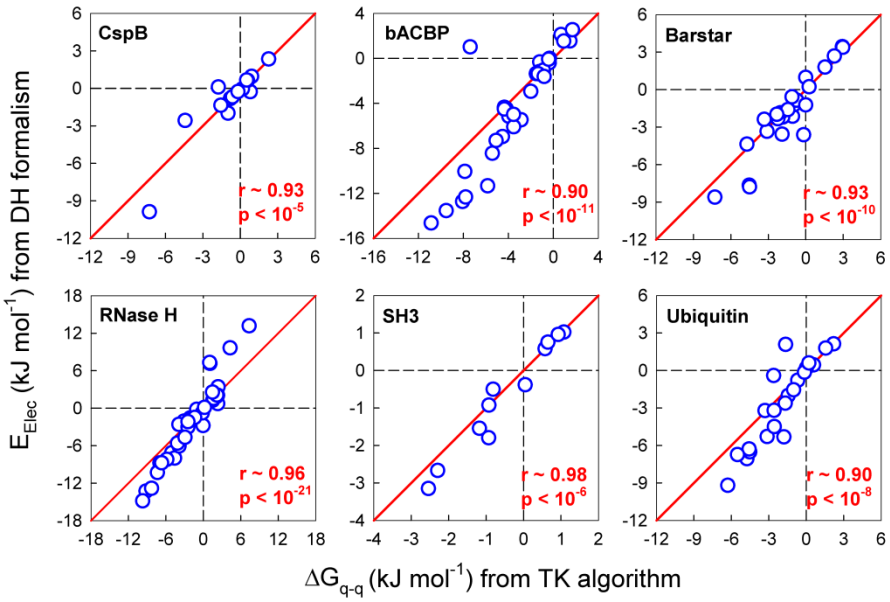
Supplementary Information

pStab: Prediction of Stable Mutants, Unfolding Curves, Mutational Hotspots and Electrostatic Frustration in Proteins

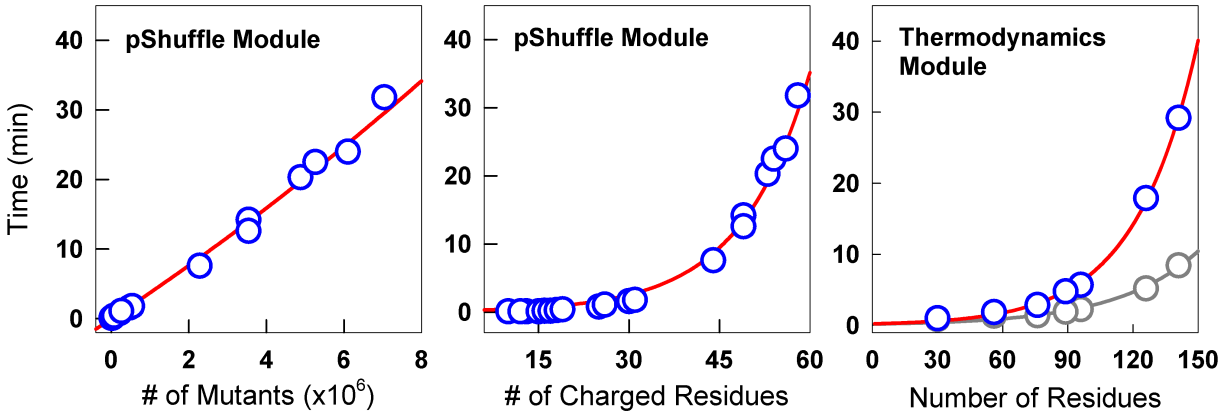
Soundhararajan Gopi, Devanshu Dev, Praveen Krishna and Athi N. Naganathan*

Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras (IITM), Chennai 600036, India.

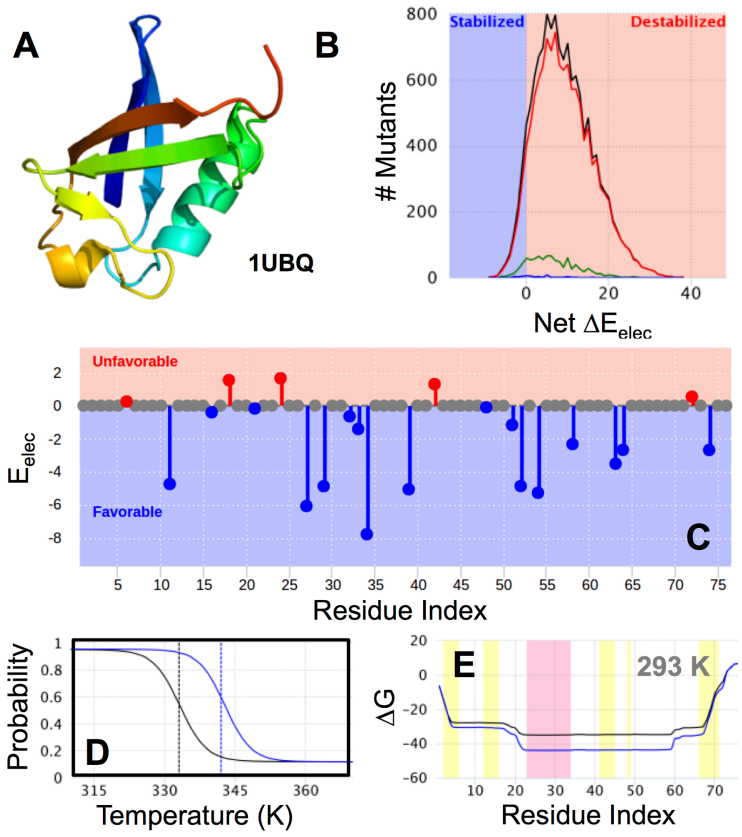
e-mail: athi@iitm.ac.in



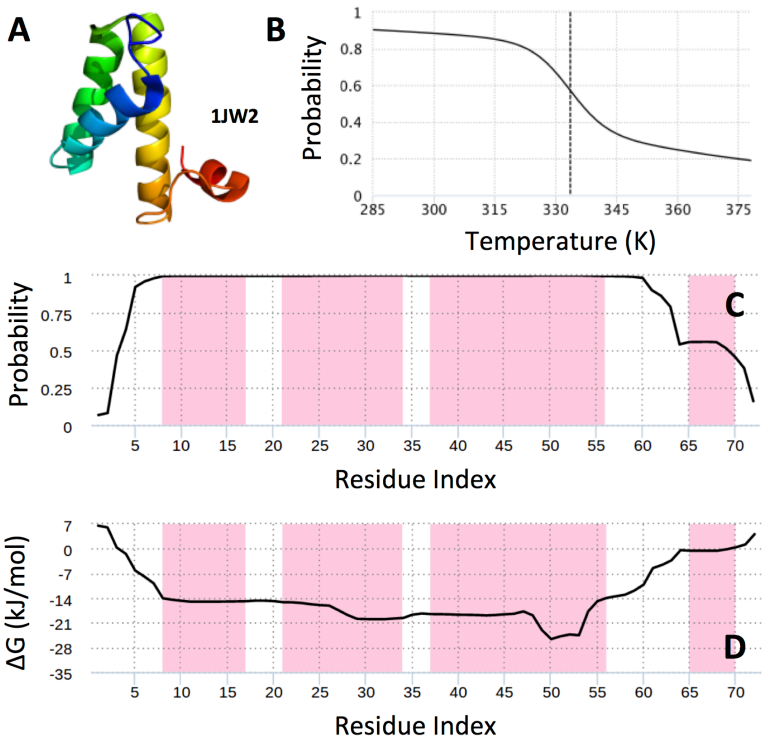
Supplementary Figure S1 Comparison of net electrostatic interaction energy (E_{elec}) of charged residues calculated using the Debye-Hückel (DH) formalism (Naganathan, 2012) and the more complex Tanford−Kirkwood (TK) algorithm (Ibarra-Molero, et al., 1999; Tanford and Kirkwood, 1957).



Supplementary Figure S2 Total time taken by the pStab server for predicting the electrostatic energy distribution for a specified number of mutants (panel A), for a specific number of charged residues (panel B) and in predicting the unfolding curves for a single protein (gray in panel C) or for 10 mutants including the WT (red in panel C).



Supplementary Figure S3 All energy units are in kJ mol^{-1}. On input of the PDB file 1UBQ (panel A), the server generates distributions of the apparent stabilities (ΔE_{elec}) for up till triple mutations (panel B; blue, green, red and black for single, double, triple and all mutations, respectively). The server also outputs the frustration due to unfavorable surface charge distributions in the WT protein providing a glimpse of potential mutations that can enhance stability (panel C). On choosing a particular mutation (E18K/E24K/R42E) for predicting additional properties with an input melting temperature of 333 K, we obtain the melting curves of both the WT and the mutant proteins (black and blue, respectively, in panel D). The local stability profile can also be calculated as a function of residue index that highlights the N- and C-terminal strands to be less stable (shown in panel E for 293 K). In panel E, the magenta and yellow shaded areas mark the sequence boundaries of α-helix and β-strands, respectively.



Supplementary Figure S4 This is a representative example of the situation when the user chooses to obtain only the thermodynamic behavior of the system (i.e. residue probability map and local stability profile) and not the effect of mutations. On input of the PDB file 1JW2 (panel A), the server generates an average unfolding curve with an input melting temperature of 333 K (panel B). The probability of folding and the local stability profile calculated as a function of residue index at 298 K (panels C and D, respectively) highlights the C-terminal helix to be less stable, in accordance with experiments(Narayan, et al., 2017). Moreover, a clear difference in stability pattern is apparent across the sequence (panel D).

Generation of Mutants (pShuffle Module)

We generate only those mutations involving charged residues, i.e. every charged residue in the protein is mutated to oppositely charged residue and a neutral charged (polar) residue of comparable size. In addition to this, large polar residues (Asn and Gln) can be mutated to both positively and negatively charged residues optionally. The candidate residues (excluding the functionally relevant residues listed by the user) are mutated in all possible combinations constrained by the maximum number of mutations allowed per mutant – a user-controlled parameter. The website allows up to four mutations per mutant; this includes all possible single, double, triple and quadruple mutants involving the charged (+Asn/Gln) residues. The number of mutants generated for a given protein can be calculated from

$$n_{mut} = \sum_{k=1}^{m} 2^k \times C_n^k$$

where, n_{mut} is the total number of mutants, m is the maximum number of mutations allowed per mutant and n is the number of candidate residues to be mutated. For example, ubiquitin (1UBQ) contains 22 charged and 8 polar (Asn and Gln) residues, and allowing four mutations per mutant results in ~0.5 million mutants. The following table shows the details of number of mutants generated for various input parameters.

Ubiquitin		
M	**No. of Mutants**	
	Only Charged Residues	**Charged + Polar Residues**
1	44	60
2	968	1800
3	13288	34280
4	130328	472760

Wako-Saitô-Muñoz-Eaton (WSME) Model

In the version of WSME model (Muñoz and Eaton, 1999; Wako and Saito, 1978) implemented in the web-server, each residue is allowed to sample only two sets of conformations: folded (native; *1*) and unfolded (non-native; *0*), resulting in 2^N microstates for N residue protein. The effective stabilization free energy contribution of a microstate (m, n) (i.e., the residues in range of m to n are structured) is described as the sum of van der Waals interactions (E_{vdW}), electrostatic potential (E_{elec}) and solvation free energy (ΔG_{solv})(Naganathan, 2012):

$$\Delta G_{m,n}^{stab} = E_{vdW} + E_{elec} + \Delta G_{solv} \quad (1)$$

The contribution to the van der Waals (vdW) interaction energy by the heavy atom pairs are defined by a cut-off (r_{cut}) for pairwise heavy atom distance.

$$E_{vdW} = \sum_{m,n} \xi_{i,j} \rho \quad (1.1)$$

where $\rho = 1$ if $r_{ij} \leq r_{cut}$ and $\rho = 0$ otherwise.

The electrostatic interaction energy between the charged residues is obtained through Debye-Hückel (DH) treatment(Naganathan, 2012; Naganathan, 2013):

$$E_{Elec} = \sum_{m,n} K_{Coulomb} \frac{q_i q_j}{\varepsilon_{eff} r_{ij}} \exp(-r_{ij}\kappa) \tag{1.2}$$

where K_{Coulomb} is the Coulomb constant (1389 kJ.Å.mol^{-1}), q_i and q_j are the charges on i^{th} and j^{th} atoms and r_{ij} is the distance between them, ε_{eff} is the effective dielectric constant and $1/\kappa$ is the Debye screening length, which is a function of ε_{eff}, solvent ionic-strength (I) and temperature (T).

$$\kappa^2 = \frac{8\pi e^2 I}{\varepsilon_{eff} k_B T}$$

The solvation free energy is given by(Naganathan, 2012),

$$\Delta G_{solv} = x_{cont}^{m,n} \Delta C_p^{cont} [(T - T_{ref}) - T \ln(T/T_{ref})] \tag{1.3}$$

where, $x_{cont}^{m,n}$ is the number of native contacts, ΔC_p^{cont} is temperature-independent heat capacity change per native contact, at a reference temperature (T_{ref}) of 385 K (Robertson and Murphy, 1997).

The total partition function (Z) is calculated through the transfer-matrix formalism of Wako and Saitô (Wako and Saito, 1978) as follows,

$$Z(T) = v_l \left[\prod_{i=1}^{N} X_i \right] v_r^{tr} \tag{2}$$

where

$$X_i = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ z & 0 & 0 & \cdots & & 0 \\ 0 & H_1^{(i)} z & 0 & & & \vdots \\ & & H_2^{(i)} z & & 0 & 0 \\ & 0 & & & H_{N-2}^{(i)} z & H_{N-1}^{(i)} z \end{pmatrix}$$

$$v_l = (1,1,1,...,1)$$

$$v_r = (1,0,0,...,0)$$

and

$$H_k^{(i)} = \exp\left(-\beta \sum_{j=1}^{k} \Delta G_{i,i+j}^{stab} \right) \qquad (k \leq N - i) \tag{3}$$

$$H_k^{(i)} = 0 \qquad (k > N - i)$$

where $\beta = 1/RT$ and z = exp($\Delta S_{conf}/R$).

Calculating the Residue Unfolding Probability (Thermodynamics Module)

The overall probability of a particular residue to be folded (χ_i) can be calculated from

$$\chi_i = Z^{-1} v_l \left[\prod_{j=1}^{i-1} X_j \right] \left[\frac{\partial X_i}{\partial \ln z} \right] \left[\prod_{j=i+1}^{N} X_j \right] v_r^{tr} \tag{4}$$

The average of χ_i over all the residues ($\langle \chi_i \rangle_T$) can be used as the proxy for global unfolding curves.

Partial Partition Functions and Local Stability

We define the residue equilibrium constant as follows

$$K_i = \frac{Z_i^f}{Z_i^{unf}} \tag{5}$$

where the numerator and denominator represent the partial partition functions of microstates in which the residue *i* is folded (*f*) and unfolded (*unf*), respectively. The local stability is then calculated as

$$\Delta G_i = -RT \ln\left(K_i \right) \tag{6}$$

Model Parameterization

The model has four parameters - the interaction energy per native contact (ξ), the heat capacity change upon fixing a native contact ($\Delta C_{p,cont}$), the entropic cost of fixing a residue in native conformation (ΔS_{conf}) and the effective dielectric constant (ε_{eff}) – of which three are fixed based on experimentally constrained analysis from previous works (see below).

In the website implementation, three different approaches are made available to assign the entropic cost of fixing residues in the native conformation:

(1) Secondary structure dependent entropic cost – entropic cost is assigned based on the secondary structure assignment by STRIDE (Heinig and Frishman, 2004). In other words, ΔS_d (-22.6 J mol^{-1} K^{-1} per residue) is assigned for residues identified as coil (disordered environment) while ΔS_o (-16.5 J mol^{-1} K^{-1} per residue) is assigned for all other residues (ordered environment;(Rajasekaran, et al., 2016)).

(2) Sequence/Structure independent entropic cost - a uniform entropic cost of ΔS_o is assigned for all residues.

(3) Sequence/Structure independent entropic cost except for glycine and proline: a uniform entropic cost of ΔS_o is assigned for all residues except for glycine and proline. For glycine and proline,

entropic costs of ΔS_{gly} (-29.47 J mol^{-1} K^{-1} per residue) and ΔS_{pro} (0 J mol^{-1} K^{-1} per residue), respectively, are assigned (Daquino, et al., 1996).

The magnitude of effective dielectric constant is fixed to 29. This estimate robustly captures the changes in stability induced by point mutations involving charged residues (Naganathan, 2013), the differences in stability between mesophilic and thermophilic protein pairs (Naganathan, 2013) and even the role of phosphorylation in a disorder-to-order protein switch (Gopi, et al., 2015). The pH conditions are simulated by assigning the associated protonation states to the atoms of the charged residues. At pH 7, atoms NE, NH1, NH2 of arginine are assigned a charge of 0.33 each, NZ of lysine a charge of 1, and OD1, OD2 of aspartate and OE1, OE2 of glutamate a charge of −0.5. In addition to the protonation states of the atoms at pH7, ND1 and NE2 of histidine are assigned a charge of 0.5 at pH 5. All-to-all electrostatic interactions are considered that eliminate assumptions implicit in using cutoffs.

The heat capacity change upon fixing a native contact is fixed to -0.36 J/(mol. K) per native contact (Naganathan, 2012). Heavy atom contacts are identified with a distance cutoff (r_{cut}) of 6 Å excluding the nearest neighbors (j>i+1). The interaction energy per native contact (ξ) is tuned to reproduce the experimental T_m provided by the user. If not provided, a default value of 333 K is reproduced.

Supporting References

Daquino, J.A.*, et al.* The magnitude of the backbone conformational entropy change in protein folding. *Proteins* 1996;25(2):143-156.

Gopi, S.*, et al.* Energetic and topological determinants of a phosphorylation-induced disorder-to-order protein conformational switch. *Phys. Chem. Chem. Phys.* 2015;17:27264-27269.

Heinig, M. and Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nuc. Acids Res.* 2004;32:W500-W502.

Ibarra-Molero, B.*, et al.* Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* 1999;38:8138-8149.

Muñoz, V. and Eaton, W.A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. U.S.A.* 1999;96(20):11311-11316.

Naganathan, A.N. Predictions from an Ising-like Statistical Mechanical Model on the Dynamic and Thermodynamic Effects of Protein Surface Electrostatics. *J. Chem. Theory Comput.* 2012;8(11):4646-4656.

Naganathan, A.N. A Rapid, Ensemble and Free Energy Based Method for Engineering Protein Stabilities. *J. Phys. Chem. B* 2013;117(17):4956-4964.

Narayan, A.*, et al.* Graded structural polymorphism in a bacterial thermosensor protein. *J. Am. Chem. Soc.* 2017;139:792-802.

Rajasekaran, N.*, et al.* Quantifying Protein Disorder through Measures of Excess Conformational Entropy. *J. Phys. Chem. B* 2016;120:4341-4350.

Robertson, A.D. and Murphy, K.P. Protein structure and the energetics of protein stability. *Chem. Rev.* 1997;97(5):1251-1267.

Tanford, C. and Kirkwood, J.G. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *J. Am. Chem. Soc.* 1957;79:5333-5339.

Wako, H. and Saito, N. Statistical Mechanical Theory of Protein Conformation .2. Folding Pathway for Protein. *J. Phys. Soc. Japan* 1978;44(6):1939-1945.