JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Identifying Allosteric Binding Sites in Proteins with a Two-State Gō Model for Novel Allosteric Effector Discovery

Yifei Qi,[‡] Qian Wang,[†] Bo Tang,[‡] and Luhua Lai*[,†,‡]

[†]BNLMS, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, and Peking−Tsinghua Center for Life Sciences at College of Chemistry and Molecular Engineering and [‡]Center for Quantitative Biology, Peking University, Beijing 100871, China

Supporting Information

ABSTRACT: Allostery is a common mechanism of controlling many biological processes such as enzyme catalysis, signal transduction, and metabolic regulation. The use of allostery to regulate protein activity is an important and promising strategy in drug discovery and biological network regulation. In order to modulate protein activity by allostery, predictive methods need to be developed to discover allosteric binding sites. In the present study, we developed a new approach to identify allosteric sites in proteins based on the coarse-grained two-state Gō model. Starting from the concept that allostery is a conformation population shift process, we first constructed an ensemble of two functional states of a protein and tuned the energy landscape to bias one state. We then added perturbations to a binding site and monitored the population distribution of the new ensemble. If population redistribution occurred, then the binding perturbed site was predicted as a potential allosteric site. Our approach successfully identified all the known allosteric sites in a set of test proteins. Several new allosteric sites in the test proteins were also predicted. By use of one of the new allosteric sites predicted from *Escherichia coli* phosphoglycerate dehydrogenase (PGDH), novel allosteric regulating molecules were screened by molecular docking and enzymatic assay. Three novel allosteric inhibitors were discovered and their binding modes were confirmed by mutation experiments and competitive assay. The IC_{50} of the strongest inhibitor discovered was 21 μM, which is comparable to that of the native allosteric inhibitor L-serine. The novel allosteric site discovered in PGDH is L-serine-independent, and inhibitors targeting this site can be used as novel regulators of the *E. coli* serine synthesis pathway. Our approach for allosteric site prediction is generally applicable and the predicted sites can be used in discovering novel allosteric regulating molecules.

INTRODUCTION

Allostery is a common mechanism of regulating biological processes such as enzyme catalysis, signal transduction, and metabolic regulation. The use of allostery to regulate the activity of proteins is an important and promising strategy in drug discovery. In order to regulate activity by allostery, predictive methods need to be developed to discover allosteric binding sites to which subsequent drug design approaches can be applied.[1] Compared to regular drug binding sites, such as catalytic sites in enzymes, allosteric binding sites offer new opportunities to control protein activity, as exemplified by the allosteric sites in the kinase family[2] and various proteases.[3] In addition, as allosteric binding sites are usually not evolutionarily conserved, the corresponding allosteric effectors are likely to have high specificity. More importantly, in the context of regulating enzyme activity, allosteric effectors can act as both activators and inhibitors.[4−6]

Many studies have tried to identify new allosteric binding sites in proteins. Experimentally, Wells and co-workers[7] developed the tethering method, and discovered an allosteric site in the caspase family.[5,8] Another example is the inhibitory site in p38 MAP kinase near the ATP binding site identified by high-throughput screening.[9] Computationally, the identification of new allosteric binding sites is frequently coupled with the prediction of allosteric pathways. Ranganathan and co-workers[10] developed a method named statistical coupling analysis to deduce residue−residue coupling in an allosteric communication network from the sequence alignment of a protein family. This method can be applied to protein families with a large amount of known sequences such as G protein-coupled receptors. Another method is the ensemble-based method which uses molecular dynamics simulations to generate an ensemble and uses features such as folding/unfolding,[11] energy flows,[12−14] and motions in Cartesian space[15−17] or internal coordinates[18,19] to get the long-distance coupling information. An example is the MutInf method developed by McClendon et al.[18] In their approach, several conventional molecular dynamics simulations are carried out and the statistically filtered correlations of the internal coordinates are calculated from the trajectories. This method is suitable for allostery without large conformational changes.[20] Other methods include machine learning approach,[21] geometric method using residue contact information,[22−24] and the widely applied normal-mode analysis method.[25−28] However, most of these methods focus on elucidating allosteric pathways and do not pay much attention on the binding property of the regulatory site. Recently, Blondel and co-workers[29] proposed a method of finding pockets with two distinct conformations (e.g., open and closed) in two functional states of a protein as allosteric sites. The use of such pockets as allosteric sites is straightforward, as the binding of a ligand to one conformation of the pocket

Received: May 16, 2012
Published: July 9, 2012

stabilizes the protein in the corresponding state and thus can allosterically regulate the activity. However, this approach can only treat cases where the binding sites have significantly different conformations in the two states.

In the present study, we used a coarse-grained two-state Gō model[30] to identify allosteric sites in proteins. This model combines two conventional single-state Gō potentials and generates an ensemble of two different functional states of a protein such as the T and R states in allostery. It has been used to study the conformational transitions of proteins,[31−33] the coupling between binding and protein conformational change,[34] and the allosteric coupling in a drug exporter.[35] Using the two-state Gō model, we first constructed an ensemble of the two functional states of a protein with the underlying concept that allostery is a population shift process.[36,37] We then added perturbations to a potential binding site and monitored the population distribution of the new ensemble. If population redistribution occurs, then the binding site perturbed is predicted as an allosteric site. We tested the method on five known allosteric proteins and successfully found all the known allosteric sites. Furthermore, several novel allosteric sites were also predicted in these proteins. We selected one of the proteins, *Escherichia coli* phosphoglycerate dehydrogenase (PGDH), as a test case to see whether novel allosteric inhibitors can be discovered that target the predicted site. Virutual screening by molecular docking was performed with a known chemical database and the top-ranking molecules were experimentally tested. Three compounds showed remarkable inhibition activities with the lowest IC_{50} value of 21.6 μM. Mutation studies confirmed that these molecules did bind to the predicted allosteric site. Compared to the known allosteric site used by L-serine, this new allosteric site is much larger in size and can be easily used to design novel chemical regulators.

MATERIALS AND METHODS

Collection of Test Cases. In order to test our approach, we collected allosteric proteins from two sources. The first was Gray's database of allosteric proteins,[38] which consists of 51 proteins with known bound and unbound structures. Since this database was published in 2007, we also searched the Protein Data Bank (www.rcsb.org)[39] for allosteric proteins deposited thereafter. Each of the resulting proteins was then manually inspected, and all those with peptide or ions as allosteric effectors were discarded. The protein had to satisfy several requirements to be applicable to our approach. First, a structural rearrangement after the binding of the allosteric effector should exist. This is because in the two-state Gō potential, the two states are mainly distinguished by their native contacts, thus two closely similar structures are likely to degenerate to a single state. Second, in order to reduce computational cost, the size of the protein was limited to below 1800 residues. Third, the protein should be complete with only a limited number of missing residues. The final test cases consisted of five enzymes: protein tyrosine phosphatase 1B (PTP1B),[40] aspartokinase III (AKIII),[41] caspase 1,[42] 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase (DAHPS),[43] and PGDH[44] (Table 1).

Coarse-Grained Simulations. For all the coarse-grained simulations, we used the software Cafemol developed by the Takata group.[45] The system was simulated with Langevin dynamics at constant temperature, which was set to 0.85 of the folding transition temperature determined by the search module of Cafemol (Table 1). The energy function consists of bond length, bond angle, dihedral angle, 12−10 Gō potential for native contact, and a repulsion term for nonnative contact. Two important parameters shape the energy surface of a system, namely, Δ, which determines the height of the energy barrier, and ΔV, which determines the relative energy of the two states. These parameters were systematically tuned so that without any perturbations, the population of the unbound state was between 75% and 95% (Table 1). Each simulation consisted of 5×10^7 steps, and the trajectory was saved every 1000 steps. From the resulting 5×10^4 snapshots, we calculated the population distribution with an order parameter χ, which indicates whether the system is in the first (unbound, negative χ) or the second state (bound, positive χ). Detailed explanations of the order parameter as well as Δ and ΔV can be found in ref 30 and the Cafemol manual. To add the perturbations, we assumed that the binding of a ligand may change the interactions between the residues of the binding site. Practically, we modified the native structure information file that was read by Cafemol, and we rescaled the native contact interactions within a binding site by a factor F. Both decreased and increased interactions were tested. Five simulations were performed for each protein with different F values at 0.2, 0.5, 1.0, 2.0, and 4.0, which were selected on the basis of perturbation−response curves in several trial simulations. If perturbations can shift the unbound state to the bound state, then the target site was predicted as an allosteric site.

Table 1. List of Five Test Proteins and the Parameters Used in Simulations

protein name[a]	bound structure	unbound structure	temp, K	Δ	ΔV
PTP1B	1T48	1PTY	307.0	98	18
AKIII	2J0X	2J0W	314.2	295	82
caspase 1	2FQQ	2HBQ	300.0	90	277
DAHPS	1KFL	1N8F	323.4	150	−247
PGDH	1PSD	1YBA	302.1	250	526

[a]PTP1B, protein tyrosine phosphatase 1B; AKIII, aspartokinase III, DAHPS, 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase; PGDH, phosphoglycerate dehydrogenase.

This approach was first performed on the known allosteric sites in the five testing proteins. Residues perturbed in the known allosteric sites were those within 5 Å of the allosteric effector in the bound structure. We then applied the approach to predict novel allosteric sites. Surface cavities were identified by use of the Cavity program.[46] All the identified new cavities were simulated by the above procedure to predict potential allosteric sites.

Calculation of Residue−Residue Interactions. To compare residue−residue interactions in an allosteric site, hydrogen bonds were defined by use of the program HBPLUS.[47] A contact was defined when two residues, separated by at least five residues, have a nearest atom−atom distance of less than 5 Å.

For the energy decomposition scheme, each state (bound and unbound) of the test proteins was first simulated in water for 20 ns. The protein was modeled with the ff99SB force field[48] and was solvated in a periodic TIP3P water box. Then counterions were added to neutralize the system. For all the bound states, we explicitly modeled the allosteric effectors, whose parameter and topology files were derived with the

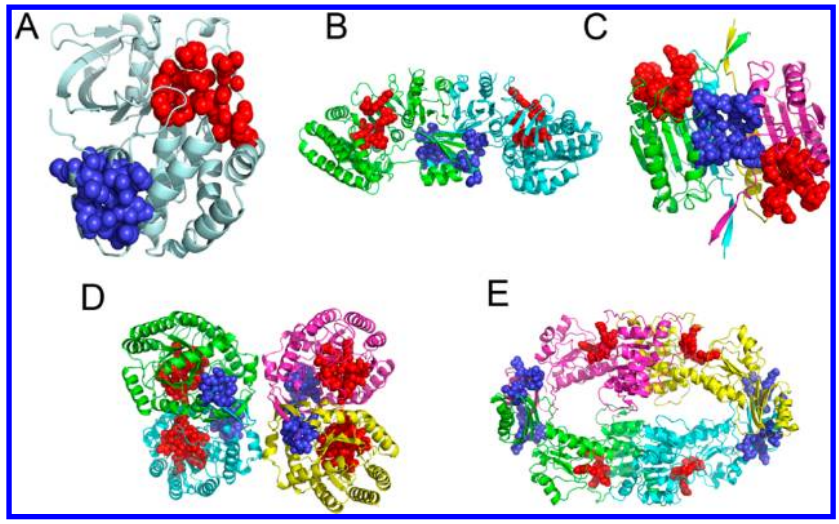

Figure 1. Structures of the five test proteins. Allosteric binding sites are shown as blue spheres. Active sites of the proteins are shown as red spheres. (A) PTP1B; (B) AKIII; (C) caspase 1; (D) DAHPS; (E) PGDH. Figure was prepared with Pymol.[65].

Antechamber package.[49] The system was minimized with 500 steps of steepest descent and 500 steps of conjugate gradient methods and was then heated to 290 K in 20 ps. In the production run, the trajectory was saved every 20 ps without water molecules and ions. The energy of each water-free snapshot was then decomposed on a per-residue basis with the MM/GBSA method. Both the atomistic simulation and the energy decomposition were carried out with Amber.[50]

Virtual Screening and Enzyme Inhibition Testing. Virtual screening against cavity 2 in PGDH was carried out with DOCK 6[51] and AutoDock Vina.[52] Specifically, DOCK 6 with default parameters was used first to screen the SPECS library, and the resulting top 10 000 compounds were further screened with AutoDock Vina under default parameters and flexible-ligand representation. The top 1000 outputs of AutoDock were then manually inspected and 178 compounds were purchased from SPECS for in vitro experiments.

For the experimental studies, reduced nicotinamide adenine dinucleotide (NADH), α-ketoglutarate (αKG), and L-serine were all from Sigma–Aldrich; ampicilin, clarithromycin, isopropyl β-thiogalactoside (IPTG), dithiothreitol (DTT), ethylenediaminetetraacetic acid (EDTA), phenylmethanesulfonyl fluoride (PMSF), and dimethyl sulfoxide (DMSO) were all from Amresco. Water was purified with a Milli-Q Reagent Water System (Millipore, Billerica, MA). The coding region of the PGDH gene was amplified by polymerase chain reaction (PCR) from the genome of *E. coli* strain K-12 by use of Phusion high-fidelity DNA pPolymerase [M0530, New England Biolabs (NEB)] and synthetic primers PGDH-*Nde*IF (5′-CGTGTACATATGATGGCAAAGGTATCGCTGGAG-3′) and PGDH-*Xho*IR (5′-CGTGTACTCGAGGTACAGCA-GACGGGCG-3′). *Nde*I and *Xho*I restriction endonucleases were purchased from NEB with codes R0111 and R0146. The amplified fragments of PGDH were ligated into the pET21a(+) vector by use of T4 DNA ligase (M0202, NEB), confirmed by DNA sequencing (Genewiz), and then transformed into the Rosetta-DE3 strain of *E. coli*. All the mutations were constructed by use of the QuikChange site-directed mutagenesis kit and confirmed by DNA sequencing (Genewiz). For expression, recombinant cells were cultivated at 37 °C in 1 L of Luria–Bertani (LB) medium containing 100 μg/mL ampicillin and 34 μg/mL clarithromycin until OD_{600} reached 0.6–0.8, and expression was then induced with 0.5 mM IPTG at 30 °C for another 4 h. The cells were harvested by centrifugation at 6000 rpm for 15 min at 4 °C and resuspended in buffer A [40 mM phosphate-buffered saline (PBS), pH 7.3, 10 mM imidazole, 100 mM NaCl, and 1 mM DTT] with 50 μM PMSF. Cell lysis was achieved by ultrasonication and then centrifugation at 17 000 rpm for 30 min at 4 °C. The supernatant was filtered by Syringe filters (0.20 μm, Minisart-N, Sartorius Stedim Biotech), applied to a nickel–nitrilotriacetic column (GE Healthcare), and equilibrated with buffer A. Then the proteins bound to the resin were eluted with a 0–100% gradient of buffer B (40 mM PBS, pH 7.3, 500 mM imidazole, 100 mM NaCl, and 1 mM DTT). The target protein, confirmed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), was concentrated by centrifugation through centrifugal filters (50 kDa, Amicon Ultra) and loaded on a gel-filtration column (Sephacryl S-200 HR, GE Healthcare) that had been pre-equilibrated with buffer C (40 mM PBS, pH 7.3, 0.5 mM EDTA, 100 mM NaCl, and 1 mM DTT). The final purity of the protein was >95% as judged by SDS–PAGE. Protein concentrations were measured via Nanodrop 2000 (Thermo Scientific).

The enzyme activity of PGDH can be measured by monitoring the disappearance of NADH, which absorbs at 340 nm.[53] As most of the compounds tested have absorption at 340 nm, the fluorescence emission of NADH was monitored instead (excitation at 338 nm and emission at 456 nm). The enzyme was incubated at room temperature in 96-well microtiter plates (Costar, Corning Inc.) in 20 mM tris-(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), pH 7.5, 1 mM DTT, 1 mM EDTA, and 0.25 mM NADH. After 6 min preincubation of a compound (dissolved in DMSO) and the enzyme, the reactions were initiated by adding the substrate α-KG. Then the fluorescence emission of NADH was monitored via a plate reader (Synergy, Biotek). End-point method was used to measure enzyme activity and inhibition. According to the linear range of enzyme reaction rate (Figure S1, Supporting Information), the signals at 30 s were recorded and used in inhibition percentage calculation. The NADH concentrations were estimated by comparing the fluorescence strengths to a standard curve (Figure S2, Supporting

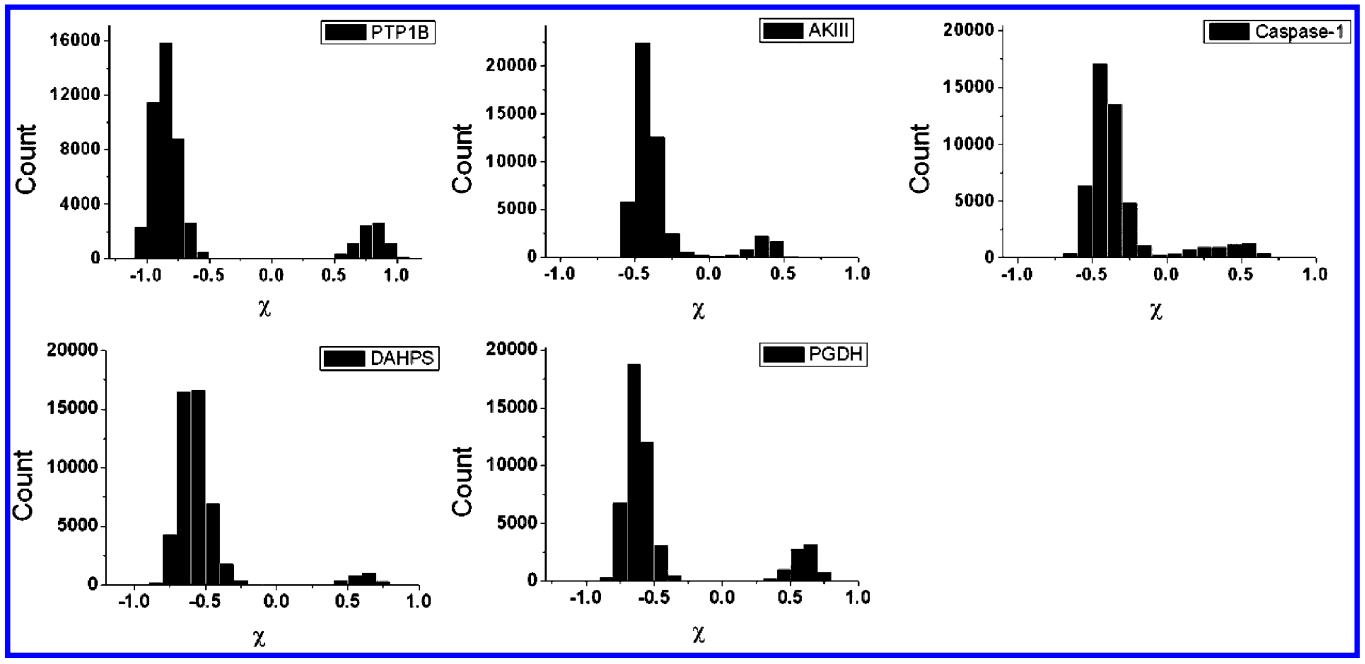

Figure 2. Population distributions of unperturbed ensembles for the five test proteins. The order parameter χ indicates whether the system is in the unbound state (negative χ) or the bound state (positive χ). The energy surfaces were tuned so that the percentage of unbound state was between 75% and 95%.

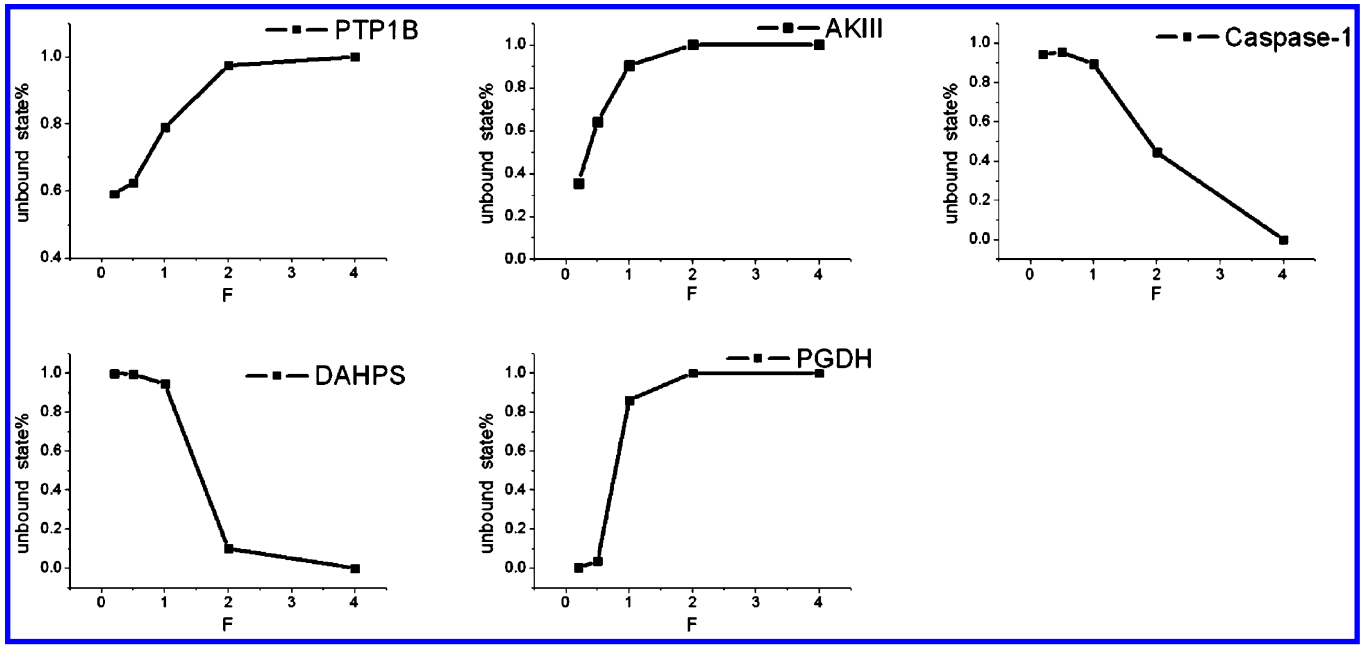

Figure 3. Prediction of known allosteric sites. The percentage of the unbound state of the five test proteins with respect to the factor F that rescales the native contact interactions within the allosteric binding sites is plotted. A potential binding site is predicted as an allosteric site if changing the interactions in this site causes a population shift from the unbound to the bound state.

Information). Percentages of inhibition were calculated according to the following equation:

$$1 - \frac{C_0 - C_t}{C_0 - C_b}$$

in which C_0 is the initial concentration of NADH, and C_t and C_b are the NADH concentrations measured in the test and blank samples. IC_{50} values were calculated from dose−response curves by fitting to a three-parameter logistic model. Data were expressed as mean ± SEM ($n \geq 3$). In competition experiments, the compounds concentrations were kept at their IC_{50} values, whereas the substrate concentration was gradually increased from 78 to 625 μM.

RESULTS

Description of Test Proteins. There are quite a few allosteric proteins found so far.[54] We limited our test proteins to those using small ligands other than peptides or ions as allosteric effectors. Moreover, as the energy of the two-state Gō model uses residue−residue contact information, the structures

dx.doi.org/10.1021/ct300395h | J. Chem. Theory Comput. 2012, 8, 2962−2971

of the bound and unbound states must have remarkably different residue−residue contact maps. The structural difference could be a local rearrangement such as loop displacement or a global structural change such as domain−domain movement. After searching through a public allosteric protein data set[38] and the Protein Data Bank (www.rcsb.org),[39] five allosteric proteins were selected as our test set: protein tyrosine phosphatase 1B (PTP1B), aspartokinase III (AKIII), caspase 1, 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase (DAHPS) and phosphoglycerate dehydrogenase (PGDH) (Figure 1 and Table 1). All five proteins are enzymes, and all the allosteric effectors are inhibitors. In PTP1B, the effector blocks the motility of a loop near the active site and thus prevents formation of the active form.[40] In AKIII, the effector causes a large domain rearrangement and blocks the ATP-binding site.[41] In caspase 1, the effector disrupts the hydrogen-bond network that connects two active sites.[42] A loop near the active site also shows different conformations in the bound and unbound structures. In DAHPS, the effector significantly changes the conformation of one active loop.[43] In PDGH, the effector causes a rearrangement of the orientations of the substrate-binding domain and the nucleotide-binding domain.[44]

Outline of the Approach. We first constructed an ensemble of the bound and unbound states of a protein with the coarse-grained model. The simulation was carried out with Langevin dynamics at constant temperature, and each residue was represented by a bead (see Materials and Methods). Two parameters, the energy barrier and the relative energy depth of the two states, determine the population distribution of the ensemble. These parameters were fine-tuned so that the unbound state was dominant in the ensemble (Table 1 and Figure 2). We then added perturbations to a binding site to mimic the binding of a ligand. The perturbation approach has been used in several studies on allosteric pathways in proteins.[12,22] In our approach, the interactions within a binding site were directly modified with the assumption that the binding of a ligand may change the interactions between the residues in that site. More specifically, we rescaled the native contact interactions within the binding site by a factor *F*. After adding the perturbation, we ran another simulation and monitored the population distribution. If the bound state became the dominant state, the perturbed site was classified as an allosteric binding site.

Prediction of Known Allosteric Sites. We first applied the approach to the five known allosteric binding sites to see whether it can give correct predictions. Five simulations for each protein were run with *F* of 0.2, 0.5, 1.0, 2.0, and 4.0. The conformation ensemble indeed changed from the unbound state to the bound state (Figure 3). Interestingly, the five proteins can be classified into two groups based on the behavior when changing *F*. For PTP1B, AKIII, and PGDH, the bound state becomes the dominant state when the interactions are decreased ($F < 1$). For caspase 1 and DAHPS, the bound state becomes the dominant state when the interactions are increased ($F > 1$). We noted that, for PTP1B, the percentage of unbound state did not fall below 50% when *F* was 0.2, in contrast to other proteins where the unbound states were <40%. This means the bound state only becomes comparable to the unbound state for PTP1B. As the allosteric inhibitor of PTP1B in the bound state was weak with an IC_{50} of 350 μM,[40] it is understandable in this case that the inhibitor does not significantly reshape the energy landscape. In comparison, the effector of PGDH has an IC_{50} of 8 μM,[55] and the transition in this protein was sharper than that in PTP1B (Figure 3). The percentage of unbound state in PGDH was 0.34% at $F = 0.2$.

In order to understand why the transitions happened at either $F > 1$ or $F < 1$, we analyzed the structural features and residue−residue interaction energies for the five test cases. From a structural point of view, we compared the number of hydrogen bonds and contacts formed within the binding sites in the bound and unbound structures. We reasoned that if the unbound-to-bound transition occurs when $F < 1$, the number of hydrogen bonds and contacts should be smaller in the bound structure. On the contrary, the number should be larger if the transition occurs when $F > 1$. The comparison indeed supported this assumption (Table 2). For PTP1B, AKIII, and PGDH, the unbound-to-bound transitions occurred when $F < 1$, and the numbers of hydrogen bonds and contacts in the bound structures were smaller. Transitions in caspase 1 and DAHPS occurred when $F > 1$, and the numbers of hydrogen bonds and contacts in the bound structures were larger than those in the unbound structures. From a dynamic point of view, we carried out molecular dynamics simulations to further elucidate the residue−residue interactions in the known allosteric sites. Specifically, the bound and unbound states of each test protein were simulated in water for 20 ns, and the residue−residue interactions in each snapshot of the trajectory were calculated with an energy decomposition scheme (see Materials and Methods). For proteins whose transition point is $F < 1$ (PTP1B, AKIII, and PGDH), the interactions in the bound state were weaker than those in the unbound states, whereas for proteins whose transition point is $F > 1$ (caspase 1 and DAHPS), the interactions in the bound state were stronger (Figure 4). Thus the residue−residue interactions from both static and dynamic structures of the protein support the assumption described above.

Table 2. Number of Hydrogen Bonds and Contacts Formed within Known Allosteric Sites from Unbound and Bound Structures of Five Test Proteins

		no. of hydrogen bonds		no. of contacts	
protein name	transition condition	unbound	bound	unbound	bound
PTP1B	$F < 1$	10	9	14	8
AKIII	$F < 1$	7	3	13	12
caspase 1	$F > 1$	2	3	7	8
DAHPS	$F > 1$	5	6	12	16
PGDH	$F < 1$	3	2	8	8

Prediction of Novel Allosteric Sites. We then applied this approach to evaluate other cavities in the five test proteins. The cavities other than the known allosteric sites were detected by use of the program Cavity.[46] Only the cavities suitable for small molecule design were kept, while those large cavities between the monomer−monomer interfaces in large proteins like DAHPS and PGDH were discarded. We calculated the percentage of the unbound state when different perturbations were applied to these cavities with the same procedure as the known allosteric sites (Figure 5). By use of the same criterion, several potential allosteric binding sites were identified. These were cavity 1 in caspase 1; cavities 3, 4, and 5 in DAHPS; and cavities 1 and 2 in PGDH (Figure 6). Cavities 1, 2, and 3 in PTP1B, cavity 2 in AKIII, and cavity 2 in caspase 1 did not have significant impact on the equilibrium and were classified as nonallosteric sites. Cavity 1 in AKIII was a plausible site

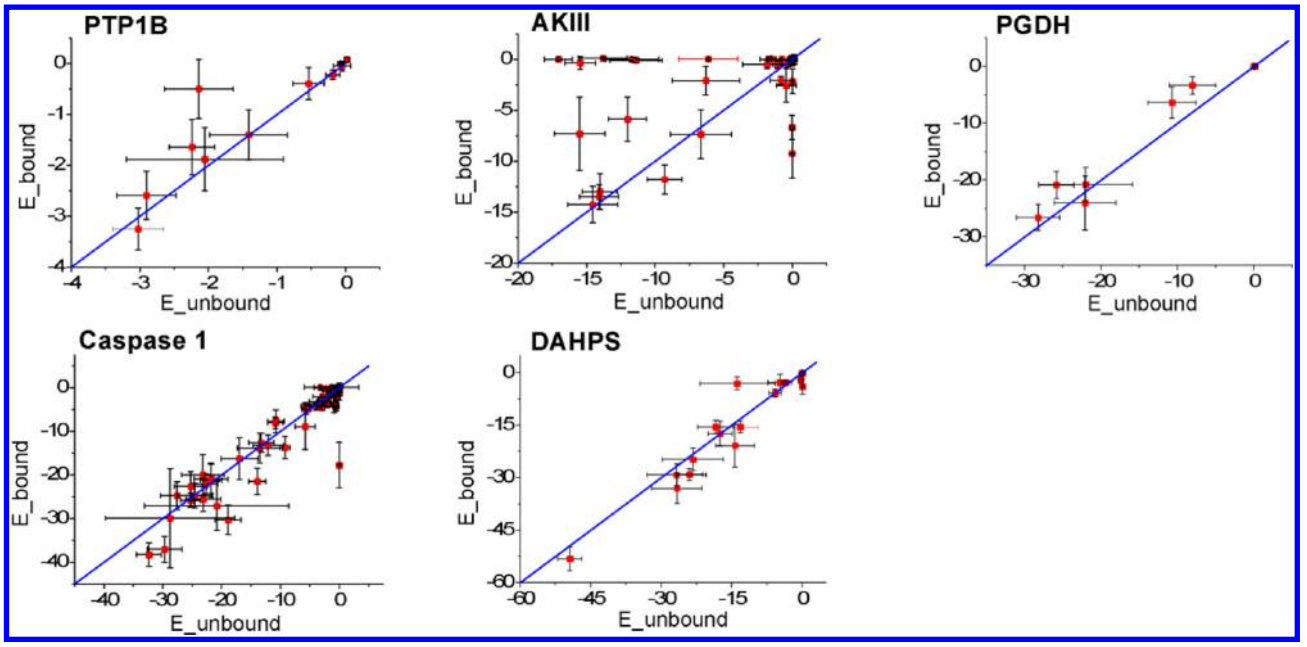

Figure 4. Residue−residue interactions in the known allosteric sites from the two states of test proteins. The energy value was decomposed and averaged over 1000 snapshots from a molecular dynamics simulation. For PTP1B, AKIII, and PGDH, the interactions in the bound state were weaker than those in the unbound states, whereas for caspase 1 and DAHPS, the interactions in the bound state were stronger. Blue lines indicate the isopleths, and crosses indicate standard error.

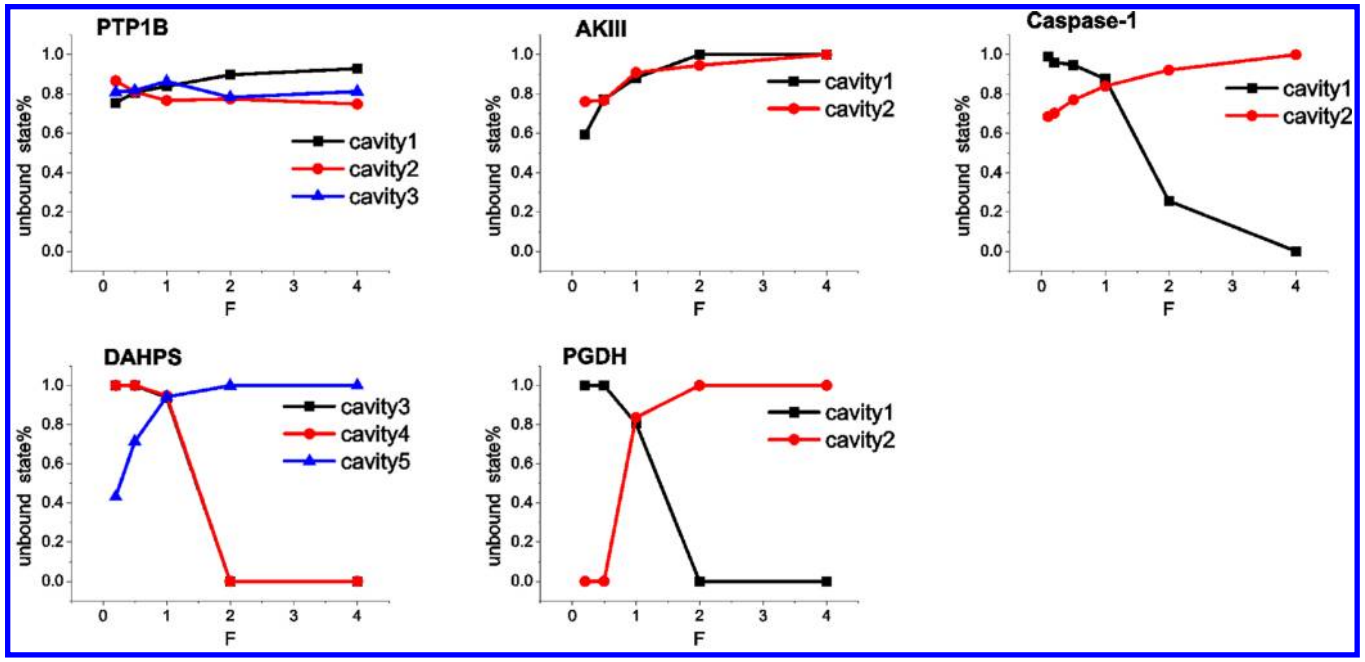

Figure 5. Prediction of new allosteric sites in test proteins. Percentages of the unbound state of the five test proteins with respect to the factor *F* that rescales the native contact interactions within each cavity are plotted.

because the percentage of the unbound state was ∼50% when *F* was small. The residue−residue interactions in these nonallosteric sites were analyzed by the same approach as the allosteric sites (Figure S3, Supporting Information). For cavities 1, 2, and 3 in PTP1B and cavity 2 in caspase 1, the residue−residue energies in the unbound and bound states are comparable. For cavity 2 in AKIII, some residue pairs have larger energies in the bound state, while some others have larger energies in the unbound state. This means that there is no remarkable residue−residue interaction energy difference in the bound and unbound structures for nonallosteric sites, while the differences are observable for allosteric sites.

Cavity 1 in caspase 1 resides on the opposite site of the known allosteric site with a volume of 346.37 Å^3. The predicted maximal pK_d is 8.77, which means the maximal binding affinity of this cavity to a ligand is $10^{-8.77}$ M or in the nanomolar range. Cavity 1 in PGDH lies between the substrate binding site and the nucleotide binding site with a volume of 1628.0 Å^3. The predicted maximal pK_d is 9.96. Cavity 2 in PGDH is near the allosteric site of the regulatory binding domain with a volume of 1277.2 Å^3. The predicted maximal pK_d is 9.63. Three

dx.doi.org/10.1021/ct300395h | J. Chem. Theory Comput. 2012, 8, 2962−2971

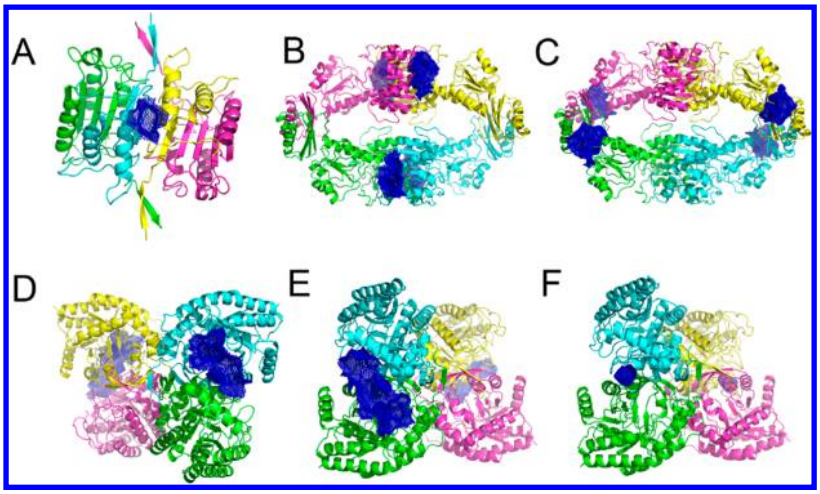

Figure 6. Structures of the cavities predicted as potential allosteric sites. Surfaces of the cavities are shown as dark blue dots. (A) Cavity 1 in caspase 1; (B, C) cavities 1 and 2 in PGDH; (D–F) cavities 3, 4, and 5 in DAHPS.

residues from cavity 2, namely, Arg339, Arg405, and Arg407, have been shown to allosterically affect the activity of PGDH.[56] These residues are involved in hydrogen-bond interactions between the regulatory and substrate-binding domains. Mutation of Arg339 to Lys causes a 10-fold reduction of k_{cat}, and mutations of Arg405 and Arg407 to Ala lower the k_{cat} and k_{cat}/K_m by 1–2 orders of magnitude.[56] A ligand that binds to cavity 2 and disturbs the conformations of these residues is likely to allosterically affect the activity in a similar way. Cavities 3, 4, and 5 from DAHPS all reside on the lateral interface of two monomers. Cavities 3 and 4 are quite large, with volumes of 1483.12 and 2067.87 Å^3, and predicted maximal pK_d of 9.55 and 9.49, respectively. Though cavity 5 is relatively small (170.25 Å^3), it is highly ligandable with a predicted maximal pK_d of 9.85, which might be suitable for binding to a small ion.

The existence of more than one allosteric site in one protein is commonly observed. For example, ribonucleotide reductase has two allosteric sites, one for activity regulation and the other for specificity regulation.[57] Another example is fructose-1,6-bisphosphatase. In addition to the allosteric site in each monomer, a second allosteric site located near the center of the tetramer was identified by Honzatko and co-workers.[58] Although it is difficult to prove that a site is nonallosteric, the proportion of allosteric sites among all cavities with appropriate binding ability is an interesting question. In our case, if we take all the predicted allosteric sites as true, the percentage is about 58%.

Discovery of Novel Allosteric Inhibitors for PGDH Based on the Predicted Novel Allosteric Site. In the previous section, cavity 2 in PGDH was predicted to be a novel allosteric site, which was partly supported by mutagenesis study reported.[56] To verify whether this site is a true allosteric site and to discover novel allosteric inhibitors, we carried out molecular docking-based virtual screening and measured enzyme inhibition activities of the top-ranking molecules. PGDH from *E. coli* is an enzyme that is involved in the serine synthesis pathway, which catalyzes the formation of 3-phosphohydroxypyruvate from D-3-phosphoglycerate with NAD/NADH as cofactor.[44] It is allosterically inhibited by L-serine with an IC_{50} of 8 μM.[55] We used DOCK 6[51] and AutoDock Vina[52] to screen the SPECS library against cavity 2, and we selected 178 compounds for experimental testing.

The assay monitored the decrease of NADH fluorescence along with the enzyme reaction. Three compounds showed remarkable inhibition effects with IC_{50} values of 207.3, 21.6, and 195.5 μM (Table 3, Figure 7, and Figure S4 in Supporting Information). To test whether these inhibitors are indeed allosteric inhibitors that do not compete with the substrate, we carried out inhibitor–substrate competition experiments. If the inhibitor and substrate shared the same binding site, increasing the substrate concentration would reduce the inhibition effect. Our experimental results indicated that the substrate concentration did not affect the inhibition effects of the three compounds (Figure S5, Supporting Information), and thus all three inhibitors are indeed allosteric inhibitors. To further confirm whether the compounds bind at the predicted site, we carried out mutagenesis studies and measured the effect of the mutations on the strongest inhibitor, **2**. Three single mutations in cavity 2, namely, K8A, S316A, and Y410A, were selected according to the docked structure of inhibitor **2** (Figure 8). All the mutants retained their enzymatic activities (Figure S6, Supporting Information) but were inhibited much less by inhibitor **2** (Table 3, Figure S7, Supporting Information). Together, these results suggest that inhibitor **2** is a novel allosteric inhibitor of PGDH and cavity 2 is a novel allosteric site.

Table 3. PGDH Inhibitors Discovered by Experiment and Verification of Their Binding Sites

			inhibition of mutants,[a] %		
inhibitor	IC_{50} (μM)	substrate competition	K8A	S316A	Y410A
1	207 ± 10	no	nd[b]	nd	nd
2	21.6 ± 6.2	no	23.5	27.1	12.0
3	195.5 ± 5.8	no	nd	nd	nd

[a]Concentrations of inhibitors were kept at their IC_{50}. [b]nd, not determined due to weak inhibition.

DISCUSSION

In the present study, we combined a two-state Gō model and a perturbation method to predict allosteric sites in proteins. Our method was based on the assumption that the allosteric site can reshape the energy landscape after binding to a ligand. All the known allosteric sites in the five test proteins were successfully

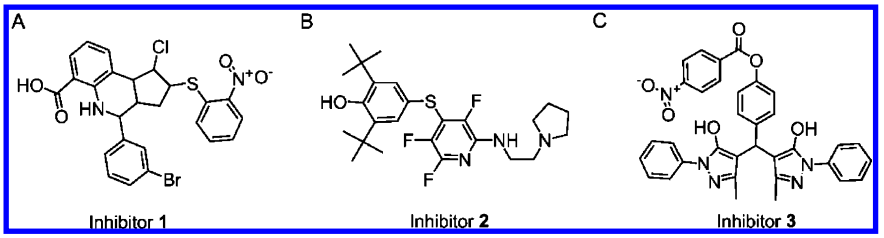

Figure 7. Structures of the three inhibitors. The SPECS IDs of inhibitors **1**−**3** are AG-205/13052034, AF-407/33312007, and AG-219/11941095.

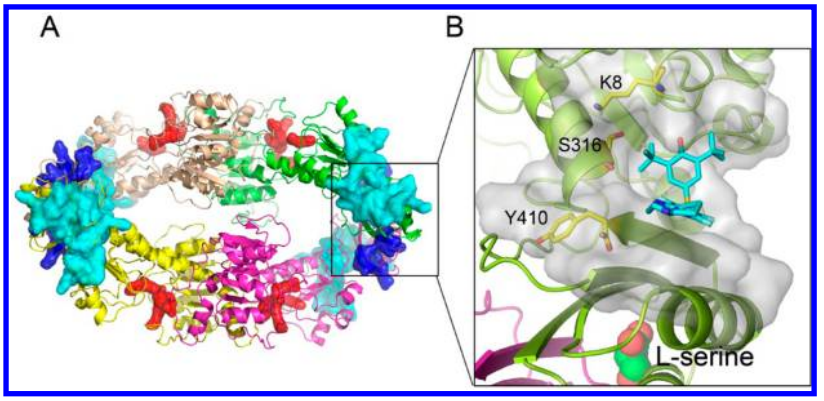

Figure 8. Structure of inhibitor **2** docked to cavity 2 in PGDH. (A) Location of cavity 2 in PGDH with active site in red, serine binding site in blue, and cavity 2 in cyan. (B) Enlargement of cavity 2. Inhibitor **2** and three mutation sites are in ball-and-stick representation; endogenous allosteric effector L-serine is shown in spheres, and cavity 2 is in surface mode.

found. Applying this method to other cavities in the test proteins led to the prediction of several new potential allosteric sites. Using one of the predicted sites in PGDH, three novel allosteric inhibitors were discovered by virtual screening and enzyme assay, confirming that the site predicted is indeed an allosteric site. Though the activity of inhibitor **2** is comparable to the endogenous allosteric effector, L-serine, the large size of its binding site (cavity 2, 1277.2 Å^3, Figure 8) is more suitable for inhibitor optimization and discovery of novel inhibitors with diverse chemical structures. Moreover, the new allosteric site allows the design of non-serine-competitive effectors, which provides a new way of regulating PGDH enzyme activity and the corresponding L-serine synthetic pathway in *E. coli*.

Many studies have been performed to elucidate the mechanism of allostery in various proteins. A major focus of these studies was to find potential allosteric pathways that connect the functional site to other regions of the protein. Less attention has been paid to the properties of the regulatory site, that is, whether the site is suitable for binding a ligand. A more practical approach is to predict a potential allosteric site that can be used for ligand binding. Once these sites are found, structure-based drug design approaches, such as virtual screening and fragment-based drug design, can be directly used to find new allosteric effectors. Our method directly addresses this problem by combining a coarse-grained model with a binding-site detection method. The advantage of using the two-state coarse-grained model is that the ensemble of the two functional states can be readily constructed. In comparison, all-atom simulation usually suffers from insufficient sampling, especially for proteins with large conformational changes. Although the atomistic details are omitted in the coarse-grained model, studies using this method, such as protein structure network and normal-mode analysis, have successfully elucidated the mechanisms of allostery in several proteins.[22,25−27,59] This means the information of allostery is still retained at residue level. The integration of atomistic models and the coarse-grained model may be implemented by multiscale methods, which use the coarse-grained model as a guide to the sampling of the atomistic model.[60,61]

The second advantage is that we implicitly incorporated the binding of the ligand by changing the interactions within the binding site, which does not require prior knowledge of the protein−ligand complex structure. This is of great importance for novel allosteric site prediction where the chemical structures of binding ligands are unknown. Several studies have also used implicit-ligand representation in combination with the two-state Gō model. In the work of Okazaki and Takada,[34] the interaction between the ligand and the protein was calculated with an additional Gaussian potential, which varies with the distance between two ligand-mediated residues. In another study, Daily et al.[62] included additional interactions in the Gō potential, with different criteria for selecting ligand-mediated residue pairs. Note that, in these two representations, the complex structure must be known first to define the ligand-mediated residue pairs. Recently, Sali and co-workers[63] used an implicit-ligand representation together with a two-state model to study different mechanisms of allostery: induced fit, population shift, and entropy-driven. In their approach, the effector binding was modeled with the size of the allosteric site. Atom pairs within a certain distance cutoff of the allosteric site have one energy minimum corresponding to the distance in the bound or unbound state, while all other pairs have two minima. By varying the distance cutoff, the system is driven from one state to another. This approach has the advantage that it can also treat entropy-driven allostery (allostery with no or little structural changes). However, it may be difficult to identify new allosteric sites in this approach, as increasing the distance cutoff in any target site will eventually change the state of the system, that is, generate an allosteric response. For explicit representation, Takagi and Kikuchi[64] used five beads to

represent a nucleotide molecule and studied the nucleotide dissociation of the myosin motor domain. The interaction between the molecule and the protein was calculated with a Gō potential including native and nonnative contact terms. Recently, Yao et al.[35] used hydrophobic energy to account for the ligand−protein interaction in a multistate Gō potential and studied the drug export process. We also tested the explicit representations and, unfortunately, adding a ligand at a known allosteric site did not generate an allosteric response at the active site in several test proteins. Neither the contact potential nor the hydrophobic energy was sufficient to represent the ligand−protein interactions in these cases.

Although our method is more generally applicable, it also has two limitations. The first is that both the bound and unbound structures must be known, and the structures must be different at the residue−residue contact level. However, in many cases we may only know the structure of one state, or little structural change occurs when the allosteric effector binds to the protein.[20] In these situations, methods that require only one structure and rely on dynamic information may be used. Second, as there is no direct way to calculate the parameters that shape the energy landscape, our method requires many trials to construct a proper two-state ensemble. This is quite time-consuming, especially for large complexes.

In summary, we have developed a method that can be used to detect potential allosteric sites in proteins. The current method requires preknowledge of the three-dimensional structures for the two functional states of a protein under investigation. Predicted allosteric sites can be used to design molecules that regulate protein functions through allostery. We have used one of the predicted sites in PGDH to discover novel allosteric inhibitors and successfully found three active compounds. Along with the rapid development of structural biology, we expect more protein structures, more cases of allostery, and more applicable examples of our approach.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Seven figures showing standard curve for PGDH activity measurement; residue−residue interactions in the predicted nonallosteric sites; inhibition and competition experiments of inhibitors **1**, **2**, and **3** to PGDH; activities of the three mutants K8A, S316A, and Y410A; and inhibition of inhibitor **2** to the three mutants. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Telephone 86-10-62757486; fax 86-10-62751725; e-mail lhlai@pku.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Xinqiu Yao for advice on the usage of Cafemol and Dr. Iain C. Bruce for revision of the manuscript. This work was supported in part by grants from the Ministry of Science and Technology of China and the National Natural Science Foundation of China (90913021, 21173013, and 11021463).

■ REFERENCES

(1) Hardy, J. A.; Wells, J. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 706−715.
(2) Bogoyevitch, M. A.; Fairlie, D. P. *Drug Discovery Today* **2007**, *12*, 622−633.
(3) Shen, A. *Mol. Biosyst.* **2010**, *6*, 1431−1443.
(4) Lupardus, P. J.; Shen, A.; Bogyo, M.; Garcia, K. C. *Science* **2008**, *322*, 265−268.
(5) Wolan, D. W.; Zorn, J. A.; Gray, D. C.; Wells, J. A. *Science* **2009**, *326*, 853−858.
(6) Zorn, J. A.; Wells, J. A. *Nat. Chem. Biol.* **2010**, *6*, 179−188.
(7) Erlanson, D. A.; Wells, J. A.; Braisted, A. C. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 199−223.
(8) Hardy, J. A.; Lam, J.; Nguyen, J. T.; O'Brien, T.; Wells, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12461−12466.
(9) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. *Nat. Struct. Biol.* **2002**, *9*, 268−272.
(10) Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. *Nat. Struct. Biol.* **2003**, *10*, 59−69.
(11) Pan, H.; Lee, J. C.; Hilser, V. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12020−12025.
(12) Ota, N.; Agard, D. A. *J. Mol. Biol.* **2005**, *351*, 345−354.
(13) Ishikura, T.; Yamato, T. *Chem. Phys. Lett.* **2006**, *432*, 533−537.
(14) Leitner, D. M. *J. Chem. Phys.* **2009**, *130*, No. 195101.
(15) Sharp, K.; Skinner, J. J. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 347−361.
(16) Morra, G.; Verkhivker, G.; Colombo, G. *PLoS Comput. Biol.* **2009**, *5*, No. e1000323.
(17) Morra, G.; Neves, M. A. C.; Plescia, C. J.; Tsustsumi, S.; Neckers, L.; Verkhivker, G.; Altieri, D. C.; Colombo, G. *J. Chem. Theory Comput.* **2010**, *6*, 2978−2989.
(18) McClendon, C. L.; Friedland, G.; Mobley, D. L.; Amirkhani, H.; Jacobson, M. P. *J. Chem. Theory Comput.* **2009**, *5*, 2486−2502.
(19) Long, D.; Bruschweiler, R. *J. Am. Chem. Soc.* **2011**, *133*, 18999−19005.
(20) Tsai, C. J.; del Sol, A.; Nussinov, R. *J. Mol. Biol.* **2008**, *378*, 1−11.
(21) Demerdash, O. N.; Daily, M. D.; Mitchell, J. C. *PLoS Comput. Biol.* **2009**, *5*, No. e1000531.
(22) del Sol, A.; Fujihashi, H.; Amoros, D.; Nussinov, R. *Mol. Syst. Biol.* **2006**, *2*, No. 2006.0019.
(23) Daily, M. D.; Upadhyaya, T. J.; Gray, J. J. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 455−466.
(24) Daily, M. D.; Gray, J. J. *PLoS Comput. Biol.* **2009**, *5*, No. e1000293.
(25) Balabin, I. A.; Yang, W.; Beratan, D. N. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 14253−14258.
(26) Tehver, R.; Chen, J.; Thirumalai, D. *J. Mol. Biol.* **2009**, *387*, 390−406.
(27) Yang, Z.; Majek, P.; Bahar, I. *PLoS Comput. Biol.* **2009**, *5*, No. e1000360.
(28) Williams, G. *BMC Struct. Biol.* **2010**, *10*, 11.
(29) Laine, E.; Goncalves, C.; Karst, J. C.; Lesnard, A.; Rault, S.; Tang, W. J.; Malliavin, T. E.; Ladant, D.; Blondel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11277−11282.
(30) Okazaki, K.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11844−11849.
(31) Best, R. B.; Chen, Y. G.; Hummer, G. *Structure* **2005**, *13*, 1755−1763.
(32) Chu, J. W.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3860−3871.
(33) Yang, S.; Roux, B. *PLoS Comput. Biol.* **2008**, *4*, No. e1000047.
(34) Okazaki, K.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11182−11187.
(35) Yao, X. Q.; Kenzaki, H.; Murakami, S.; Takada, S. *Nat. Commun.* **2010**, *1*, 117.
(36) Swain, J. F.; Gierasch, L. M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 102−108.

(37) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 433−443.
(38) Daily, M. D.; Gray, J. J. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 385−399.
(39) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(40) Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randal, M.; McDowell, R. S.; Hansen, S. K. *Nat. Struct. Mol. Biol.* **2004**, *11*, 730−737.
(41) Kotaka, M.; Ren, J.; Lockyer, M.; Hawkins, A. R.; Stammers, D. K. *J. Biol. Chem.* **2006**, *281*, 31544−31552.
(42) Datta, D.; Scheer, J. M.; Romanowski, M. J.; Wells, J. A. *J. Mol. Biol.* **2008**, *381*, 1157−1167.
(43) Shumilin, I. A.; Zhao, C.; Bauerle, R.; Kretsinger, R. H. *J. Mol. Biol.* **2002**, *320*, 1147−1156.
(44) Schuller, D. J.; Grant, G. A.; Banaszak, L. J. *Nat. Struct. Biol.* **1995**, *2*, 69−76.
(45) Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W. F.; Okazaki, K.; Yao, X. Q.; Takada, S. *J. Chem. Theory Comput.* **2011**, *7*, 1979−1989.
(46) Yuan, Y.; Pei, J.; Lai, L. *J. Chem. Inf. Model.* **2011**, *51*, 1083−1091.
(47) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777−793.
(48) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(49) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph Model* **2006**, *25*, 247−260.
(50) Case, D. A.; Darden, T. A. ; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; , Kollman, P. A. *AMBER 11*, University of California, San Francisco, 2010.
(51) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. *RNA* **2009**, *15*, 1219−1230.
(52) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455−461.
(53) Zhao, G.; Winkler, M. E. *J. Bacteriol.* **1996**, *178*, 232−239.
(54) Huang, Z.; Zhu, L.; Cao, Y.; Wu, G.; Liu, X.; Chen, Y.; Wang, Q.; Shi, T.; Zhao, Y.; Wang, Y.; Li, W.; Li, Y.; Chen, H.; Chen, G.; Zhang, J. *Nucleic Acids. Res.* **2011**, *39*, D663−669.
(55) Al-Rabiee, R.; Zhang, Y.; Grant, G. A. *J. Biol. Chem.* **1996**, *271*, 23235−23238.
(56) Grant, G. A.; Hu, Z.; Xu, X. L. *J. Biol. Chem.* **2001**, *276*, 1078−1083.
(57) Eriksson, M.; Uhlin, U.; Ramaswamy, S.; Ekberg, M.; Regnstrom, K.; Sjoberg, B. M.; Eklund, H. *Structure* **1997**, *5*, 1077−1092.
(58) Choe, J. Y.; Nelson, S. W.; Arienti, K. L.; Axe, F. U.; Collins, T. L.; Jones, T. K.; Kimmich, R. D.; Newman, M. J.; Norvell, K.; Ripka, W. C.; Romano, S. J.; Short, K. M.; Slee, D. H.; Fromm, H. J.; Honzatko, R. B. *J. Biol. Chem.* **2003**, *278*, 51176−51183.
(59) Panjkovich, A.; Daura, X. *BMC Struct. Biol.* **2010**, *10*, 9.
(60) Messer, B. M.; Roca, M.; Chu, Z. T.; Vicatos, S.; Kilshtain, A. V.; Warshel, A. *Proteins: Struct., Funct., Bioinf.* **2009**, *78*, 1212−1227.
(61) Li, W.; Takada, S. *Biophys. J.* **2010**, *99*, 3029−3037.
(62) Daily, M. D.; Phillips, G. N., Jr.; Cui, Q. *J. Mol. Biol.* **2010**, *400*, 618−631.
(63) Weinkam, P.; Pons, J.; Sali, A. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 4875−4880.
(64) Takagi, F.; Kikuchi, M. *Biophys. J.* **2007**, *93*, 3820−3827.
(65) The PyMOL Molecular Graphics System, Version 1.2r1, Schrodinger, LLC, 2009.