PNAS
www.pnas.org

Supplementary Information for

Selection for cooperativity causes epistasis predominately between native contacts and enables epistasis based structure reconstruction

R. Charlotte Eccleston, David D. Pollock and Richard A. Goldstein

Richard A. Goldstein.
E-mail: r.goldstein@ucl.ac.uk

This PDF file includes:

Supporting Information Text

A random coil model for the generation intermediate structures. We used a random coil model (1, 2) to generate a large number of structures for each of the partially folded and unfolded states. For a fully unfolded structure, a point is generated at random with respect to the world origin (0,0,0) and the world coordinate system $\widehat{x_0} = [1, 0, 1]$, $\widehat{y_0} = [0, 1, 0]$ and $\widehat{z_0} = [0, 0, 1]$. A point is then chosen at random from within the sphere of radius r, where r=7Å, by chosing a value between $3 - 7$Å to define the bond length l_1 from the world origin to the location of the first beta carbon atom $C_{\beta,1}$, where the random distribution is biased outwards to account for the volume effects of the sphere. A value for the azimuthal angle ϕ_1 is chosen from the uniform distribution between 0 and 2π. This angle is defined as the angle between positive y-axis and the projection of the bond vector l_1 in the y-z plane. A value for the polar angle θ_1, defined as the angle from the positive x-axis and the bond vector l_1, is chosen at random between 0 and π from a distribution tapering at the poles.

These three values (l_1, θ_1, ϕ_1) define the point p_1, which is the location of the first beta carbon atom $C_{\beta,1}$ of the random protein structure. To define the location of the second beta carbon atom $C_{\beta,2}$, the unit vector in the direction of the vector l_1 is set to be the unit vector in the direction of the x-axis ($\widehat{x_1}$) of a rotated coordinate system S_1. The y-axis of this new coordinate system is found by taking the cross product of $\widehat{x_1}$ and a random vector in the plane $\widehat{x_0} - \widehat{x_1}$, conditional on the angle between this new y-axis and $\widehat{x_0}$ being acute. The vector y_1 can then be normalised to find the unit vector in the direction of y_1, $\widehat{y_1}$. The unit vector in the direction of the z-axis of this new coordinate system can be found by normalising the vector obtained as a result of the cross-product of $\widehat{x_1}$ and $\widehat{y_1}$. In the same way as before, the values of r_2, θ_2 and ϕ_2 are found at random to define the location p_2 of $C_{\beta,2}$ with respect to the new coordinate system S_1. To get the coordinates of this point in the world coordinate system we can use the rotation matrix $\mathbf{R_{1,0}}$, defined as:

$$\mathbf{R_{1,0}} = \begin{bmatrix} R_{XX} & R_{XY} & R_{XZ} \\ R_{YX} & R_{YY} & R_{YZ} \\ R_{ZX} & R_{ZY} & R_{ZZ} \end{bmatrix}$$

R_{XX} is the component of $\widehat{x_1}$ in the direction of the world coordinate system x-axis, R_{XY} is the component in the direction of the world coordinate system y-axis and R_{XZ} is the component in the direction of the world coordinate system z-axis. R_{Yi} and R_{Zi} represent the components of $\widehat{y_1}$ and $\widehat{z_1}$ in the direction of the three world axis. The coordinates of the point p_2 in the world coordinate system p_2^* can then be found as follows:

$$p_2^* = p_2\mathbf{R_{1,0}} + C_{1,0}, \qquad [1]$$

where $C_{1,0}$ is the vector joining the world origin to the origin of the coordinate system S_1. When generating structures of fully denatured proteins, this process is repeated $N - 2$ times, where N is the number of amino acids in the protein, to obtain the coordinates of each amino acid in the world coordinate system. When generating structures for each of the intermediate states, this process if carried out for the section(s) of the protein which are allowed to unfold. At the end of each unfolded segment, the coordinates of the fixed sections from the PDB file coordinates are translated relative to the position of the final unfolded residue. Additionally, each time a new point is generated, it is only accepted if the does not overlap with any of the previous points (i.e. if the distance between the new point and all other points in the protein is greater than 3Å), otherwise a new point is chosen until the condition is satisfied and only then can the next point be generated.

References

1. PJ Flory, *Principles of Polymer Chemistry.* (Cornell University Press), (1953).
2. PJ Flory, *Statistical Mechanics of Chain Molecules.* (Hanser Gardner Publications), (1989).