactions (5 kcal/mol)
- Hydrogen-bond Interactions (3-7 kcal/mol)
- Van Der Waals Interactions (1 kcal/mol)
- Hydrophobic Interactions ($<$ 10 kcal/mol)

The total inter-atomic force acting between two atoms is the sum of all the forces they exert on each other.

Energy Profile



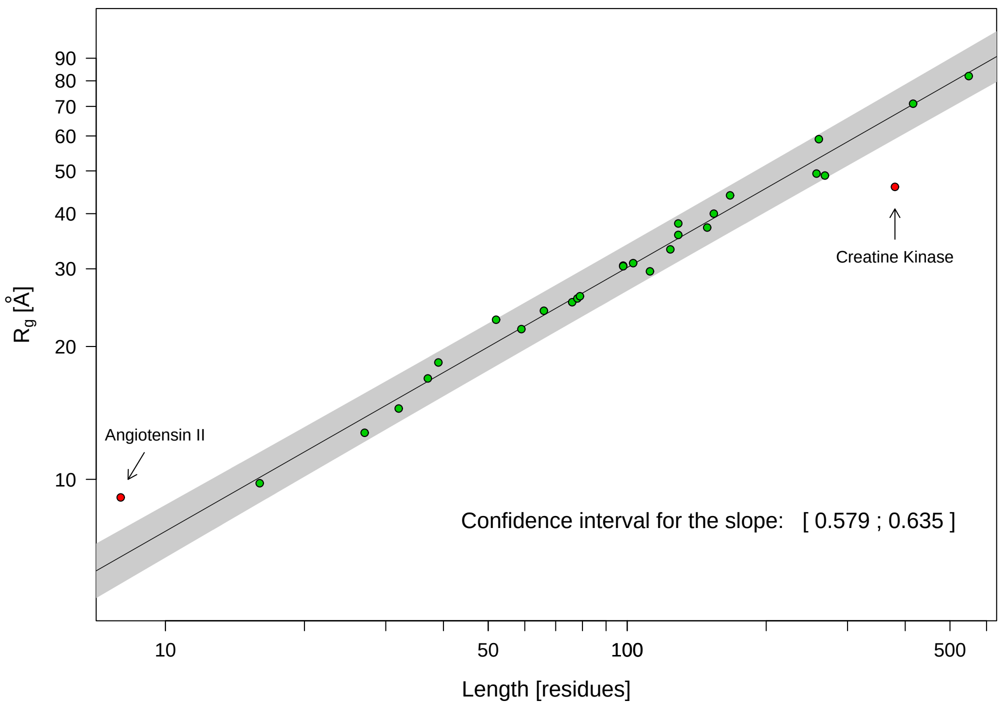
Radius of Gyration of Denatured Proteins

Do chemically denatured proteins behave as random coils?

- The radius of gyration R_g of a protein is defined as the root mean square distance from each atom of the protein to their centroid.
- For an ideal (infinitely thin) random-coil chain in a solvent, the average radius of gyration of a random coil is a simple function of its length n: $R_g \propto n^{0.5}$.
- For an excluded volume polymer (a polymer with non-zero thickness and non-trivial interactions between monomers) in a solvent, the average radius of gyration, we have $R_g \propto n^{0.588}$ (Flory 1953).

$\longrightarrow$ The radius of gyration can be measured using small angle x-ray scattering.

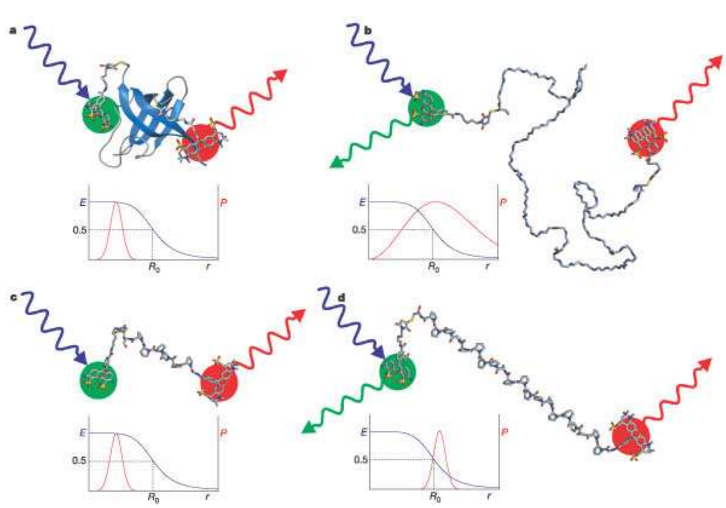
Radius of Gyration of Denatured Proteins



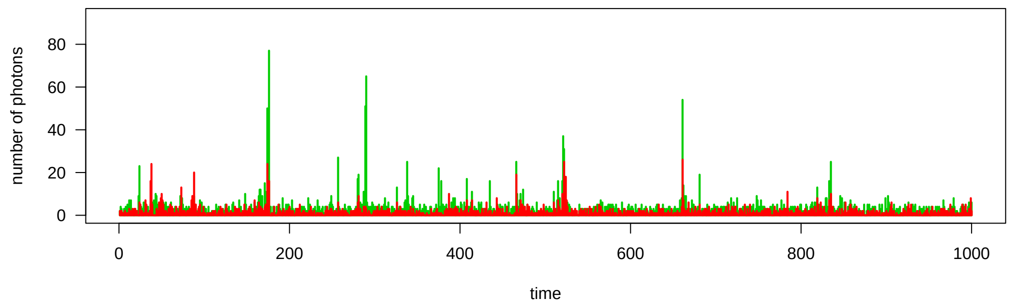
Deviations from Random Coil Behaviour

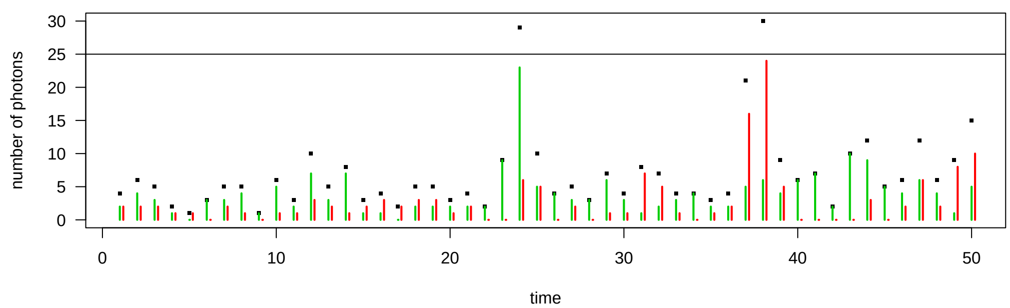
Are there site-specific deviations from random coil dimensions?

Förster Resonance Energy Transfer enables us to measure the distance between two dye molecules within a certain range. This can be used to study site-specific deviations from random coil dimensions in highly denatured peptides.

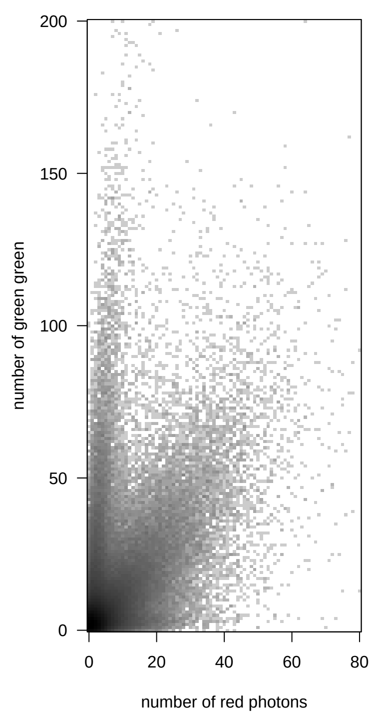


Deviations from Random Coil Behaviour





Deviations from Random Coil Behaviour



We have two underlying distributions for the green and red photons:

- One stemming from a peptide only having a donor dye.
- One stemming from a peptide being properly tagged with a donor and an acceptor dye.

Assume a photon has probability p_0 of being red in the former situation, and p_1 in the latter.

Deviations from Random Coil Behaviour

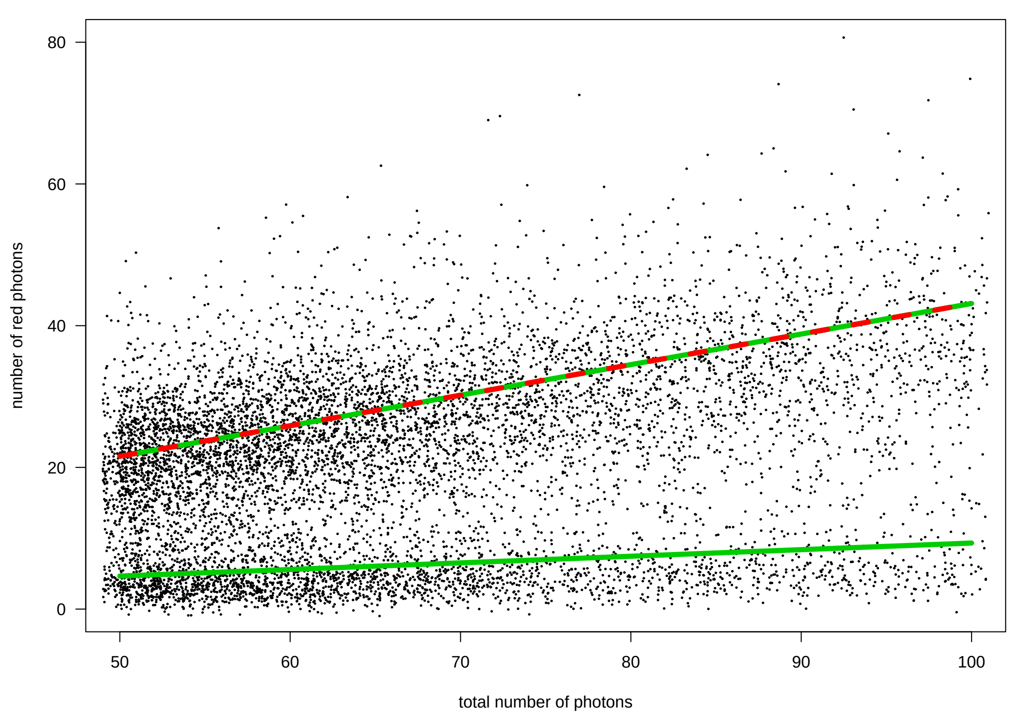
Assume we observe n_i photons at time point i. Then the number of red photons is simply Bernoulli(n_i, p_i), where p_i is either p_0 or p_1 . Assume that the probability of observing photons from a peptide without an acceptor dye at any time is p, independent of the total number of photons observed. Let X be the number of red photons. Then

$$\begin{aligned} P(X = x_i | n_i) &= P(X = x_i | n_i, p_0) \times p + P(X = x_i | n_i, p_1) \times (1-p) \\ &= \binom{n_i}{x_i} p_0^{x_i} (1-p_0)^{n_i - x_i} \times p + \binom{n_i}{x_i} p_1^{x_i} (1-p_1)^{n_i - x_i} \times (1-p), \end{aligned}$$

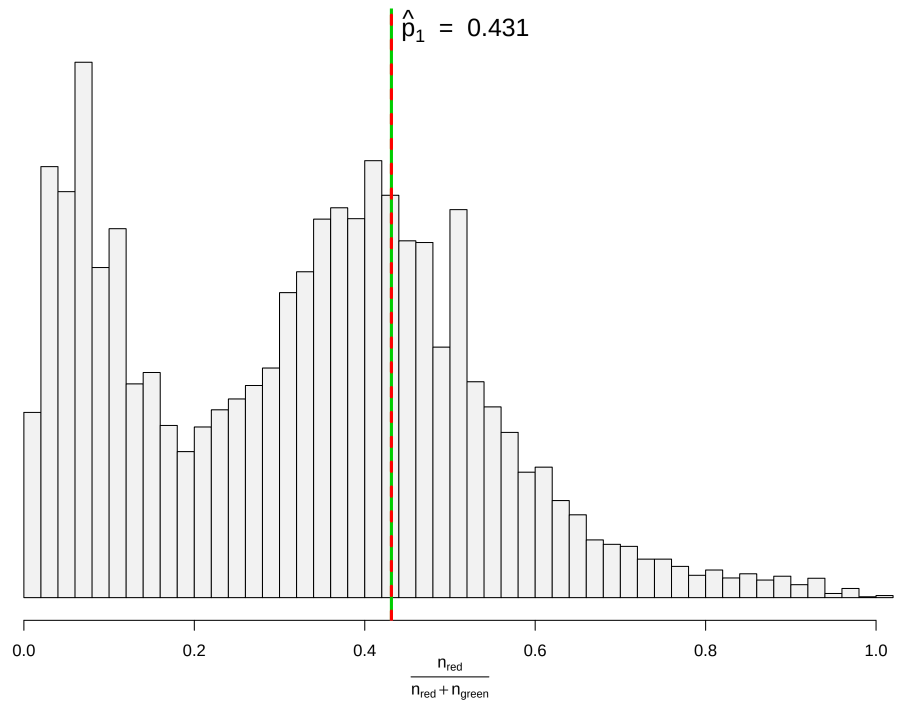
and hence

$$L(p, p_0, p_1) = \prod_{i=1}^{N} \left[\binom{n_i}{x_i} p_0^{x_i} (1-p_0)^{n_i - x_i} \times p + \binom{n_i}{x_i} p_1^{x_i} (1-p_1)^{n_i - x_i} \times (1-p) \right].$$

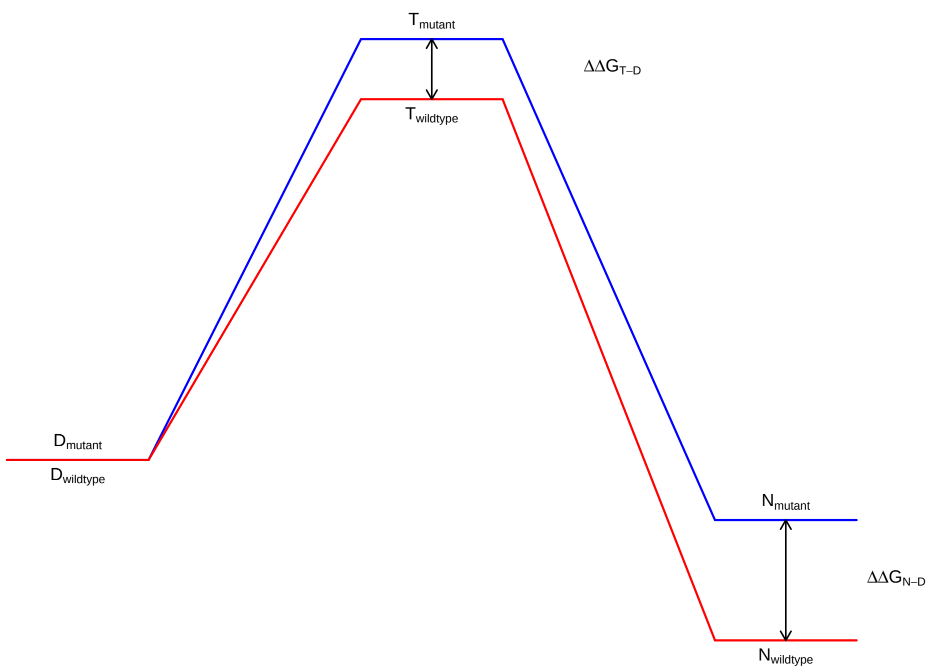
Deviations from Random Coil Behaviour



Deviations from Random Coil Behaviour

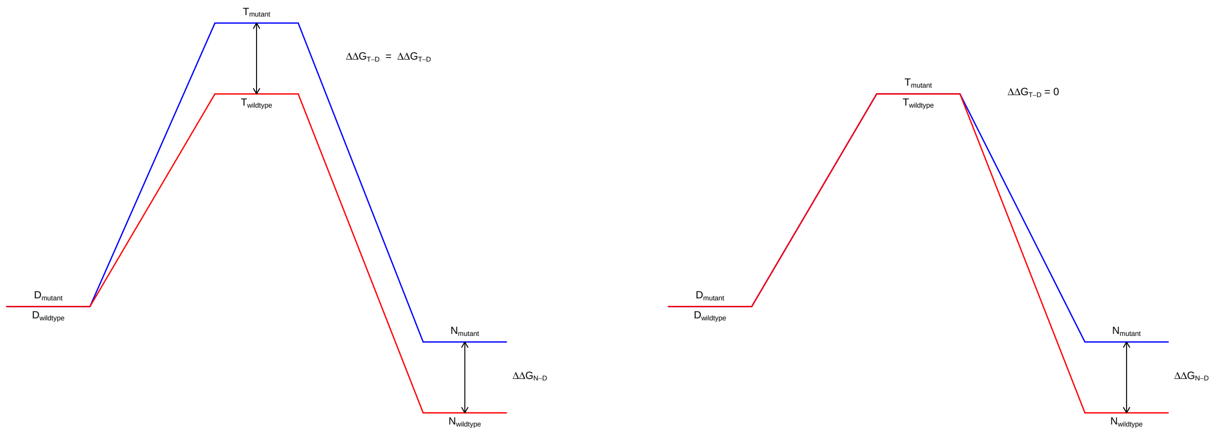


Energy Profile



$\longrightarrow$ The Φ-value is defined as the ratio $\Delta\Delta G_{T\text{-}D}/\Delta\Delta G_{N\text{-}D}$.

Energy Profile

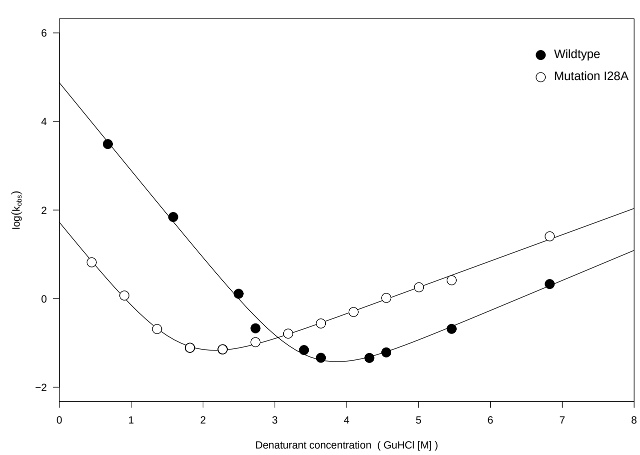


- If the part of the protein that contains the mutant amino acid is fully structured in the transition state, we have $\Delta\Delta G_{T\text{-}D} = \Delta\Delta G_{N\text{-}D}$, and hence $\Phi = 1$.
- If the part of the protein that contains the mutant amino acid is equal in denatured and the transition state, we have $\Delta\Delta G_{T\text{-}D} = 0$, and hence $\Phi = 0$.

Chevron Plots

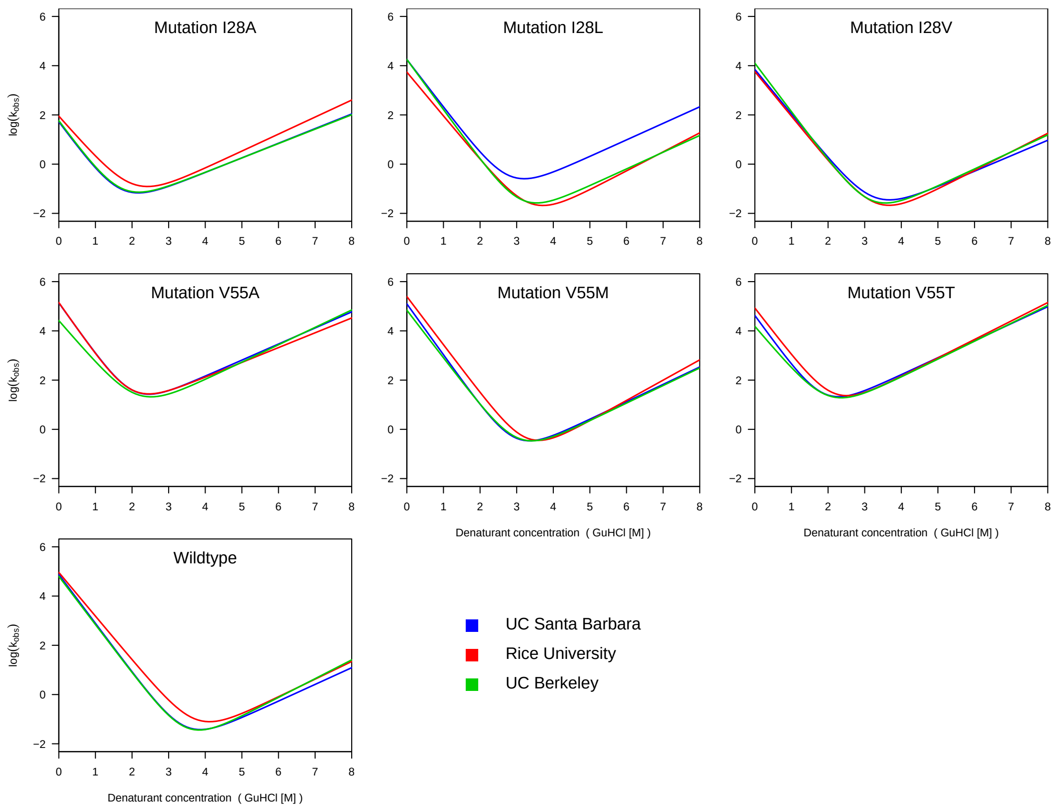
$$\Delta\Delta G_{T\text{-}D} = RT \times \left[\log(k_f^{wildtype}) - \log(k_f^{mutant}) \right]$$

$$\Delta\Delta G_{N\text{-}D} = RT \times \left[\log(k_f^{wildtype}) - \log(k_u^{wildtype}) - \log(k_f^{mutant}) + \log(k_u^{mutant}) \right]$$

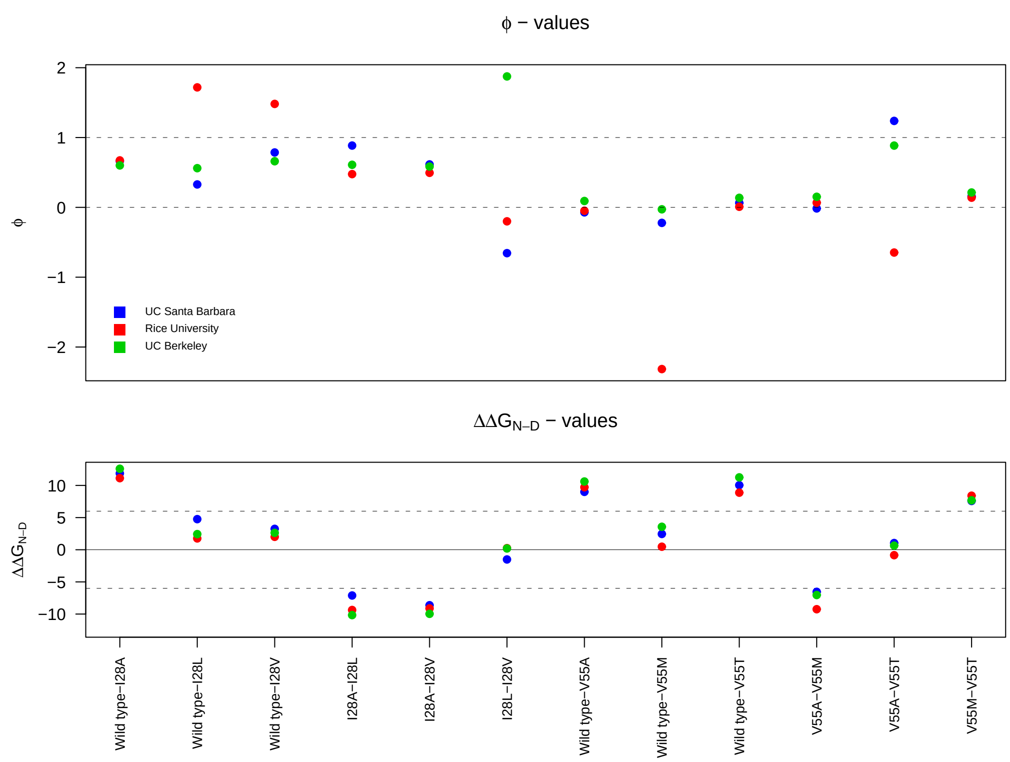


$$\log(k_{obs}) = \log\left(\exp\left[\log(k_f) + m_f \times \frac{C_{GuHCl}}{RT}\right] + \exp\left[\log(k_u) + m_u \times \frac{C_{GuHCl}}{RT}\right] \right)$$

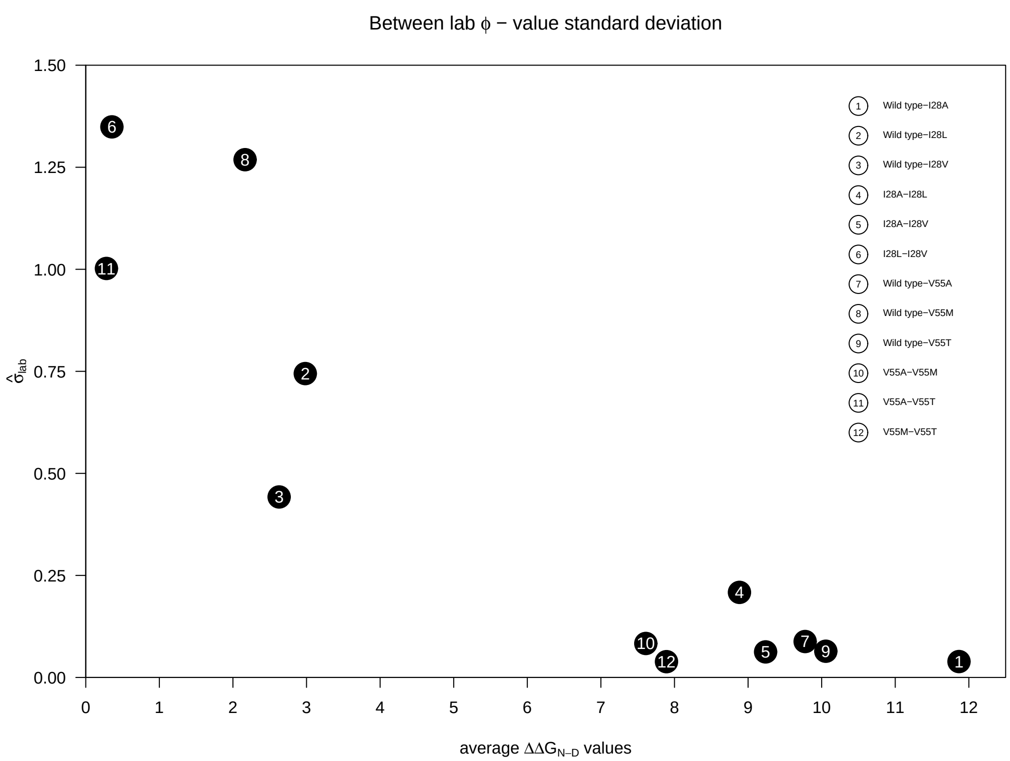
More Chevron Plots



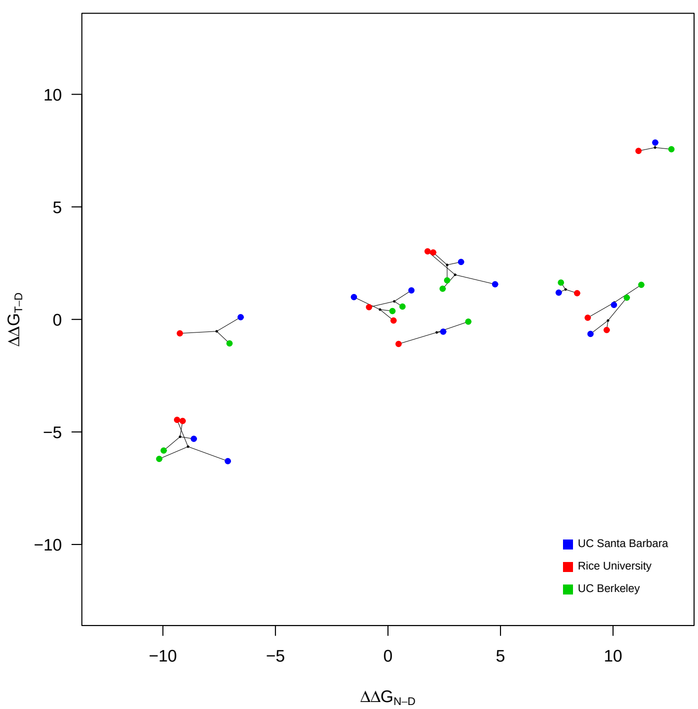
Variability



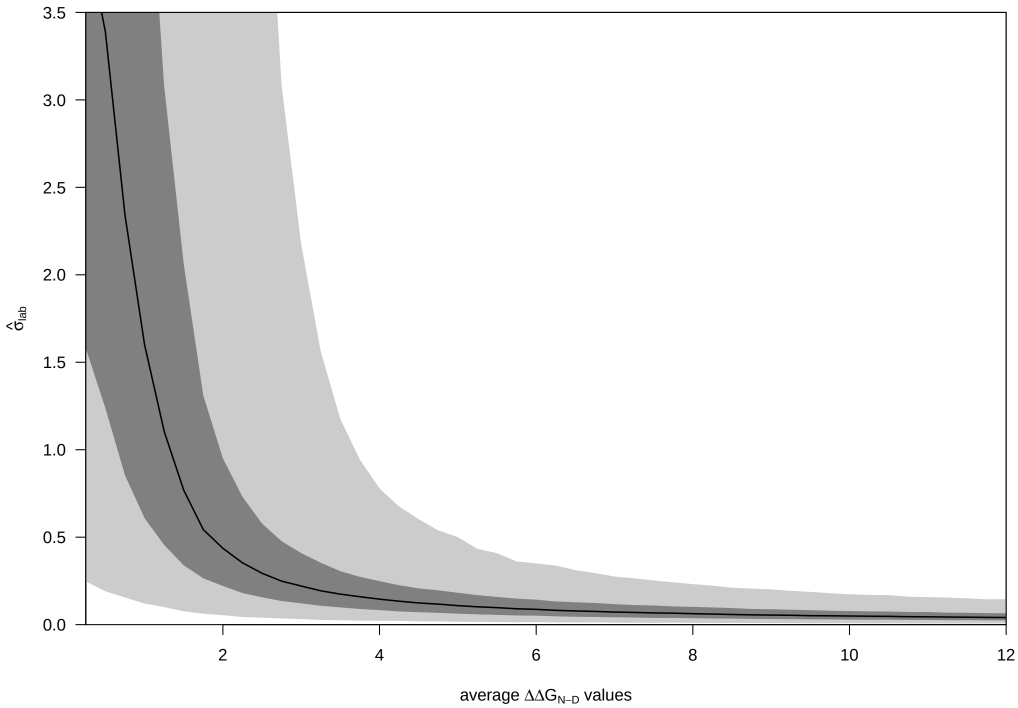
Variability



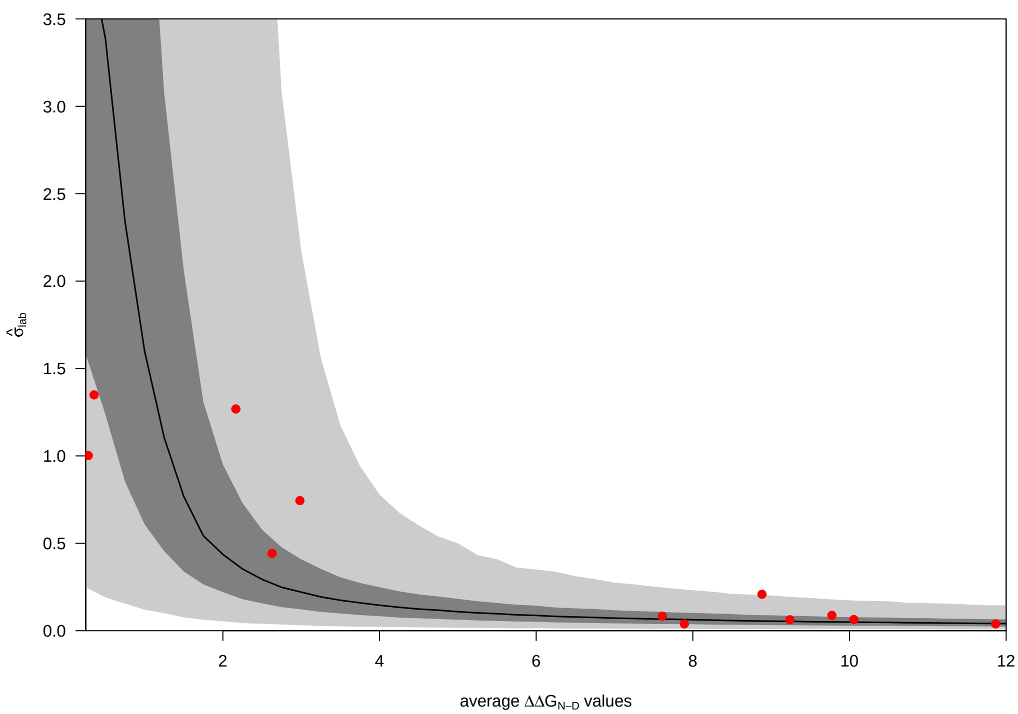
Variability



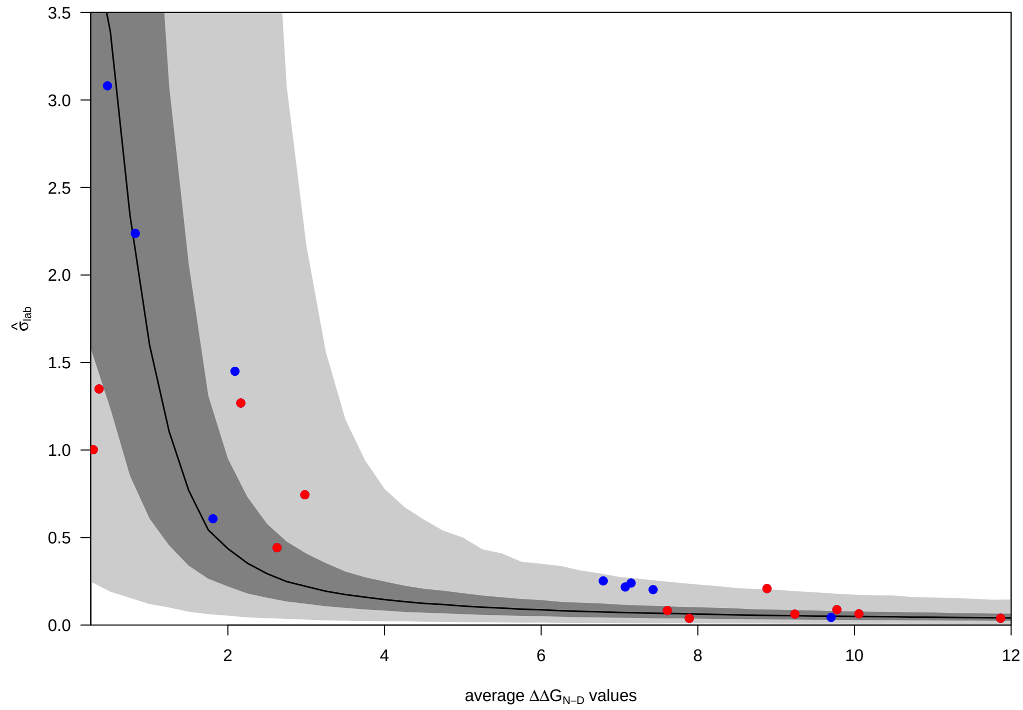
Some Simulation



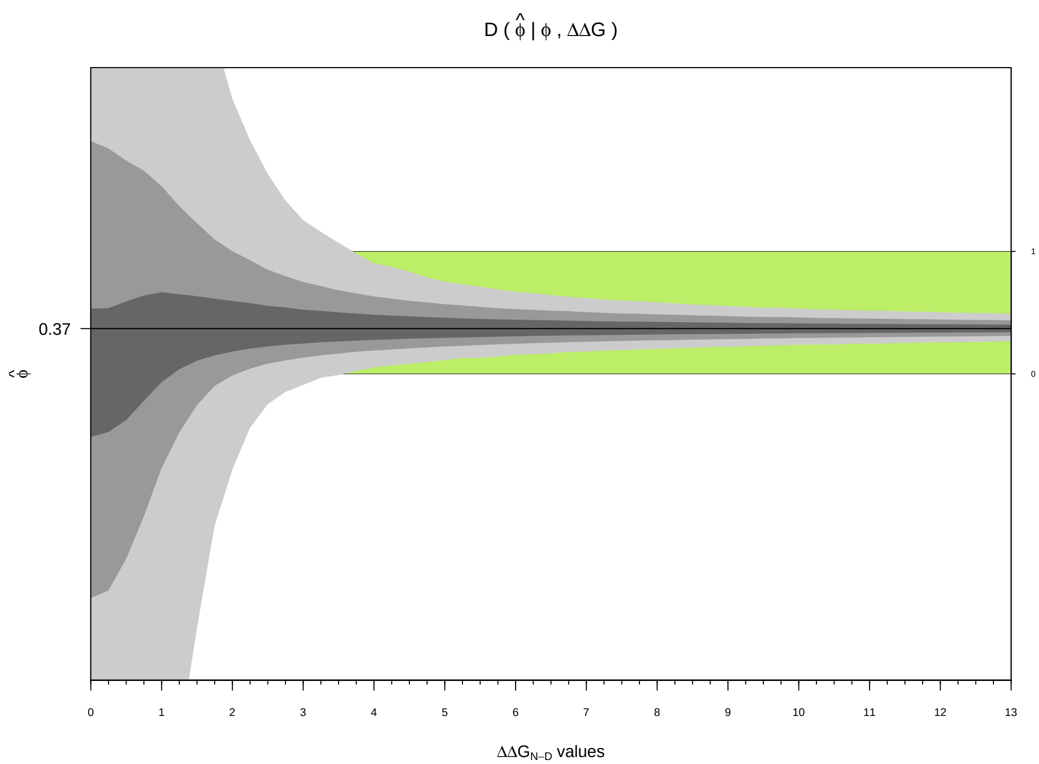
Some Simulation



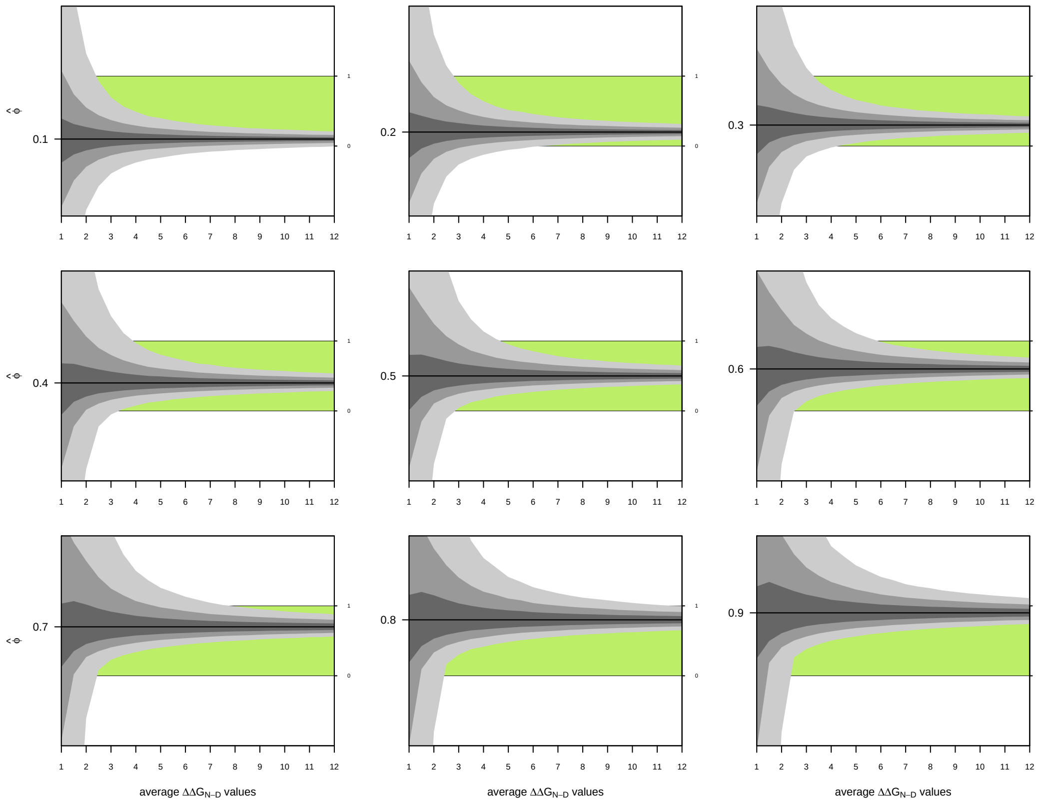
Some Simulation



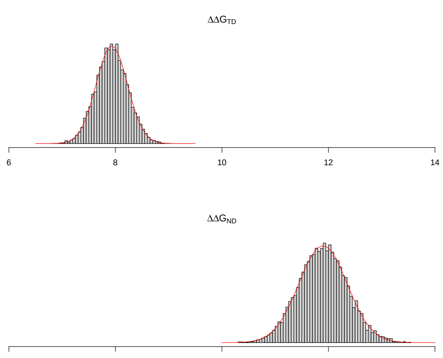
Some More Simulations

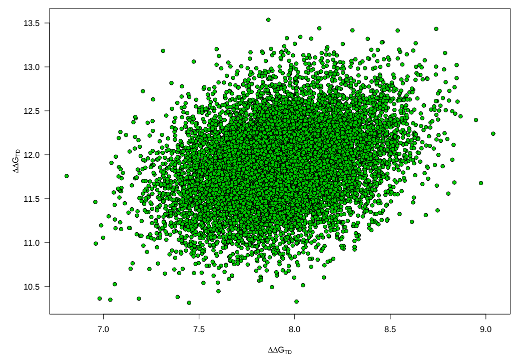


Some More Simulations



Phi-Value Estimation





$$\begin{bmatrix} \widehat{\Delta\Delta G}_{\text{TD}} \\ \widehat{\Delta\Delta G}_{\text{ND}} \end{bmatrix} \sim N\left(\begin{bmatrix} \Delta\Delta G_{\text{TD}} \\ \Delta\Delta G_{\text{ND}} \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_3^2 \\ \sigma_3^2 & \sigma_2^2 \end{bmatrix} \right)$$

with

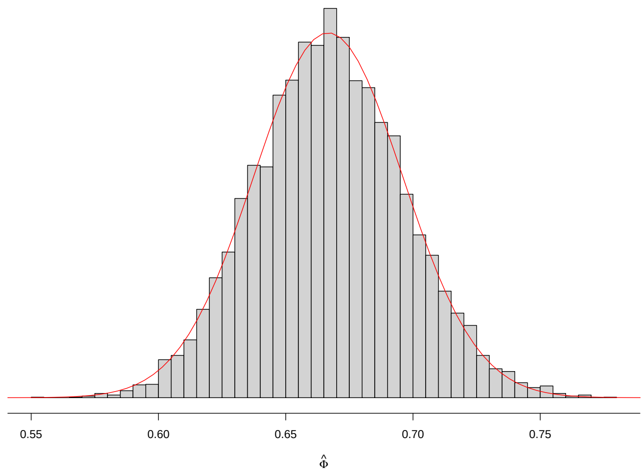
$$\begin{aligned} \sigma_1^2 &= \sigma_{\text{F}_W}^2 + \sigma_{\text{F}_M}^2 \\ \sigma_2^2 &= \sigma_{\text{F}_W}^2 + \sigma_{\text{F}_M}^2 + \sigma_{\text{U}_W}^2 + \sigma_{\text{U}_M}^2 - 2\rho_{\text{W}}\sigma_{\text{F}_W}\sigma_{\text{U}_W} - 2\rho_{\text{M}}\sigma_{\text{F}_M}\sigma_{\text{U}_M} \\ \sigma_3^2 &= \sigma_{\text{F}_W}^2 + \sigma_{\text{F}_M}^2 - \rho_{\text{W}}\sigma_{\text{F}_W}\sigma_{\text{U}_W} - \rho_{\text{M}}\sigma_{\text{F}_M}\sigma_{\text{U}_M} \end{aligned}$$

Phi-Value Estimation

$$\widehat{\Phi} = \frac{\widehat{\Delta\Delta G}_{\text{TD}}}{\widehat{\Delta\Delta G}_{\text{ND}}} \approx N(\Phi, B)$$

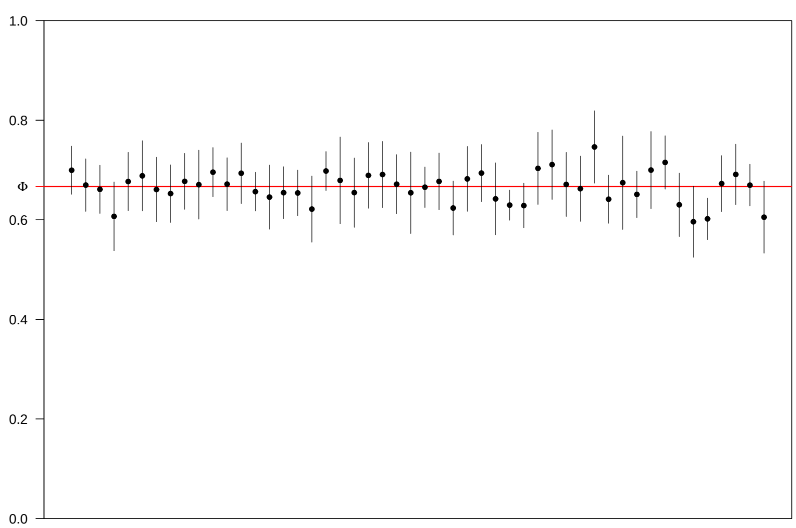
with

$$B = \frac{1}{(\widehat{\Delta\Delta G}_{\text{ND}})^4}(\sigma_1^2(\widehat{\Delta\Delta G}_{\text{ND}})^2 - 2\sigma_3^2\widehat{\Delta\Delta G}_{\text{TD}}\widehat{\Delta\Delta G}_{\text{ND}} + \sigma_2^2(\widehat{\Delta\Delta G}_{\text{TD}})^2).$$



Phi-Value Estimation

$$I = \left[\hat{\Phi} - t_{n_1+n_2-10}^{0.975} \times \sqrt{B}\ ;\ \hat{\Phi} + t_{n_1+n_2-10}^{0.975} \times \sqrt{B}\right] \quad ?$$



Evolution and Folding Kinetics

Are amino acids in proteins conserved because of folding kinetics?

To what extent does natural selection act to optimize the details of protein folding kinetics? Is there a relationship between an amino acid's evolutionary conservation and its role in protein folding kinetics?

Some comments:

- Our studies of sequence conservation among residues known to participate in the folding nuclei of all of the appropriately characterized proteins reported to date have not provided any evidence that highly conserved residues are more likely to participate in the protein folding nucleus than poorly conserved residues.
- This is in contrasts to some of the beliefs stemming from theoretical considerations (good science, good people).
- This is also in contrast to the conclusions certain people drew from experimental data (really aweful statistics).
- These people do not like us.