

# Exploring the Energy Landscapes of Protein Folding Simulations with Bayesian Computation

Nikolas S. Burkoff,<sup>†</sup> Csilla Várnai,<sup>†</sup> Stephen A. Wells,<sup>‡</sup> and David L. Wild<sup>†\*</sup>

<sup>†</sup>Systems Biology Centre and <sup>‡</sup>Department of Physics and Centre for Scientific Computing, University of Warwick, Coventry, United Kingdom

# Supporting Material

## Figures Referenced in Main Text

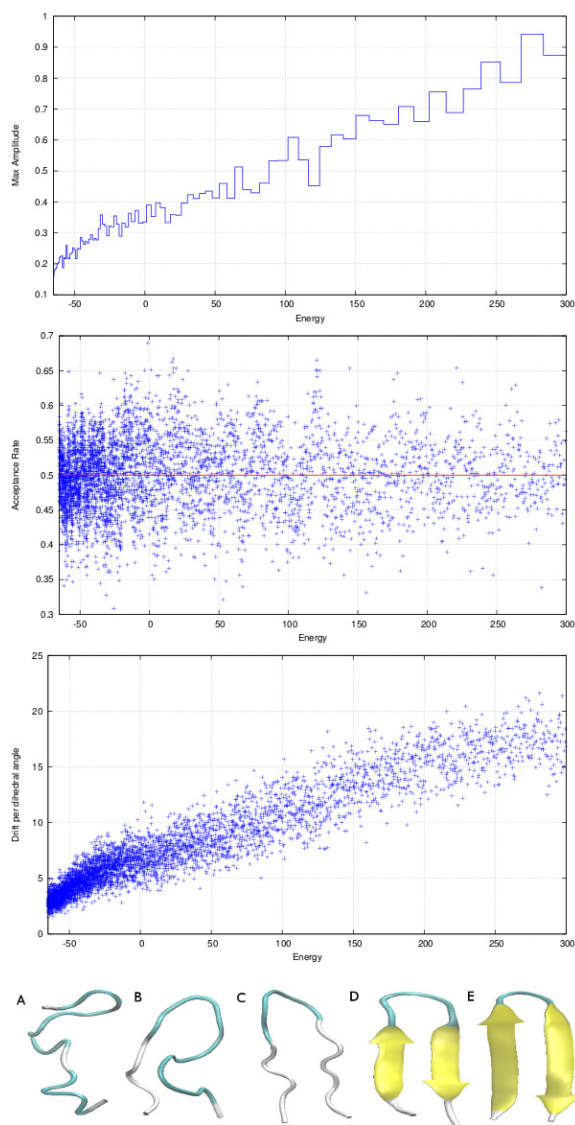


Figure S1: Top graph: the maximum allowed amplitude of the crankshaft rotations  $\alpha_0$  (in radians); Middle graph: the acceptance rate of the MC chains; Bottom graph: the drift per dihedral angle (the distance between the start and end conformations of a single MC chain). All with respect to the current energy threshold, for a 16 residue polyalanine  $\beta$ -hairpin; see text for more details. Bottom: Five snapshots from a single nested sampling simulation of a  $\beta$ -hairpin with  $K = 1000$  and  $m = 2500$ . The snapshots are equally spaced along the  $\log(X)$  axis and have energies 3567, 190, 0, -46 and -66 (A-E). For comparison, the expectation of the internal energy at room temperature is -43.

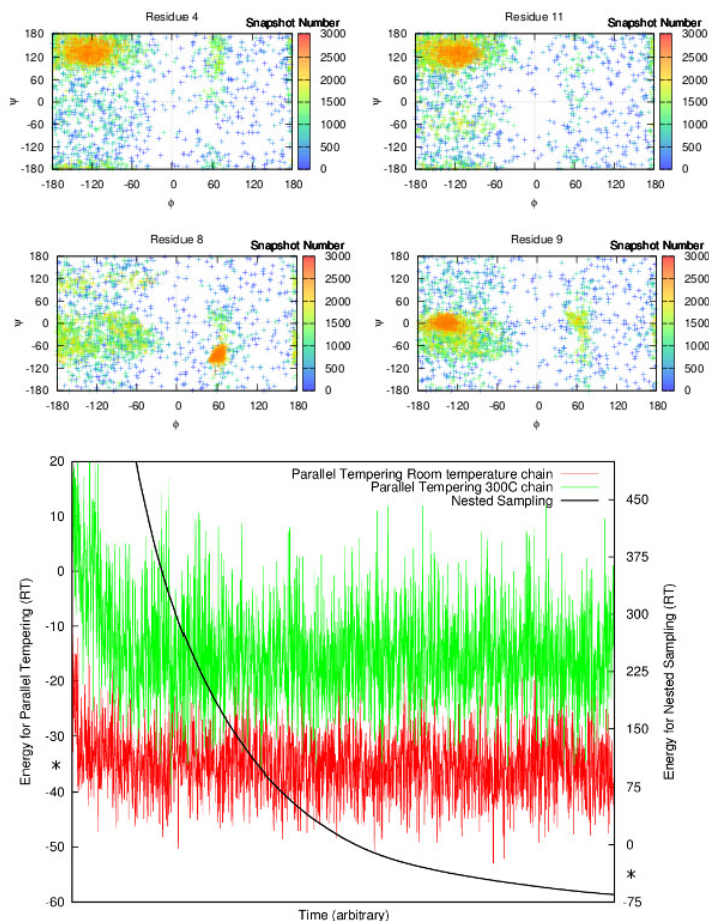


Figure S2: Top: Dihedral angle evolution for residues 4, 8, 9 and 11 of the 16 residue polyalanine nested sampling simulation. In the later snapshots, residues 4 and 11 are distributed in the standard  $\beta$ -sheet region of the Ramachandran plot. Residues 8 and 9 contain the turn of the polypeptide. The dihedral angles of the turn residues, 8 ( $60 \pm 15, -90 \pm 30$ ) and 9 ( $-150 \pm 30, 0 \pm 30$ ), are closest to the values of type II' turn ( $(60, -120)$  and  $(-80, 0)$ ) Bottom: Energy v Time graph for nested sampling (right hand axis) and two of the chains from a parallel tempering simulation (room temperature and  $300^\circ\text{C}$ ; both left hand axis.). On both vertical axes a star marks the expected thermodynamic energy at room temperature.

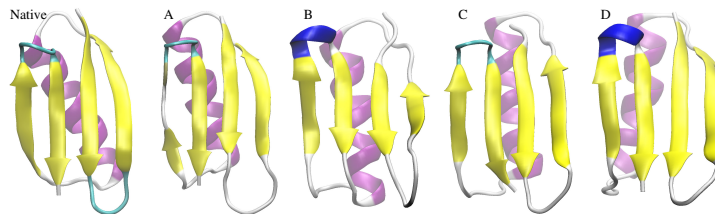


Figure S3: The native (crystal) structure of protein G (left) with a sample of conformations accessible at room temperature from a simulation with  $K = 20000$  and  $m = 15000$ . All figures of protein G in this paper have been oriented so that the first  $\beta$ -strand is the second strand from the right and the N-terminal residue is at the top. The thermodynamic energy at room temperature, estimated from the simulation, is  $-190$  and conformers A,B,C and D have have energies  $-189$ ,  $-190$ ,  $-191$  and  $-190$ , respectively. The backbone RMSDs from the crystal structure are  $1.93 \text{ \AA}$ ,  $2.96 \text{ \AA}$ ,  $3.97 \text{ \AA}$  and  $5.22 \text{ \AA}$ , respectively. The angle between the helix projected onto the sheet and the first  $\beta$ -strand is  $17.9^\circ$ ,  $8.6^\circ$ ,  $-4.7^\circ$  and  $-15.1^\circ$ , respectively, compared to  $21.9^\circ$  of the crystal structure.

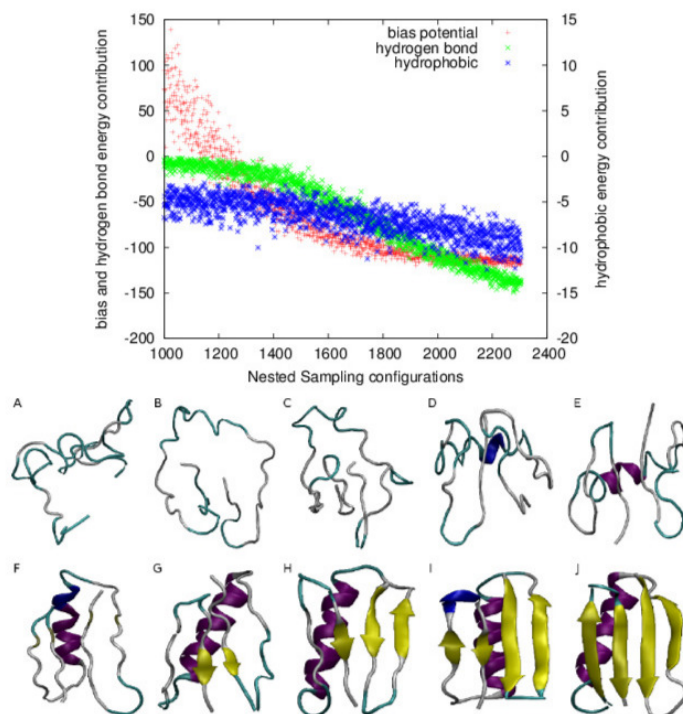


Figure S4: Top: Energy contributions of the  $G\bar{o}$ -type bias potential (red), hydrogen bonds (green) and hydrophobic interactions (blue) in the second half of a nested sampling simulation of protein G with  $K = 20000$  and  $m = 15000$ . Units of energy are in  $RT$  corresponding to temperature. Note the different scale on the vertical axes. Bottom: Ten conformations of protein G from the same simulation, in order of decreasing energy.

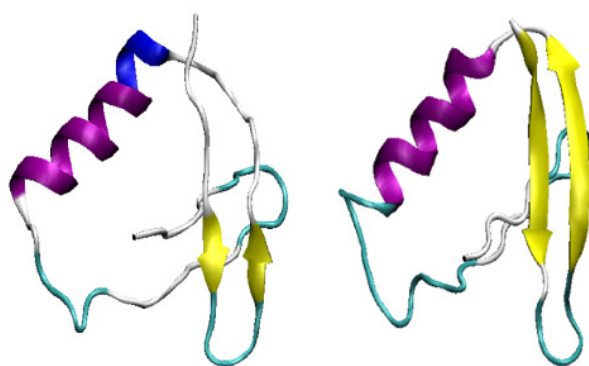


Figure S5: The topological analysis places similar conformations in the same basin. For example, these two conformations (which both have the 3rd and 4th  $\beta$ -strands aligned incorrectly) are placed in the same basin.

# 1 The Protein Model

We modelled the polypeptide as a chain of peptide groups elastically connected at the  $C_\alpha$  atoms, with the valence angles constrained to  $111.5^\circ \pm 2.8^\circ$ . The positions of all backbone and  $C_\beta$  atoms, including hydrogen, were specified by the orientations of the peptide bonds. We fixed the peptide bond lengths and angles at standard values (1–3). The distance between  $C_\alpha$  atoms separated by *trans* peptide bonds was fixed at 3.8 Å. The  $C_\beta$  positions were stipulated by the tetrahedral geometry of the  $C_\alpha$  atoms and corresponded to L-amino acids. Most of the conformational variability of polypeptides comes from relatively free rotation around N– $C_\alpha$  and  $C_\alpha$ –C bonds characterised, respectively, by dihedral angles  $\phi$  and  $\psi$  (Fig. 1 in (4)). These rotations are least restricted in glycine that lacks  $C_\beta$ . The dihedral angles  $\phi$  in proline were elastically constrained to  $-60^\circ \pm 7^\circ$  by covalent bonding (5). We introduced a harmonic potential  $E_i^B$  to impose these and other elastic constraints. A more detailed description of the model is given in our previous work (4).

In this work, we represented other side chain atoms by one, or in the case of branched side chains, two pseudo-atoms, following (6). The side chain dihedral angles  $\chi$  were permitted to vary, and take the values  $\{\pm 60^\circ, 180^\circ\}$ , or in the case of proline  $\{\pm 30^\circ\}$ , with probabilities dependent on residue type, with values corresponding to the distribution of the  $\chi$  angles in the same ASTRAL PDB database (7) that was used in (4), and here, to learn the potential parameters by a statistical machine learning procedure, contrastive divergence (8).

We modelled van der Waals repulsions so that there is a prohibitively large energetic cost of overlaps between atoms. We used values of atomic radii close to a lower limit of the range found in the literature (9–12):  $r(C_\alpha) = r(C_\beta) = 1.57$  Å,  $r(C) = 1.42$  Å,  $r(O) = 1.29$  Å,  $r(N) = 1.29$  Å. We adopted values of the contact radii for the pseudo-atoms from (6).

Hydrogen bonding is a major polar interaction between NH and CO groups of polypeptide backbone. Based on surveys of the Protein Data Bank (PDB) (13), important reviews of hydrogen bonding in globular proteins have formulated the basics of the current understanding of hydrogen bond geometry and networking (14–17). We considered the hydrogen bond formed when three distance and angular conditions were satisfied:  $r(O, H) < \delta$ ,  $\angle OHN > \Theta$ , and  $\angle COH > \Psi$ , where  $r(O, H)$  is the distance between oxygen and hydrogen, and symbol  $\angle$  denotes the angle between the three atoms (see Fig. 1A in (18)). The lower bound on the separation between the atoms ( $r(O, H) > 1.8$  Å) was implicitly set by the hard-sphere collision between oxygen and nitrogen. We used the same hydrogen bond potential regardless of the secondary structure adopted by the peptide backbone. The energy of the hydrogen bond (Fig. 1B in (18)) was described in (18) by a square-well potential,

$$E_{ij}^{HB} = -n_h H \quad (1)$$

where  $H$  is the strength of each hydrogen bond, and  $n_h$  is the number of hydrogen bonds between the amino acids  $i$  and  $j$ . The strength of the hydrogen bonds,  $H$ , as well as the three cutoff parameters,  $\delta$ ,  $\Theta$ ,  $\Psi$  was determined by a machine learning procedure, contrastive divergence (8). We found that softening the hard cutoffs  $\{\delta, \Psi, \Theta\}$  improved the results, and hence we used a steep continuous approximation to the square well.

We modelled hydrophobic interactions in a manner consistent with (6). The hydrophobic interaction contribution between hydrophobic atoms  $A$  and  $B$  of amino acids  $i$  and  $j$  ( $|i - j| \geq 2$ ) is

$$E_{AB}^{hyd} = \begin{cases} f k_h & r_{AB} < r_{cut,AB} \\ f k_h \left( \frac{r_{AB} - r_{cut,AB}}{\Delta} \right) & r_{cut,AB} \leq r_{AB} < r_{cut,AB} + \Delta \end{cases} \quad (2)$$

where  $k_h$  is a constant parameter proportional to the Kauzmann parameter (19), the cutoff distance  $r_{cut,AB}$  is the sum of the vdW radii of atoms  $A$  and  $B$  listed in (6),  $\Delta = 2.8$  Å is a smoothing range beyond the cutoff distance, and the multiplicative factor  $f$  takes the value 2, if both amino acids are hydrophobic, 1, if one is hydrophobic and the other one is amphipathic, and 0, if neither are hydrophobic. Hydrophobic amino acids are cysteine, isoleucine, leucine, methionine, phenylalanine, tryptophan and valine; amphipathic residues are alanine, histidine, threonine and tyrosine.

The sequence-*dependent* part of the potential (the negative log-likelihood) was approximated in our model by pair-wise interactions between side-chains, as described in (20). Our main focus was on the resulting effect of these interactions and how they stabilise secondary structural elements. We did not consider the detailed physical nature of these forces, or how they depend on the amino acid types. We introduced these interactions between the polypeptide side chains as an effective G $\bar{o}$ -type potential (21) dependent on the distance between C $\beta$  atoms,

$$E_{ij}^{SC} = \kappa C_{ij} (r_{ij} - r)^2 \quad (3)$$

where  $r_{ij}$  is a distance between non-adjacent C $\beta$  atoms,  $|i - j| > 1$ ;  $r$  a constant and  $\kappa$  is a force constant. In (20) we introduced a ‘regularised contact map’,  $C_{ij}$ . In this binary matrix, two types of contacts were defined in the context of protein secondary structure. First, only lateral contacts in the parallel and anti-parallel  $\beta$ -sheets were indicated by 1’s. Second, the contacts between amino acids  $i$  and  $i + 3$  in  $\alpha$ -helices were also represented by 1’s. These contacts typically have the closest C $\beta$ -C $\beta$  distance among non-adjacent contacts in native proteins. The force constants and  $r$  depend on the secondary structure type, introducing positive  $\kappa_\alpha$ ,  $\kappa_\beta$ ,  $r_\alpha$  and  $r_\beta$ . Non-adjacent contacts in secondary structural elements were, therefore, stabilised by attracting potentials.

We also modelled interactions between sequential residues. This interaction was defined by the mutual orientation of adjacent residues that are involved in secondary structural elements,

$$E_{i,i+1}^{SC} = \eta \cos \gamma_{i,i+1} \quad (4)$$

where  $\gamma_{i,i+1}$  is the dihedral angle N $_i$ -C $_{\alpha,i}$ -C $_{\alpha,i+1}$ -C $_{i+1}$  between the adjacent residues. The purpose of this interaction is to bias the conformation towards the naturally occurring orientations of residues in secondary structural elements. In  $\alpha$ -helices, adjacent residues adopt a conformation with positive  $\cos \gamma$ . In  $\beta$ -sheets,  $\cos \gamma$  is negative. We, therefore, used two values of the force constant: negative  $\eta_\alpha$  and positive  $\eta_\beta$ .

As in (20), all parameters were determined by a statistical machine learning procedure, contrastive divergence (8) and in this work  $\delta = 2.06$ ,  $-\cos \Theta = 0.89$ ,  $-\cos \Psi = 0.766$ ,  $H = 4.35$ ,  $\eta_\beta = 3.5$ ,  $\eta_\alpha = -4.9$ ,  $\kappa_\alpha = 3.3$ ,  $\kappa_\beta = 1.2$ ,  $r_\alpha = 5.66$ ,  $r_\beta = 5.35$  and  $k_h = 0.08$ , where the unit of energy is  $RT$  at room temperature. With the improved model and force field described in this paper, contrastive divergence provided good parameters without the need of further adjustments, as had been the case in (20).

To summarise, the total energy of a polypeptide chain with conformation  $\Omega$  was calculated as follows

$$E(R, \Omega) = \sum_{i=1}^N E_i^B + \sum_{i=1}^N \sum_{j=1}^i (E_{ij}^{vdW} + E_{ij}^{HB} + E_{ij}^{SC} + E_{ij}^{hyd}) \quad (5)$$

where we consider harmonic valence elasticity,  $E_i^B$ , van der Waals repulsions,  $E_{ij}^{vdW}$ , hydrogen bonding,  $E_{ij}^{HB}$  and hydrophobic packing,  $E_{ij}^{hyd}$ . The valence elasticity, van der Waals repulsions, and hydrogen bonding that contribute to this potential have a clear physical meaning and are analogous to traditional *ab initio* approaches. The side-chain interactions,  $E_{ij}^{SC}$  in this model were introduced as a long-range quadratic G $\bar{o}$ -type potential based on the contact map and secondary structure assignment. This pseudo-potential had two purposes: it was needed to stabilise the secondary structural elements, and to provide a biasing force that allows reconstruction of the backbone conformation in the course of Metropolis Monte Carlo simulations (4, 20).

## 2 Additional Results

### Src Tyrosine Kinase SH3 Domain.

Src Tyrosine Kinase SH3 Domain is a 56-residue protein, comprising a 5-stranded  $\beta$ -barrel. The last strand is interrupted by a single turn of a  $3_{10}$ -helix, which was not included in the ‘regularised’ contact map used to define the native state in these simulations. The native (crystal) structure is shown in Fig. S6.

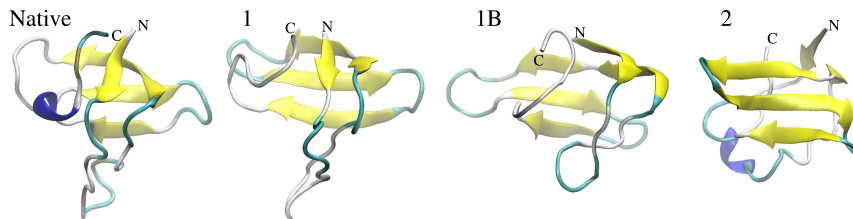


Figure S6: The native (crystal) structure of the SH3 Domain, with conformers from the two main funnels accessible at room temperature from a simulation with  $K = 15000$  and  $m = 15000$ . Conformers 1 and 1B have the correct backbone topology, unlike conformer 2. The backbone RMSDs are 4.89 Å, 5.28 Å and 11.51 Å, respectively. The N and C termini are marked; see the text for more details.

The energy landscape for a simulation with  $K = 15000$  and  $m = 15000$ , depicted in Fig. S7, shows two main funnels (funnels 1 and 2) that further split into sub-funnels (funnels 1 and 1A, and funnels 2 and 2A). At room temperature, conformations representative of funnels 1 and 2 are accessible, and most of the posterior mass is in funnel 1, which contains conformers with the correct backbone topology. Funnel 1 splits further at  $-43$  energy units. The sub-funnels 1 and 1B are connected at room temperature, and the probability of moving from one to the other is non-zero, due to the non-zero probability of adopting a conformation that is indistinguishable from those typical to funnels 1 and 1B. The estimated energy at room temperature is  $-39$  energy units.

Fig. S6 shows low energy conformers from the same simulation. Conformers 1 and 1B (from funnels 1 and 1B in Fig. S7) have the correct backbone topology and a backbone RMSD of 4.61 Å and 5.47 Å with respect to the crystal structure. Moreover, although the N-terminal loop is not included in the ‘regularised’ contact map used in the simulation, the packing of the loop is in reasonable agreement with the crystal structure. Conformer 2, taken from the bottom of the other major funnel, adopts a conformation with an incorrect backbone topology – note the relative positions of the N and C termini with respect to the sheet. The conformers shown in Fig. S6 correspond to lower energies than the estimated energy at room temperature, but have been shown as it is clearer to see the differences between them once the  $\beta$ -strands have fully formed. Conformations available at room temperature typically have shorter  $\beta$ -strands. The estimated backbone RMSD at room temperature is  $\mathbb{E}(\text{RMSD}|\beta = 1) = 6.46$  Å.

## Chymotrypsin Inhibitor 2.

Chymotrypsin inhibitor 2 is a 65-residue protein which contains a four-stranded  $\beta$ -sheet and an  $\alpha$ -helix, but differs in topology from protein G. However, it also possesses a similar symmetry with regard to the packing of the  $\alpha$ -helix against the  $\beta$ -sheet.

As with protein G, the energy landscape shows two main folding funnels. Fig. S9 shows the prior energy landscape chart and the posterior energy landscape chart at room temperature for a simulation with  $K = 15000$  and  $m = 15000$ . Sample conformers from the main funnels at different energy levels are also included. The posterior energy landscape chart shows that virtually all the posterior mass is in funnel 1 (including funnel 1A) at room temperature, with funnel 2 being insignificant. Funnel 1 splits at  $-75$  energy units into funnels 1 and 1A, which are connected at room temperature; the probability of adopting a conformation that is indistinguishable from the ones in funnels 1 and 1A is non-zero. The estimated value of the energy at  $\beta = 1$  is  $-82$  units.

The native structure and sample conformers from the main funnels are also compared in Fig. S8. As with protein G, in our model, the helix can be packed on either side of the sheet, and in Fig. S8 conformers 1 and 2 are taken from the bottom of the two funnels. Conformer 1 has the correct backbone topology, whereas conformer 2 has the helix packed on the incorrect side of the sheet. Conformer 1A from funnel 1A has the correct backbone topology, but the hydrophobic residues of the C-terminal  $\beta$ -strand are packed on the wrong



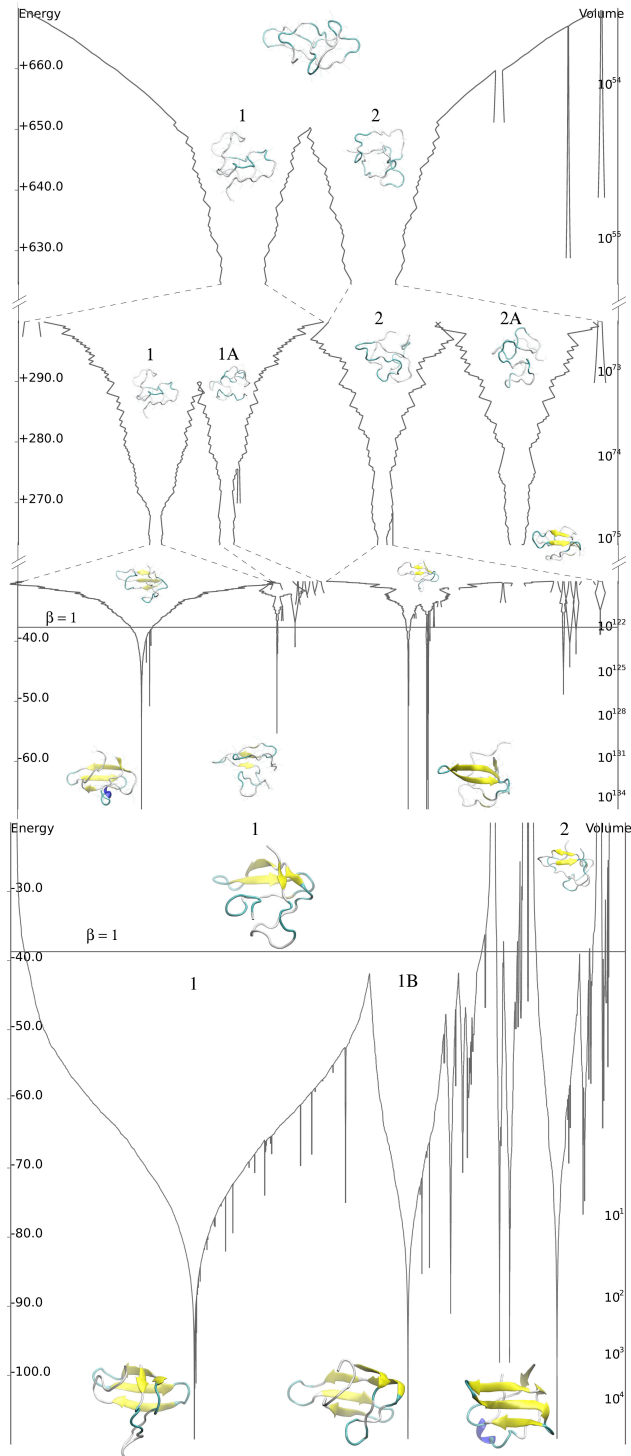


Figure S7: Top: prior (potential) energy landscape chart, Bottom: posterior energy landscape chart at  $\beta = 1$  for a simulation of the SH3 Domain using  $K = 15000$  and  $m = 15000$ . Sample conformers of funnels 1, 1A, 1B, 2 and 2A are marked on the chart.

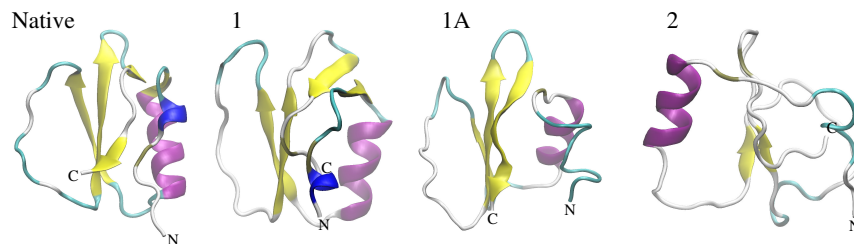


Figure S8: The native (crystal) structure for chymotrypsin inhibitor 2 (left) and 3 conformers from a simulation with  $K = 15000$  and  $m = 15000$ . 1 and 1A: conformers from near the bottom of the funnel which contains conformers with the correct topology with backbone RMSD of 4.86 Å and 4.91 Å, respectively. The 1A conformer has the hydrophobic residues of the C-terminal  $\beta$ -strand on the wrong side. 2: a conformer from near the bottom of the other funnel with backbone RMSD of 11.18 Å. Note that the helix is packed on the wrong side of the sheet in conformer 2. The N and C termini are shown.

side of the sheet. Both conformers 1 and 1A have much lower energy than found at room temperature, and it is interesting to note that in both of these conformers the C-terminal  $\beta$ -strand has formed spontaneously without contact bias. In the crystal structure this strand is actually a large coil. In simulations, we find that the secondary structure that is defined by the regularised contact map forms first, and, since the model allows a large amount of freedom for residues which do not have contact bias, nested sampling then tries to place the remaining residues in the lowest energy position possible. The backbone RMSD of conformers 1, 1A and 2, from the crystal structure, are 4.86 Å, 4.91 Å and 11.18 Å, respectively, while the estimated backbone RMSD at room temperature is  $\mathbb{E}(\text{RMSD}|\beta = 1) = 5.55$  Å.

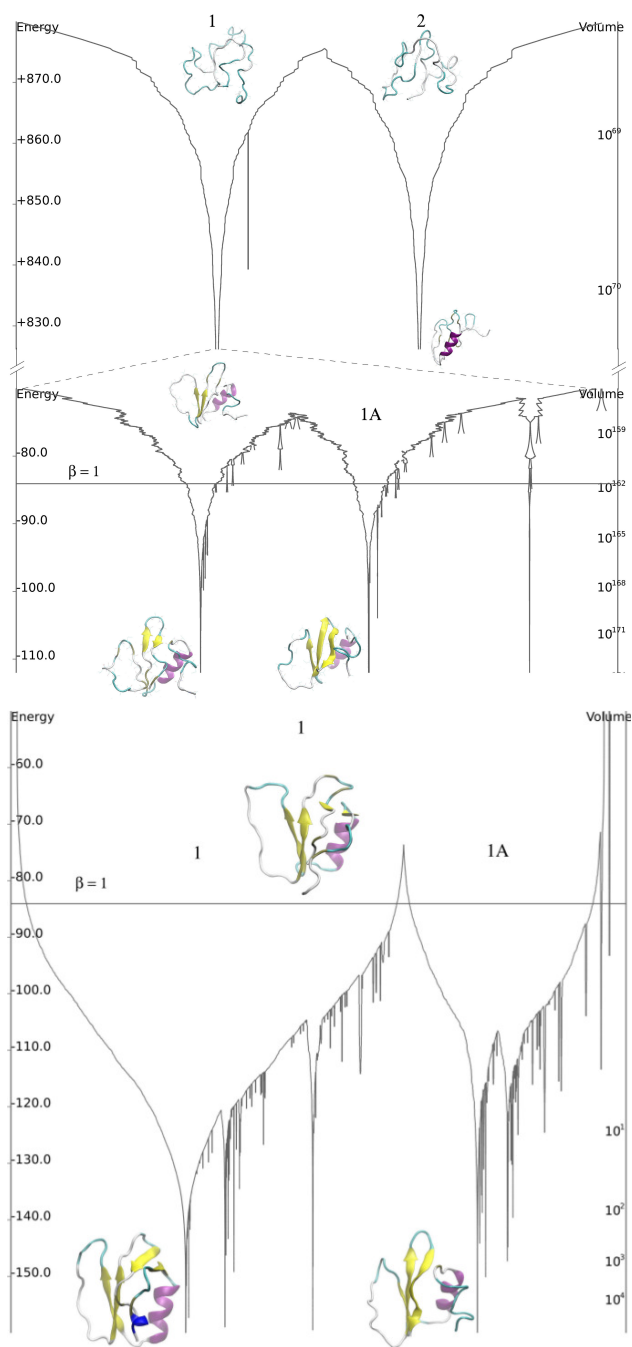


Figure S9: Top: prior (potential) energy landscape chart, Bottom: posterior energy landscape chart at  $\beta = 1$  for a simulation of the chymotrypsin inhibitor 2 using  $K = 15000$  and  $m = 15000$ . Sample conformers of funnels 1, 1A and 2 are marked on the charts.

### 3 Rigidity and Flexibility of Protein G

We have carried out rigidity analysis on Protein G using the software “FIRST” (22). Our input was the 1PGA.pdb structure with hydrogens added using the “Reduce” software (23). Rigidity analysis balances the degrees of freedom of the atoms against the constraints introduced by covalent bonding, hydrophobic tethers, salt bridges and hydrogen bonds. The result is a decomposition of the structure into rigid and flexible regions, known as a rigid cluster decomposition. The set of hydrogen bonds included is determined by a (negative) cutoff energy  $E_{cut}$ . An  $E_{cut}$  value near zero will include a large number of weak bonds and largely rigidify a structure; progressively lowering  $E_{cut}$  eliminates hydrogen bonds in an order from weakest to strongest, a process known as “rigidity dilution”. The progress of this loss of rigidity can be mapped in a dilution plot (Figure S10a) in which the rigid cluster membership of each residue is mapped onto a 1-D representation of the protein backbone. A new line is plotted for each cutoff energy at which the rigidity of the mainchain changes.

At cutoff energies above  $-1.844$  kcal/mol, the helix and the beta-sheet form a single rigid cluster (Figure S10b), while at lower energies the helix is a rigid body but the beta-sheet has become flexible (Fig. S10c). We stress that the backbone-backbone hydrogen bonding in both the helix and the sheet persists to much lower cutoff values. Once the helix and the sheet are not a single rigid cluster, motion of the helix with respect to the sheet becomes possible. The amplitude of such motion will be constrained by covalent and non-covalent interactions, in particular the many hydrophobic tethers between helix and sheet residues.

We obtain an eigenvector for flexible motion using a coarse-grained (one site per residue) elastic network model as implemented in the software “ElNeMo” (24). The lowest-frequency non-trivial mode, mode 7, corresponds to a rotation of the helix about an axis perpendicular to the beta-sheet. Linear projection of the structure along this mode would rapidly introduce unphysical distortions such as elongation of the helix. In order to project the motion to finite non-zero amplitude, we make use of geometric simulation using the “FRODA” module (25) included in FIRST. FRODA generates new conformations of the protein structure by repeatedly introducing small perturbations of the atomic positions and reimposing the constraints. We use the elastic-network mode eigenvector to bias the perturbations (26); this allows us to project the motion to large amplitudes while maintaining covalent, non-covalent and steric constraints.

The mode can be projected to a  $C_\alpha$  RMSD of more than  $3 \text{ \AA}$  from the initial structure (Figure S11a) while maintaining the network of hydrophobic tethers that are present in the original crystal structure. During this projection (Fig. S11b-d) the helix rotates from its initial position diagonal to the sheet to lie parallel to the beta-sheet strands. The projected structures are very similar to conformations from the folding simulation, shown in the main article.

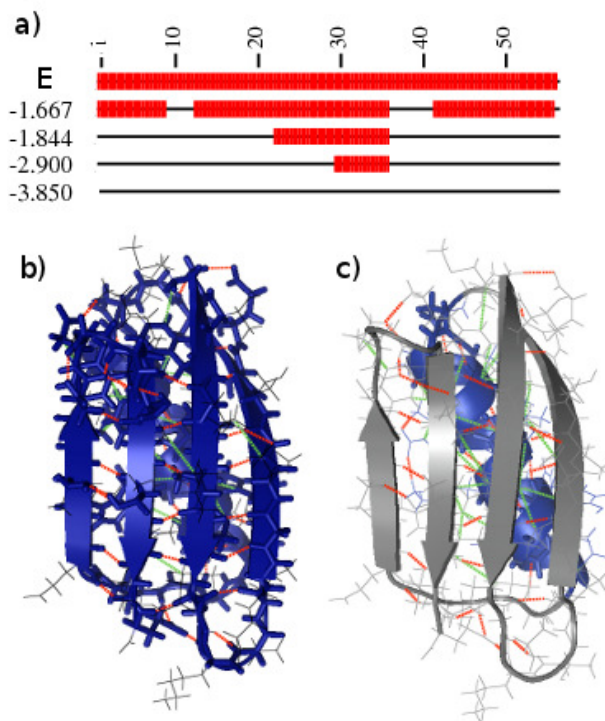


Figure S10: (a) Rigidity dilution of Protein G structure 1PGA as hydrogen bond energy cutoff is lowered. Each line represents the protein backbone; a thin line represents a flexible region while a thick line indicates membership of a large rigid cluster. The beta-sheets lose their rigidity at a cutoff of  $-1.844$  kcal/mol while the helix (residues 22-35) remains rigid to much lower cutoffs. (b) Rigid cluster decomposition of 1PGA at a cutoff of  $-1.0$  kcal/mol. (c) Rigid cluster decomposition at a cutoff of  $-1.9$  kcal/mol. Green and red dashed lines represent hydrophobic tethers and hydrogen bonds.

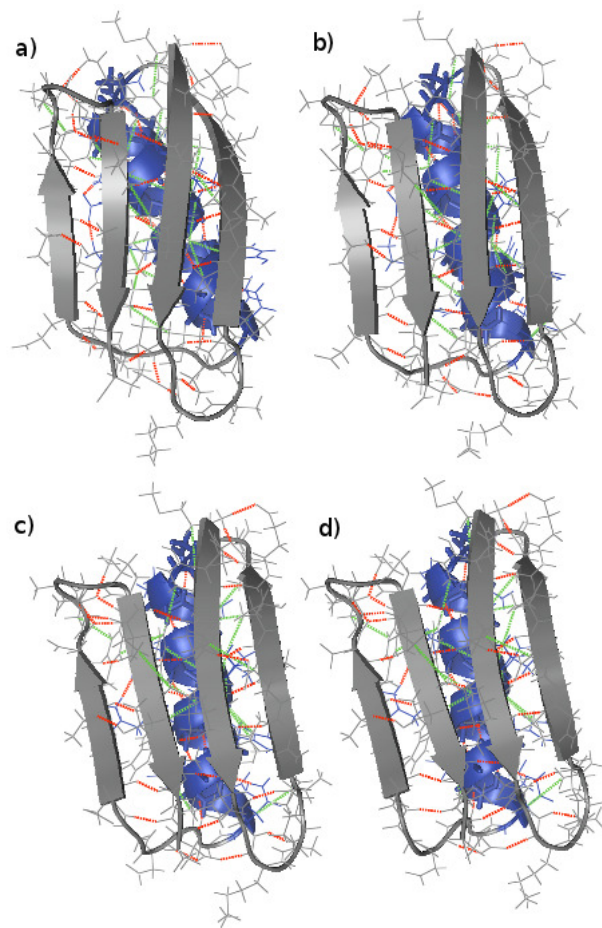


Figure S11: Projection of lowest-frequency non-trivial elastic network mode from initial structure (a) to 3Å RMSD (b–d). Green and red dashed lines represent hydrophobic tethers and hydrogen bonds.

## References

- [1] Engh, R. A., and R. Huber, 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallographica Section A* 47:392–400.
- [2] Engh, R. A., and R. Huber, 2001. Structure quality and target parameters. In M.G. Rossman, E. Arnold (eds.) *International Tables for Crystallography*, vol. F: Crystallography of biological macromolecules, 1st edn. Kluwer Academic Publishers for the International Union of Crystallography, Dordrecht Boston London.
- [3] Brünger, A., 1992. X-PLOR, Version 3.1: a system for X-ray crystallography and NMR. Yale University Press, New Haven.
- [4] Podtelezchnikov, A. A., and D. L. Wild, 2005. Exhaustive Metropolis Monte Carlo sampling and analysis of polyalanine conformations adopted under the influence of hydrogen bonds. *Proteins* 61:94–104.
- [5] Ho, B. K., E. A. Coutsiias, C. Seok, and K. A. Dill, 2005. The flexibility in the proline ring couples to the protein backbone. *Protein Sci* 14:1011–8.
- [6] Srinivasan, R., and G. D. Rose, 1999. A physical basis for protein secondary structure. *Proc Natl Acad Sci U S A* 96:14258–14263.
- [7] Brenner, S. E., P. Koehl, and M. Levitt, 2000. The ASTRAL compendium for protein structure and sequence analysis. *Nucl. Acids Res.* 28:254–256.
- [8] Carreira-Perpinan, M., and G. Hinton, 2005. On contrastive divergence learning. In: *Artificial Intelligence and Statistics*. The Savannah Hotel, Barbados.
- [9] Ramachandran, G. N., C. Ramakrishnan, and V. Sasisekharan, 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 95–9.
- [10] Hopfinger, A. J., 1973. *Conformational properties of macromolecules*. Academic Press, New York.
- [11] Word, J. M., S. C. Lovell, T. H. LaBean, H. C. Taylor, M. E. Zalis, B. K. Presley, J. S. Richardson, and D. C. Richardson, 1999. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* 285:1711–33.
- [12] Pappu, R. V., R. Srinivasan, and G. D. Rose, 2000. The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc Natl Acad Sci U S A* 97:565–70.
- [13] Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, 2000. The protein data bank. *Nucleic. Acids. Res.* 28:235–42.
- [14] Baker, E. N., and R. E. Hubbard, 1984. Hydrogen bonding in globular proteins. *Proc. Biophys. Mol. Biol.* 44:97–179.
- [15] Savage, H. J., C. J. Elliott, C. M. Freeman, and J. M. Finney, 1993. Lost hydrogen bonds and buried surface area: rationalising stability in globular proteins. *J. Chem. Soc. Faraday Trans.* 89:2609–2617.
- [16] Stickle, D. F., L. G. Presta, K. A. Dill, and G. D. Rose, 1992. Hydrogen bonding in globular proteins. *J. Mol. Biol.* 226:1143–59.
- [17] McDonald, I. K., and J. M. Thornton, 1994. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238:777–93.
- [18] Podtelezchnikov, A. A., Z. Ghahramani, and D. L. Wild, 2007. Learning about Protein Hydrogen Bonding by Minimizing Contrastive Divergence. *Proteins* 66:588–99.

- [19] Lee, B., and F. M. Richards, 1971. The interpretation of protein structures estimation of static accessibility. *J. Mol. Biol.* 55:379–400.
- [20] Podtelezhnikov, A. A., and D. L. Wild, 2009. Reconstruction and Stability of Secondary Structure Elements in the Context of Protein Structure Prediction. *J. Biophys.* 96:4399–4408.
- [21] Gō, N., 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* 12:83–210.
- [22] Thorpe, M. F., M. Lei, A. J. Rader, D. J. Jacobs, and L. A. Kuhn, 2001. Flexible and rigid regions in proteins. *J. Mol. Graph. Model* 19:60–69.
- [23] Word, J. M., S. Lovell, J. S. Richardson, and D. C. Richardson, 1999. Asparagine and Glutamine: Using Hydrogen Atoms Contacts in the Choice of Side-Chain Amide Orientation. *J. Mol. Biol.* 285:1735–1747.
- [24] Suhre, K., and Y.-H. Sanejouand, 2004. ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research (Web Issue)* 32:610–614.
- [25] Wells, S. A., S. Menor, B. M. Hespeneide, and M. F. Thorpe, 2005. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.* 2:S127–S136.
- [26] Jimenez-Roldan, J. E., R. B. Freedman, R. A. Roemer, and S. A. Wells, 2011. Protein flexibility explored with normal modes and geometric simulation. *Phys. Biol.* Accepted.