

# Optimal Allosteric Stabilization Sites Using Contact Stabilization Analysis

Alex Dickson

*Department of Biochemistry & Molecular Biology,  
Michigan State University, East Lansing, MI 48824 and  
Department of Computational Mathematics,  
Science and Engineering,  
Michigan State University,  
East Lansing, MI 48824;  
correspondence to alexrd@msu.edu*

Christopher T. Bailey

*Department of Biochemistry & Molecular Biology,  
Michigan State University, East Lansing, MI 48824*

John Karanicolas

*Department of Molecular Biosciences,  
University of Kansas, Lawrence, KS, 66045*

(Dated: July 28, 2016)

## Abstract

Proteins can be destabilized by a number of environmental factors such as temperature, pH and mutation. The ability to restore function by small molecule stabilizers, or the introduction of disulfide bonds, would be a very powerful tool, but the physical principles that drive this stabilization are not well understood. The first problem lies is in choosing an appropriate binding site or disulfide bond location that will 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<sub>0</sub>-like model, we examine the dynamics of the glycosidase enzyme beta-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. This is done for three different scaling strengths for the intramolecular contacts, and we examine how the calculated stabilization factors depend on the makeup of the ensemble of destabilized conformations. We further examine a locally destabilized mutant of beta-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.

## I. INTRODUCTION

The conformational state 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 the 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 storage<sup>1</sup>. Much effort has also been devoted to the design of small molecule drugs that can stabilize enzymes and transcription factors<sup>2–10</sup>. 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 p53<sup>3</sup>. More compounds were discovered by in silico screening that bound to different sites on p53<sup>8</sup>, and were targeted to a specific mutation. In 2004, it was shown that the formation of transthyretin amyloid fibrils could be kinetically stabilized by analogs of an FDA-approved drug (diflunisal)<sup>5</sup>. These compounds worked by stabilizing the ground state structure and raising the barrier to unfolding and aggregation. Despite successes in the area of small molecule stabilization, general principles that can guide the design of these molecules in new cases is 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<sup>11,12</sup>. Recent studies employing this technique have shown that ligands can be both stabilizing and destabilizing<sup>13</sup>, the result being a delicate balance between the stabilizing ligand-protein interactions and instabilities within the protein that are induced by small molecule binding. The design of molecules that can stabilize an active site via an allosteric mechanism is particularly challenging, as it is difficult to predict which allosteric binding sites would 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<sup>14–19</sup>. Simple functional forms, and the ability to implicitly approximate solvation effects allow for the observation of long time-

scale molecular events<sup>20</sup>. Here we use the Karanicolas-Brooks sequence-flavored Gō model<sup>16</sup>, which has been applied to many protein folding systems<sup>21–26</sup>. 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 hours<sup>26</sup>, using Window-Exchange Umbrella Sampling (WEUS<sup>27</sup>) 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 “reaction coordinate”. Here the reaction coordinates is the fraction of contacts that are satisfied by the protein structure ( $\phi$ ), and the windows span from  $\phi = 0$  to  $\phi = 1$ . Periodically, conformational exchanges are attempted between adjacent windows, which encourages relaxation orthogonal to the reaction coordinate<sup>26</sup>.

Here we use WEUS and the Karanicolas-Brooks Gō model to simulate protein unfolding for the enzyme beta glucuronidase (referred to hereafter as  $\beta$ -gluc), and use this data to measure the correlation of the formation of each contact wit structural changes in the active site.  $\beta$ -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 beta-rich domain (residues 224-342) and a C-terminal domain that forms a TIM-barrel motif and contains the active site (residues 343-632)<sup>28</sup>. 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 the function of the protein in a concentration-dependent manner<sup>29,30</sup>. 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 allosteric stabilizing effect.

## II. METHODS

### A. Coarse-grained simulation

We use the coarse-grained, structure-based model of Karanicolas and Brooks<sup>16</sup>, 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. The model is built using the  $\beta$ -gluc structure from chain A of PDBID 3HN3 and the Gō model builder of the MMTSB toolset<sup>16</sup>. A dihedral correction term is employed, following De Sancho and Best<sup>31</sup>, 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  $k_\phi = -1.16$  kcal mol<sup>-1</sup> and  $\delta = 297.35^\circ$ .

The nonbonded interactions are uniformly scaled by the multiplier  $\alpha$ , 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 CHARMM<sup>32</sup>, using a 20 fs time step with a Langevin integrator and a friction coefficient of 0.2 ps<sup>-1</sup>.

### B. Window exchange umbrella sampling

To facilitate the unfolding of  $\beta$ -gluc, we use Umbrella Sampling<sup>33</sup>, with the number of intramolecular contacts involving residues in the C-terminal domain that contains the active site (residues 310-608) as the reaction coordinate, achieved using the Distance Matrix COnstraint (DMCO) facility in CHARMM<sup>34</sup>. We define 50 windows along the reaction coordinate 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 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, and the constant  $k$  is set to 20 Å<sup>-1</sup>.

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<sup>26,35</sup>. 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 I. In total we run over 800 μs of sampling.

### C. 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)<sup>36–38</sup>.

TABLE I: Details of WEUS trajectory set.

Alteration	Intramolecular scaling ( $\alpha$ )	Number of runs	Cycles	Total simulation time ( $\mu$ s)
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
$\Delta W_{504}$	1.05	3	3754,3869,3730	33,33,32
$\Delta W_{504}$	1.1	3	3694,3719,3678	27,27,27

Given the biasing energies in each window ( $W_i$ ), 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  $k$ :

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

The set of  $f_k$  values are determined self-consistently by iteratively solving Eqs. 7, using freely available software<sup>38</sup>. 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.

#### D. 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:

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

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, and  $a$  is a constant, set here to

10. 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 the percent change to the expected RMSD upon reweighting for a given contact.

### III. RESULTS

#### A. $\beta$ -gluc unfolding

Following previous work, we investigated intramolecular scaling strengths ( $\alpha$ ) in the neighborhood of 1.0 to 1.1<sup>24–26,31</sup>.  $\alpha$  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  $\alpha$  values, and convergence of the PMFs is shown in Figure S1. The curves here are each averaged over three simulations. Figure S1 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  $\alpha$ , 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  $\alpha$  value. Clustering is performed in MSMBuilder<sup>39</sup> 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  $\alpha$  value. For representative structures of state “A” we randomly pick structures from the highest

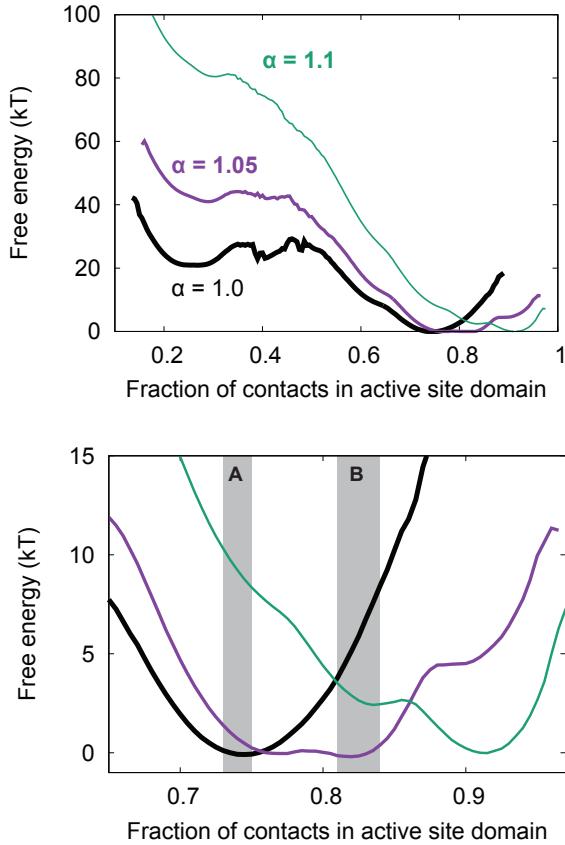


FIG. 1: Potentials of mean force as a function of the fraction of contacts made by the C-terminal domain. Each curve is an average over three simulations, discarding the first 60% of the data in each. (Top) The complete range from  $\Phi = 0$  to 1, showing the free energy of complete unfolding. (Bottom) 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.

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 with a largely folded C-terminal domain while the remainder of protein dissociates from the core (Figure 2). Across all  $\alpha$ -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  $\alpha$ , with  $\alpha = 1.1$  losing the least.

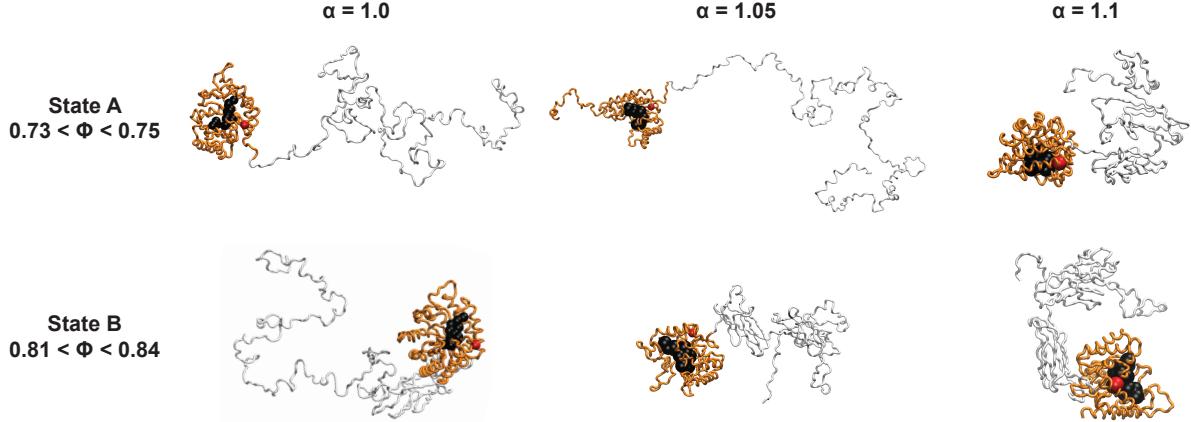


FIG. 2: Rendering of  $\beta$ -gluc *A*- and *B*-type conformations for all intramolecular contact strengths. The black spheres represent the active site, the red sphere represents Trp504, orange represents the active site domain, and white represents the remainder of the protein.

### B. 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  $\beta$ -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  $\alpha$  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  $\langle R_a \rangle \approx \langle R_a e^{-aC} \rangle$ , 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 Å. 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  $\alpha$  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  $\langle R_a \rangle$  and  $\langle R_a e^{-\alpha C} \rangle$ , 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 S2). 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

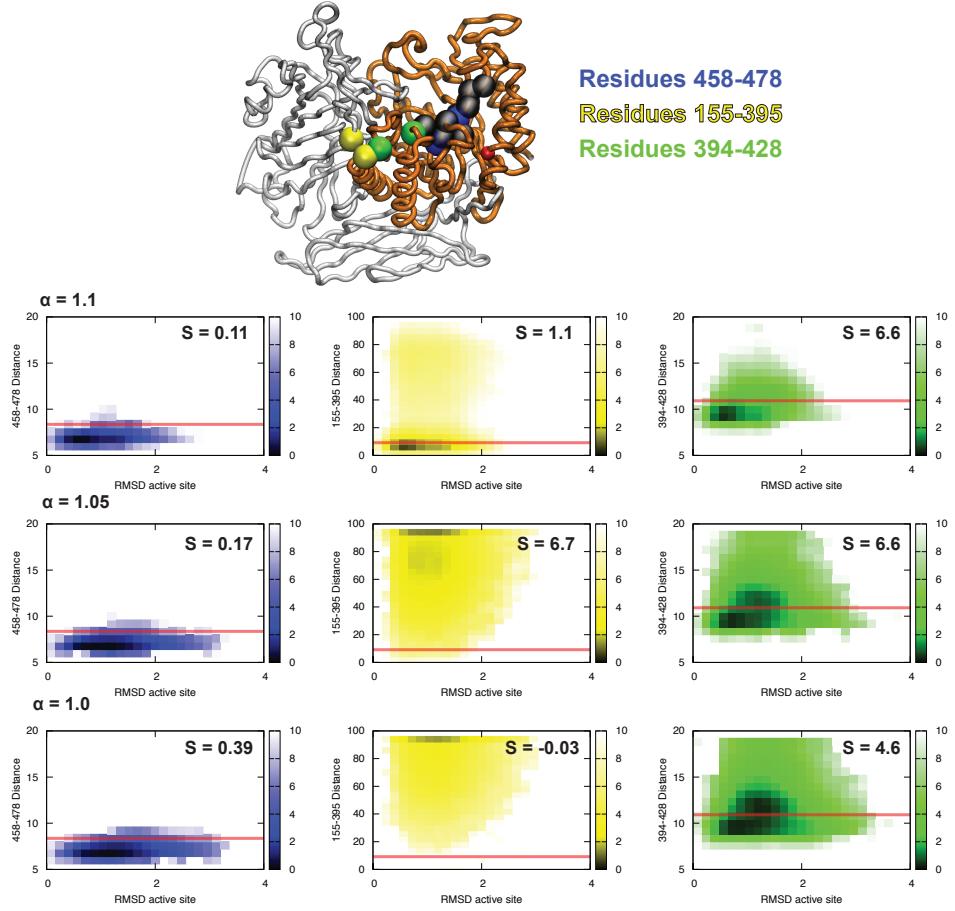


FIG. 3: Free energy maps of active site RMSD and the inter-residue distance for three representative contacts. (Top) The location of each residue pair is shown on the  $\beta$ -gluc molecule. Residues 458 and 478 are shown in blue, and residue within the active site. Residues 155 and 395 are on the interface 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 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 beta 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  $\alpha$  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  $\alpha$  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 II). The average stabilization factor for the 13 contacts involving W504 increases from 0.45 to 0.64 upon  $\Delta W504$  destabilization for  $\alpha = 1.05$ , and increases from 0.38 to 0.74 upon  $\Delta W504$  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.

#### IV. DISCUSSION

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

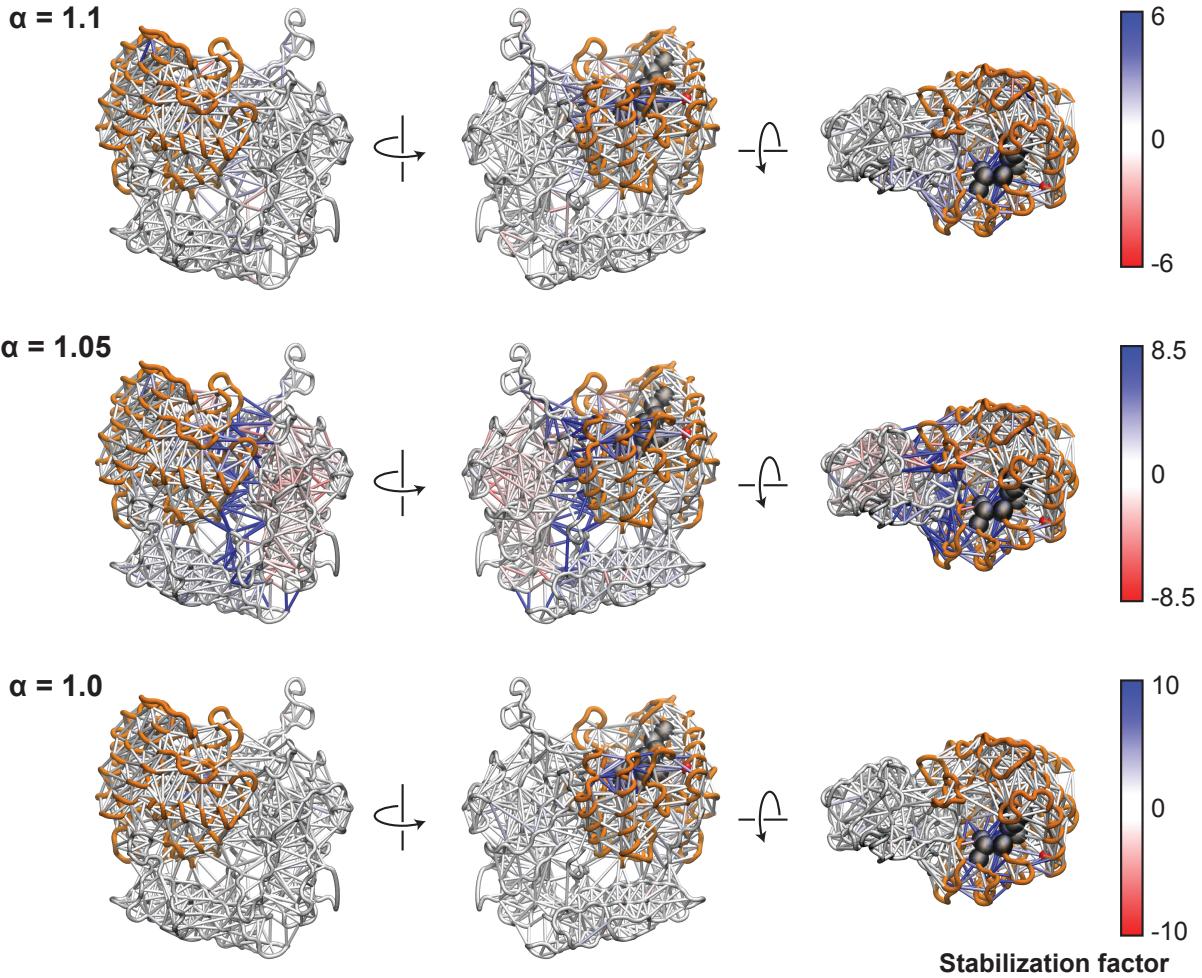


FIG. 4: Stabilization factors for each contact are shown for all three  $\alpha$  values. Each contact is shown on the molecular surface as a cylinder, which is colored according to its stabilization factor: red is destabilizing, white is neutral, and blue is stabilizing.  $S$ -values from the three 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.

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  $a$  in Eq. 8, with higher values

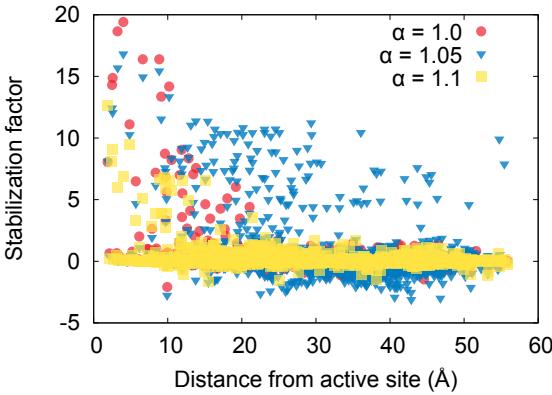


FIG. 5: Scatter plot of stabilization factors against the distance from the active site for both  $\alpha = 1.05$  and  $\alpha = 1.1$ . The distance from the active site is determined using the distance between the geometric center of the selection of active site residues shown in Figure 4, and the midpoint of each contact.

increasing the tolerance for high free-energy contacts.

As shown above, the stabilization power of a given contact depends 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  $\alpha$  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 Gō-like model, although the framework is easily generalizable to higher resolution models. Indeed, a limiting factor of a Gō-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  $\beta$ -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 WExplore<sup>40</sup>, or Temperature-Accelerated Molecular Dynamics<sup>41</sup>.

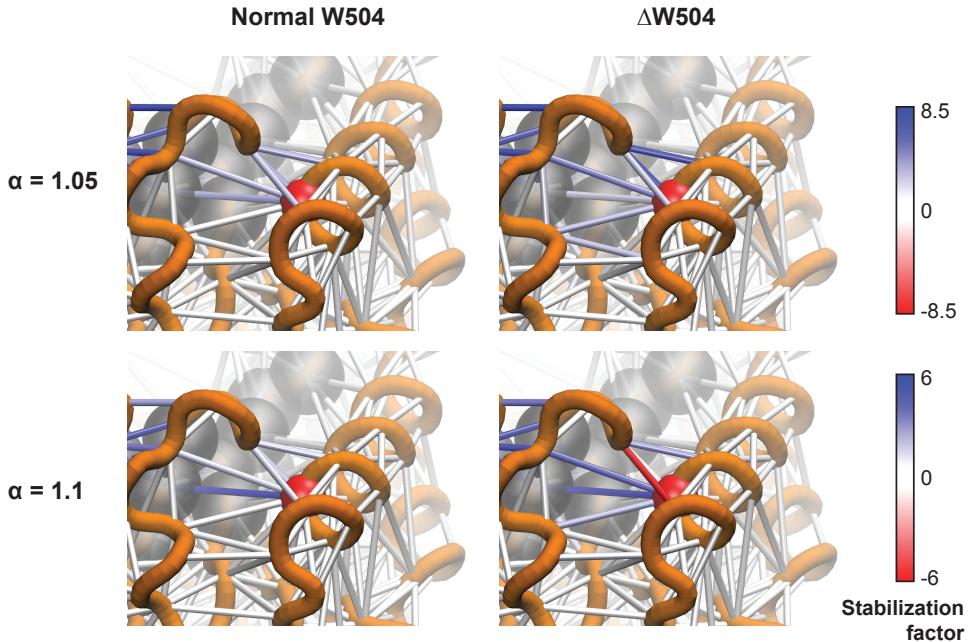


FIG. 6: Trp504-knockout mimics change the local stabilization factors. Interactions involving Trp504 (red sphere) show an increase in their stabilization factor in simulations where their contact strength is reduced ( $\Delta W504$ ). Color scales are the same as in Figure 4, and all stabilization factors are shown on top of the crystal structure conformation.

As small molecule stabilizers bind, their impact is multifaceted. Although they can provide a direct stabilizing effect by forming interactions that bridge different parts of the protein, they can also have a destabilizing effect by disrupting hydrogen bond networks. Thus, different compounds binding to a given site could have opposing effects on stability, depending on the details of the interactions. Indeed, this phenomenon has already been observed experimentally<sup>13</sup>. Contact stabilization factors could be used in conjunction with atomic resolution binding models to score potential ligands and predict their stabilizing or destabilizing effects.

This example used 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

TABLE II: Stabilization factors for contacts involving residue 504, with ( $\Delta W504$ ) and without (Normal W504) destabilization of the Trp504 contacts.

		$\alpha = 1.05$		$\alpha = 1.1$	
From	To	Normal W504	$\Delta W504$	Normal W504	$\Delta W504$
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

that are compatible with complex formation as a proxy. Thus, this presents a strategy to search for allosteric protein-protein interaction stabilizers as well<sup>7</sup>.

## V. ACKNOWLEDGEMENTS

ARD and CB acknowledge support from the High Performance Computing Center at Michigan State University, and Sam Lotz for useful discussions.

---

<sup>1</sup> T. J. Kamerzell, R. Esfandiary, S. B. Joshi, C. R. Middaugh, and D. B. Volkin, Advanced Drug Delivery Reviews **63**, 1118 (2011).

- <sup>2</sup> S. J. Prestrelski, N. Tedeschi, T. Arakawa, and J. F. Carpenter, *Biophysical journal* **65**, 661 (1993).
- <sup>3</sup> B. A. Foster, H. A. Coffey, M. J. Morin, and F. Rastinejad, *Science (New York, N.Y.)* **286**, 2507 (1999).
- <sup>4</sup> T. M. Rippin, V. J. N. Bykov, S. M. V. Freund, G. Selivanova, K. G. Wiman, and A. R. Fersht, *Oncogene* **21**, 2119 (2002).
- <sup>5</sup> S. L. Adamski-Werner, S. K. Palaninathan, J. C. Sacchettini, and J. W. Kelly, *Journal of Medicinal Chemistry* **47**, 355 (2004).
- <sup>6</sup> R. L. Wiseman, S. M. Johnson, M. S. Kelker, T. Foss, I. A. Wilson, and J. W. Kelly, *Journal of the American Chemical Society* **127**, 5540 (2005).
- <sup>7</sup> P. Block, N. Weskamp, A. Wolf, and G. Klebe, *Proteins* **68**, 170 (2007).
- <sup>8</sup> F. M. Boeckler, A. C. Joerger, G. Jaggi, T. J. Rutherford, D. B. Veprintsev, and A. R. Fersht, *Proceedings of the National Academy of Sciences of the United States of America* **105**, 10360 (2008).
- <sup>9</sup> S. Connelly, S. Choi, S. M. Johnson, J. W. Kelly, and I. A. Wilson, *Current Opinion in Structural Biology* **20**, 54 (2010).
- <sup>10</sup> X. Liu, R. Wilcken, A. C. Joerger, I. S. Chuckowree, J. Amin, J. Spencer, and A. R. Fersht, *Nucleic Acids Research* **41**, 6034 (2013).
- <sup>11</sup> M. W. Pantoliano, E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. a. Springer, P. Lane, et al., *Journal of Biomolecular Screening* **6**, 429 (2001).
- <sup>12</sup> D. Matulis, J. K. Kranz, F. R. Salemme, and M. J. Todd, *Biochemistry* **44**, 5258 (2005).
- <sup>13</sup> R. Dai, T. W. Geders, F. Liu, S. W. Park, D. Schnappinger, C. C. Aldrich, and B. C. Finzel, *Journal of Medicinal Chemistry* **58**, 5208 (2015).
- <sup>14</sup> M. Levitt and a. Warshel, *Nature* **253**, 694 (1975).
- <sup>15</sup> A. Liwo, S. Ołdziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, *Journal of Computational Chemistry* **18**, 849 (1997).
- <sup>16</sup> J. Karanicolas, C. L. Brooks, and C. L. Brooks III, *Journal of Molecular Biology* **334**, 309 (2003).
- <sup>17</sup> L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tielemans, and S. J. Marrink, *Journal of Chemical Theory and Computation* **4**, 819 (2008).
- <sup>18</sup> S. M. Gopal, S. Mukherjee, Y. M. Cheng, and M. Feig, *Proteins: Structure, Function and*

- Bioinformatics **78**, 1266 (2010).
- <sup>19</sup> J. Feng, N. G. Walter, and C. L. Brooks, Journal of the American Chemical Society **133**, 4196 (2011).
- <sup>20</sup> R. D. Hills and C. L. Brooks III, International Journal of Molecular Sciences **10**, 889 (2009).
- <sup>21</sup> J. Karanicolas and C. L. Brooks III, Protein Science **11**, 2351 (2002).
- <sup>22</sup> R. D. Hills and C. L. Brooks, Journal of Molecular Biology **382**, 485 (2008).
- <sup>23</sup> R. D. Hills, S. V. Kathuria, L. A. Wallace, I. J. Day, C. L. Brooks III, and C. R. Matthews, Journal of Molecular Biology **398**, 332 (2010).
- <sup>24</sup> L. S. Ahlstrom, A. Dickson, and C. L. Brooks III, Journal of Physical Chemistry B **117**, 13219 (2013).
- <sup>25</sup> L. S. Ahlstrom, S. M. Law, A. Dickson, and C. L. Brooks III, Journal of Molecular Biology **427**, 1670 (2015).
- <sup>26</sup> A. Dickson, L. S. Ahlstrom, and C. L. Brooks III, Journal of Computational Chemistry **37**, 587 (2015).
- <sup>27</sup> S. Park, T. Kim, and W. Im, Physical Review Letters **108**, 108102 (2012).
- <sup>28</sup> S. Jain, W. B. Drendel, Z. W. Chen, F. S. Mathews, W. S. Sly, and J. H. Grubb, Nature structural biology **3**, 375 (1996).
- <sup>29</sup> K. Deckert, S. J. Budiardjo, L. C. Brunner, S. Lovell, and J. Karanicolas, Journal of the American Chemical Society **134**, 10055 (2012).
- <sup>30</sup> Y. Xia, N. Diprimio, T. R. Keppel, B. Vo, K. Fraser, K. P. Battaile, C. Egan, C. Bystroff, S. Lovell, D. D. Weis, et al., Journal of the American Chemical Society **135**, 18840 (2013).
- <sup>31</sup> D. De Sancho and R. B. Best, Molecular BioSystems **8**, 256 (2012).
- <sup>32</sup> B. Brooks, C. Brooks III, A. MacKerell Jr, L. Nilsson, R. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., Journal of Computational Chemistry **30**, 1545 (2009).
- <sup>33</sup> J. M. Torrie and J. P. Valleau, Journal of Computational Physics **23**, 187 (1977).
- <sup>34</sup> F. B. Sheinerman and C. L. Brooks III, Journal of Molecular Biology **278**, 439 (1998).
- <sup>35</sup> A. Dickson, M. Maienschein-Cline, A. Tovo-Dwyer, J. R. Hammond, and A. R. Dinner, Journal of Chemical Theory and Computation pp. 2710–2720 (2011).
- <sup>36</sup> S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, Journal of Computational Chemistry **13**, 1011 (1992).
- <sup>37</sup> B. Roux, Computer Physics Communications **91**, 275 (1995).

- <sup>38</sup> A. Grossfield (version 2.0.9), URL <http://membrane.urmc.rochester.edu/content/wham/>.
- <sup>39</sup> K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque, and V. S. Pande, Journal of Chemical Theory and Computation **7**, 3412 (2011).
- <sup>40</sup> A. Dickson, C. Brooks, and C. L. Brooks III, The Journal of Physical Chemistry B **118**, 3532 (2014).
- <sup>41</sup> C. F. Abrams and E. Vanden-Eijnden, Proceedings of the National Academy of Sciences of the United States of America **107**, 4961 (2010).