Published in final edited form as:

J Comput Chem. 2017 June 05; 38(15): 1138–1146. doi:10.1002/jcc.24517.

Optimal Allosteric Stabilization Sites Using Contact Stabilization Analysis

Alex Dickson*, Christopher T. Bailey[†], and John Karanicolas[‡]

*Department of Biochemistry & Molecular Biology and the Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824

[†]Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI 48824

[‡]Department of Molecular Biosciences, University of Kansas, Lawrence, KS, 66045

Abstract

Proteins can be destabilized by a number of environmental factors such as temperature, pH and mutation. The ability to subsequently restore function under these conditions by adding small molecule stabilizers, or by introducing disulfide bonds, would be a very powerful tool, but the physical principles that drive this stabilization are not well understood. The first problem lies in choosing an appropriate binding site or disulfide bond location to best confer stability to the active site and restore function. Here we present a general framework for predicting which allosteric binding sites correlate with stability in the active site. Using the Karanicolas-Brooks G⁻-like model, we examine the dynamics of the enzyme β -glucuronidase using an Umbrella Sampling method to thoroughly sample the conformational landscape. Each intramolecular contact is assigned a score termed a “stabilization factor” that measures its correlation with structural changes in the active site. We have carried out this analysis for three different scaling strengths for the intramolecular contacts, and we examine how the calculated stabilization factors depend on the ensemble of destabilized conformations. We further examine a locally destabilized mutant of β -glucuronidase that has been characterized experimentally, and show that this brings about local changes in the stabilization factors. We find that the proximity to the active site is not sufficient to determine which contacts can confer active site stability.

Graphical abstract

Proteins are prone to unfold and aggregate, especially when mutated or under stress. Stabilizing a proteins' functional state, either using engineered disulfide bonds or designed small-molecule stabilizers, is an important goal, but first we must determine an optimal stabilization site to impact activity. We use simulation to measure the stabilization capacity of every contact in the β -glucuronidase enzyme, and show that the stabilization capacity of a contact is only loosely related to its proximity from the active site.

Keywords

allostery; protein stabilization; coarse-grained model; molecular dynamics; chemical rescue; protein re-activation

INTRODUCTION

The conformational ensemble of a protein is critical to its function. As proteins naturally undergo motions on a wide range of timescales, in their lifetimes they exhibit conformations that are both compatible and incompatible with function. Stressors such as temperature, pH, ionic stress and mutation can shift the equilibrium between the functional and nonfunctional structural ensembles, destabilizing the protein. The ability to manipulate protein function by shifting this equilibrium exogenously would be valuable for both biotechnological and therapeutic purposes.

Excipients are widely used in pharmaceutical and biotechnology applications to stabilize proteins for long-term storage¹. Much effort has also been devoted to designing small molecule drugs to stabilize enzymes and transcription factors^{2–10}. As early as 1999, compounds were identified by high-throughput screening that could stabilize and restore tumor suppressor activity to the DNA binding domain of p53³. More compounds were discovered by in silico screening that bound to different sites on p53⁸, and were targeted to a specific mutation. In 2004, it was shown that formation of transthyretin amyloid fibrils was

decelerated by analogs of an FDA-approved drug (diflunisal)⁵. These compounds worked by stabilizing the ground state structure and raising the free-energy barrier to unfolding and aggregation. Despite important successes in this arena, general principles to guide design of such molecules in new cases are lacking.

The link between small molecule binding and protein stability is not straightforward. Measurements via differential scanning fluorimetry can reveal changes in the melting temperature of a protein as a function of ligand concentration^{11,12}. Recent studies employing this technique have shown that ligands can be both stabilizing and destabilizing¹³, the result of a delicate balance between the stabilizing ligand-protein interactions and instabilities within the protein that are induced by small molecule binding. Designing molecules that can stabilize an active site via an allosteric mechanism is particularly challenging, as it is difficult to predict which allosteric binding sites might best confer stability to the active site.

Coarse-grained sampling approaches have been used for many years to elucidate the fundamental physics governing protein and RNA folding^{14–19}. Simple functional forms, and the ability to implicitly approximate solvation effects, allow for the observation of long time-scale molecular events²⁰. Here we use the Karanicolas-Brooks sequence-flavored G model¹⁶, which has been applied to many protein folding systems^{21–26}. Using coarse-grained (CG) sampling in combination with other enhanced sampling techniques we can observe extremely long time-scale events: as an example, Dickson, Ahlstrom and Brooks characterized the unfolding and unbinding of the HdeA homodimer at pH 7.0, a process that normally occurs on the time-scale of hours²⁶, using Window-Exchange Umbrella Sampling (WEUS²⁷) and a CG Hamiltonian. The WEUS method uses a set of windows, each with a harmonic biasing force with a minimum at a different position along a “progress variable”. Here the progress variable is the fraction of contacts that are satisfied by the protein structure (ϕ), and the windows span from $\phi = 0$ to $\phi = 1$. Periodically, conformational exchanges are attempted between adjacent windows, which encourages relaxation orthogonal to the progress variable²⁶.

Here we use WEUS and the Karanicolas-Brooks G model to simulate protein unfolding for the enzyme β -glucuronidase (referred to hereafter as β -gluc), and use this data to measure the correlation of the formation of each contact with structural changes in the active site. β -gluc is an enzyme from the glycosidase family that catalyzes the breakdown of complex carbohydrates; it is a tetramer under physiological conditions and is comprised of three domains: a highly distorted barrel structure (residues 22–223), a β -rich domain (residues 224–342) and a C-terminal domain that forms a TIM-barrel motif and contains the active site (residues 343–632)²⁸. Using the RMSD of the active site as a proxy for the activity of the enzyme, we measure the correlation between protein activity and the formation of specific residue-residue contacts. We introduce a scoring function that reports on the strength of the correlation (the “stabilization factor”), and use this to identify and visualize the group of contacts that best re-activate protein activity given a particular destabilization. The residue W502, adjacent to the active site, has been previously mutated to glycine to demonstrate an engineered indole-rescue strategy, where the addition of indole can restore enzyme activity in a concentration-dependent manner^{29,30}. We mimic this destabilization by modulating residue-residue contact strengths, and show how this perturbation alters the local

stabilization factors. Spatial properties of the group of best stabilizing contacts are compared across different simulation conditions, and in all cases we find that proximity to the active site is not sufficient to determine the maximal potentially available allosteric stabilizing effect.

METHODOLOGY

Coarse-grained simulation

We use the coarse-grained, structure-based model of Karanicolas and Brooks¹⁶, which uses a single bead for each residue, and incorporates protein sequence information into the residue-residue contact strengths as well as the angles and dihedral angle potentials between beads in the protein chain. Solvent effects are implicit in this model. The model is built using the β -gluc structure from chain A of PDBID 3HN3 and the G₁ model builder of the MMTSB toolset¹⁶. A dihedral correction term is employed, following De Sancho and Best³¹, to correct for the over-representation of alpha helices in the data set used to build the statistical potential. We thus add $V(\phi) = k_{\phi} \cos[\phi - \delta]$ to all dihedral angle terms, where $k_{\phi} = -1.16 \text{ kcal mol}^{-1}$ and $\delta = 297.35^{\circ}$.

The nonbonded interactions are uniformly scaled by the multiplier α , which takes on values between 1.0 and 1.1 in this work. The SHAKE algorithm constrains all bonds in the system. We use a 400 Å cubic box, which is large enough to ensure the molecule does not interact with itself even when it is completely unfolded. A cutoff of 25 Å is used for the nonbonded interactions. Dynamics are performed in CHARMM³², using a 20 fs time step with a Langevin integrator and a friction coefficient of 0.2 ps^{-1} .

Window exchange umbrella sampling

To facilitate the unfolding of β -gluc, we use Umbrella Sampling³³, with the number of intramolecular contacts involving residues in the C-terminal domain that contains the active site (residues 310–608) as the progress variable, achieved using the Distance Matrix COntstraint (DMCO) facility in CHARMM³⁴. We define 50 windows along the progress variable that bias the system to exhibit a certain fraction of intramolecular contacts in the C-terminal domain (f_i), and equally space these windows between 0 (no contacts formed) and 1 (all contacts formed). The energy of the biasing force is defined:

$$W_i(\mathbf{X}) = \frac{F_B}{2} (\Phi(\mathbf{X}) - f_i)^2 \quad (1)$$

where F_B is a constant that controls the strength of the bias in units of kcal/mol, and $\Phi(\mathbf{X})$ is the observed fraction of contacts satisfied in a given conformation $\mathbf{X}$. The biasing strength used here is $F_B = 500 \text{ kcal/mol}$, which allows for a reasonable amount of overlap between adjacent windows (Figure S1). The contact fraction $\Phi(\mathbf{X})$ is defined as follows:

$$\Phi(\mathbf{X}) = \frac{1}{N_c} \sum_{i=1}^{N_c} C_i(\mathbf{X}) \quad (2)$$

where $C_i(\mathbf{X})$ measures the formation of contact i :

$$C_i(\mathbf{X}) = \frac{1}{1 + e^{k(d_i(\mathbf{X}) - c_i)}}, \quad (3)$$

and N_c is the total number of contacts in the C-terminal domain (here, 922), $d_i(\mathbf{X})$ is the distance between the residues of contact i in the conformation $\mathbf{X}$, c_i is equal to the cutoff for that contact plus 1.25 Å. The constant k is set to 20 Å⁻¹ following previous work³⁴, and results in a rapid transition between $C_i(\mathbf{X}) = 0$ and $C_i(\mathbf{X}) = 1$.

We begin the simulations with only one region active, which corresponds to the fully folded structure, and progressively initialize the remaining regions over the course of the simulation, as introduced in previous work^{26,35}. In a sampling “cycle”, we run dynamics in each active window for 200 ps, and then attempt exchanges between adjacent windows as follows. An active window (A) and a neighboring window (B) are randomly chosen. If B is active, the conformations $\mathbf{X}_A$ and $\mathbf{X}_B$ are swapped with probability:

$$P_{\text{acc}} = \min(1, \exp[\beta(V_A(\mathbf{X}_A) + V_B(\mathbf{X}_B) - V_A(\mathbf{X}_B) - V_B(\mathbf{X}_A))]) \quad (4)$$

where $V_i(\mathbf{X}_j)$ is the potential energy of conformation j in window i . Inactive regions can be activated by adjacent windows by swapping moves if the window B is not active, with probability:

$$P_{\text{act}} = \min(1, \exp[-\beta(V_A(\mathbf{X}_A) - V_B(\mathbf{X}_A))]). \quad (5)$$

We run WEUS simulations for different strengths of the intramolecular contacts, and runs where the contacts made by Trp504 are destabilized. These simulations are all performed in triplicate, and the length of individual simulations ranges from 3600 to 8700 cycles, or 27 to 84 μs of combined simulation time. The details of the entire trajectory set are summarized in Table 1. In total we run over 800 μs of sampling.

Weighted histogram analysis method

Following the WEUS simulations, the results from all windows are synthesized into global probability distributions using the Weighted Histogram Analysis Method (WHAM)^{36–38}. Given the biasing energies in each window (W_j), unbiased probabilities for each frame are then obtained using the relationship

$$P(\mathbf{X}_{i,l}) = \frac{1}{\sum_{j=1}^N n_j e^{-\beta[W_j(\mathbf{X}_{i,l}) - f_j]}} \quad (6)$$

where $\mathbf{X}_{i,l}$ is the l -th frame of the i -th simulation, N is the number of windows, n_j is the number of frames sampled in window j , and f_j is a biasing free energy for window j , obtained using the following equations for each window k :

$$e^{\beta f_k} = \frac{\sum_{i=1}^N \sum_{l=1}^{n_i} e^{-\beta W_k(\mathbf{X}_{i,l})}}{\sum_{j=1}^N n_j e^{-\beta [W_j(\mathbf{X}_{i,l}) - f_j]}}. \quad (7)$$

The set of f_k values are determined self-consistently by iteratively solving Eqs. 7, using freely available software³⁸. The probabilities for each frame from Eq. 6 can then be used to compute expectation values for observables, such as the root mean squared deviation (RMSD) of the active site and the stabilization factors, introduced below.

Stabilization factors

To quantify the impact of each residue-residue contact on maintaining the stability of the active site, we define the “stabilization factor” for contact i as:

$$S_i = \frac{100}{\langle R_a \rangle} \left(\langle R_a \rangle - \frac{\sum_k R_a(\mathbf{X}_k) p_k e^{-\alpha C_i(\mathbf{X}_k)}}{\sum_k p_k e^{-\alpha C_i(\mathbf{X}_k)}} \right) \quad (8)$$

where

$$\langle R_a \rangle = \frac{\sum_k R_a(\mathbf{X}_k) p_k}{\sum_k p_k}, \quad (9)$$

and $R_a(\mathbf{X}_k)$ is active site RMSD in conformation $\mathbf{X}_k$. The sums are over all observed frames k , p_k is the probability of frame k , C_i is defined in Eq. 3. The S_i quantities compare the expectation value of R_a with a modified expectation value that gives a higher weight to structures where contact i is formed (i.e. $C_i = 1$). Stabilization factors can be positive or negative; positive values indicate the formation of a contact contributes to the stability of the active site, and a negative value indicates that the contact formation decreases active site stability. The factor of 100 allows us to view the S_i values as the percent change to the expected RMSD upon reweighting for a given contact. The constant α controls the strength of the weighting of $C_i = 1$ states, set here to 10 which amplifies the population of states where the contact is formed by a factor of $e^{10} \approx 20000$. This can also be seen as a 10 kT boost in free energy to the $C_i = 1$ states, which was chosen to be comparable to the free energy of ligand binding³⁹, such that these states could be feasibly rescued by a ligand. In applications that are concerned with rescue of destabilized states via engineered disulfide bonds, it would be feasible to use a much larger value of the constant α .

RESULTS

β -gluc unfolding

Following previous work, we investigated intramolecular scaling strengths (α) in the neighborhood of 1.0 to 1.1^{24–26,31}. α is used to scale all nonbonded interactions in the molecule that govern both secondary and tertiary structure. The potential of mean force along the number of active site contacts is shown in Figure 1, showing three different α values, and convergence of the PMFs is shown in Figure S2. The curves here are each averaged over three simulations. Figure S2 shows that the PMF converges by 50% of the simulation length in the region $0.6 < \Phi < 1.0$, although there is still some slight movement in the $\Phi \approx 0.3$ region for $\alpha = 1.05$, which is inconsequential for our analysis below. The globally stable states for $\alpha = 1.0$ and $\alpha = 1.05$ ($\Phi = 0.745$ and $\Phi = 0.820$) show that a significant number of contacts made by the C-terminal domain are broken.

In all PMFs there is a sharp increase in free energy as more contacts are lost at $\Phi = 0.7$. For all values of α , structures at $\Phi = 0.7$ have the C-terminal domain largely intact, and have lost only the contacts between the C-terminal domain and the rest of the protein (Figure 2). The cooperative loss of structure in the C-terminal domain is strongly unfavored, with a free energy cost of 29, 44 and 81 kT, for $\alpha = 1.0$, 1.05 and 1.1, respectively. This set of PMFs reveals three globally stable basins, which we label “A”, “B” and “F”, the latter of which, globally stable at $\alpha = 1.1$ is the folded structure. We examine states “A” and “B” by clustering the results from the three simulations at each α value. Clustering is performed in MSMBuilder⁴⁰ using the complete set of contacts in the coarse-grained model. We use the KCenters algorithm to identify 2000 clusters from over > 400000 frames in each α value. For representative structures of state “A” we randomly pick structures from the highest weighted cluster with $0.73 < \Phi < 0.75$. Similarly for B we use the highest weighted cluster with $0.81 < \Phi < 0.84$.

These states are characterized by a largely folded C-terminal domain while the remainder of protein dissociates from the core (Figure 2). Across all α -values, B-type conformations preserve more secondary structure outside of the C-terminal domain than their respective A-type conformations. Only state A from $\alpha = 1.05$ displayed a loss of secondary structure in the C-terminal domain. The magnitude of secondary structure loss in the rest of the protein varies with α , with $\alpha = 1.1$ losing the least.

Contact stabilization analysis

With our diverse ensemble of both functional and non-functional conformations in hand, we can now investigate the relative stabilizing effects of each intramolecular contact. As a proxy for the activity of β -gluc, we use the RMSD of the active site residues Glu451, Tyr504 and Glu540 (R_a), as well as others in close proximity (Figure 2). Contact-specific stabilization factors (S_i) are calculated by comparing the expectation values of R_a in structures where a given contact is formed with the unaltered expectation value of R_a (see Methods). As our PMFs converge after 50% of the data is generated, we calculate the S values using only the last 50% of the simulation.

Figure 3 shows 2D probability maps along the RMSD of the active site and the distance between a given contact pair for three different contacts. The contact between residues 458 and 478 (blue) is inside the active site, but is given a low stabilization score for all three α values: $S = 0.11, 0.17$ and 0.39 for $\alpha = 1.1, 1.05$ and 1.0 , respectively. This is because in all sets of simulations, this contact is never broken, so $\langle R_d \rangle \approx \langle R_d e^{-\alpha C} \rangle$, and $S \approx 0$. In other words, the formation of this contact is not correlated with active site stability, and would thus not be a good target for stabilization. The contact between residues 155 and 395 (yellow) is broken in all sets of simulations, but its S -value varies. For $\alpha = 1.05$ the probability distribution of the RMSD of the active site skews lower when the contact is formed (at or below the red line) than when it is not formed (above the red line), resulting in a high, positive stabilization score. For $\alpha = 1.1$ there is a slight skew to the RMSD of the active site, but the magnitude of the stabilization factor is limited by the low probability of the broken state, as in the case of the blue contact. $S \approx 0$ for this contact for $\alpha = 1.0$, as the probability of forming the contact (below the red line) is extremely low. The contact between residues 394 and 428 (green) breaks in all three simulation sets, although the inter-residue distance does not exceed $20 \text{ \AA}$. In all cases, the RMSD of the active site is lower when the contact is formed. This indicates that strengthening this contact would confer stabilization to the active site. We thus predict that stabilization of this contact, either by targeting this region with a small molecule stabilizer or by an engineered disulfide bond, would improve the stability of the active site.

Figure 4 shows the stabilization factors for all contacts, mapped onto the molecular structure using colored cylinders. All three α values show some stabilizing contacts away from the active site, although this is most pronounced in $\alpha = 1.05$, where the entire interface between the C-terminal domain (orange) and the rest of the molecule is scored as highly stabilizing. This shows how collective motions can give rise to very long length-scale allosteric stabilization effects, and points to a large set of potential sites for stabilizer design. As exemplified by the yellow contact in Figure 3, this is due to both the reasonable probability of A-type and B-type structures, and the reasonable probability of structures where these contacts are formed. This results in differences between $\langle R_d \rangle$ and $\langle R_d e^{-\alpha C} \rangle$, which result in high S values. It is worth noting that these structures exist in the $\alpha = 1.1$ ensemble as well, although they are of much lower probability, and do not contribute as strongly to the expectation values used to calculate the stabilization factors. This highlights that the stabilizing potential of given contact is dependent upon not just the structures in the destabilized ensemble, but on their relative probabilities as well.

High S contacts are clustered closer to the active site for $\alpha = 1.0$ and $\alpha = 1.1$, and there is moderate correlation between the two sets of S -values (Figure S3). Interestingly, there is a set of contacts with high S under $\alpha = 1.0$, and low S under $\alpha = 1.1$. These connect the tops alpha helices along the “rim” of the C-terminal domain, most visible in the middle column of Figure 4. There is also a set of contacts that have higher S values in the $\alpha = 1.1$ state than the $\alpha = 1.0$ state. These connect the top of the active domain to the rest of the molecule, and are best visible in the rightmost column of Figure 4. These contacts also have high stabilization scores in $\alpha = 1.05$, and their importance suggests that the connection between the C-terminal domain and the distorted β -barrel domain plays a role in stabilizing the active site.

We also find that the stabilization factor (S) is poorly predicted by the active site proximity (r_a) alone. Figure 5 shows a scatter plot of these two quantities for all three α values. For $\alpha = 1.05$ there are highly stabilizing contacts that are up to 45 Å away from the active site. Although the stabilizing effect of contacts for $\alpha = 1.0$ and $\alpha = 1.1$ falls off more quickly: for $r_a > 20$ Å these S values are approximately zero. However, in all cases the fraction of contacts with low r_a that have high S values is small. To illustrate this, there are 257 contacts that are within 15 Å of the active site, and the number of these that have $S > 2.0$ is equal to 29, 35 and 21 for $\alpha = 1.0$, 1.05 and 1.1 respectively.

The difference between the S values as a function of α show that the nature of the destabilized ensemble has an impact on the relative and absolute values of the stabilization factors. To investigate this effect further we conduct simulations where all interactions involving residue 504 are destabilized, accomplished by multiplying the depths of the corresponding van der Waals energy wells by 0.01. These simulations are again performed in triplicate for $\alpha = 1.05$ and $\alpha = 1.1$. Figure 6 shows the local effects of destabilization around residue 504. In both cases there is a modest effect on the average S value of the contacts involving residue 504 (Table 2). The average stabilization factor for the 13 contacts involving W504 increases from 0.45 to 0.64 upon W504 destabilization for $\alpha = 1.05$, and increases from 0.38 to 0.74 upon W504 destabilization for $\alpha = 1.1$. This confirms the intuitive result that the importance of stabilizing a given region increases as it becomes more destabilized. It also provides another example of how the nature of the destabilizing force can affect the distribution of stabilization factors.

DISCUSSION AND CONCLUSIONS

The functional form of the stabilization factor was chosen to take into account both the free energy of contact formation and the active site RMSD probability distributions in both the contact-formed and contact-broken states. By design it will give low stabilization factors to contacts that have an extremely low chance of forming in the first place, as it is unlikely that a ligand could stabilize such a contact using only non-covalent interactions. An example of this is the low stabilization factor given to the “yellow” contact under $\alpha = 1.0$ in Figure 3. This screening effect can be tuned using the parameter α in Eq. 8, with higher values increasing the tolerance for high free-energy contacts.

Contact stabilization analysis is essentially a prediction for an allosteric mode of action that connects a given contact to the active site. There are many existing methods that can predict such allosteric couplings, which can be guided by normal mode analysis^{41,42}, energy dissipation pathways⁴³, relative configurational entropies of residue-immobilized states⁴⁴, or amino-acid sequence covariance⁴⁵. To our knowledge this is the first method that determines allosteric coupling based on explicit models of the destabilized ensemble, which has the unique ability to predict how the allosteric coupling can change with the nature of the destabilizing force

As shown above, there can be significant changes to the stabilization power of a given contact, depending on the type and strength of the destabilizing force. For a given problem, this underscores the need to accurately model the conditions of interest. This said, we have

found contacts in our model that have high stabilization factors across the set of α values examined here (e.g. the “green” contact in Figure 3). This motivates the search for contacts in other models that have allosteric stabilizing effects across a broad range of destabilizing forces. Alternatively, it could be beneficial to take advantage of selective stabilizing effects to engineer stabilizers that work only under certain sets of conditions.

In this work we use a G-like model, although the framework is easily generalizable to higher resolution models. Indeed, a limiting factor of a G-like model is that we can only look at the effects of native contacts. It would be very interesting to see stabilization factors for non-native contacts in comparison. Although here we exhaustively sample the complete unfolding landscape of β -gluc using WEUS, only those structures with a free energy $< \sim 15$ kT contribute meaningfully to the analysis. An alternative strategy using a higher resolution model could be to enhance motions along only along a set of contacts of interest, using a method such as WExplore⁴⁶, or Temperature-Accelerated Molecular Dynamics⁴⁷.

When small molecules bind to proteins, their impact is multifaceted. They can provide a direct stabilizing effect by forming interactions that bridge different parts of the protein, but they can also have a destabilizing effect by disrupting other networks of interactions. Thus, different compounds binding to a given site can have opposing effects on stability, depending on the details of the interactions: indeed, this phenomenon has already been observed experimentally¹³. Contact stabilization factors could be used in conjunction with atomic resolution binding models to score potential ligands and predict their stabilizing or destabilizing effects.

In this study we chose an enzyme with a defined active site, and we used the RMSD of the active site as a proxy for enzyme activity. Alternative strategies could involve other activity proxies, such as the solvent accessibility of certain residues, or the distance between catalytic residues. If the activity of the protein relies on the formation of an intermolecular complex with another protein, or DNA, then one could also use the free energy of states that are compatible with complex formation as a proxy. Thus, this approach may also represent an extensible strategy to search for allosteric protein-protein interaction stabilizers⁷.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

ARD and CB acknowledge support from the High Performance Computing Center at Michigan State University, and Sam Lotz for useful discussions. This work was also supported by a grant from the National Institute of General Medical Sciences of the National Institutes of Health (R01GM099959).

References

1. Kamerzell TJ, Esfandiary R, Joshi SB, Middaugh CR, Volkin DB. *Advanced Drug Delivery Reviews*. 2011; 63:1118. [PubMed: 21855584]
2. Prestrelski SJ, Tedeschi N, Arakawa T, Carpenter JF. *Biophysical journal*. 1993; 65:661. [PubMed: 7693001]
3. Foster BA, Coffey HA, Morin MJ, Rastinejad F. *Science (New York, NY)*. 1999; 286:2507.

4. Rippin TM, Bykov VJN, Freund SMV, Selivanova G, Wiman KG, Fersht AR. *Oncogene*. 2002; 21:2119. [PubMed: 11948395]
5. Adamski-Werner SL, Palaninathan SK, Sacchettini JC, Kelly JW. *Journal of Medicinal Chemistry*. 2004; 47:355. [PubMed: 14711308]
6. Wiseman RL, Johnson SM, Kelker MS, Foss T, Wilson IA, Kelly JW. *Journal of the American Chemical Society*. 2005; 127:5540. [PubMed: 15826192]
7. Block P, Weskamp N, Wolf A, Klebe G. *Proteins*. 2007; 68:170. [PubMed: 17393392]
8. Boeckler FM, Joerger AC, Jaggi G, Rutherford TJ, Veprintsev DB, Fersht AR. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:10360. [PubMed: 18650397]
9. Connelly S, Choi S, Johnson SM, Kelly JW, Wilson IA. *Current Opinion in Structural Biology*. 2010; 20:54. [PubMed: 20133122]
10. Liu X, Wilcken R, Joerger AC, Chuckowree IS, Amin J, Spencer J, Fersht AR. *Nucleic Acids Research*. 2013; 41:6034. [PubMed: 23630318]
11. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer Ba, Lane P, et al. *Journal of Biomolecular Screening*. 2001; 6:429. [PubMed: 11788061]
12. Matulis D, Kranz JK, Salemme FR, Todd MJ. *Biochemistry*. 2005; 44:5258. [PubMed: 15794662]
13. Dai R, Geders TW, Liu F, Park SW, Schnappinger D, Aldrich CC, Finzel BC. *Journal of Medicinal Chemistry*. 2015; 58:5208. [PubMed: 26068403]
14. Levitt M, Warshel a. *Nature*. 1975; 253:694. [PubMed: 1167625]
15. Liwo A, Ołdziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *Journal of Computational Chemistry*. 1997; 18:849.
16. Karanicolas J, Brooks CL III. *Journal of Molecular Biology*. 2003; 334:309. [PubMed: 14607121]
17. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. *Journal of Chemical Theory and Computation*. 2008; 4:819. [PubMed: 26621095]
18. Gopal SM, Mukherjee S, Cheng YM, Feig M. *Proteins: Structure, Function and Bioinformatics*. 2010; 78:1266.
19. Feng J, Walter NG, Brooks CL. *Journal of the American Chemical Society*. 2011; 133:4196. [PubMed: 21375305]
20. Hills RD, Brooks CL III. *International Journal of Molecular Sciences*. 2009; 10:889. [PubMed: 19399227]
21. Karanicolas J, Brooks CL III. *Protein Science*. 2002; 11:2351. [PubMed: 12237457]
22. Hills RD, Brooks CL. *Journal of Molecular Biology*. 2008; 382:485. [PubMed: 18644380]
23. Hills RD, Kathuria SV, Wallace LA, Day IJ, Brooks CL III, Matthews CR. *Journal of Molecular Biology*. 2010; 398:332. [PubMed: 20226790]
24. Ahlstrom LS, Dickson A, Brooks CL III. *Journal of Physical Chemistry B*. 2013; 117:13219.
25. Ahlstrom LS, Law SM, Dickson A, Brooks CL III. *Journal of Molecular Biology*. 2015; 427:1670. [PubMed: 25584862]
26. Dickson A, Ahlstrom LS, Brooks CL III. *Journal of Computational Chemistry*. 2015; 37:587. [PubMed: 26250657]
27. Park S, Kim T, Im W. *Physical Review Letters*. 2012; 108:108102. [PubMed: 22463457]
28. Jain S, Drendel WB, Chen ZW, Mathews FS, Sly WS, Grubb JH. *Nature structural biology*. 1996; 3:375. [PubMed: 8599764]
29. Deckert K, Budiardjo SJ, Brunner LC, Lovell S, Karanicolas J. *Journal of the American Chemical Society*. 2012; 134:10055. [PubMed: 22655749]
30. Xia Y, Diprimio N, Keppel TR, Vo B, Fraser K, Battaile KP, Egan C, Bystroff C, Lovell S, Weis DD, et al. *Journal of the American Chemical Society*. 2013; 135:18840. [PubMed: 24313858]
31. De Sancho D, Best RB. *Molecular BioSystems*. 2012; 8:256. [PubMed: 21892446]
32. Brooks B, Brooks C III, MacKerell A Jr, Nilsson L, Petrella R, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al. *Journal of Computational Chemistry*. 2009; 30:1545. [PubMed: 19444816]
33. Torrie JM, Valleau JP. *Journal of Computational Physics*. 1977; 23:187.

34. Sheinerman FB, Brooks CL III. *Journal of Molecular Biology*. 1998; 278:439. [PubMed: 9571063]
35. Dickson A, Maienschein-Cline M, Tovo-Dwyer A, Hammond JR, Dinner AR. *Journal of Chemical Theory and Computation*. 2011; 7:2710. [PubMed: 26605464]
36. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. *Journal of Computational Chemistry*. 1992; 13:1011.
37. Roux B. *Computer Physics Communications*. 1995; 91:275.
38. A. Grossfield (version 2.0.9). <http://membrane.urmc.rochester.edu/content/wham/>
39. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. *Journal of Molecular Biology*. 2007; 371:1118. [PubMed: 17599350]
40. Beauchamp KA, Bowman GR, Lane TJ, Maibaum L, Haque IS, Pande VS. *Journal of Chemical Theory and Computation*. 2011; 7:3412. [PubMed: 22125474]
41. Panjkovich A, Daura X. *BMC Bioinformatics*. 2012; 13:273. [PubMed: 23095452]
42. Kav B, Öztürk M, Kabakçiolu A. *Proteins: Structure, Function and Bioinformatics*. 2016; 84:655.
43. Ota N, Agard DA. *Journal of Molecular Biology*. 2005; 351:345. [PubMed: 16005893]
44. Kalescky R, Zhou H, Liu J, Tao P. *PLoS Computational Biology*. 2016; 12:1.
45. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:E1293. [PubMed: 22106262]
46. Dickson A, Brooks CL III. *The Journal of Physical Chemistry B*. 2014; 118:3532. [PubMed: 24490961]
47. Abrams CF, Vanden-Eijnden E. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:4961. [PubMed: 20194785]

Figure 1.

Potentials of mean force as a function of the fraction of contacts made by the C-terminal domain. Each curve is an average over three simulations, discarding the first 60% of the data in each. (Top) The complete range from $\Phi = 0$ to 1, showing the free energy of complete unfolding. (Bottom) The range from $\Phi = 0.65$ to 1.0, which shows global and metastable minima around the completely folded state. The “A” and “B” regions are shown using grey boxes.

Figure 2.
Rendering of β -gluc *A*- and *B*-type conformations for all intramolecular contact strengths.
The black spheres represent the active site, the red sphere represents Trp504, orange
represents the active site domain, and white represents the remainder of the protein.

Figure 3.

Free energy maps of active site RMSD and the inter-residue distance for three representative contacts. (Top) The location of each residue pair is shown on the β -gluc molecule. Residues 458 and 478 are shown in blue, and reside within the active site. Residues 155 and 395 are on the interface between the C-terminal domain and the rest of the protein, and are shown in yellow. Residues 394 and 428 are approximately 15 Å from the active site, but within the C-terminal domain, and are shown in green. (Bottom) Free energy maps for the three contact pairs using $\alpha = 1.0$, $\alpha = 1.05$ and $\alpha = 1.1$. Each map is averaged over three simulations. The color scales show the free energy in units of kT . In each plot the equilibrium distance of the contact is shown by a horizontal red line, and the stabilization factor is given in the top right of each plot. The colors of the plots correspond to the colors of the residue pairs above. In each map, the probability of any states outside the range of the map is added to the probability of the boxes at the border. The vertical black lines show $\langle R_a \rangle$, and the dashed red lines show the modified averages $\langle R_a e^{-aC} \rangle$.

Figure 4.

Stabilization factors for each contact are shown for all three α values. Each contact is shown on the molecular surface as a cylinder, which is colored according to its stabilization factor: red is destabilizing, white is neutral, and blue is stabilizing. S -values from the three sets of simulations are shown using different color scales in order to better see the variation in the results. The maximum in each color scale is chosen to be 50% of the highest S -value in each case. In each structure, the C-terminal domain is colored orange, and the selection of active site residues used to compute the RMSD are shown as black spheres, inside the C-terminal domain. Trp504 is shown as a red sphere. In each frame the molecule is in its crystallographic conformation.

Figure 5.

Scatter plot of stabilization factors against the distance from the active site for both $\alpha = 1.05$ and $\alpha = 1.1$. The distance from the active site is determined using the distance between the geometric center of the selection of active site residues shown in Figure 4, and the midpoint of each contact.

Figure 6.

Trp504-knockout mimics change the local stabilization factors. Interactions involving Trp504 (red sphere) show an increase in their stabilization factor in simulations where their contact strength is reduced (Δ W504). Color scales are the same as in Figure 4, and all stabilization factors are shown on top of the crystal structure conformation.

Table 1

Details of WEUS trajectory set.

Alteration	Intramolecular scaling (α)	Number of runs	Cycles	Total simulation time (μ)
None	1.0	3	8545,8730,8620	82,84,83
None	1.05	3	7672,7621,7425	69,70,69
None	1.1	3	6910,6880,7274	57,58,57
W504	1.05	3	3754,3869,3730	33,33,32
W504	1.1	3	3694,3719,3678	27,27,27

Stabilization factors for contacts involving residue 504, with () W504) and without (Normal W504) destabilization of the Trp504 contacts.

Table 2

		$\alpha = 1.05$			$\alpha = 1.1$		
From	To	Normal W504	W504	Normal W504	W504	Normal W504	W504
504	469	0.08	1.73	0.16	1.16		
504	466	1.68	2.65	0.80	2.31		
504	463	2.04	3.95	1.01	0.53		
504	508	-0.03	0.31	0.00	0.12		
504	509	-0.05	0.13	0.03	0.19		
504	475	0.06	0.07	0.03	0.25		
504	499	0.01	-0.02	0.01	0.18		
504	507	0.04	-0.10	0.01	0.26		
504	501	0.00	-0.13	0.00	0.23		
504	459	1.93	0.97	2.86	3.83		
504	500	0.04	-0.20	0.00	0.18		
504	512	0.02	-0.40	0.03	0.39		
504	477	0.11	-0.59	0.04	0.02		
Average		0.45	0.64	0.38	0.74		