Title

Discovery of multiple hidden allosteric sites by combining Markov state models and experiments

Journal

Proceedings of the National Academy of Sciences of the United States of America, 112(9)

ISSN

0027-8424

Authors

Bowman, Gregory R
Bolin, Eric R
Hart, Kathryn M
et al.

Publication Date

2015-03-03

DOI

10.1073/pnas.1417811112

Peer reviewed

Discovery of multiple hidden allosteric sites by combining Markov state models and experiments

Gregory R. Bowman[a,b,1], Eric R. Bolin[c], Kathryn M. Hart[a,b], Brendan C. Maguire[b], and Susan Marqusee[a,b,1]

[a]Department of Molecular and Cell Biology, [b]Institute for Quantitative Biosciences, and [c]Biophysics Graduate Program, University of California, Berkeley, CA 94720

Edited by Robert L. Baldwin, Stanford University, Stanford, CA, and approved January 27, 2015 (received for review September 16, 2014)

The discovery of drug-like molecules that bind pockets in proteins that are not present in crystallographic structures yet exert allosteric control over activity has generated great interest in designing pharmaceuticals that exploit allosteric effects. However, there have only been a small number of successes, so the therapeutic potential of these pockets—called hidden allosteric sites—remains unclear. One challenge for assessing their utility is that rational drug design approaches require foreknowledge of the target site, but most hidden allosteric sites are only discovered when a small molecule is found to stabilize them. We present a means of decoupling the identification of hidden allosteric sites from the discovery of drugs that bind them by drawing on new developments in Markov state modeling that provide unprecedented access to microsecond- to millisecond-timescale fluctuations of a protein's structure. Visualizing these fluctuations allows us to identify potential hidden allosteric sites, which we then test via thiol labeling experiments. Application of these methods reveals multiple hidden allosteric sites in an important antibiotic target—TEM-1 β-lactamase. This result supports the hypothesis that there are many as yet undiscovered hidden allosteric sites and suggests our methodology can identify such sites, providing a starting point for future drug design efforts. More generally, our results demonstrate the power of using Markov state models to guide experiments.

thiol labeling | antibiotic resistance | molecular dynamics

A hidden allosteric site is a binding pocket that is not present in the crystal structure of a protein, but becomes available as the protein fluctuates and is capable of controlling the protein's function by communicating with the active site (Fig. 1) (1). Ligands that bind these sites exert control over the protein's function by perturbing the ensemble of structures the protein adopts (2, 3). Such sites could have unknown biological functions and serve as valuable targets for drug design, particularly for proteins that are currently considered undruggable because known structures lack pockets that are suitable for drug design. Unfortunately, it has been difficult to explore either of these possibilities because identifying hidden allosteric sites and molecules that bind them remains a profound challenge. For example, most structure-function studies focus on a single representative structure of a protein and give little hint as to where hidden allosteric sites might occur or what sort of molecules might bind them.

Detecting hidden allosteric sites experimentally is difficult because most of the available methods couple the identification of such sites with the drug discovery process. Given the difficulties inherent to drug design, this coupling likely produces many false negatives, leaving hidden allosteric sites undiscovered. For example, high-throughput screening can reveal hidden allosteric sites (4, 5) but failure to identify allosteric modulators does not disprove the existence of allosteric sites. Tethering provides a site-directed screen that is useful for specifically searching for allosteric sites but will still suffer from false negatives if the library being screened does not include small molecules that will bind an allosteric site tightly enough (6, 7). Understanding the full ensemble of structures a protein can adopt would overcome these limitations (8), but such an understanding remains elusive. For instance, room temperature crystallography and NMR have the potential to reveal alternative structures containing hidden allosteric sites, but further developments are required to make such measurements routine for any given protein target (9–14).

A number of computational techniques have been developed to aid in the discovery of hidden allosteric sites. For example, there are a variety of methods for understanding how information flows from one region of a protein to another (15–19), as well as for identifying potential binding pockets (20–24). More recent work has accounted for both of these ingredients (25, 26). Although these methods are important developments, many are only applicable to small proteins operating on fast timescales. Therefore, new approaches are needed for addressing many biologically relevant systems. Furthermore, these methods only partially decouple the discovery of hidden allosteric sites from the identification of allosteric ligands because identifying a compound that binds to a predicted site is still the primary means of testing computationally predicted allosteric sites.

Here, we use a combination of computation and experiment to identify hidden allosteric sites without requiring the simultaneous discovery of ligands that bind and modulate them. Decoupling the discovery of allosteric sites from the identification of allosteric ligands should facilitate drug discovery by providing more information to base design decisions on. For example, rather than performing blind screens, computationally generated structures of potential allosteric sites can be used as a starting point for rational design. The locations of potential hidden allosteric sites can also be used to direct tethering screens, potentially providing a less resource-intensive means of discovering hidden allosteric

Significance

Rational drug design efforts typically focus on identifying inhibitors that bind to protein active sites. Pockets that are not present in crystallographic structures yet can exert allosteric (i.e., long-range) control over distant active sites present an exciting alternative. However, identifying these hidden allosteric sites is extremely challenging because one usually has to simultaneously find a small molecule that binds to and stabilizes the open conformation of the pocket. Here, we present a means of combining advances in computer modeling—using Markov state models to capture long timescale dynamics—with biophysical experiments to identify hidden allosteric sites without requiring the simultaneous discovery of drug-like compounds that bind them. Using this technology, we discover multiple hidden allosteric sites in a single protein.

Author contributions: G.R.B., E.R.B., K.M.H., B.C.M., and S.M. designed research; G.R.B., E.R.B., K.M.H., and B.C.M. performed research; G.R.B., E.R.B., K.M.H., B.C.M., and S.M. contributed new reagents/analytic tools; G.R.B., E.R.B., K.M.H., B.C.M., and S.M. analyzed data; and G.R.B., E.R.B., K.M.H., B.C.M., and S.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: bowman@biochem.wustl.edu or marqusee@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1417811112/-/DCSupplemental.

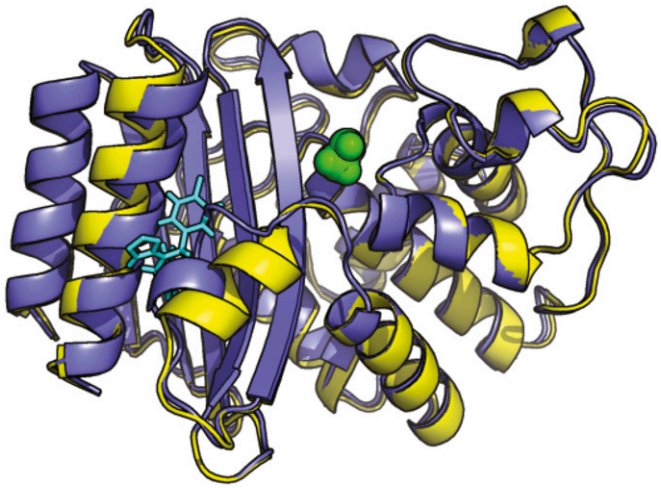

Fig. 1. Crystal structures of TEM-1 β-lactamase in the absence of any ligand (blue) (49) and with an inhibitor (cyan) bound in a hidden allosteric site (yellow) (34). A key catalytic serine (S70) is in green spheres.

sites than applying tethering to every possible location in a protein. Our approach can be applied to most soluble proteins, but could be of particular value for the large number of proteins that are currently considered undruggable because their active sites or binding interfaces are not considered viable drug targets (27).

The first step in our approach is to build a Markov state model of a protein of interest, which is essentially an atomically detailed map of the ensemble of conformations the protein can adopt. Such models are of great value because they can capture relatively slow conformational changes that are typically beyond other computational methods and they allow the user to learn what degrees of freedom are important rather than requiring the user to select them a priori (28, 29). Potential allosteric sites are then identified by querying the Markov model for local fluctuations that form pockets that are surrounded by residues whose rotameric orientations are correlated with those of the active site. These correlations are a property of the ensemble of structures the protein can adopt and give insight into where perturbations, such as ligand binding, are likely to exert allosteric control over distant sites, such as the active site (26).

Predicted pockets are tested experimentally with thiol labeling. In thiol labeling, a cysteine is introduced at a site buried within a pocket and a chemical reagent is introduced, such as 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB, also called Ellman's reagent). This reagent will react covalently with the cysteine thiol if the thiol is sufficiently exposed and available to solvent. DTNB is also attractive for our purposes because it is a drug-sized molecule, so its ability to bind within a pocket suggests there may be enough room for a typical drug to bind. Thiol labeling is analogous to hydrogen-deuterium exchange but is sensitive to changes in the exposure of side-chains rather than the backbone. Therefore, it is better suited to detecting pockets that form when elements of secondary structure separate from one another, as is observed in many pocket-opening fluctuations, without requiring unfolding and exposure of the backbone amides. This approach closely parallels thiol-exchange techniques for studying protein folding/unfolding and can provide opening rates for the various pockets as well as the fraction of time the pocket is open (30–32). Communication between a pocket and the active site is tested by assaying whether covalently linking a small molecule within a pocket has any detectable effect on the protein's activity, as in tethering (6). A number of controls are also performed to ensure that the labeling we observe is due to local fluctuations captured by our computational models rather than global unfolding.

To test this approach, we apply it to TEM-1 β-lactamase with the M182T substitution, a change commonly found in antibiotic-resistant variants that is known to stabilize the native state. This protein is an important drug target because of its role in antibiotic resistance (33). It already has one known hidden allosteric site (34), so it is an excellent system for testing whether our approach can discriminate allosteric sites from nonallosteric sites. Finally, β-lactamase is a large enough protein that it is reasonable to ask if there are other hidden allosteric sites, especially given our recent computational prediction that such sites exist (26). Newly predicted allosteric sites that are similar in nature to the known site could be attractive drug targets given that we already know ligands that bind the known hidden allosteric site are capable of modulating β-lactamase's activity.

Results and Discussion

Computation and Experiment Successfully Detect the Known Hidden Allosteric Site. Our previous work demonstrated that our computational approach successfully identifies the known hidden allosteric site in TEM-1 β-lactamase (26). The known site is the largest of the unexpected pockets formed in this protein and is also in the open state more than any other pocket. The conformations of residues surrounding this pocket are also correlated with the conformations of active site residues, providing a means for communication between these sites. Therefore, we expect perturbations at this site can modulate the ensemble of structures β-lactamase adopts in a manner that alters the protein's activity. These results suggest an important role for conformational selection in the function of this hidden allosteric site. That is, the pocket is present even in the absence of a small molecule to bind it rather than being created by interactions between the protein and small molecule, which is often called induced fit. Therefore, the known hidden allosteric site should be detectable in our thiol labeling experiments because the pocket will be present even in the absence of a specific allosteric modulator.

If the known hidden allosteric site opens as a result of the fluctuations predicted by our computational model, then we expect to see labeling of residues whose side-chains line this pocket. To test this prediction, we applied our thiol labeling technique to L286. This residue was chosen because it has negligible solvent accessible surface area in the apo structure of β-lactamase, but our computational model predicts that it becomes exposed when the known hidden allosteric site opens (Fig. 2 *A* and *B*). The L286C mutation required for thiol labeling is also one of the more conservative mutations we could have chosen in the known hidden allosteric site, so this mutation minimizes potential perturbation to the protein.

Indeed, our labeling studies indicate that the known hidden allosteric site opens as β-lactamase fluctuates. The L286C variant labels at a rate of $\sim 5.9 \times 10^{-4} \pm 5.7 \times 10^{-5}$ s^{-1} in 1 mM DTNB (Fig. 2*C* and Fig. S1). The expected rate of labeling for a fully exposed residue is about 1 s^{-1} (31, 35), so the observed labeling cannot be attributed to a reorganization of the protein's structure that exposes this residue in the ground state.

As in hydrogen exchange, we interpret the observed labeling rate with the Linderstrom–Lang model (36). This model assumes the protein is in equilibrium between conformations where a pocket is either closed or is open and available to react with DTNB (Scheme 1). Given an opening rate (k_{op}), closing rate (k_{cl}), and rate of labeling from the open state (k_{int}), the observed rate of labeling is $k_{obs} = (k_{op}k_{int})/(k_{op} + k_{cl} + k_{int})$. In the limit where $k_{op} \ll k_{cl}$ and $k_{cl} \ll k_{int}$, this reduces to $k_{obs} = k_{op}$. This scenario is called the EX1 regime and can be identified because the observed rate of labeling will be independent of the concentration of labeling reagent. In the limit where $k_{op} \ll k_{cl}$ and $k_{cl} \gg k_{int}$, then $k_{obs} = K_{op}k_{int}$, where $K_{op} = k_{op}/k_{cl}$ is the equilibrium constant for the pocket being open. This scenario is called the EX2 regime and can be identified because the observed rate of labeling will be linearly dependent on the concentration of labeling reagent (k_{int}).

$$\text{Closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{Open} \xrightarrow{k_{int}} \text{Labeled}$$

To determine whether the observed rate of labeling is providing information about the opening rate or the fraction of time a pocket is open or exposed, we measured the rate of labeling with varying concentrations of DTNB. Fig. 2*D* shows that the labeling rate is independent of [DTNB], which is consistent with the EX1 regime. Therefore, we conclude that the observed rate of labeling captures the opening rate of this pocket.

Pockets Are Clearly Distinguishable from Nonpockets. Our experimental approach might give false positives if the cysteine mutations cause significant destabilization of the protein. For example, introducing a cysteine could globally destabilize the protein such that labeling occurs directly from global unfolding rather than transient exposure of the pocket within the native state ensemble. If this was true for the L286C variant, we would expect the labeling rate to approximate the rate of global unfolding because labeling is in the EX1 regime.

To test whether labeling is due to global unfolding, we determined the unfolding rate of our cysteine variant and compared it with the measured labeling rate (Table S1). Following previous work on the unfolding of β-lactamase (37, 38), we measured the unfolding rate of the L286C variant by monitoring the change in the circular dichroism (CD) signal as a function of the final urea concentration (Fig. 3). Extrapolating back to 0 M urea (the labeling conditions), we find that the rate of unfolding is about 20-fold smaller than the observed rate of labeling. Therefore, labeling must be due to a fluctuation across a barrier from the native state that is lower than the barrier to global unfolding.

As a control, we created cysteine variants at buried sites not predicted to form a pocket. Residues L190 and I260 are both buried in the ligand-free structure of β-lactamase, and our model predicts that there are no pockets that expose these residues to drug-sized molecules. Consistent with this prediction, we do not observe any labeling of cysteines at these positions over the course of a 12-h labeling reaction. Therefore, we conclude that these residues remain buried in the native-state ensemble and that introducing a cysteine does not cause a local destabilization that creates an unpredicted pocket or local unfolding. This result, in combination with the lack of observed labeling for the two endogenous cystines in the protein that are oxidized in a disulfide bond, also confirms that the labeling we observe for other residues is not due to a reaction with the two cysteines that naturally form a disulfide in β-lactamase. The fact that our computational model successfully discriminates where labeling will and will not occur also adds significant weight to our conclusion that labeling is due to the formation of a pocket rather than a large-scale unfolding event.

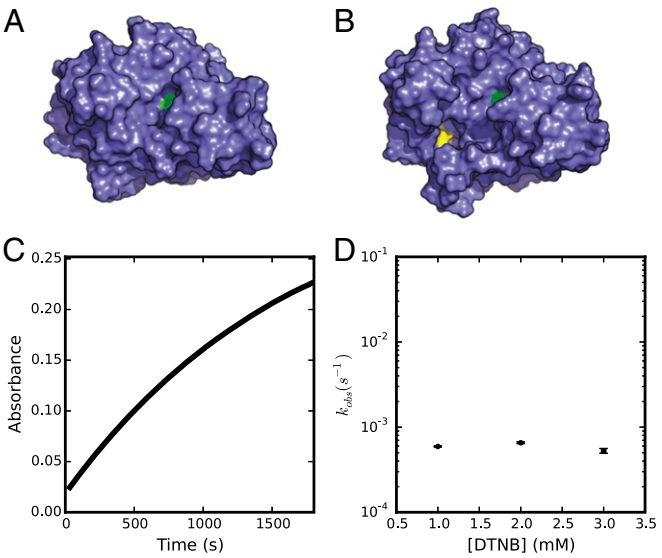

Fig. 2. Thiol labeling of the known hidden allosteric site. (*A* and *B*) Surface representations of the closed and open states of the known hidden allosteric site, respectively. L286 (yellow) is only visible in the open state. A key catalytic serine (S70) is shown in green as a reference point. (*C*) Labeling of L286C in 1 mM DTNB. (*D*) The dependence of the labeling rate of L286C on the concentration of the labeling reagent (DTNB) with error bars representing the SD from three replicates.

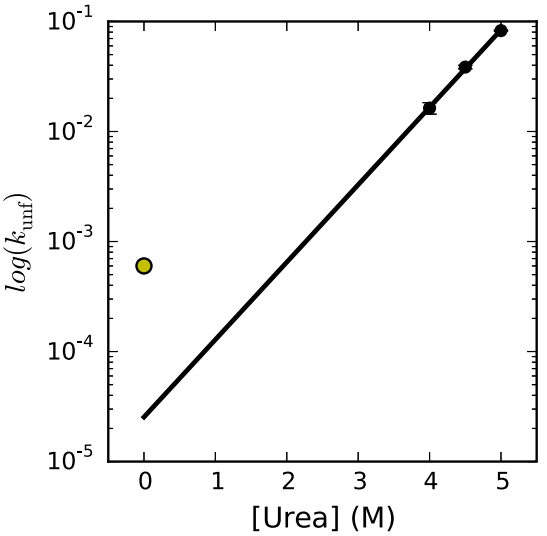

Fig. 3. Thiol labeling is not due to unfolding. Log of the unfolding rate of L286C as monitored by CD for different urea concentrations with a linear fit (black line) used for extrapolating back to the unfolding rate at 0 M urea. The labeling rate (yellow circle) is considerably faster than unfolding, so it must correspond to a fluctuation within the native state.

Given the proximity of the known hidden allosteric site to two of the four tryptophan residues in β-lactamase, we reasoned that opening of this pocket may expose these tryptophans to solvent and lead to a change in the protein's fluorescence. Indeed, opening of this pocket in our computational model increases the solvent accessible surface area of Trp229's side-chain from 36% in the ligand-free structure to 69 ± 9% when the pocket is open. The solvent accessible surface area of Trp290's side-chain increases from 43% in the ligand-free structure to 85 ± 8% when the pocket is open. Because pocket opening precedes global unfolding and might be on the pathway to global unfolding, we hypothesized that monitoring unfolding by fluorescence should detect pocket opening and yield a faster rate than monitoring unfolding by CD. To test this prediction experimentally, we measured the rate of change in fluorescence of the L286C variant as a function of the final urea concentration and used linear extrapolation to find the rate of change in the absence of denaturant. This procedure yielded a rate of $5.5 \times 10^{-4} \pm 2.4 \times 10^{-4}$ s^{-1}, in reasonable agreement with the rate of labeling with DTNB of $5.9 \times 10^{-4} \pm 5.7 \times 10^{-5}$ s^{-1}. The fact that these rates are 20-fold larger than the rate of unfolding demonstrates that labeling precedes unfolding and, therefore, occurs from a rare state on the native side of the rate-limiting barrier to unfolding. Interestingly, this state appears to be distinct from previously characterized high-energy states on the unfolded side of the rate-limiting barrier to unfolding that were also detected by fluorescence (37, 38).

There Is Communication Between the Known Allosteric Site and Active Site. We also exploited our thiol labeling to test whether there is communication between the pockets we detect and the active site, as indicated by correlations in the ensemble of structures β-lactamase adopts. We previously predicted that almost any hidden binding pocket should also serve as an allosteric site due to coupling of many residues to different portions of the active site (26). To test whether a given pocket communicates with the active site, we measured the activity of proteins with and without TNB (one half of DTNB) covalently bound within the pocket. A measurable change in activity would demonstrate that there is communication, although it should not be used as a quantitative measure of potential inhibition because other molecules that bind the same site could be more potent inhibitors or even enhance the protein's activity (39). Using this approach, we find that the specific activity of L286C is reduced from 361 ± 29 to 97 ± 6 nmol product/μg/min. These results suggest that there is communication between the site of modification and the active site. Although DTNB is a drug-sized molecule (*SI Materials and Methods*), TNB is significantly smaller than typical drugs and has not been optimized for binding this hidden allosteric site. Therefore, it is entirely possible that an allosteric modulator specifically designed to bind this site could be a much stronger β-lactamase inhibitor. Identifying such noncovalent inhibitors (or activators) would serve as the ultimate verification of the existence of our hidden allosteric sites and remains an important future direction. Although we have not yet discovered new molecules that bind the hidden allosteric sites revealed by our approach, we note that the allosteric inhibitor discovered by Horn et al. (34) demonstrates that it is possible for small molecules to bind such hidden allosteric sites strongly enough to stabilize the open form of a pocket and alter an enzyme's activity.

As a control, we tested whether thiol labeling of residues that our computational model predicts should have little communication with the active site alters β-lactamase's activity. Specifically, we chose to perform thiol labeling of A150 because it is a surface residue that is available for labeling and surrounding residues have weak correlations with the active site in our computational model. Complete labeling of the A150C variant has a negligible effect on the protein's activity (Table 1), consistent with our prediction that it does not communicate with the active site.

Discovery of Previously Unidentified Hidden Allosteric Sites. Now we can begin testing whether our model successfully predicts novel hidden allosteric sites. Toward this end, we chose to focus on pockets that our simulations predict will expose residues that are completely buried in the static, apo-crystal structure. Such pockets are the most amenable to our thiol labeling experiments, although there may be other pockets that are equally potent allosteric sites but lack residues with this differential exposure.

Table 1. Specific activities (nmol product/μg protein/min) of labeled and unlabeled proteins

Pocket	Mutation	Unlabeled activity	Labeled activity
	Background (M182T)	532 ± 5	
1	L286C	361 ± 29	97 ± 6
2	A232C	545 ± 75	255 ± 21
2	A249C	193 ± 13	150 ± 18
3	S203C	520 ± 13	345 ± 17
Surface	A150C	235 ± 25	249 ± 28
Buried	L190C	213 ± 22	
Buried	I260C	267 ± 38	

The reduction in activity on labeling demonstrates that there is communication between the proposed allosteric sites and the active site. The pockets are (1) the known allosteric site, (2) the first predicted site, and (3) the second predicted site.

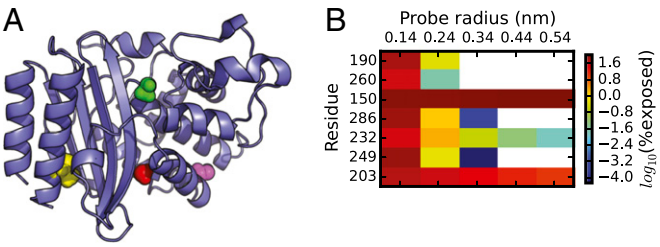

Fig. 4. Residues selected for labeling in each pocket. (*A*) Ribbon diagram of β-lactamase highlighting residues in the known hidden allosteric site (L286, yellow), the first predicted site (A232, red), and the second predicted site (S203, magenta). A key catalytic serine (S70) is shown in green. (*B*) Average percent of residues' surface area that is accessible to a variety of probe sizes. L190 and I260 are buried, whereas A150 is on the surface.

Residues appropriate for our thiol labeling experiments were selected by examining the fraction of each residue's surface area that is typically accessible to probes of varying sizes (Fig. 4). Specifically, we chose probe radii of 0.14, 0.24, 0.34, 0.44, and 0.54 nm. This range mimics molecules like water at the smallest scale and more drug-like molecules at the largest scale. We calculated the accessible surface area for the side-chain of each residue for a given probe size as follows: (*i*) computationally roll a sphere with the given probe radius across the surface of a representative structure for each state in the Markov model and calculate the accessible surface area of every residue's side-chain; (*ii*) calculate the average accessible surface area of each side-chain by taking the average across all states, weighted by their equilibrium populations; and (*iii*) divide the result for each residue by the total possible accessible surface area of its side-chain. One result of this analysis is that the fluctuations β-lactamase undergoes make basically every residue's side-chain accessible to water, as previously observed in other proteins (40). However, many residues are not accessible to larger molecules. We selected residues that are accessible to probes with radii of at least 0.34 nm because this is consistent with the size of DTNB and residues that we have already shown to label are accessible at this probe size, whereas residues that we have shown do not label are not. Of the residues that meet this criterion, we selected the smallest residues possible to minimize the perturbation caused by mutating to a cysteine. Based on these criteria, residues A232 and A249 point into the most promising pocket. Residue S203 also points into a second pocket.

We performed thiol labeling of A232C to test our first predicted pocket (Fig. 5). Fig. 5 *C* and *D* shows that the A232C variant labels at a rate of $3.6 \times 10^{-3} \pm 1.0 \times 10^{-3}$ s^{-1} independent of the concentration of labeling reagent, and therefore this is the rate at which the residue becomes exposed. The rate of unfolding is also 200-fold slower than the rate of labeling (Table S1), so the observed labeling is not due to global unfolding. Complete labeling of the protein reduces the activity of the protein by 1.5-fold (Table 1). In addition to our labeling experiments, we again reasoned that opening of this pocket could lead to a change in fluorescence by exposing Trp229 to solvent. Indeed, opening of this pocket in our computational model increases the solvent accessible surface area of Trp229's side-chain from 36% in the ligand-free structure to 56 ± 12% when the pocket is open. Experimentally monitoring unfolding by fluorescence, as described previously, yielded a rate of $1.9 \times 10^{-3} \pm 1.2 \times 10^{-3}$ s^{-1}, in reasonable agreement with the DTNB labeling rate for this variant. We also performed separate experiments on an A249C variant. The side-chain of this residue points into the same pocket but has less exposure than residue 232 because it is not exposed in the exact same set of structural states where this pocket is open as residue 232. Indeed, we observe labeling of the A249C variant at a rate fivefold less than the A232 variant (Table S1), consistent with the residue at position 249 being exposed on opening of the pocket. Based on all of

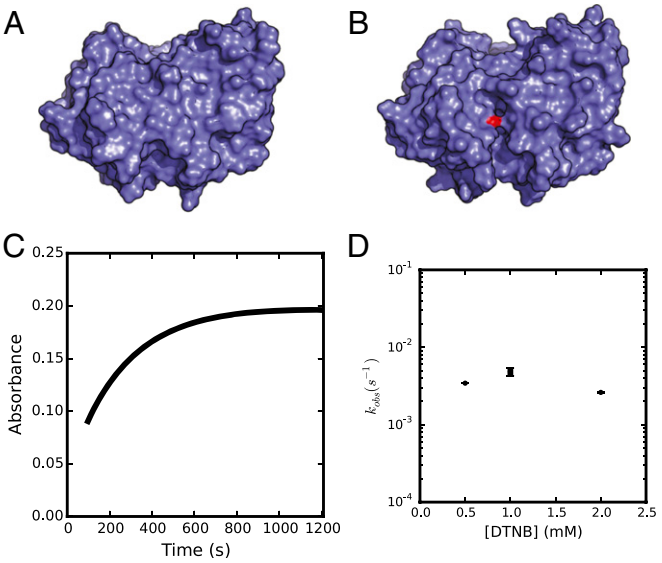

Fig. 5. Thiol labeling of the first predicted hidden allosteric site. (*A* and *B*) Surface representations of the closed and open states of the known hidden allosteric site, respectively. A232 (red) is only visible in the open state. (*C*) Labeling of residue A232C in 1 mM DTNB. (*D*) The dependence of the labeling rate on the concentration of the labeling reagent (DTNB) with error bars representing the SD from three replicates.

the results for these two variants, we conclude that this site is a hidden allosteric site, consisting of an unexpected pocket with the ability to communicate with the active site. An allosteric modulator of this site would need to have a greater effect on activity but, as explained before, this is entirely possible.

We also tested the second predicted pocket via thiol labeling of S203C. This residue labels at a rate of $1.5 \times 10^{-2} \pm 3.4 \times 10^{-3}$ s^{-1}, again independent of the concentration of labeling reagent (Fig. S2). This rate is 50-fold faster than the rate of global unfolding (Table S1), so it captures the rate of exposure of the residue. There is also communication between this site and the active site, as demonstrated by an ~1.5-fold reduction in activity on labeling. Therefore, we conclude that this second predicted site is also a hidden allosteric site.

Conclusions

We developed an approach that combines computation and experiment to detect hidden allosteric sites arising from the ensemble of structures a protein can adopt. Importantly, our approach does not require the simultaneous discovery of small molecules that bind and modulate these sites, so our methodology can be used to guide subsequent drug design efforts.

Using this approach, we have demonstrated that a single protein—TEM-1 β-lactamase—accommodates multiple hidden allosteric sites. This result is surprising because TEM-1 β-lactamase has been studied extensively without observing these sites. Furthermore, there may even be other hidden allosteric sites in this single protein that are not amenable to the experimental methodology we describe here.

Our results suggest there are many as yet undiscovered hidden allosteric sites and that our techniques should provide a means of detecting them. Once discovered, these allosteric sites can then be targeted with rational drug design or followed up on to discover their biological relevance. In the case of TEM-1, the hidden allosteric sites we discovered could be valuable targets for antibiotic development.

These results lay an important foundation for future work on hidden allosteric sites. For example, an important next step will be to discover new allosteric modulators that bind the hidden allosteric sites revealed by our methodology. Furthermore, our results demonstrate the value of our advanced computational methods and argue for further developments to make an even more quantitative comparison between computation and experiment.

Materials and Methods

Simulations were run with Gromacs (41, 42), and the Markov state model was built with MSMBuilder (43, 44), as described previously (26). This particular model provides a statistically reliable description of dynamics on tens of microsecond timescales. Further details are given in *SI Materials and Methods*. Accessible surface areas were measured with Gromacs and a Voronoi-based method (45). Structures were visualized with PyMOL (46).

TEM-1 β-lactamase with the M182T stabilizing mutation and all cysteine variants of this background sequence were purified from the periplasmic fraction of BL21(DE3) cells as described in previous work (47) and *SI Materials and Methods*. We measured enzyme activities following previous work (48), as described in *SI Materials and Methods*. Cysteine mutations were introduced with Quik-Change mutagenesis. Thiol labeling experiments were run with 30 μM protein and varying concentrations of DTNB (also called Ellman's reagent) in 20 mM Tris, pH 8.0. The time course for labeling was monitored by following the absorbance at 412 nm in a Cary 100Bio UV-Vis spectrophotometer (27 °C) after starting the reaction via manual mixing. Rates were interpreted using the Linderstrom–Lang model (36). Unfolding rates and equilibrium melts were monitored by circular dichroism, as described in *SI Materials and Methods*. Further details are given in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Norbert Lindow, Daniel Baum, and Hans-Christian Hege for helpful discussion on pocket detection. This work was funded by National Institutes of Health Grant R01-GM050945. G.R.B. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and was also supported by the Miller Institute.

1. Hardy JA, Wells JA (2004) Searching for new allosteric sites in enzymes. *Curr Opin Struct Biol* 14(6):706–715.
2. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. *Nature* 508(7496):331–339.
3. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins* 57(3):433–443.
4. Lundqvist T, et al. (2007) Exploitation of structural and regulatory diversity in glutamate racemases. *Nature* 447(7146):817–822.
5. Ceccarelli DF, et al. (2011) An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. *Cell* 145(7):1075–1087.
6. Erlanson DA, et al. (2000) Site-directed ligand discovery. *Proc Natl Acad Sci USA* 97(17):9367–9372.
7. Arkin MR, et al. (2003) Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci USA* 100(4):1603–1608.
8. Wrabl JO, et al. (2011) The role of protein conformational fluctuations in allostery, function, and evolution. *Biophys Chem* 159(1):129–141.
9. Fischer M, Coleman RG, Fraser JS, Shoichet BK (2014) Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat Chem* 6(7):575–583.
10. Fraser JS, et al. (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462(7273):669–673.
11. Manley G, Loria JP (2012) NMR insights into protein allostery. *Arch Biochem Biophys* 519(2):223–231.
12. Wand AJ (2013) The dark energy of proteins comes to light: Conformational entropy and its role in protein function revealed by NMR relaxation. *Curr Opin Struct Biol* 23(1):75–81.
13. Boehr DD, McElheny D, Dyson HJ, Wright PE (2006) The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* 313(5793):1638–1642.
14. Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. *Nature* 450(7172):964–972.
15. Hilser VJ, Dowdy D, Oas TG, Freire E (1998) The structural distribution of cooperative interactions in proteins: Analysis of the native state ensemble. *Proc Natl Acad Sci USA* 95(17):9903–9908.
16. McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP (2009) Quantifying correlations between allosteric sites in thermodynamic ensembles. *J Chem Theory Comput* 5(9):2486–2502.
17. Dubay KH, Bothma JP, Geissler PL (2011) Long-range intra-protein communication can be transmitted by correlated side-chain fluctuations alone. *PLOS Comput Biol* 7(9):e1002168.
18. Süel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 10(1):59–69.

19. Feher VA, Durrant JD, Van Wart AT, Amaro RE (2014) Computational approaches to mapping allosteric pathways. *Curr Opin Struct Biol* 25:98–103.
20. Eyrisch S, Helms V (2007) Transient pockets on protein surfaces involved in protein-protein interaction. *J Med Chem* 50(15):3457–3464.
21. Johnson DK, Karanicolas J (2013) Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface. *PLOS Comput Biol* 9(3):e1002951.
22. Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X (2011) MDpocket: Open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* 27(23):3276–3285.
23. Frembgen-Kesner T, Elcock AH (2006) Computational sampling of a cryptic drug binding site in a protein receptor: Explicit solvent molecular dynamics and inhibitor docking to p38 MAP kinase. *J Mol Biol* 359(1):202–214.
24. Schames JR, et al. (2004) Discovery of a novel binding trench in HIV integrase. *J Med Chem* 47(8):1879–1881.
25. Grant BJ, et al. (2011) Novel allosteric sites on Ras for lead generation. *PLoS ONE* 6(10):e25711.
26. Bowman GR, Geissler PL (2012) Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc Natl Acad Sci USA* 109(29):11681–11686.
27. Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* 1(9):727–730.
28. Bowman GR, Huang X, Pande VS (2010) Network models for molecular kinetics and their initial applications to human health. *Cell Res* 20(6):622–630.
29. Chodera JD, Noé F (2014) Markov state models of biomolecular conformational dynamics. *Curr Opin Struct Biol* 25:135–144.
30. Bernstein R, Schmidt KL, Harbury PB, Marqusee S (2011) Structural and kinetic mapping of side-chain exposure onto the protein energy landscape. *Proc Natl Acad Sci USA* 108(26):10532–10537.
31. Sridevi K, Udgaonkar JB (2002) Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates. *Biochemistry* 41(5):1568–1578.
32. Jha SK, Marqusee S (2014) Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci USA* 111(13):4856–4861.
33. Bush K (2013) Proliferation and significance of clinically relevant β-lactamases. *Ann N Y Acad Sci* 1277:84–90.
34. Horn JR, Shoichet BK (2004) Allosteric inhibition through core disruption. *J Mol Biol* 336(5):1283–1291.
35. Jocelyn PC (1987) Spectrophotometric assay of thiols. *Methods Enzymol* 143:44–67.
36. Berger A, Linderstrom-Lang K (1957) Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys* 69:106–118.
37. Vanhove M, Raquet X, Frère JM (1995) Investigation of the folding pathway of the TEM-1 beta-lactamase. *Proteins* 22(2):110–118.
38. Kather I, Jakob RP, Dobbek H, Schmid FX (2008) Increased folding stability of TEM-1 beta-lactamase by in vitro selection. *J Mol Biol* 383(1):238–251.
39. Sadowsky JD, et al. (2011) Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc Natl Acad Sci USA* 108(15):6056–6061.
40. Damjanović A, García-Moreno B, Lattman EE, García AE (2005) Molecular dynamics study of water penetration in staphylococcal nuclease. *Proteins* 60(3):433–449.
41. Van Der Spoel D, et al. (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26(16):1701–1718.
42. Pronk S, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854.
43. Bowman GR, Huang X, Pande VS (2009) Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* 49(2):197–201.
44. Beauchamp KA, et al. (2011) MSMBuilder2: Modeling conformational dynamics at the picosecond to millisecond scale. *J Chem Theory Comput* 7(10):3412–3419.
45. Lindow N, Baum D, Hege H-C (2011) Voronoi-based extraction and visualization of molecular paths. *IEEE Trans Vis Comput Graph* 17(12):2025–2034.
46. Schrödinger LLC (2002) The PyMOL Molecular Graphics System, Version 1.6.0.0.
47. Hanes MS, Ratcliff K, Marqusee S, Handel TM (2010) Protein-protein binding affinities by pulse proteolysis: Application to TEM-1/BLIP protein complexes. *Protein Sci* 19(10):1996–2000.
48. O'Callaghan CH, Morris A, Kirby SM, Shingler AH (1972) Novel method for detection of beta-lactamases by using a chromogenic cephalosporin substrate. *Antimicrob Agents Chemother* 1(4):283–288.
49. Wang X, Minasov G, Shoichet BK (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol* 320(1):85–95.