Exposons exploit cooperative changes in solvent exposure to detect cryptic allosteric sites and other functionally-relevant conformational transitions

Justin R. Porter*[1], Katelyn E. Moeder*[1], Carrie A. Sibbald[1], Maxwell I. Zimmerman[1], Kathryn M. Hart[2], Michael J. Greenberg[1], Gregory R. Bowman[1,3,†]

* these authors contributed equally to this work

[1] Department of Biochemistry & Molecular Biophysics, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110, United States

[2] Department of Chemistry, Williams College, 880 Main Street, Williamstown, MA 01267

[3] Department of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in St. Louis, One Brookings Drive, St. Louis, Missouri 63130, United States

[†]g.bowman@wustl.edu

Abstract

Conformational changes can dramatically alter a protein's function by changing the surfaces that are accessible to interact with binding partners. However, it is often difficult to hone in on the most relevant conformational changes from the cartesian coordinates of atoms on the protein's surface. Instead, we describe a protein's surface in terms of groups of residues that undergo cooperative changes in their solvent exposure. We term these groups *exposons*. We demonstrate that Markov state models (MSMs) elegantly identify the conformational transitions that give rise to an exposon, enabling users to rapidly find the most interesting conformational changes in their system. For example, this approach readily identifies previously-known cryptic allosteric sites and other functionally-relevant conformational transitions. Moreover, it predicts a cryptic allosteric site in an important target for combating antibiotic resistance that lacks known cryptic pockets. Experimental tests confirm that targeting this site reduces catalytic efficiency 15-fold.

Introduction

Proteins are dynamic molecules that can undergo conformational changes to a broad spectrum of different excited states. A growing body of evidence suggests that reweighting the relative populations of these different structural states by post-

translational modification, protein-protein interaction, or the addition of a ligand is a common regulatory mechanism in biological systems.[1] Manipulating these conformational equilibria presents powerful opportunities for drug design, including by the design of small-molecule modulators of protein-protein interfaces[2] or cryptic pockets that open when a protein fluctuates to an excited state can exert allosteric control over distant functional sites (Fig. 1a and 1b). Identifying these cryptic sites could offer new druggable sites on established drug targets, provide a means to inhibit proteins that are currently considered undruggable, or enable the enhancement of desirable activities.[3,4] Raltegravir, for example, is a first-line treatment for HIV infection[5] that inhibits HIV integrase by binding a cryptic pocket. However, the population of a given excited state is often too low to detect experimentally unless a binding partner that stabilizes it is present. Therefore, a systematic means to identify functionally-

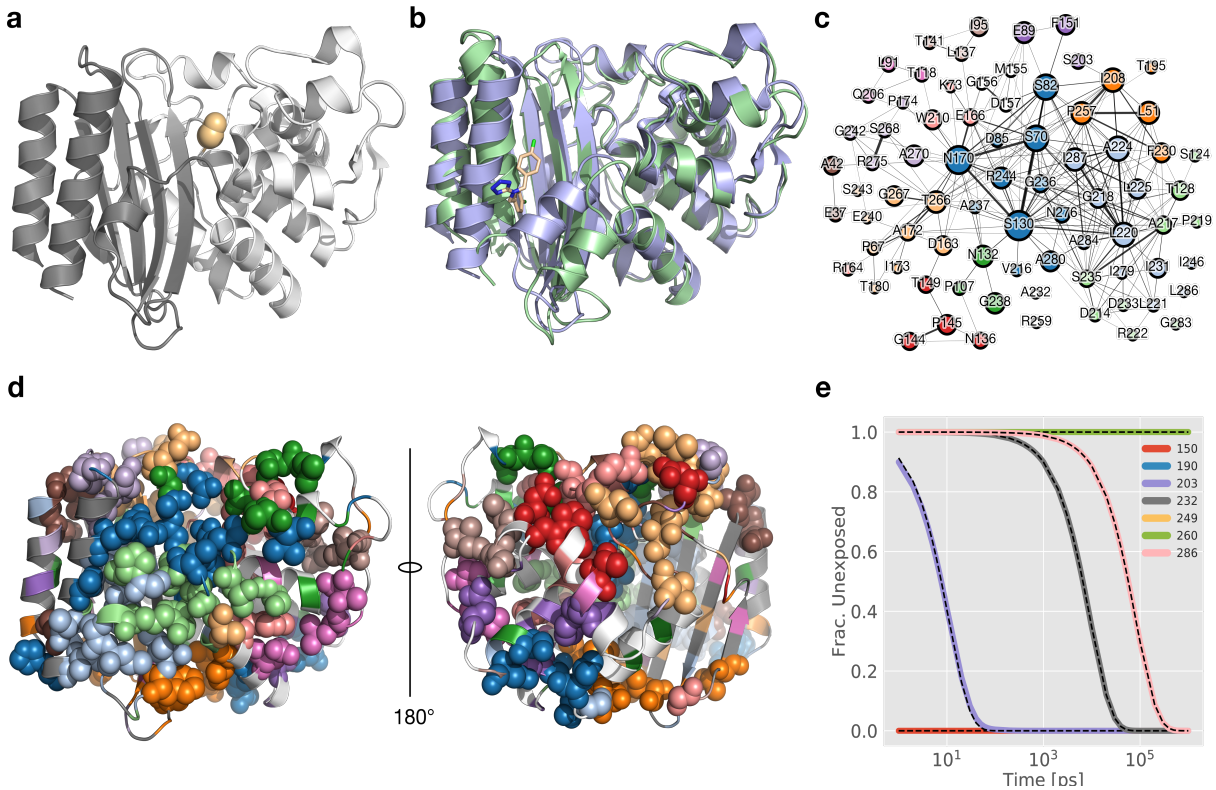

Figure 1: Exposons capture cryptic pockets in TEM-1 β-lactamase. a) The crystal structure 1JWP[48] of unliganded TEM-1 β-lactamase. For reference, the alpha helical domain is shown in white and the beta sheet domain is shown in dark grey. The catalytic serine is shown in gold spheres. b) A comparison of the crystal structure 1PZO[18] (blue) including the inhibitor CBT (yellow) to the open structure for the relevant exposon (green) extracted from our MSM. c) An embedding of the graph derived from the mutual information matrix. Each color denotes a specific exposon. For readability, edges with low mutual information or low normalized mutual information are filtered out. d) The location of exposons superimposed on the unliganded structure of TEM-1. Residues with high total information (sum of the MI matrix row) are shown in spheres. Colors match Fig 1c. e) Synthetic labeling traces showing the fraction of equilibrium population that has never exposed over time in our MSM (solid lines) along with fits to a single decaying exponential (dashed lines) for key residues.

relevant conformational transitions to excited states in the absence of stabilizing interactions could provide many biological insights and new therapeutic opportunities.

Computer simulations are a promising means to identify functionally-relevant conformational changes, including fluctuations to excited states with cryptic pockets,[6,7] but they have historically faced severe limitations. Indeed, the design of raltegravir was based on simulations that revealed the cryptic trench where the drug binds.[8] However, the difficulty of simulating rare events, such as the opening of many cryptic pockets, has hampered the routine discovery of functionally-relevant conformational changes. Moreover, even when simulations can access the relevant timescales, it is often unclear how to analyze the data to extract the most useful information. For example, the leading pocket detection algorithms over the past two decades have defined pockets as concave openings on the surface of a protein.[7,9,10] This approach is intuitive and logical, especially for the original application of analyzing individual crystal structures. However, applying them to trajectories requires numerous subjective choices. Examples include determining if openings in different structures are the same pocket[10] and deciding whether two openings separated by a thin channel constitute one or two pockets. As a result, these algorithms often miss valuable features, including cryptic extensions of enzyme active sites, which may be extremely useful for drug design. Similar difficulties arise when searching for other types of conformational changes—although it must be noted that substantial progress has been made in recent years in developing tools that leverage clustering algorithms[10] or dynamical information[11,12] to address this challenge with increasing levels of sophistication.

Here, we reasoned that conformational changes that dramatically alter the surfaces available to interact with different binding partners are likely to be functionally-relevant. To identify these conformational changes, we characterize a protein's surface in terms of groups of residues that undergo cooperative changes in their solvent exposure. We call these sets of residues *exposons*. In the case of cryptic pockets, exposons arise because pocket opening concomitantly increases the solvent exposure of surrounding residues while closing a pocket simultaneously reduces their exposure. Focusing on exposons greatly increases the efficiency and robustness of pocket identification. For example, there is rarely a need to adjust parameters, so new pockets can be identified quickly with next to no manual intervention. Because exposons exist in sequence space, the results are insensitive to structural alignments and can be easily compared with experimental techniques that provide a read-out at the primary structural level, like hydrogen-deuterium exchange or side-chain labeling experiments. Exposons also identify features that are often missed by alternative methods, such as cryptic extensions to enzyme active sites.

To demonstrate the value of this new perspective on characterizing proteins' surface chemistry, we present an integrative method that combines simulations and experiments to identify cryptic allosteric sites. First, we use Markov state models (MSMs)[13] to model a protein's conformational heterogeneity. MSMs allow us to capture slow conformational changes by integrating many independent simulations into a map

of a protein's conformational space consisting of a set of structural states and the probabilities of hopping between them.[14,15] We then use these models to identify exposons and the structural fluctuations that create them. Finally, we test our predictions by using our models to design thiol labeling experiments, which we conduct using a stopped-flow instrument to ensure access to the relevant timescales.

We demonstrate the utility of this approach for identifying cryptic allosteric sites by applying it to two homologous β-lactamase enzymes, TEM-1 and CTX-M-9. TEM-1 and CTX-M-9 are of great medical importance because bacterial expression of β-lactamases is one of the most common sources of antibiotic resistance.[16] In the United States alone, over 2 million people contract antibiotic-resistant infections every year, resulting in tens of thousands of deaths and billions of dollars in healthcare costs.[17] Frighteningly, resistance is on the rise, and there are few new antibiotics in the development pipeline, because many pharmaceutical companies are abandoning antibiotic development.[18] Therefore, there is a great need for novel antibiotic strategies to avoid a post-antibiotic era in which routine infections are life-threatening events.[19] Besides their medical relevance, we chose these systems because TEM has known cryptic pockets to test our approach on[20-22] while CTX-M-9 presents a new challenge, as it has no known cryptic pockets.

To demonstrate the broad applicability of our approach, we also present evidence that our methodology detects other functionally-relevant conformational changes in the Ebola virus' nucleoprotein (NP). Ebola is an RNA virus responsible for a severe hemorrhagic fever with high case-fatality rates.[23] NP coats the viral genome to protect it from the host, but must release it for transcription of the viral genome to occur. Recent work suggests conformational dynamics controls NP's RNA-binding affinity, presenting a potential target for pharmacotherapy.[24]

Results

Exposons identify known cryptic pockets. To identify exposons, we first featurize each conformation in a set of simulations by expressing each of these data points as a binary vector with each element representing whether or not a sidechain is exposed. Extending previous work,[22] sidechains are classified as exposed if their SASA to a drug-sized probe exceeds 2 $Å^2$. We quantify the correlations between the solvent exposure of each pair of residues by calculating the mutual information (MI) between them. Mutual information—defined in Eq 1 of *Methods*—is a non-linear measure of the correlation between two random variables that has been previously used in studies of protein allostery.[25,26] The matrix of all pairwise mutual informations is sparse (Fig S1-S3) and can be represented as a graph (Fig 1c). Finally, we cluster this MI matrix using affinity propagation.[27] The resulting clusters are termed *exposons*.

In support of the validity of our approach, we identify exposons corresponding to known cryptic pockets in TEM-1 β-lactamase (Fig 1d). For example, the pale blue

exposon corresponds to the binding site of a known allosteric ligand.[20] The orange exposon corresponds to a second cryptic pocket we previously discovered.[22]

MSMs identify the motions that give rise to exposons. An MSM's top eigenmodes capture how much each conformational state participates in the slowest motions observed in a simulation, so we reasoned that they would provide a facile means to identify the dominant open form of each exposon. To identify which eigenmode reports on a particular exposon, we first compute the degree to which changes in an exposon's solvent accessible surface area (SASA) are correlated with each eigenmotion (Fig S4) in an approach similar to dynamical fingerprinting.[28] We choose the eigenmode that maximizes this correlation and extract its extreme structures for visual examination (Fig 1b, green).

To test that our strategy for extracting high-population, realistic configurations for open exposon conformations is reasonable, we compared the known cryptic pocket's ligand-bound crystal structure[20] to the open state for that exposon. We find that the configuration identified by this approach is even more open than the bound crystal structure (Fig 1b), consistent with previous evidence that the pocket opens even further in solution than is seen in the crystal structure.[21] Examining structures along the highest flux pathway[29] to the open conformation reveals structures that are more similar to the ligand-bound crystal structure (Fig S5).

MSMs retrodict trends in thiol labeling rates. We leverage our model's residue-level predictions to choose a small number of specific positions to test with our thiol labeling technique, which we have improved upon from previous work[22] by the use of a stopped-flow instrument. Our thiol labeling assay uses a drug-sized labeling reagent, DTNB (Ellman's reagent) that changes absorbance upon covalently reacting with solvent-exposed cysteines. Thus, it provides a time-resolved measurement of residue-level solvent exposure with millisecond resolution. To make the comparison between our MSM and our thiol labeling experiment, we also developed a 'synthetic labeling' calculation. Synthetic labeling gives, as a function of time, the fraction of the population that is exposing the relevant sidechain to solvent for the first time. This is analogous to the way absorbance reports the amount of protein first exposing to the thiol label over time in thiol labeling experiments.

We find that synthetic labeling consistently discriminates pocket and non-pocket residues, matching *in vitro* results. As predicted, residues that remain buried in our experiments (L190, I260) do not show synthetic labeling, a surface control (A150) labels immediately in both our experiments and synthetic labeling, and experimentally-confirmed pocket residues (S203, A232, L286) label at intermediate rates in our model (Fig 1e). Furthermore, amongst residues that label at an intermediate rate, residues that label faster *in vitro* also label faster *in silico*. The main discrepancy between our synthetic and experimental labeling occurs at S249, which labels *in vitro* but not *in silico*, likely because finite sampling prevented us from ever observing an exposed state for this residue. Our predicted rates are also substantially faster than the *in vitro*

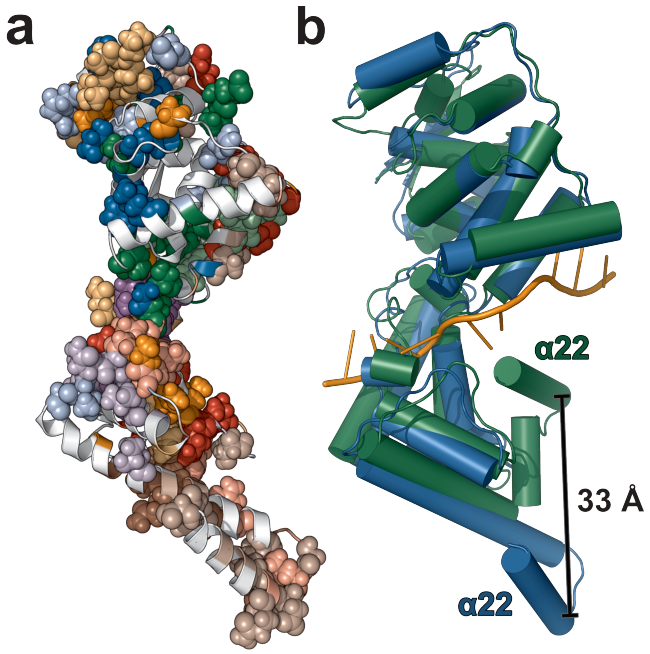

Figure 2: Exposons recapitulate conformational transitions in the NP protein from Ebola. a) The exposons of NP superimposed on a crystal structure. The pink and dark brown exposons are dispersed in space. b) The extremes of the eigenmotion best correlated with residues in the pink exposon. This motion of alpha helix 22 (indicated as α22) is thought to be important for RNA binding .

results. This discrepancy is partly due to the fact that only a sub-population of open structures will have an appropriate geometry for covalently reacting with our labeling reagent. Systematic force field errors may also contribute. Regardless, the ability to predict these trends and distinguish pockets from non-pockets is a strong testament to the predictive power of our models.

Exposons identify other functionally-relevant conformational transitions at protein surfaces. To demonstrate that exposons are valuable for a broad range of applications, we applied our analysis pipeline to the NP protein from Ebola, which has no homology to β-lactamase whatsoever. In Figure 2a, we identified an exposon (pink) that, unlike the cryptic allosteric sites identified above, does not consist of a contiguous group of residues in the available x-ray structure.[30] As shown in Figure 2b, this exposon reports not on a cooperative opening motion, but on a collective curling of the terminal helices into the RNA-binding cleft. This motion was recently identified as important for controlling NP's RNA-binding affinity.[24] Crucially, this dynamic process is consistent with hydrogen-deuterium exchange data that cannot be accounted for using available cryoelectron microscopy structures.[24] Manipulating this conformational equilibria with small molecules or peptides could provide a powerful means to inhibit Ebola. Indeed, a peptide that binds this interface has already been found to inhibit viral replication.[24,30]

Discovery of a novel cryptic allosteric site in TEM-1. To demonstrate our method's predictive power, we evaluated whether other exposons could reveal novel cryptic pockets in TEM-1. One novel exposon (Fig. 1d, beige) was particularly appealing because it is in the vicinity of the Ω-loop, a structural element containing residues absolutely required for enzymatic activity.[31] Therefore, we expect a small molecule that

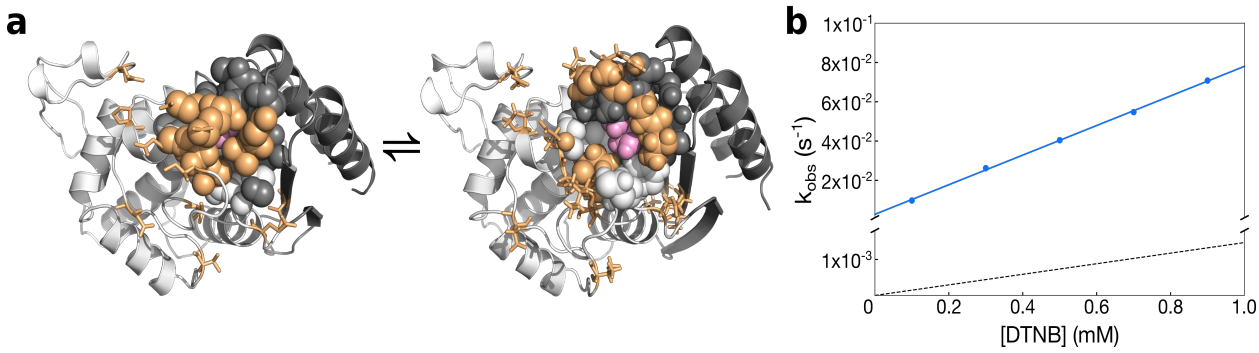

Figure 3: Exposons predict a novel cryptic pocket in TEM-1. a) The pocket opening motion for the S243 pocket (left buried, right exposed) as determined by the MSM eigenmode maximally correlating with S243 SASA changes. To highlight the change in packing, residues within 7Å are shown in spheres. S243 is shown in pink, and other residues within the exposon are colored beige (matching Fig 1d). Residues in the exposon but outside the 7Å radius are shown in sticks. b) TEM-1 S243C labeling (solid circles) fit to the EX2 model (solid line) is much faster than can be accounted for by unfolded enzyme alone (dashed line). Error bars were on the order of 10^{-4} and are not included for visual clarity.

binds this pocket and displaces the Ω-loop would be a potent inhibitor while a drug that stabilizes the closed conformation would increase activity. The formation of a cryptic pocket is also consistent with work from our lab demonstrating that the Ω-loop has significant conformational heterogeneity.[32] In our previous work, we were unable to detect this pocket because it frequently forms a channel-like connection with the active site, causing it to be spuriously combined with the active site pocket by pocket clustering methods.[21,22]

Our MSM revealed that S243 is significantly exposed by the opening of this pocket (Fig 3a). The open form also appears well-structured and druggable. This conclusion is supported by quantitative druggability scores from fpocket[33] (Fig S6). Together, these results prompted us to experimentally characterize the labeling of the S243C variant. Consistent with our prediction, the introduced cysteine labels at an intermediate rate that cannot be explained by protein unfolding (Fig 3b). This rate also clearly differs from the labeling rates of deeply-buried residues, which label on the order of hours, and surface residues, which label in less than seconds for this protein.[22] Then, to model the allosteric effect of a putative drug binding at this site, we assay the catalytic function of the labeled enzyme. The TNB covalently attached to the sidechain of residue 243 during the labeling reaction prevents complete closure of the pocket, much like a drug might, giving rise to a 3.75-fold increase in catalytic efficiency, driven primarily by a ~4-fold decrease in K_m (Fig. S7, Table S1). However, because of TNB's hydrophilicity and small size, the effect measured in this state may be less than the effect of a true drug. This is consistent with recent evidence suggesting that discovering allosteric activators may be easier than one might assume.[34] We are especially heartened that this method is able to identify a new cryptic allosteric site even in TEM-1, which is a protein that has been studied for multiple years by many groups, including specifically by our group for the purpose of locating these sites. The discovery of this new cryptic site strengthens

our argument that the paucity of previously-identified cryptic allosteric pockets may stem from technical limitations in locating them rather than a low prevalence.

Discovery of the first known cryptic pocket in CTX-M. As a more stringent test of the predictive power of our approach, we applied our method to CTX-M-9 β-lactamase. This enzyme has less than 40% sequence identity with the TEM-1 variant of the protein and lacks previously known cryptic pockets. Fortuitously, CTX-M-9 contains a native cysteine at position 69 that is buried in available x-ray structures but participates in an exposon and is therefore expected to label (Figs 4a and 4b). Consequently, applying thiol labeling does not require the introduction of a mutation, so we can directly test our model's prediction without any concern that a mutation created a pocket where none existed before. To test our prediction, we measured the solvent exposure kinetics of C69 with our thiol labeling assay and found that, consistent with the *in silico* prediction, it participates in a cryptic pocket. We find that C69's rate of labeling is inconsistent with the global unfolding process (Fig 4c) and is much faster than the labeling rate of deeply-buried residues. In contrast, S123C, a position that is buried in crystal structures of CTX-M-9,[35] showed only limited labeling during a three-hour reaction, implying a labeling rate consistent with labeling due

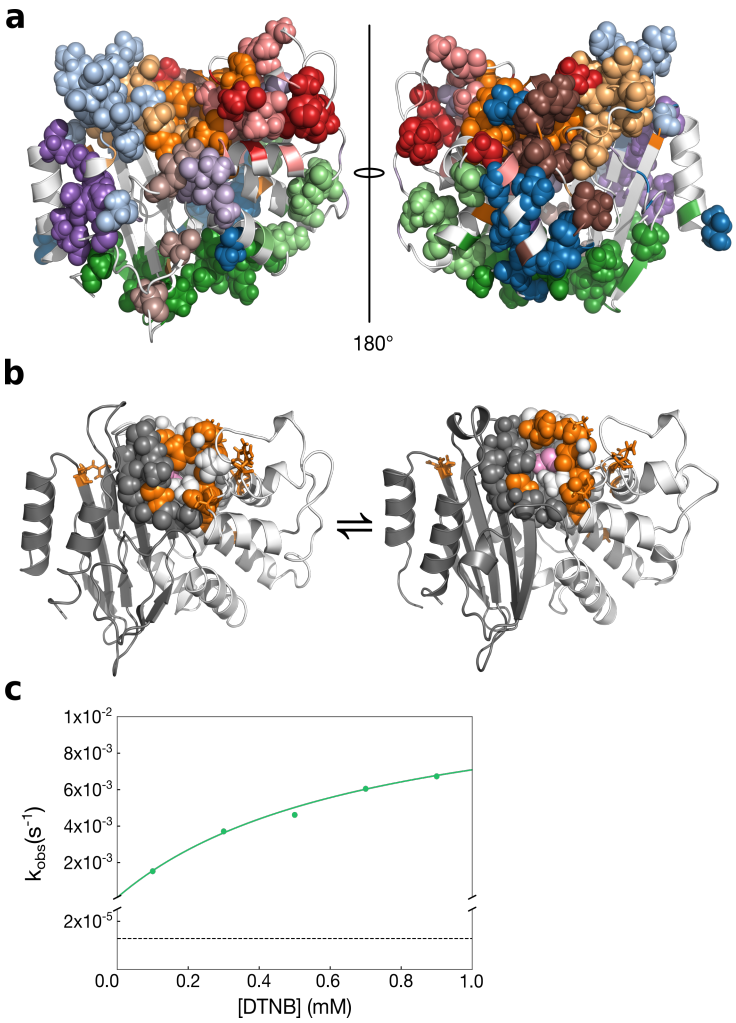

Figure 4: Exposons predict novel cryptic pockets in TEM-1 and CTX-M-9. a) The exposons of CTX-M-9 superimposed on the crystal structure 2P74; each color denotes a specific exposon. b) The pocket opening motion for the CTX-M-9 C69 pocket (left buried, right exposed) as determined by the MSM eigenmode maximally correlating with C69 SASA changes. Coloring and representation choices are the same as Fig 2a. c) CTX-M-9 C69 labeling (solid circles) fit to the EXX model (solid line) is much faster than can be accounted for by the global unfolding process alone (dashed line). Error bars were on the order of 10^{-5} and are omitted for visual clarity.

exclusively to the global unfolding process. As with TEM-1, we once again tested the allosteric activity of our proposed allosteric site by measuring the kinetic function of the C69-labeled enzyme. We found an approximately 15-fold reduction in the catalytic efficiency (Fig S7). This is the first cryptic pocket to be identified in CTX-M-9. By comparison, this same assay applied to previously-identified cryptic pockets in TEM-1 showed less than a threefold change in activity.[22] The pocket in CTX-M also appears to be well-structured (Fig 4b), as opposed to disordered, making it a viable drug target. Taken together, this newly-predicted pocket is the most attractive cryptic drug target found to date in either TEM or CTX-M-9.

Discussion

We have demonstrated that exposons provide a powerful conceptual framework for identifying cryptic allosteric sites and other functionally-relevant conformational motions. Exposons retrodict known cryptic pockets, retrospectively identify important surface rearrangements, and predict new pockets in two enzymes that confer bacteria with antibiotic resistance. Our previous methods based on traditional pocket detection tools failed to detect the new allosteric sites we present here. Importantly, we are unaware of any evidence for cryptic pockets in CTX-M-9 prior to our work. Our integrated computational and experimental approach reveals that novel cryptic pockets predicted with exposons have more influence over enzymatic activity than previously identified pockets and can induce both inhibition and activation of enzymatic function. Therefore, these newly predicted sites are valuable targets for designing drugs that could be used to restore the efficacy of existing antibiotics. More generally, we expect our methodology to serve as a powerful first step for a drug development pipeline targeting cryptic sites. Finally, because many proteins' most biologically interesting behaviors involve their surface chemistry, we expect that exposons will also prove useful for dissecting other types of functionally relevant conformational changes.

Methods

Molecular dynamics simulations

As described previously,[21] simulations were run at 300 K with the GROMACS software package[36] using the Amber03 force field[37] and TIP3P[38] explicit solvent, deployed on the Folding@home distributed computing platform[39] for β-lactamase and deployed on a local cluster for NP. In aggregate, we ran 90.5 μs of simulation for TEM-1 with the M182T substitution, 76.0 μs of simulation for CTX-M-9, and 28.0 μs of simulation for NP.

Assignment of residues to exposons

Each residue is assigned to at most one exposon. First, we compute the solvent accessibility to a drug-sized probe using the Shrake-Rupley[40] algorithm, as implemented in MDTraj.[41] Although the size of the spherical probe is, in principle, a tunable parameter in our method, we did not need to deviate in this work from the chosen 2.8 Å, even to model nucleoprotein. 2.8 Å was chosen because previous work suggests this value identifies pockets that can accommodate a drug-sized molecule.[22] Jug[42] was used to organize the parallel execution of solvent accessibility calculations.

Using solvent accessibilities for each atom, we classify each sidechain as exposed or buried using a fixed threshold. We chose a fixed threshold rather than a continuous threshold to reduce the number of parameters (a sigmoid, for example, would require a step width and a step midpoint) and because previous work suggested that mutual information performs better when a smaller number of bins are used[26] Our fixed threshold in this paper was 2.0 $Å^2$, but choices in the range 2.0–5.0 $Å^2$, as well as values formulated as a fraction of maximum possible sidechain exposure in the 1-3% range, gave similar results. This invariance ensures that our algorithm does not erroneously favor larger residues due to their larger maximum possible SASA. The end result is a featurization of the original trajectory, wherein each snapshot is represented by a binary vector with one entry per residue that contains a one for exposed residues and a zero for buried residues.

We then calculate the mutual information between each pair of residues. Mutual information (MI) is a measure of the statistical interdependence of two random variables. It is given by the equation,

$$MI(X,Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) log \left(\frac{p(x,y)}{p(x)p(y)} \right) \quad (1)$$

where X and Y are any pair of residues and x and y represent the solvent accessibility states (i.e. buried, exposed) of the corresponding residue. The probabilities $p(x)$ and $p(y)$ represent the probability that residues are observed in state x and $p(x,y)$ is the joint probability of x and y.

Exposons are then the cluster assignments computed by affinity propagation.[27] We use the affinity propagation implemented in scikit-learn 0.19.0[43] with damping 0.75 and zero initial affinities. Affinity propagation is robust to the choice of damping parameter, giving similar values for much of the range of valid choices (Fig S8).

Markov state models and *in silico* labeling

We define MSM microstates by clustering the sidechain SASA featurized representation. We find using this featurized space rather than a geometrical criterion like RMSD to do our state space discretization better captures the behaviors we are interested in. We discover microstates using the hybrid k-centers/k-medoids algorithm[44,45] with a Euclidean distance metric and fit transition probabilities with the

transpose method. The lag time and clustering stopping condition of 2.6 nm^2 and 4 ns (TEM-1), and 3.0 nm^2 and 0.6 ns (CTX-M-9) were chosen by the implied timescales test (Fig. S9). Ten rounds of k-medoids updates were used.

We predict time-dependent labeling behavior using the MSM we fit as described above. Synthetic labeling experiments are performed first by identifying all states in which the residue of interest is exposed, and converting them to sink states by zeroing out the rows in the transition probability matrix. Then, iteratively multiplying the equilibrium probability distribution by this new matrix gives a monotonically decaying fraction of 'unlabeled' probability density as density flows into the sink states and 'disappears.' Finally, we fit the unlabeled fraction as a function of time to a single exponential to yield a rate. In the limit of a perfectly good fit, this rate is equivalent to a mean first passage time. [35,36] An implementation of this simple procedure is provided in a Jupyter notebook (see 'Code Availability'). We used SciPy[46] version 0.19.1 for curve fitting.

Protein expression and purification

TEM-1 was purified from the periplasmic fraction of BL21(DE3) cells (Agilent Technologies) using both cation exchange and size exclusion chromatography. The full protocol is described in previous work.[32]

CTX-M-9 was purified from the insoluble cytoplasmic fraction. We subcloned the gene into the multiple cloning site of pET9-a vector. Plasmids were transformed into BL21(DE3) Gold cells (Agilent Technologies) for expression under T7 promoter control. Cells were induced with 1mM IPTG at OD=0.6 and grown for 5 hours at 37°C. The cells were then centrifuged and the pellet was frozen.

CTX-M-9 cells were resuspended in 20 mM sodium acetate, pH 5.5 and sonicated. After centrifugation, the pellet was unfolded in 9 M urea 20 mM sodium acetate, pH 5.5 and centrifuged. CTX-M-9 was then refolded in 20 mM sodium acetate, pH 5.5 and purified and stored similarly to TEM-1.

Thiol labeling

Our previous work with thiol labeling was performed on a standard UV-Vis spectrophotometer,[22] but in this work we used a stopped-flow instrument that gives access to faster timescale motions and improves the quality of fits because it offers a short, consistent, and known dead time. It also allows for the use of lower DTNB and protein concentrations.

We observe the change in absorbance over time of DTNB (Ellman's reagent), a small molecule that covalently binds free cysteines. We used a SX20 stopped-flow instrument (Applied Photophysics) with a dead time of 1.5 ms. Measurements were taken over time in 20 mM Tris, pH 8 1% DMSO at an absorbance of 412 nm (ε_{412} = 14,150 M^{-1} cm^{-1}) and fit by a single exponential (Fig S10).

The labeling rate at a given DTNB concentration can be fit with a Linderstrøm-Lang model, originally derived for hydrogen exchange[47]:

$$\text{(2)}\quad \text{closed} \underset{k_{\text{cl}}}{\overset{k_{\text{op}}}{\rightleftharpoons}} \text{open} \xrightarrow{k_{\text{int}}[\text{DTNB}]} \text{labeled}$$

In general, the observed rate can be described as

$$\text{(3)}\quad k_{\text{obs}} = \frac{k_{\text{op}} k_{\text{int}}[\text{DTNB}]}{k_{\text{op}} + k_{\text{cl}} + k_{\text{int}}[\text{DTNB}]}$$

which is a nonlinear function that approaches a linear dependence on [DTNB] at low concentrations and [DTNB] independence at high concentrations. In the limiting case where $k_{\text{cl}} \ll k_{\text{int}}[\text{DTNB}]$ called the EX1 regime, the observed rate of labeling reduces to

$$\text{(4)}\quad k_{\text{obs}}^{(\text{EX1})} = k_{\text{op}} \qquad (k_{\text{cl}} \ll k_{\text{int}}[\text{DTNB}])$$

In the limiting case where $k_{\text{cl}} \gg k_{\text{int}}[\text{DTNB}]$ called the EX2 regime, the observed rate of labeling reduces to

$$\text{(5)}\quad k_{\text{obs}}^{(\text{EX2})} = \frac{k_{\text{op}}}{k_{\text{cl}}} k_{\text{int}}[\text{DTNB}] = K k_{\text{int}}[\text{DTNB}] \qquad (k_{\text{cl}} \gg k_{\text{int}}[\text{DTNB}])$$

where K is the equilibrium constant between the open and closed forms. In the intermediate regime where $k_{\text{cl}} \approx k_{\text{int}}[\text{DTNB}]$, called the EXX regime, one must fit to the full expression (given in equation 3). We found that TEM-1 S243C was in the EX2 regime (linear dependence on [DTNB]) and CTX-M-9 was in the EXX regime (nonlinear dependence on [DTNB])

The three regimes differ in terms of the controls required to demonstrate that the rates of labeling are not reporting on global unfolding. In the EX1 regime, the observed labeling rate for a pocket must be faster than the rate of global unfolding. Neither of the pockets we test in this paper were in this regime, but we have previously observed pockets that showed this behavior.[22] In the EX2 regime, the equilibrium constant for a pocket must be greater than that for the unfolded state (Equation 5). To determine these quantities for TEM-1 S243C, we measured the K of unfolding (Table S2) and the intrinsic rate of labeling. To determine the intrinsic rate of labeling, our labeling assay was repeated with the addition of 6 M Urea (Fig. S11). In the EXX regime, the observed labeling rate for a pocket must be greater than the minimum of the expected labeling rates of the unfolded state in either the EX1 or EX2 regimes (derivation in SI). Thus, to test that CTX-M-9's labeling rate is not consistent with global unfolding alone, we measured both the rate of unfolding (Fig S12) and its thermodynamic stability (Table S3). We then combined that with the fit value of k_{int} (Fig 4c, Table S2) to produce a

piecewise function that is an upper bound for labeling from the unfolded state (Fig 4c). In this case, however, we found that the unfolding rate is the relevant control for all DTNB concentrations that we gathered data for—the population of unfolded enzyme is relevant only at DTNB concentrations less than about 170 nM, the DTNB concentration where $k_{\mathrm{cl}} = k_{\mathrm{int}}$.

Urea melts and unfolding

Equilibrium data and unfolding kinetics were acquired on a Chirascan circular dichroism spectrometer (Applied Photophysics) at a temperature of 25°C. Protein denaturation was observed by measuring the average ellipticity over 60 seconds at 222 nm as a function of urea concentration (Fig S13, Table 3). Samples of 35 μg/ml protein were equilibrated in 50 mM potassium phosphate pH 7 and urea overnight prior to data collection.

To determine the global unfolding rate, we fit the log observed unfolding rates as a function of urea concentration at concentrations above the concentration at which it is half folded and half unfolded (the C_m, Table S3) and extrapolated back to 0 M urea (Fig S12). Concentrations were between 4 and 5.5 M urea for M182T and between 1.8 and 2.8 M urea for CTX-M-9.

Activity measurements

Activity measurements were performed on both labeled and unlabeled protein. In order to measure the activities of the labeled proteins, 10 μM S243C and 5 μM CTX-M-9 were each incubated with excess DTNB for one hour, giving ample time for both proteins to fully label prior to the activity measurements. The proteins were then separated from excess DTNB using size exchange chromatography.

Enzyme activities against nitrocefin (Cayman Chemical Company) were monitored at 482 nm (ε_{482} = 15,000 M^{-1} cm^{-1}) using a Cary 100 UV-vis spectrophotometer (Agilent Technologies). Reactions were measured in 50 mM potassium phosphate, 10% glycerol (v:v), 2% DMSO pH 7.0 at 25 °C using 2 nM enzyme. Initial velocities were plotted as a function of nitrocefin concentration and fit to a Michaelis–Menten model to extract k_{cat} and K_m values (Table S1).

Visualizations

Protein structures were visualized using PyMOL 1.6[48]. Graphs were embedded with the ForceAtlas 2 algorithm and visualized using Gephi.[49]

Code availability

Code is available on GitHub as `gbowman/enspara`. The analysis described in this manuscript is reproduced in miniature, complete with example parameter choices and

visualization recommendations, in a Jupyter notebook in the `enspara/examples` directory.

Acknowledgements

We are grateful to the Folding@home users for computing resources. This work was funded by National Institutes of Health Grants R01GM12400701, U19AI109664, and T32GM02700, as well as by the National Science Foundation CAREER Award MCB-1552471. G.R.B. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and a Packard Fellowship for Science and Engineering from The David & Lucile Packard Foundation. M.I.Z. holds a Monsanto Graduate Fellowship and a Center for Biological Systems Engineering Fellowship.

References

1. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5,** 808–814 (2009).
2. Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **3,** 301–317 (2004).
3. Hardy, J. A. & Wells, J. A. Searching for new allosteric sites in enzymes. *Curr. Opin. Struct. Biol.* **14,** 706–715 (2004).
4. Cimermancic, P. *et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J. Mol. Biol.* **428,** 709–719 (2016).
5. Günthard, H. F. *et al.* Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2016 Recommendations of the International Antiviral Society-USA Panel. *Journal of the American Medical Association* **316,** 191–210 (2016).
6. Johnson, D. K. & Karanicolas, J. Computational Screening and Design for Compounds that Disrupt Protein-protein Interactions. *Curr. Top. Med. Chem.* **17,** 2703–2714 (2017).
7. Wagner, J. R. *et al.* Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **116,** 6370–6390 (2016).
8. Schames, J. R. *et al.* Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **47,** 1879–1881 (2004).
9. Hendlich, M., Rippmann, F. & Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* **15,** 359–63– 389 (1997).
10. Wagner, J. R. *et al.* POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J. Chem. Theory Comput.* **13,** 4584−4592 (2017).
11. La Sala, G., Decherchi, S., De Vivo, M. & Rocchia, W. Allosteric Communication Networks in Proteins Revealed through Pocket Crosstalk Analysis. *ACS Cent. Sci.* **3,** 949–960 (2017).

12. Amaro, R. E. Toward Understanding 'the Ways' of Allosteric Drugs. *ACS Central Science* **3,** 925–926 (2017).
13. Bowman, G. R., Pande, V. S. & Noé, F. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. (Springer Science & Business Media, 2013).
14. Voelz, V. A., Bowman, G. R., Beauchamp, K. A. & Pande, V. S. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *J. Am. Chem. Soc.* **132,** 1526–1528 (2010).
15. Paul, F. *et al.* Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* **8,** 1095 (2017).
16. Salverda, M. L. M., De Visser, J. A. G. M. & Barlow, M. Natural evolution of TEM-1 β-lactamase: experimental reconstruction and clinical relevance. *FEMS Microbiol. Rev.* **34,** 1015–1036 (2010).
17. Obama, B. *Executive Order—combating antibiotic-resistant bacteria*. (2014).
18. Cooper, M. A. & Shlaes, D. Fix the antibiotics pipeline. *Nature* **472,** 32–32 (2011).
19. World Health Organization. *Antimicrobial resistance: global report on surveillance*. (2014).
20. Horn, J. R. & Shoichet, B. K. Allosteric inhibition through core disruption. *J. Mol. Biol.* **336,** 1283–1291 (2004).
21. Bowman, G. R. & Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 11681–11686 (2012).
22. Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C. & Marqusee, S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 2734–2739 (2015).
23. WHO Ebola Response Team *et al.* West African Ebola epidemic after one year--slowing but not yet under control. *N. Engl. J. Med.* **372,** 584–587 (2015).
24. Su, Z. *et al.* Electron Cryo-microscopy Structure of Ebola Virus Nucleoprotein Reveals a Mechanism for Nucleocapsid-like Assembly. *Cell* **172,** 966–978.e12 (2018).
25. McClendon, C. L., Friedland, G., Mobley, D. L., Amirkhani, H. & Jacobson, M. P. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *J. Chem. Theory Comput.* **5,** 2486–2502 (2009).
26. Singh, S. & Bowman, G. R. Quantifying Allosteric Communication via Both Concerted Structural Changes and Conformational Disorder with CARDS. *J. Chem. Theory Comput.* acs.jctc.6b01181 (2017). doi:10.1021/acs.jctc.6b01181
27. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315,** 972–976 (2007).
28. Noé, F. *et al.* Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 4822–4827 (2011).
29. Metzner, P., Sch tte, C. & Vanden-Eijnden, E. Transition Path Theory for Markov Jump Processes. *Multiscale Model. Simul.* **7,** 1192–1219 (2009).

30. Leung, D. W. *et al.* An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. *Cell Rep* **11,** 376–389 (2015).
31. Adachi, H., Ohta, T. & Matsuzawa, H. Site-directed mutants, at position 166, of RTEM-1 beta-lactamase that form a stable acyl-enzyme intermediate with penicillin. *J. Biol. Chem.* **266,** 3186–3191 (1991).
32. Hart, K. M., Ho, C. M. W., Dutta, S., Gross, M. L. & Bowman, G. R. Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* **7,** 12965 (2016).
33. Schmidtke, P. & Barril, X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.* **53,** 5858–5867 (2010).
34. Hart, K. M. *et al.* Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. *PloS one* **12,** e0178678 (2017).
35. Chen, Y., Bonnet, R. & Shoichet, B. K. The acylation mechanism of CTX-M beta-lactamase at 0.88 a resolution. *J. Am. Chem. Soc.* **129,** 5378–5380 (2007).
36. Berendsen, H. J., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91,** 43–56 (1995).
37. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24,** 1999–2012 (2003).
38. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
39. Pande, V. S. *et al.* Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* **68,** 91–109 (2003).
40. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79,** 351–371 (1973).
41. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109,** 1528–1532 (2015).
42. Coelho, L. P. Jug: Software for Parallel Reproducible Computation in Python. *Journal of Open Research Software* **5,** 022109 (2017).
43. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12,** 2825–2830 (2011).
44. Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A. & Swope, W. C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **126,** 155101 (2007).
45. Beauchamp, K. A. *et al.* MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **7,** 3412–3419 (2011).
46. Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng.* **9,** (2007).

47. Ha, J. H. & Loh, S. N. Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nat. Struct. Biol.* **5,** 730–737 (1998).
48. DeLano, W. L. The PyMOL molecular graphics system. (2002).
49. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Icwsm* 361–362 (2009).