Modulation of Allostery with Multiple Mechanisms by Hotspot Mutations in TetR

Jiahua Deng*,[†,§] Yuchen Yuan,[†,§] and Qiang Cui*,[†,‡,¶]

†*Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, United States*

‡*Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, United States*

¶*Department of Biomedical Engineering, Boston University, 44 Cummington Mall, Boston, Massachusetts 02215, United States*

§*J. D. and Y. Y. contributed equally to this work*

E-mail: jhdeng@bu.edu,qiangcui@bu.edu

Abstract

Modulating allosteric coupling offers unique opportunities for biomedical applications. Such efforts can benefit from efficient prediction and evaluation of allostery hotspot residues that dictate the degree of co-operativity between distant sites. We demonstrate that effects of allostery hotspot mutations can be evaluated qualitatively and semi-quantitatively by molecular dynamics simulations in a bacterial tetracycline repressor (TetR). The simulations recapitulate the effects of these mutations on abolishing the induction function of TetR and provide a rationale for the different degrees of rescuability observed to restore allosteric coupling of the hotspot mutations. We demonstrate that the same non-inducible phenotype could be the result of perturbations in distinct structural and energetic properties of TetR. Our work underscore the value of explicitly computing the functional free energy landscapes to effectively evaluate and rank hotspot mutations despite the prevalence of compensatory interactions, and therefore provide quantitative guidance to allostery modulation for therapeutic and engineering applications.

Introduction

Allostery couples distant sites by binding an allosteric effector at the allosteric site, and propagation of structural or dynamical perturbations to the functional site to modulate its activity. It is essential to the function of proteins and nucleic acids in many biological processes.[1] Controlling the allosteric coupling and communications is of broad therapeutic value and offers unique opportunities[2] for biomedical and biotechnological applications. Designing molecules that target an allosteric site rather than the active site can potentially lead to higher levels of selectivity and specificity, especially for drug targets that feature highly conserved binding pockets such as kinases.[3,4] Numerous serious human diseases such as various cancers are caused by mutations that either disrupt or activate allosteric communications,[5] and resistance to drugs also often arises due to mutations that perturb allostery in the target protein.[6] The understanding of allosteric communications also makes it feasible to design and engineer proteins with desired purposes by mutations, such as controlling enzyme activities for catalysis, constructing regulatory systems for gene expression, and developing biosensors and biological logic gates.[7–13]

To effectively design allosteric therapeutics and engineer/re-engineer allosteric proteins, it is useful to first identify and target allostery hotspots. Hotspots are the set of key residues that dictate the co-operativity between the effector site and the active site. In addition to their locations, the specific sequence features of allostery hotspot residues may also be important to their functional significances, since their contributions are likely sensitive to the local steric or polar interactions they are engaged in. Allostery hotspots can be explored by various computational techniques, such as statistical analysis of residual co-evolution,[14–16] motional network and community analysis,[17–20] mapping propagation of conformational distortions,[21] and perturbation analysis of collective normal modes.[22–24] Experimentally, site-directed mutagenesis,[25] solution NMR chemical shift perturbations and relaxation measurements,[26,27] surface plasmon resonance,[28] and *in vitro* and *in vivo* biochemical essays/kinetics[29] are commonly used approaches to probe allostery hotspots. As each methodology has its own

limitation(s) in either cost or accuracy, it is still challenging to identify a complete set of allosteric hotspots, which prevents the evaluation of the individual role of each hotspot and a clear understanding of their mechanisms. A high-throughput methodology which could exhaustively and systematically analyze the sequence-function relationship is of great promise. Deep mutational scanning (DMS) analysis[30–32] has emerged as a powerful technique for functional screening of allostery hotspots in proteins in a comprehensive and unbiased fashion. As demonstrated by recent applications,[33–38] DMS analysis readily reveals both the location and sequence variability of allostery hotspots, leading to new insights into the unique characteristics of allosteric networks in various systems. However, DMS studies are mostly limited to the identification of phenotypes through the binary functional readout due to its high-throughput nature. Establishing the effect of a given mutation on protein function and other quantitative properties requires additional molecular details.

Another feature that is commonly observed in allosteric systems is the network of compensatory interactions,[39] which leads to near-degenerate allosteric communication pathways.[40] This feature makes allostery highly adaptable during evolution and ensures robustness of functionally essential co-operativities;[34] it can also facilitate re-wiring of allostery communication pathways in engineering applications.[12,41] Nevertheless, these compensatory effects have major implications to the design of allosteric inhibitors, as the effect of design might be easily abolished by additional mutations.[4,6] Therefore, robust designs should target allostery communications that, once disrupted, are difficult to restore. Such properties are costly to evaluate using experiments due to the need of sampling a large number of sequences. Providing insights into how easily perturbations due to allostery hotspot mutations are reversed using computational approaches[42] is of tremendous value but has not been widely attempted.

To comparatively evaluate the effect of hotspot mutations, it is important to employ computations to quantify the functional outcome of different mutations. Such analysis will greatly facilitate the development of specific strategies for either enhancing or reversing the effect of mutation by introducing additional mutations or interactions with other molecules.[43]

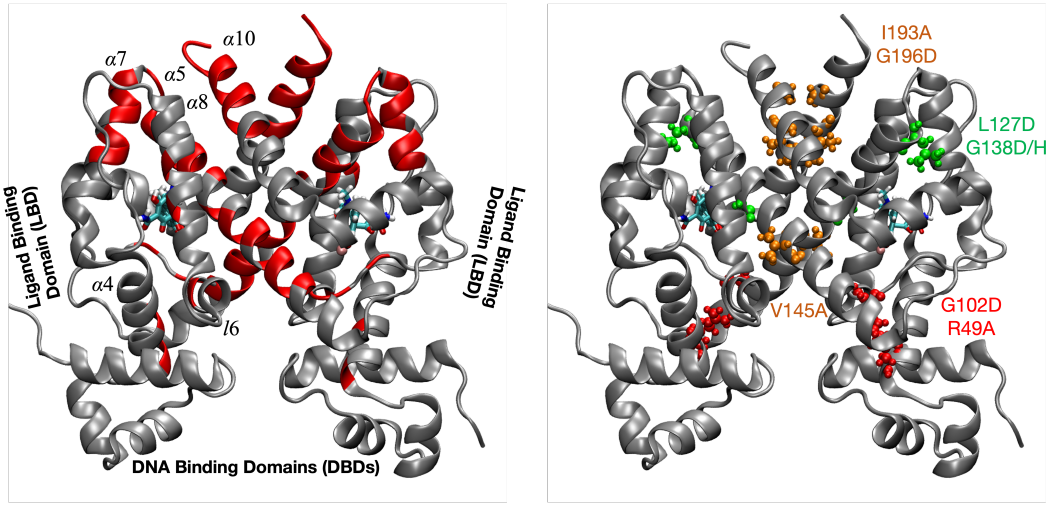
For example, allosteric communication can be tuned by modulating the stability of altered loop conformations,[44,45] targeting new surface binding pockets via small molecules,[16,43,46,47] or mutations of key residues along allostery pathways.[12,48]

A fitting example in this context is a bacterial transcription repressor, tetracycline repressor (TetR). TetR's normal function relies on the allosteric coupling between the ligand binding domains (LBDs) and the DNA binding domains (DBDs), which are separated by more than 20 Å (Fig. 1a). Binding of the inducer (ligand-Mg^{2+} complex) leads to the dissociation of the repressor from the promoter DNA sequence, allowing the access of RNA polymerase and therefore transcription of the repressed gene. Firmly establishing the contributions of hotspot residues has fundamental values to many aspects of prokaryotic physiology, including antibiotic resistance, and provides guidance to rational engineering of the TetR family for gene regulation.[49] Therefore, allostery in TetR has been extensively studied with structural,[50–53] mutagenesis[54–56] and more recently, DMS studies.[34,35] Based on propagation of structural distortions, a set of allostery hotspots abolishing the induction function in TetR were proposed[57] and verified[58,59] to localize at the domain interface of the two binding sites and relay the effect of inducer binding to the DBDs, whose reorientation impacts the DNA binding affinity. By sharp contrast, DMS analysis observed that allostery hotspots in the TetR family members are broadly distributed over a major portion of the protein structures,[34,35] underscoring that a mechanical view of allostery,[21,47,60,61] especially in TetR is incomplete. Moreover, the DMS experiments found that the loss of induction in hotspot mutants could be restored by additional mutations.[34] These observations again highlighted that the allostery network has a considerable degree of degeneracy and redundancy, and not all allostery hotspots are of equal weights during evolution.

These considerations motivated our recent atomistic molecular dynamics (MD) simulations of the wild type (WT) TetR.[62] Multiple local (e.g., backbone dihedral and residual contacts) and global (e.g., motional coupling and free energy landscapes) analyses established the molecular and energetic basis of the strong anti-cooperativity between the ligand

and DNA binding sites through a conformational selection mechanism that underlies the allostery function of TetR. The multifaceted computational framework we established captured a major fraction of the hotspot residues identified in the DMS study and demonstrated that these residues contribute to allostery in distinct ways. Nevertheless, our previous simulations focused entirely on the WT protein, thus the roles and features of hotspot residues had to be inferred indirectly. Moreover, the general strategy for hotspot identification also generated a considerable number of "false positives" and missed hotspot residues in two regions, the middle of the α8 helix and the C-terminus of the protein.[62] Evidently, the molecular mechanism of the disruption of allostery in the hotspot mutants, the non-inducible phenotype, remains unclear.

In the current study, the major goal is to establish a computational framework that enables us to evaluate the functional outcome of hotspot mutations in TetR. We show that such evaluation can be accomplished at qualitative (whether a mutant is non-inducible) and semi-quantitative (rank order the impact of the mutation) levels by judiciously combining extensive unbiased (6-60 μs) and enhanced sampling molecular dynamics (MD) simulations. We select several representative mutants that involve hotspot residues in distinct regions of the protein (see Fig. 1b): domain interface (G102D, R49A), near the ligand binding site (L127D, G138D/H), and near the monomer-monomer interface, either in the middle of the α8 helix (V145A) or in the C-terminal helix (I193A and G196D). We note that several residues (V145A, I193A and G196D) were missed entirely in our previous computational analyses on the WT protein[62] and thus are chosen here to explicitly probe their mutation effects.



(a) (b)

Figure 1: **Overview of the TetR structure and experimentally-identified allostery hotspot distributions.** (a) DMS studies[34] revealed a broad distribution of allostery hotspots (in red) throughout the majority of the TetR structure. In the ligand-bound crystal structure (PDB code: 4ac0), the ligand, minocycline, is bound to a Mg^{2+} ion. The ligand is converted to anhydrotetracycline in all simulations. (b) The hotspot mutations studied in this work are grouped into three categories based on their locations: the domain interface (red), near the ligand binding site (green), and near the monomer-monomer interface (orange).

Computational Methods

Unbiased MD Simulations of Different Functional States

Extensive unbiased MD simulations (6-60 μs) are conducted to compare hotspot mutants and WT TetR in terms of key structural properties and construct free energy landscapes in multiple functional (apo, ligand-bound and DNA-bound) states on local NVIDIA P100 GPUs and the Anton 2 machine. WT trajectories were generated in the previous study.[62] The mutants studied in this work include G102D, R49A, L127D, G138D, G138H, V145A, I193A, and G196D. All mutations are generated *in silico* based on the corresponding coordinates of the WT. Unbiased MD setup for the mutants, the definitions of various order parameters and structural properties are the same as our previous work on the WT.[62] Simulation details and data analysis for the unbiased trajectories are presented in the SI.

For the ligand-bound state, we start with the mutant ligand-bound structures based on the crystal structure of the WT (PDB code: 4AC0). For the DNA-bound state, the initial structure is modeled based on the crystal structure of the TetR(D)-DNA complex (PDB code: 1QPI), in which the protein has a high level of sequence homology to TetR(B) (66%). 1 μs and 5 μs trajectories are generated for the bound states on local GPUs and Anton 2, respectively.

For the apo state, three independent models are used as the initial structures to reduce bias and facilitate the exploration of the conformational space. The first structure is the same as the ligand-bound crystal structure but without the ligand-Mg^{2+} complex; the second structure is the same as the DNA-bound structure but without the DNA; the third structure is modeled using the TetR(D) apo crystal structure (PDB code: 1BJZ) as the template. Independent simulations of apo states are carried out both on local GPUs and Anton 2 starting from different initial structures for each system. For apo state simulations on local GPUs, 3 independent trajectories of 1 μs or 3 μs are carried out starting from the aforementioned three initial structures, respectively. For the apo state simulations on Anton

2, 3 independent trajectories of 15 μs, 15 μs, and 20 μs are carried out for the WT starting from the three initial structures, respectively; 2 independent trajectories of 15 μs each are carried out for G102D starting from the first two initial structures of the apo state, respectively; 3 independent trajectories of 10 μs each are carried out for the rest mutants starting from the three initial structures, respectively.

Metadynamics Simulations of Ligand-bound States

Multiple-walker well-tempered metadynamics simulations[63,64] are carried out using PLUMED[65,66] interfaced with OpenMM[67] on NVIDIA A100 GPUs. The collective variable (CV), DNA-binding domain (DBD) distance, is the center-of-mass (COM) separation between the Cα of residues 37-44 in each monomer. The bias factor is set to be 10. The simulation temperature is 303.15 K. A new Gaussian biasing potential is added every 1 ps with an initial height of 1.2 kJ/mol and a width of 0.2 nm. Ten walkers with different initial velocities are used per system in parallel while sharing hill history with other walkers every 1 ps. A harmonic restraint is applied to the CV with a force constant of 150 kJ/mol when the value of the CV is larger than 5 nm (upper wall) or smaller than 3 nm (lower wall). Other basic MD parameters are the same as the unbiased simulations. Convergence is monitored by the block-average analysis of which the details are presented in the next subsection. The simulation is stopped when adequate convergence has reached, which typically takes 110 - 140 ns per walker in each independent metadynamics run.

Data Analysis for Metadynamics Simulations

Metadynamics trajectories are analyzed using the metadynamics reweighting algorithm[68] implemented in the PLUMED package (version 2.8)[66] to obtain the unbiased distribution of the DBD distance, which is then converted into the 1D PMF. Block-average analysis is applied to assess convergence of metadynamics simulations and the associated statistical errors for each system. The average errors of the block-averaged PMF as a function of the

block size are plotted as the convergence check. The final 1D PMF of each system is the block average of every 20 ns trajectories in each walker, and the shaded area in the PMF plot represents standard error of the mean (SEM) of the block-averaged PMF. The reweighted distributions of other order parameters, structural properties, as well as the 2D free energy landscapes are obtained using the same reweighting algorithm.

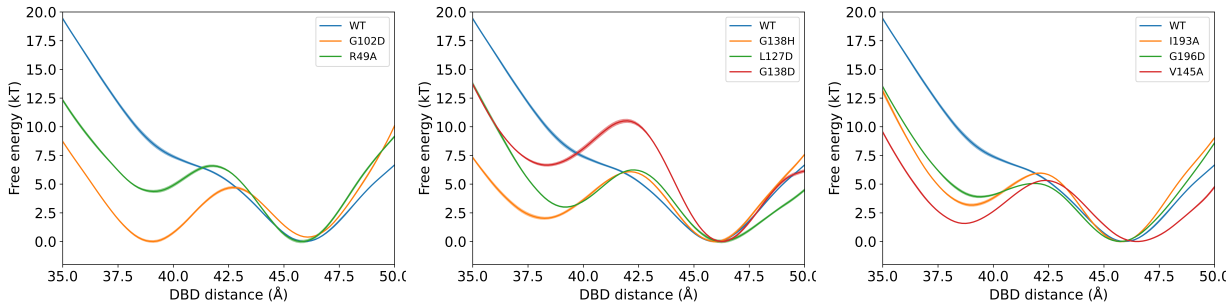
Results

Doubly-Bound States Are Better Stabilized in Hotspot Mutants

According to the induction function of TetR, ligand binding in the WT protein was proposed to rigidify the protein structure such that the two DBDs remain too far apart to allow favorable binding to the DNA. Thus, we reason that the loss of the induction function in some, if not all, hotspot mutants is due to adopting favorable conformations in the DBDs for DNA binding even in the ligand-bound state; i.e., these mutants are able to bind tightly to both the ligand and DNA in such conformations, referred to as the doubly-bound state. To test this hypothesis, we conduct metadynamics simulations to explicitly probe how hotspot mutations modulate the stability of the doubly-bound state, which is not expected to be stable in the WT protein. Under the condition of the DMS experiments, the ligand concentration is such that most TetR proteins are expected to be ligand bound. Therefore, free energy profiles for the DBD distance separation are computed in the ligand-bound state. The use of the DBD distance as the reaction coordinate is supported by previous structural studies of TetR[50,52] and MD simulations on the order of 50-100 ns,[69] which observed that the DBD distance can be used to distinguish induced (ligand-bound) and non-induced (DNA-bound) conformations, and thus a key indicator for the DNA binding affinity.

We note that the DBDs are perturbed by the hotspot mutations in both the DNA-bound (Fig. S1) and apo (Fig. S2) states in unbiased simulations. As shown in Fig. S2, the DNA-bound state of the mutants tends to span broader DBD distances compared to

the WT, leading to Kullback-Leibler (KL) divergence values in the range of 0.8-1.8 for the corresponding DBD distance distributions (Table S2). These values reflect modest differences in the DBD conformations between the mutants and WT. Regarding the DBDs in the apo state, their conformations in the hotspot mutants are overall similar to the WT, except that the DBD separations are slightly shifted towards longer distances (Fig. S2), most notably in G138D and G138H.



(a) (b) (c)

Figure 2: **Potentials of mean force (PMF) for the DBD distance separation in the ligand-bound states from metadynamics simulations.** The reaction coordinate is the center of mass separation between the two DBDs in each monomer. See the precise definition in Materials and Methods. All PMFs show a basin at $\sim$ 46 Å, referred to as the "ligand-only basin". All mutant PMFs also show another basin at $\sim$ 38 Å, referred to as the "doubly-bound basin". Subfigures are grouped by the regions shown in Fig. 1b. The solid lines are the average values and the shaded areas show the statistical error estimated from block analysis. The statistical errors are all on the order of 0.1 k_BT. For additional analysis of statistical error and convergence, see Fig. S3 in the SI.

In the WT, the computed free energy profile (Fig. 2) indicates that there is a single free energy basin at $\sim$ 46 Å along the DBD distance coordinate. The basin (ligand-only basin) corresponds to the crystal structure of the ligand-bound TetR, in which the DBD distance is $\sim$ 42 Å.[50] As the DBD distance decreases, the free energy increases monotonically. For the DBD distance reaching $\sim$38 Å, the corresponding conformation is expected for favorable DNA binding,[52] which has a free energy cost about 10 k_BT. These observations are consistent with the normal function of TetR. The low DNA binding affinity upon ligand binding suggests that compressing the DBDs in the ligand-bound state is energetically unfavorable in the WT. In all the hotspot mutants studied, by remarkable contrast, the computed free energy profiles

Table 1: Overlap of DBD distance distributions between the ligand-bound and DNA-bound states characterized by the Kullback-Leibler (KL) divergence [a]

	WT	G102D	R49A	G138H	L127D
reweighted	6.8	1.2	3.0	1.7	2.8
doubly-bound basin[b]	6.0	0.4	0.5	0.1	0.6
ligand-only basin[b]	54.4	56.7	57.3	56.7	57.5
		G138D	V145A	I193A	G196D
reweighted		3.8	2.8	2.0	3.3
doubly-bound basin[b]		0.4	1.0	0.8	1.0
ligand-only basin[b]		56.7	57.0	56.1	57.2

a. In the KL divergence calculations, the reference is taken to be the distribution from the ligand-bound simulation. DNA-bound state simulations are all unbiased simulations (see Table S1). For the ligand-bound states, reweighted distributions from metadynamics trajectories are used. The distributions are shown in Fig. S4-S5.

b. Reweighted distributions from metadynamics simulations are split into "doubly-bound basin" and "ligand-only basin" components at the DBD distance of 42.5 Å as the boundary.

are double-well in nature (Fig. 2a-c). The ligand-only basin is consistently observed at the large DBD distance of ∼46 Å, similar to the WT. In addition, a second free energy basin is observed at a shorter DBD distance expected for favorable DNA binding in the ligand-bound state. The location of this inner basin is largely consistent (∼38 Å) among all the mutants studied here.

To further verify the nature of the second basin in the PMFs along the DBD distance in the mutants, we measure the KL divergence of the DBD distance distributions between the ligand-bound and DNA-bound state simulations (see Figs. S4-S5). As shown in Table 1, the WT simulations lead to larger KL divergence values of the DBD distance distributions between the ligand-bound and DNA-bound states. A smaller KL divergence value indicates a larger similarity between the two distributions under comparison, suggesting that the DBD distance distributions in the ligand-bound state of the mutants are closer to the corresponding distributions in the DNA-bound state such that the mutants are more favorable for DNA binding than the WT in the ligand-bound state. It is also worthwhile decoupling the contributions of the two basins from metadynamics simulations to better illustrate the origin of more favorable DNA binding in the ligand-bound state of the mutants. Accordingly, we fur-

ther split the reweighted distributions from metadynamics simulations into two components at the DBD distance of 42.5 Å as the boundary. The DBD distance distributions of the inner basin in the ligand-bound state of the hotspot mutants overlap well with the corresponding DNA-bound state with KL divergence values no larger than 1.0, showing that the DBDs in the inner basin indeed favor DNA-binding. The large KL divergence values in the basin at the longer DBD distance suggest that the DNA is unfavorable to bind, verifying the nature of this ligand-only basin. Another structural variable thought to be important to DNA binding (affinity) is the relative angle of the $\alpha 4$ helices at the domain interface connecting the DBDs and LBDs[50,52] (see definitions of key DBD order parameters in SI). KL divergence of the $\alpha 4$ angle distributions in the two bound states is also substantially lower in the hotspot mutants than in the WT (Table S3 and Fig. S6-S7), also showing that the ligand-bound states of the mutants have higher populations in the DNA-binding-competent conformation than the ligand-bound state of the WT. Collectively, these results quantitatively demonstrate that the free energy basin at the short DBD distance sampled in the metadynamics is indeed favorable for DNA binding in the ligand-bound state. Therefore, although only the ligand is explicitly included in the metadynamics simulations, we refer to this basin as the "doubly-bound basin" since the corresponding conformational ensemble is expected to bind tightly to both the ligand and DNA.

Among the hotspot mutants studied, the doubly-bound basin is higher in free energy than the ligand-only basin at the longer DBD distance by 2-7 $k_B T$. The only exception is G102D, in which the two basins are similar in stability (Fig. 2a), suggesting that G102D is highly poised to bind both the ligand and the DNA. While the R49A mutation is also close to the domain interface and involves a similar magnitude change in net charge, the mutation effect on the DBD energetics is notably smaller. Interestingly, G138D and G138H exhibit rather different stabilities of the doubly-bound basins (Fig. 2b) despite implicating mutation of the same residue. These variations underscore the importance of the unique local interactions associated with specific mutations. For example, in G102D, the introduced aspartate D102 is

engaged in new interactions (compared to the WT) with several cationic residues nearby, such as K155' and R158' in α8', and R49 in α4 (Fig. S8). By contrast, in R49A, the interaction between R49 and E157' in α8' in the WT[62] is lost. In G138D, the introduced aspartate D138 is engaged in hydrogen-bonding interactions with the hydroxyl group in the ligand and therefore has rather static orientations. While in G138H, the lack of such hydrogen-bonding interactions leaves the introduced histidine H138 flexible, with occasional flip of the sidechain to push on α5 (Fig. S8). Finally, it is worth noting that while previous WT simulations did not identify any distinct properties of the hotspot residues in the middle of α8 and C-terminal region (α10),[62] the current metadynamics simulations of the α8 mutant (V145A) and C-terminal mutants (I193A and G196D) have clearly captured the effect of the mutations on the DBDs (Fig. 2c). In these three mutants, the doubly-bound basin is distinctly visible and is only higher in free energy by $\sim$2-4 k_BT than the ligand-only basin.

Inter-Domain Coupling Is Reduced in Hotspot Mutants

The normal induction function of the WT TetR relies on the strong anti-cooperativity between the LBDs and DBDs. The double-well feature of the computed DBD free energy profiles of the hotspot mutants in Fig. 2 suggests that the magnitude of inter-domain coupling is substantially reduced compared to the WT. As a result, these mutants are more inclined to bind tightly to DNA even in the ligand-bound state. Importantly, the reduction of inter-domain coupling is also evident from several additional lines of analyses, further supporting the robustness of the conclusion.

First, we note that, in the ligand-bound state simulations, the ligand binding pocket in the hotspot mutants exhibits notable changes depending on the DBD distance. At a longer DBD distance in the ligand-only basin, the l6-α8 distance (the Cα distance between R104 and L134) that characterizes the compactness of the binding pocket exhibits overall similar distributions as the WT (Fig. 3a-c), with merely shifted peak positions or modified distribution widths. In the doubly-bound basin, which features a compressed DBD distance, much

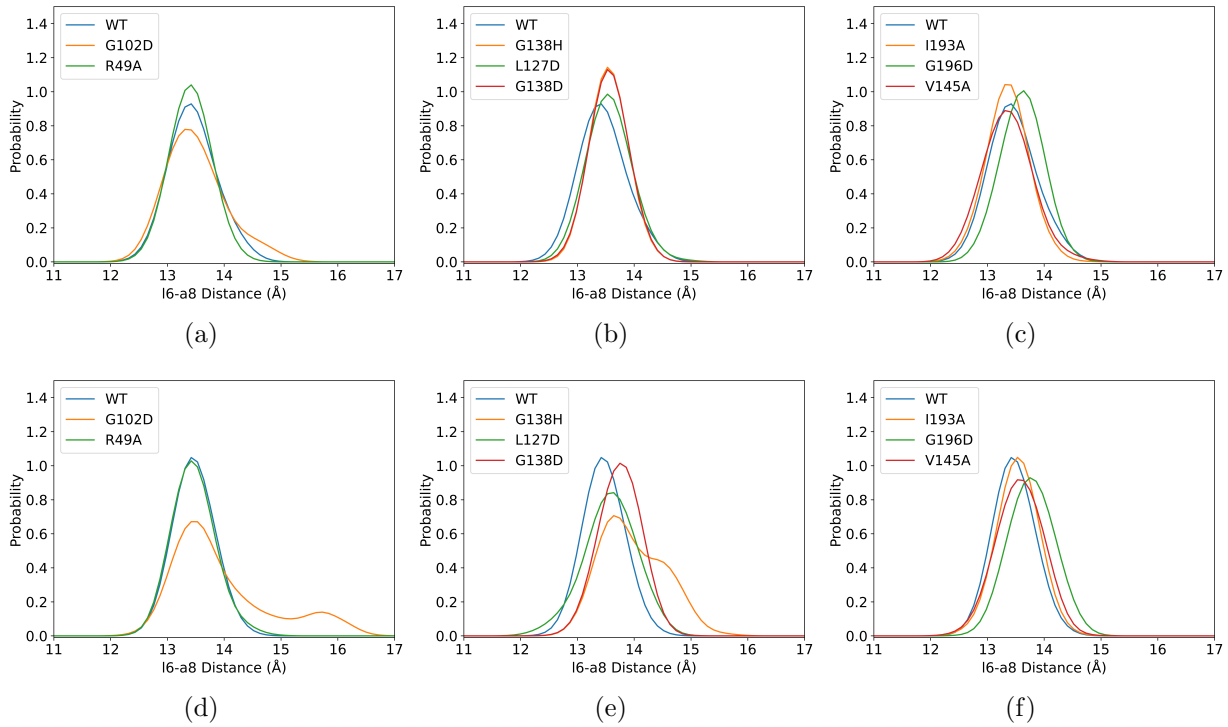


Figure 3: **Binding pocket distance distributions in different conformational states of the ligand-bound states from metadynamics simulations.** Subfigures are grouped by the regions shown in Fig. 1b. The reweighted distributions of the Cα distance between R104 and L134 (l6-α8 distance. l6 stands for the linker region of α6-α7.) (a-c) in the "ligand-only basin" (DBD distance range is between 46.5 Å and 47 Å), and (d-f) in the "doubly-bound basin" (DBD distance range is between 37.5 Å and 38 Å) from metadynamics simulations.

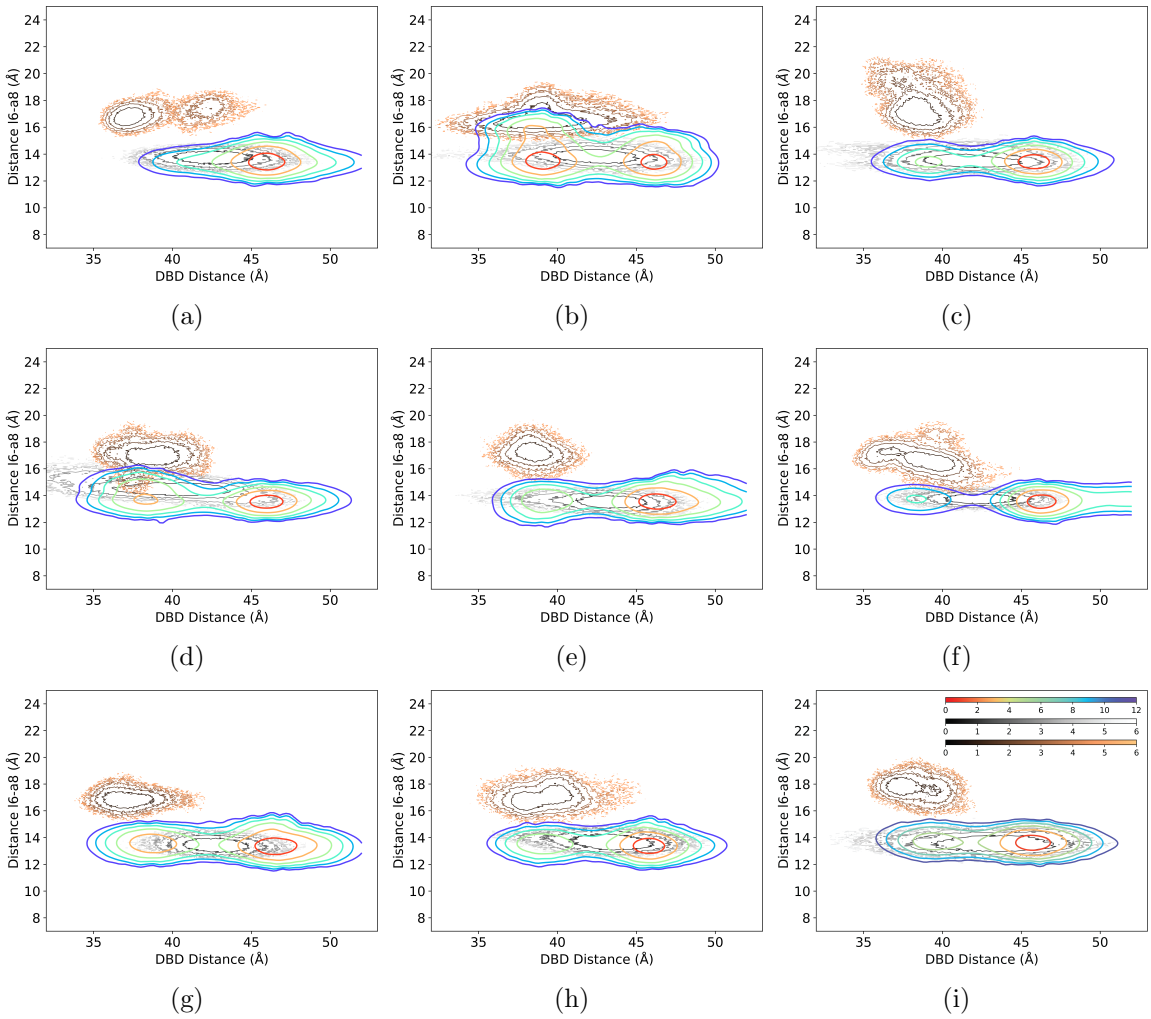


Figure 4: **Two-dimensional free energy landscapes of the DNA-bound and ligand-bound states**: (a)WT, (b) G102D, (c) R49A, (d) G138H, (e) L127D, (f) G138D, (g) V145A, (h) I193A, (i) G196D. The horizontal axis is the center-of-mass distance between the DBDs, and the vertical axis is the Cα distance between R104 and L134 (l6-α8 distance), which characterizes the compactness of the ligand binding pocket. For the DNA-bound states, the results are based on unbiased MD simulations (see Table S1 for simulation length summary). For the ligand-bound states, results based on unbiased MD simulations and reweighted distribution from metadynamics simulations are shown for comparison. While the unbiased and metadynamics simulations sample mostly similar conformational ensembles, due to the free energy barrier along the DBD distance (Fig. 2), the metadynamics simulations better capture the underlying free energy landscape and structural properties, especially for the "doubly-bound basin", which is hardly sampled in the unbiased simulations at the microsecond time scale. Color scheme: rainbow - ligand-bound state metadynamics simulations; gray - unbiased ligand-bound state simulations; copper - DNA-bound simulations. Color bars are in the unit of k_BT.

larger changes in the ligand binding pocket distributions relative to the WT are observed for several mutants (Fig. 3d-f). For example, both G102D and G138H exhibit much broader distributions. The peak position is also notably shifted in L127D, the two G138 mutants G138H/D, as well as V145A and G196D (also see Fig. S9 and Table S4). These observations suggest that the coupling between the binding pocket and the DBDs is not altered much in the ligand-only basin in the hotspot mutants, but the coupling is more significantly modified in the doubly-bound basin of the mutants. In particular, the substantially expanded pair-wise distance distributions in G102D and G138H in the doubly-bound basin suggest that the coupling between LBDs and DBDs is much reduced relative to the WT. The reduced inter-domain couplings at compressed DBD distance are consistent with the higher stability of the doubly-bound basin in these hotspot mutants in comparison to the WT (Fig. 2).

The variation of inter-domain coupling among the hotspot mutants relative to the WT is also directly visualized in the free energy landscapes of the ligand-bound states projected onto the two-dimensional space spanned by the DBD distance and the l6-α8 distance (Fig. 4). In the WT, free energy landscapes of the two bound states are distinct and exhibit limited overlap along either collective variable in the unbiased simulations as well as in the reweighted landscapes from metadynamics simulations, reflecting the strong anti-cooperativity between the LBDs and DBDs. In the free energy landscapes of the mutants, especially in the ligand-bound simulations, the DBD distance distributions are much broader than the WT. Two basins are clearly observed along the DBD distance coordinate in the reweighted landscapes from metadynamics simulations, while they share very similar equilibrium l6-α8 distances. These observations reflect weaker inter-domain anti-cooperativity in the mutants. Furthermore, along the l6-α8 distance coordinate for G102D and G138H (Fig. 4b, d), the binding pocket distributions are much broader with substantial degrees of overlapping at the compressed DBD distance, suggesting that the ligand binding pocket becomes even more flexible when the DBDs favor DNA binding and thus showing even weaker inter-domain coupling in the doubly-bound basin in these mutants.

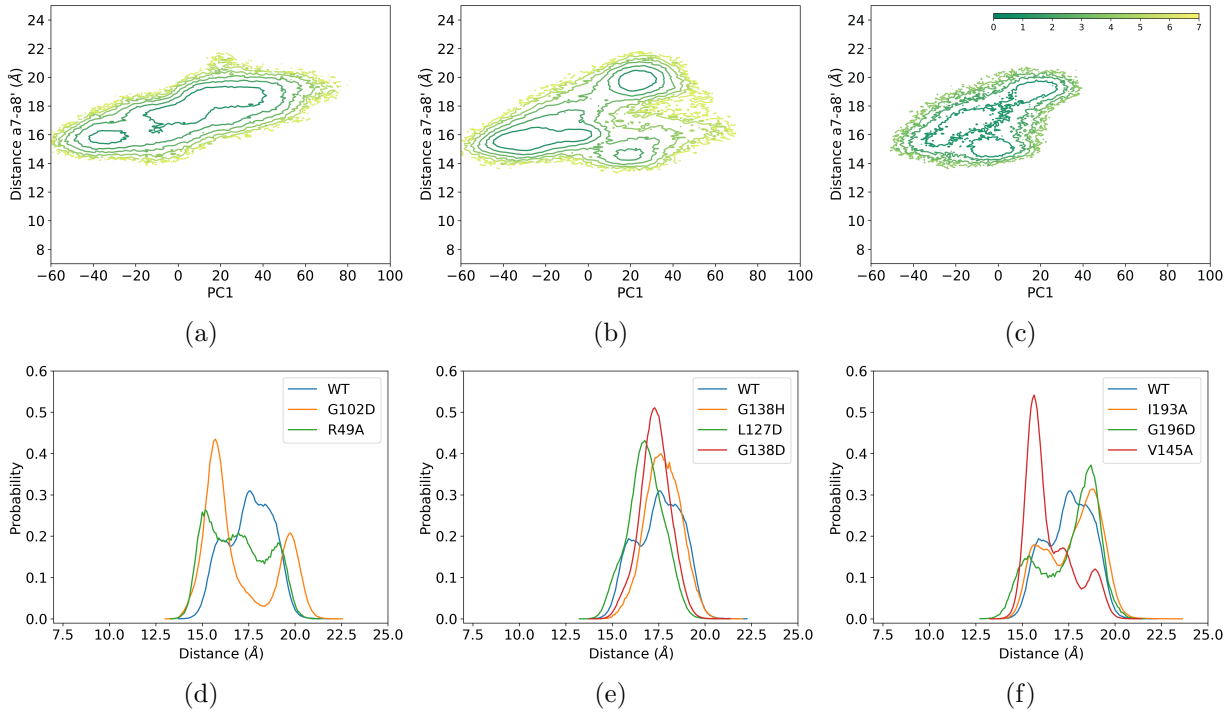


Figure 5: **Free energy landscapes and ligand binding pocket distance distributions of the apo states.** Two-dimensional free energy landscapes of the (a) WT, (b) G102D and (c) R49A variants. Results of the remaining mutants are qualitatively similar to the WT and shown in Fig. S10. The color bar is in the unit of k_BT. The horizontal axis is the first principal component (PC1), and the vertical axis is the Cα distance between Q109 and E147' (α7-α8' distance). PC1 describes the complex motions of the DBDs, while the inter-helical distance characterizes the compactness of the ligand binding pocket. The WT and mutants generally feature diagonal shape landscapes, while an extra basin is also observed in the mutants (G102D and R49A). Two-dimensional free energy landscapes projected onto different order parameters are shown in Fig. S11, which also exhibit similar changes for the two domain-interface mutants (G102D and R49A). (d-f)Distributions of the α7-α8' distance in different mutants. The results are based on extensive unbiased samplings (9-60 μs) for each variant. Subfigures are grouped by the regions shown in Fig.1b.

Finally, we examine the free energy landscapes of the apo states projected onto the two-dimensional space spanned by the collective DBD motion (represented by the first principle component PC1 used in our previous work[62]), and $\alpha 7 - \alpha 8'$ distance (the Cα distance between Q109 and E147') of the ligand binding pocket (Fig. 5a-c). Analysis of the apo landscape is complementary to the bound-state simulations discussed so far, as the shapes of the apo landscapes reflect the intrinsic features of inter-domain coupling in the absence of any local stabilization by the binding event of the ligand or the DNA. The two-dimensional free energy landscape of the WT shown in Fig. 5a is much broader than those computed of the bound states,[62] pointing to a conformational selection mechanism for TetR function. Moreover, the diagonal shape of the landscape is consistent with the strong anti-cooperativity expected for the LBDs and DBDs. For the studied hotspot mutants, the apo free energy landscapes are similarly broad, indicating that the conformational selection mechanism is still applicable. The diagonal shapes of the landscapes remain qualitatively similar to the WT for most mutants (Fig. S10), suggesting that the anti-cooperativity between the LBDs and DBDs is not significantly perturbed in the apo state of these mutants. Exceptions are the two mutations at the domain interface, G102D and R49A, of which the free energy landscapes deviate significantly from the diagonal shape with an additional basin that features a positive PC1 value and a short $\alpha 7 - \alpha 8'$ distance (compare Figs. 5b,c and 5a), which corresponds to a conformational state that has low binding propensity to both the ligand and DNA. Importantly, the most evident changes in the free energy landscapes associated with decreased inter-domain coupling in G102D and R49A are also observed when the free energy landscapes are projected onto other two-dimensional spaces spanned by the DBD distance and the $\alpha 7 - \alpha 8'$ distance (Fig. S11). As explained in the SI using a revised thermodynamic model (Fig. S12) for allostery in TetR,[62] such changes in the free energy landscape are strong signatures for the decrease of inter-domain coupling. The apo landscape in the WT features two basins that correspond to the induced (ligand-bound, $L_B \cdot D_E$) and non-induced (DNA-bound, $L_E \cdot D_B$) conformations. The inter-domain coupling free energy, γ explicitly

controls the anti-cooperativity which disfavors the two domains simultaneously adopting either the high-binding-affinity ($L_B \cdot D_B$) or the low-binding-affinity ($L_E \cdot D_E$) conformations in the WT. With reduced γ by mutations (Fig. S12c), a third basin representing the $L_E \cdot D_E$ conformation with a lower free energy is observed in the apo landscapes of G102D and R49A. Note that the $L_B \cdot D_B$ conformation still remains higher in free energy in the apo state, although the doubly-bound state is further stabilized by the binding of ligand and DNA and therefore expected to become the dominant population for the corresponding mutants in the presence of ligand and DNA (Fig. S12c). Indeed, the doubly-bound conformation is clearly visible as a free energy basin in the computed landscapes for the ligand-bound state of hotspot mutants (Fig. 4b-i).

Binding Pocket Is Perturbed to Decrease Ligand Binding in Hotspot Mutants

The free energy results (Fig. 2) and analysis of inter-domain coupling suggest that the studied hotspot mutants have much stronger propensities to bind both the ligand and DNA, which is consistent with their non-inducible phenotype observed in the DMS experiment.[34] However, a higher population of the doubly-bound state is not the only potential cause for the loss of induction function. Reducing TetR's apparent ligand binding affinity to be lower than the binding affinity for the DNA promoter sequence, or enhancing the DNA binding affinity would also lead to the non-inducible phenotype (see Fig. S12). In this context, we note that the apparent binding affinity in general contains two contributions, the free energy cost of transitioning from the inactive (binding-incompetent) to the active (binding-competent) conformation, and the intrinsic binding affinity of the active conformation. Moreover, the apparent ligand or DNA binding affinity can be modified by hotspot mutations through perturbations in either the ligand/DNA-bound state or the apo state, or both. Perturbation of the bound conformational ensemble modifies the intrinsic binding affinity. For the apo state, changes in the active conformation also modify the intrinsic binding affinity, while

population shifts in the equilibrium between the inactive and active conformations lead to variations in the activation free energy component of the apparent binding affinity. Therefore, we evaluate the impact of the hotspot mutations on the properties of both bound states and the apo state.

For the DNA-bound states, the ligand binding pocket remains loosely packed in both WT and the hotspot mutants (Fig. S13-15). For the ligand-bound states, at long DBD distances (in the ligand-only basin), the ligand binding pocket remains tightly packed and also experiences limited perturbation by mutations (Fig. 3a-c). With compressed DBD distances (in the doubly-bound basin), however, considerably broader binding pocket distributions are observed in several mutants (Fig. 3d-f), especially in G102D and G138H. Therefore, upon DNA binding, the intrinsic binding affinity to the ligand is expected to be further weakened in these mutants as compared to the WT, which contributes to the diminished induction activities.

In the apo states, many hotspot mutations impact the conformational ensembles of the ligand binding pocket. As shown in Fig. 5d, the WT ligand binding pocket favors the binding-competent conformation (long α7-α8' distances) in the apo state. In the two mutants at the domain interface (G102D and R49A), by contrast, the ligand binding pocket is dominated by the binding-incompetent conformation (short α7-α8' distances) that is favored in the DNA-bound state (see Fig. S13); a similar trend is also observed for V145A (Fig. 5f), in which the mutation is close to both the domain interface and the monomer-monomer interface (see Fig. 1b). Therefore, ligand binding in these mutants requires a notable free energy cost for the binding pocket to adopt the binding-competent conformation, leading to a lower apparent ligand binding affinity relative to the WT. In several cases, such as G102D and C-terminal mutants (I193A and G196D), the peak positions of the distributions favoring ligand-binding are substantially shifted (Fig. 5f), suggesting notable change in the binding-competent conformation and therefore variation in the intrinsic binding affinity. In other cases, such as L127D (Fig. 5e), the α7-α8' distance distribution becomes unimodal

and the peak position shifts towards a short distance that is favored in the DNA-bound state (also see Fig. S13), also suggesting considerable conformational rearrangements in the binding pocket and reduction in the ligand binding affinity.

Discussion

Evaluation of Functional Outcome of Hotspot Mutations in TetR

In our recent study,[62] motivated by the distinct contributions from allostery hotspots, we emphasized the importance of analyzing multiple properties at both the local and global scales, which was demonstrated to be essential to identifying multiple hotspots. However, since only the WT protein was investigated and contributions of residues to allostery can only be inferred indirectly, those analyses generated considerably numbers of "false positives", while missing residues in certain regions, such as the C-terminal helix and the middle of $\alpha 8$. By conducting extensive unbiased and enhanced simulations of multiple functional states of the hotspot mutants, we establish a computational framework that is able to evaluate the impact of mutations on the allostery function of TetR.

The free energy profiles for the DBD separation in the ligand-bound state show that the hotspot mutations stabilize the doubly-bound state relative to the WT and therefore compromise the induction function of TetR. Encouragingly, all of the mutant simulations have clearly captured the effects of the hotspot mutations, including those not identified in previous WT simulations.[62] The successful recapitulation of the hotspot mutation effects underscores the value of conducting explicit mutant simulations for the evaluation of hotspot candidates, since mutant simulations are able to capture the cascade of changes in structure and energetics not evident in the WT. Additionally, compensation effects of nearby residues are included, and therefore "false positives" can be minimized.

The current work demonstrates that functionally relevant PMF simulations are effective at capturing mutation effects in various hotspot mutants. This is due in part to the fact

that conformational sampling remains a general challenge, especially for the bound states, which tend to be structurally more rigid and therefore feature higher free energy barriers for conformational rearrangements. Thus, free energy simulations along the relevant coordinate(s) play critical roles for the evaluation of allostery modulation. This further highlights the importance of mechanistic analysis of allostery, which is required for the identification of proper coordinate(s) for functional free energy landscape computations. To aide the latter task, it is essential to continue the development of creative computational methodologies, such as those based on machine learning analysis of conformational ensembles associated with distinct functional states.[71–74] As complements to enhanced sampling, extensive unbiased simulations are important for providing a valid reference for the analysis of the nature of free energy basins emerged in enhanced sampling simulations (e.g., see Table 1).

Rescuability of Allostery Hotspots

In the DMS study of TetR,[34] another remarkable observation was that the loss of induction function in hotspot mutants could be rescued by additional mutation(s), also in a distributed and degenerate fashion. At the qualitative level, since the induction function can be disrupted in multiple ways by perturbing either inter-domain coupling or intra-domain energetics,[62] we expect that restoration of the normal function can also be accomplished in degenerate fashions through reversing changes in these properties. At the quantitative level, the degree of rescuability was defined as the ability of restoring the TetR function on non-inducible (dead) mutants. The DMS experiments found that among the five hotspots analyzed experimentally (R49A, D53V, G102D, N129D and G196D), the degree of rescuability follows the order of N129D>G196D>D53V>R49A>G102D.[34] As discussed in our earlier work,[62] successful design of allostery inhibition[2,4] needs to target allostery hotspots that are particularly difficult to rescue, since otherwise allostery can be readily restored with additional mutations, leading to abolishment of inhibition. Therefore, being able to evaluate the rescuability of allostery hotspots using computation is of great value.

In this context, a closely related quantity is the magnitude of perturbation on the free energy landscape. Specifically, it is interesting to compare the impact of hotspot mutations on the DBD distance free energy profile in the ligand-bound state with their degree of rescuability. As shown by the PMF calculations (Fig. 2), the impact of mutation on DBD energetics is the largest for the G102D mutant, in which the doubly-bound basin is similar in stability to the ligand-only basin; G102D was also observed to be most difficult to rescue in the DMS analysis. Although R49A is also a mutation at the domain interface and features an identical change in the net charge, the magnitude of the effect on DBD energetics is notably smaller, which is consistent with the higher degree of rescuability of R49A observed in the DMS experiment.[34] Moreover, the relative stability of the doubly-bound basin and the ligand-only basin is comparable in G196D and R49A. This is qualitatively consistent with the similar degrees of rescuability of these two mutants observed in the DMS experiment. Therefore, these observations support that mutations that lead to a larger change in the DBD free energy landscapes are more difficult to rescue.

The general consistency between computational and experimental results regarding mutant rescuability suggests that it is feasible and meaningful to use functional free energy simulations to rank hotspot mutations for inferring rescuability, which is a property of interest in the context of realistic applications, such as the design of allosteric inhibitors.[2,4] Indeed, the considerable level of redundancy for the allostery network identified in DMS studies raised the questions of whether engineering allosteric modulators can be a successful path. The current work helps compare the rescuability of hotspot mutations and supports that engineering designs to modulate allostery could be tested through MD simulations. Our computational strategies take advantage of multiple, modest-length enhanced sampling simulations and therefore are scalable to other allostery systems.

The Mechanism of Allostery Modulation by Hotspot Mutations

Our free energy results and structural analyses explicitly illustrate that the loss of induction function in the hotspot mutants is due to either reduced inter-domain coupling, which leads to stabilization of the doubly-bound state, or perturbation of the ligand binding pocket conformation, which reduces apparent ligand binding affinity. Therefore, the current observations generally support the thermodynamic model proposed in our previous work.[62] It classifies allostery hotspots into two distinct types that dictate inter-domain coupling and intra-domain energetics, respectively. The current analysis of the mutants at the atomic level highlights that a single mutation may lead to notable perturbation in both types of properties; for example, G102D clearly impacts both inter-domain coupling and conformational distributions of the ligand-binding pocket. Therefore, the type 1 & 2 hotspots discussed in the context of the thermodynamic model[62] are *idealized* classifications and the realistic situations tend to mix both effects. The PMF results for the DBD separation in the ligand-bound states (Fig. 2) clearly demonstrate that hotspot mutation effects can propagate from many regions in the protein to the DBDs, including the C-terminus, which is at the opposite end of the protein structure to the DBDs. This observation highlights the global nature of energetic coupling in proteins,[70] and is consistent with the distributed nature of allosteric network as observed in the DMS study[34] and our previous MD analysis.[62] These general features are broadly applicable to many allosteric systems, especially those sharing the two-domain structural topology.

The induction function of TetR relies critically on the anti-cooperativity between the ligand and DNA binding domains. In the hotspot mutants, we observe various signatures based on energetic and structural properties that consistently point to the reduction of inter-domain coupling in comparison to the WT protein. Our analysis of multiple functional states or free energy basins further reveals that the inter-domain coupling is not simply a constant value but dependent on the particular conformational and binding state of the protein. For example, in G102D and G138H, the coupling is observed to further decrease

upon compression of the DBDs in the ligand-bound state. Moreover, significant decrease of inter-domain coupling in the apo state is most evident only in the two domain-interface mutants (G102D and R49A), while all studied mutants exhibit reduced anti-cooperativity in the ligand-bound states as reflected by the stabilization of the doubly-bound basin.

Hotspot mutations can also perturb ligand binding properties by altering the conformations of the binding pocket. The apparent binding affinity can be modulated by perturbing either the intrinsic binding affinity associated with the binding-competent conformations, or the activation free energy component, which depends solely on the conformational distribution of the binding site in the apo state. The observations suggest that ligand binding properties such as affinity and specificity can be tuned by residues beyond the ligand-contacting residues. In other words, mutation of hotspot residues can perturb beyond the coupling between distant sites, a notion that has implications in applications. For example, this realization offers insights into how distal mutations might affect inhibitor potency in the context of drug resistance development.[6] Moreover, it also highlights additional flexibility in the strategies that can be used to fine-tune the various properties relevant to the allostery function, including effector specificity.[41]

The essence of our dissection of allostery modulation in TetR concerns the analysis of inter-domain coupling and intra-domain properties that affect the apparent binding affinity of effector (inducer) on an equal footing. The magnitude of inter-domain coupling explicitly dictates the degree of anti-cooperativity between ligand and DNA binding, while the apparent effector binding affinity also influences the competition between ligand-bound and DNA-bound populations and therefore the degree of inducibility of TetR. Since the proper function of most allostery systems relies on co-operativity between distinct domains and tight binding of the effector, the strategies and insights we have established through the analysis of TetR have major implications to the general prediction and evaluation of allostery hotspots in proteins.

Conclusions

Evaluation of hotspot mutations is of major interest to applications that aim to modulate allostery because hotspot residues dictate the degree of allosteric coupling and therefore function of allostery systems. The overarching objective of this study is to establish a computational framework to evaluate the functional outcome of hotspot mutations in TetR, both qualitatively (whether a mutant is non-inducible) and semi-quantitatively (rank order the impact of the mutation and infer the degree of rescuability). These goals have been rarely demonstrated in previous computational studies of allosteric systems. The broad distribution of the hotspot residues identified by DMS experiments in TetR also suggested that their contributions are unlikely explained by a single mechanism and we need multifaceted computational strategies. Therefore, we combined extensive (up to 60 μs) atomistic MD simulations, systematic analysis of structural and dynamical properties and multiple modest-length ($\sim \mu s$) enhanced sampling simulations to study WT and several representative mutants of TetR.

We identify the DBD distance as a collective variable that can be used to distinguish inducible and non-inducible TetR conformations, which is supported by previous structural and simulation studies. Accordingly, free energy profiles along this collective variable are highly correlated with the DNA binding affinity and the induction function of TetR. Encouragingly, all mutant simulations convincingly captured the mutation effects, including those missed in previous WT simulations. Consistently, a second basin at a short DBD distance is only observed in the hotspot mutants. KL divergence analysis clearly indicated that this second basin corresponds to a conformation that favors DNA binding in the ligand-bound state; i.e., the conformation is capable of binding to both ligand and DNA (the doubly-bound state). Stabilization of the doubly-bound state in these hotspot mutants explains their loss of induction function. Moreover, the free energy simulations revealed different degrees of stabilization of the doubly-bound state and underscored the importance of the unique local interactions associated with specific mutations. In terms of the scope of the hotspot mutants,

we have explored numerous mutations that are located in distinct structural regions of the protein and shown consistent trends in the shift of the free energy profile, further supporting the relevance of such computations to the evaluation of functional outcome. These PMF results explicitly show that hotspot mutation effects can propagate from many regions in the protein to the DBDs. In addition to metadynamics simulations, unbiased MD simulations establish important references for the evaluation of the nature of free energy basins based on KL divergence analysis in the bound states. Extensive unbiased apo state MD simulations reflect the intrinsic energetic coupling between the ligand and DNA binding domains to demonstrate the conformational selection mechanism of allostery in both the WT and mutants.

In the DMS experiments, another remarkable observation was that the loss of induction function in the mutants could be rescued by additional mutation(s). The degree of rescuability was defined as the ability of restoring the normal induction function in non-inducible (dead) mutants. A closely related quantity emerged from our simulations is the magnitude of perturbation on the computed free energy profiles relative to the WT, which correlates well with experimental rescuability data. Thus, it is feasible and meaningful to use functionally relevant free energy simulations to rank and evaluate hotspot mutations. Successful design of allostery inhibition needs to target allostery hotspots that are particularly difficult to rescue, since otherwise allostery can be readily restored with additional mutations, leading to abolishment of inhibition. From a technical point of view, free energy simulations with enhanced sampling at modest length are much more effective than unbiased molecular dynamics simulation for capturing the impact of mutation on free energy landscapes. Therefore, such a computational strategy can be readily applied to many allostery systems.

The extensive analyses of structural features of different functional states of the mutants and 2D free energy landscapes provided further insights into the molecular mechanisms through which the induction function is perturbed by the hotspot mutations. The analyses show that the hotspot mutations perturb inter-domain coupling and/or intra-domain con-

formational properties relative to the WT, leading to decrease in the inducibility. Reduced anti-cooperativity between distant binding sites is found in the mutants to reflect the loss of the induction function. Multiple mutations also alter the conformational ensemble of the ligand binding pocket for weakened binding, suggesting that hotspot mutations can be used to fine-tune allostery properties including effector specificity. A single mutation may lead to notable perturbations in both types of properties, such as in G102D. Therefore, the same non-inducible phenotype could be the result of perturbations on distinct structural and energetic properties. The computational framework and mechanistic insights are expected to be applicable to many allostery systems, especially those featuring the two-domain structural topology.

Acknowledgement

This work is supported by NIH Grant R35-GM141930. Computational resources from ACCESS (project BIO230101) are greatly appreciated. Part of the computational work was performed on the Shared Computing Cluster which is administered by Boston University's Research Computing Services (URL: www.bu.edu/tech/support/research/). Anton 2 computer time (award MCB200062P) was provided by the Pittsburgh Supercomputing Center (PSC) through NIH Grant R01GM116961. The Anton 2 machine at PSC was generously made available by D.E. Shaw Research.

Supporting Information Available

Additional simulation details, results and schematics for a revised thermodynamic model are included.

References

(1) S. J. Wodak et al. Allostery in Its Many Disguises: From Theory to Applications. *Structure* **2019**, *27*, 566–578.

(2) Chatzigoulas, A.; Cournia, Z. Rational design of allosteric modulators: Challenges and successes. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1529.

(3) Nussinov, R.; Tsai, C. J. Allostery in Disease and in Drug Discovery. *Cell* **2013**, *153*, 293–305.

(4) Ni, D.; Chai, Z.; Wang, Y.; Li, M.; Yu, Z.; Liu, Y.; Lu, S.; Zhang, J. Along the allostery stream: Recent advances in computational methods for allosteric drug discovery. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1585.

(5) Nussinov, R.; Zhang, M. Z.; Maloney, R.; Liu, Y. L.; Tsai, C. J.; Jang, H. Allostery: Allosteric Cancer Drivers and Innovative Allosteric Drugs. *J. Mol. Biol.* **2022**, *434*, 167569.

(6) Lu, S.; Qiu, Y.; Ni, D.; He, X.; Pu, J.; Zhang, J. Emergence of allosteric drug- resistance mutations: new challenges for allosteric drug discovery. *Drug Disc. Today* **2020**, *25*, 177–184.

(7) Osuna, S. The challenge of predicting distal active site mutations in computational enzyme design. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, e1502.

(8) Taylor, N. D.; Garruss, A. S.; Moretti, R.; Chan, S.; Arbing, M. A.; Cascio, D.; Rogers, J. K.; Isaacs, F. J.; Kosuri, S.; Baker, D.; Fields, S.; Church, G. M.; Raman, S. Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **2016**, *13*, 177–183.

(9) Sloas, D. C.; Tran, J. C.; Marzilli, A. M.; Ngo, J. T. Tension-tuned receptors for

synthetic mechanotransduction and intercellular force detection. *Nat. Biotech.* **2023**, DOI10.1038/s41587–022–01638–y.

(10) Dagliyan, O.; Dokholyan, N. V.; Hahn, K. M. Engineering proteins for allosteric control by light or ligands. *Nat. Protoc.* **2019**, *14*, 1863–1883.

(11) Shaaya, M.; Fauser, J.; Zhurikhina, A.; Conage-Pough, J. E.; Huyot, V.; Brennan, M.; Flower, C. T.; Matsche, J.; Khan, S.; Natarajan, V.; Rehman, J.; Kota, P.; White, F. M.; Tsygankov, D.; Karginov, A. V. Light-regulated allosteric switch enables temporal and subcellular control of enzyme activity. *eLife* **2020**, *9*, e60647.

(12) Chen, J.; Vishweshwaraiah, Y. L.; Dokholyan, N. V. Design and engineering of allosteric communications in proteins. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102334.

(13) Vishweshwaraiah, Y. L.; Chen, J.; Chirasani, V. R.; Tabdanov, E. D.; Dokholyan, N. V. Two-input protein logic gate for computation in living cells. *Nat. Commun.* **2021**, *12*, 6615.

(14) Suel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **2003**, *10*, 59–69.

(15) O'Rourke, K. F.; Gorman, S. D.; Boehr, D. D. Biophysical and computational methods to analyze amino acid interaction networks in proteins. *Comput. Struct. Biotech. J.* **2016**, *14*, 245–251.

(16) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **2011**, *147*, 1564–1575.

(17) VanWart, A.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. Exploring Residue Component Contributions to Dynamical Network Models of Allostery. *J. Chem. Theory Comput.* **2012**, *8*, 2949–2961.

(18) Feher, V. A.; Durrant, J. D.; A. T. Van Wart; Amaro, R. E. Computational approaches to mapping allosteric pathways. *Curr. Opin. Struct. Biol.* **2014**, *25*, 98–103.

(19) Rivalta, I.; Batista, V. In *Allostery: Methods and Protocols*; DiPaola, L., Giuliani, A., Eds.; Methods in Molecular Biology; Springer, 2021; Vol. 2253; Chapter Community Network Analysis of Allosteric Proteins, pp 137–151.

(20) Doshi, U.; Holliday, M. J.; Eisenmesser, E. Z.; Hamelberg, D. Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 4735–4740.

(21) Wang, J.; Jain, A.; McDonald, L. R.; Gambogi, C.; Lee, A. L.; Dokholyan, N. V. Mapping allosteric communications within individual proteins. *Nat. Commun.* **2020**, *11*, 3862.

(22) Zheng, W. J.; Brooks, B. R.; Thirumalai, D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc. Acad. Natl. Sci. U.S.A.* **2006**, *103*, 7664–7669.

(23) Guarnera, E.; Berezovsky, I. N. Toward Comprehensive Allosteric Control over Protein Activity. *Struct.* **2019**, *27*, 866–878.

(24) Lake, P. T.; Davidson, R. B.; Klem, H.; Hocky, G. M.; M. McCullagh Residue-Level Allostery Propagates through the Effective Coarse-Grained Hessian. *J. Chem. Theory Comput.* **2020**, *16*, 3385–3395.

(25) Ngan, K. C. H.; Hoenig, S. M.; Kwok, H. S.; Lue, N. Z.; Gosavi, P. M.; Tanner, D. A.; Garcia, E. M.; Lee, C.; Liau, B. B. Activity-based CRISPR scanning uncovers allostery in DNA methylation maintenance machinery. *eLife* **2023**, *12*, e80640.

(26) Lisi, G. P.; Loria, J. P. Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chem. Rev.* **2016**, *116*, 6323–6369.

(27) Holliday, M. J.; Camilloni, C.; Armstrong, G. S.; Vendruscolo, M.; Eisenmesser, E. Z. Networks of Dynamic Allostery Regulate Enzyme Function. *Struct.* **2017**, *25*, 276–286.

(28) Köhler, C.; Carlström, G.; Gunnarsson, A.; Weininger, U.; Tångefjord, S.; Ullah, V.; Lepistö, M.; Karlsson, U.; Papavoine, T.; Edman, K.; Akke, M. Dynamic allosteric communication pathway directing differential activation of the glucocorticoid receptor. *Sci. Adv.* **2020**, *6*, eabb5277.

(29) Zhuravleva, A.; Clerico, E. M.; Gierasch, L. M. An Interdomain Energetic Tug-of-War Creates the Allosterically Active State in Hsp70 Molecular Chaperones. *Cell* **2012**, *151*, 1296–1307.

(30) Fowler, D. M.; Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **2014**, *11*, 801–807.

(31) Raman, S. Systems Approaches to Understanding and Designing Allosteric Proteins. *Biochemistry* **2018**, *57*, 376–382.

(32) Araya, C. L.; Fowler, D. M. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* **2011**, *29*, 435–442.

(33) Faure, A. J.; Domingo, J.; Schmiedel, J. M.; Hidalgo-Carcedo, C.; Diss, G.; Lehner, B. Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **2022**, *604*, 175.

(34) Leander, M.; Yuan, Y.; Meger, A.; Cui, Q.; Raman, S. Functional Plasticity and Evolutionary Adaptation of Allosteric Regulation. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 25445–25454.

(35) Leander, M.; Liu, Z.; Cui, Q.; Raman, S. Deep mutational scanning and machine learning reveal structural and molecular rules governing allosteric hotspots in homologous proteins. *eLife* **2022**, *11*, e79932.

(36) Meier, G.; Thavarasah, S.; Ehrenbolger, K.; Hutter, C. A. J.; Hurlimann, L. M.; Barandun, J.; Seeger, M. A. Deep mutational scan of a drug efflux pump reveals its structure-function landscape. *Nat. Chem. Biol.* **2022**, *19*, 440.

(37) Garruss, A. S.; Collins, K. M.; Church, G. M. Deep representation learning improves prediction of LacI-mediated transcriptional repression. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2022838118.

(38) Hanning, K. R.; Minot, M.; Warrender, A. K.; Kelton, W.; Reddy, S. T. Deep mutational scanning for therapeutic antibody engineering. *Trends Pharmacol. Sci.* **2022**, *43*, 123–135.

(39) Azbukina, N.; Zharikova, A.; Ramensky, V. Intragenic compensation through the lens of deep mutational scanning. *Biophys. Rev.* **2022**, *14*, 1161–1182.

(40) Groseclose, T. M.; Rondon, R. E.; Hersey, A. N.; Milner, P. T.; Kim, D.; Zhang, F. M.; Realff, M. J.; Wilson, C. J. Biomolecular Systems Engineering: Unlocking the Potential of Engineered Allostery via the Lactose Repressor Topology. *Annu. Rev. Biophys.* **2021**, *50*, 303–321.

(41) Hersey, A. N.; Kay, V. E.; Lee, S.; Realff, M. J.; Wilson, C. J. Engineering allosteric transcription factors guided by the LacI topology. *Cell Sys.* **2023**, *14*, 645–655.

(42) Dokholyan, N. V. Controlling allosteric networks in proteins. *Chem. Rev.* **2016**, *116*, 6463–6487.

(43) Cruz, M. A.; Frederick, T. E.; Mallimadugula, U. L.; Singh, S.; Vithani, N.; Zimmerman, M. I.; Porter, J. R.; Moeder, K. E.; Amarasinghe, G. K.; Bowman, G. R. A cryptic pocket in Ebola VP35 allosterically controls RNA binding. *Nat. Commun.* **2022**, *13*, 2269.

(44) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116*, 6391–6423.

(45) Galdadas, I.; Qu, S.; Oliveira, A. S. F.; Olehnovics, E.; Mack, A. R.; Mojica, M. F.; Agarwal, P. K.; Tooke, C. L.; Gervasio, F. L.; Spencer, J.; Bonomo, R. A.; Mulholland, A. J.; Haider, S. Allosteric communication in class A β-lactamases occurs via cooperative coupling of loop dynamics. *eLife* **2021**, *10*, e66567.

(46) Greener, J. G.; Sternberg, M. J. E. Structure-based prediction of protein allostery. *Curr. Opin. Struct. Biol.* **2018**, *50*, 1–8.

(47) Lee, J.; Natarajan, M.; Nashine, V. C.; Socolich, M.; Vo, T.; Russ, M. P.; Benkovic, S. J.; Ranganathan, R. Surface sites for engineering allosteric control in proteins. *Science* **2008**, *322*, 438–442.

(48) Kityk, R.; Vogel, M.; Schlecht, R.; Bukau, B.; Mayer, M. P. Pathways of allosteric regulation in Hsp70 chaperones. *Nat. Commun.* **2015**, *6*, 8308.

(49) Stanton, B. C.; Siciliano, V.; Ghodasara, A.; Wroblewska, L.; Clancy, K.; Trefzer, A. C.; Chesnut, J. D.; Weiss, R.; Voigt, C. A. Systematic Transfer of Prokaryotic Sensors and Circuits to Mammalian Cells. *ACS Synth. Biol.* **2014**, *3*, 880–891.

(50) Hinrichs, W.; Kisker, C.; Düvel, M.; Müller, A.; Tovar, K.; Hillen, W.; Saenger, W. Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science* **1994**, *264*, 418–420.

(51) Kisker, C.; Hinrichs, W.; Tovar, K.; Hillen, W.; Saenger, W. The complex formed between Tet repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *J. Mol. Biol.* **1995**, *247*, 260–280.

(52) Orth, P.; Schnappinger, D.; Hillen, W.; Saenger, W.; Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor–operator system. *Nat. Struct. Biol.* **2000**, *7*, 215–219.

(53) Yu, Z.; Reichheld, S. E.; Savchenko, A.; Parkinson, J.; Davidson, A. R. A Comprehensive Analysis of Structural and Sequence Conservation in the TetR Family Transcriptional Regulators. *J. Mol. Biol.* **2010**, *400*, 847–864.

(54) Hecht, B.; Müller, G.; Hillen, W. Noninducible Tet Repressor Mutations Map from the Operator Binding Motif to the C Terminus. *J. Bacteriol.* **1993**, *175*, 1206–1210.

(55) Müller, G.; Hecht, B.; Helbl, V.; Hinrichs, W.; Saenger, W.; Hillen, W. Characterization of non-inducible Tet repressor mutants suggests conformational changes necessary for induction. *Nat. Struct. Biol.* **1995**, *2*, 693–703.

(56) Seedorff, J. E.; Rodgers, M. E.; Schleif, R. Opposite allosteric mechanisms in TetR and CAP. *Protein Sci.* **2009**, *18*, 775–781.

(57) Saenger, W.; Orth, P.; Kisker, C.; Hillen, W.; Hinrichs, W. The tetracycline repressor – a paradigm for a biological switch. *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 2042–2052.

(58) Sevvana, M.; Goetz, C.; Goeke, D.; Wimmer, C.; Berens, C.; Hillen, W.; Muller, Y. A. An Exclusive α/β Code Directs Allostery in TetR–Peptide Complexes. *J. Mol. Biol.* **2012**, *416*, 46–56.

(59) Werten, S.; Schneider, J.; Palm, G. J.; Hinrichs, W. Modular organisation of inducer recognition and allostery in the tetracycline repressor. *FEBS J.* **2016**, *283*, 2102–2114.

(60) Daily, M. D.; Gray, J. J. Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. *PLoS Comput. Biol.* **2009**, *5*, e1000293.

(61) Walker, A. S.; Russ, W. P.; Ranganathan, R.; Schepartz, A. RNA sectors and allosteric function within the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 19879–19887.

(62) Yuan, Y.; Deng, J.; Cui, Q. Molecular Dynamics Simulations Establish the Molecular Basis for the Broad Allostery Hotspots Distribution in the Tetracycline Repressor. *J. Am. Chem. Soc.* **2022**, *144*, 10870–10887.

(63) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67*, 159–184.

(64) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(65) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comp. Phys. Commun.* **2009**, *180*, 1961–1972.

(66) The PLUMED consortium Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **2019**, *16*, 670–673.

(67) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; C. D. Stern et al. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

(68) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem. B* **2015**, *119*, 736–742.

(69) Haberl, F.; Lanig, H.; Clark, T. Induction of the tetracycline repressor: Characterization by molecular-dynamics simulations. *Proteins: Struct. Funct. & Bioinf.* **2009**, *77*, 857–866.

(70) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and energetic basis of allostery. *Annu. Rev. Biophys.* **2012**, *41*, 585–609.

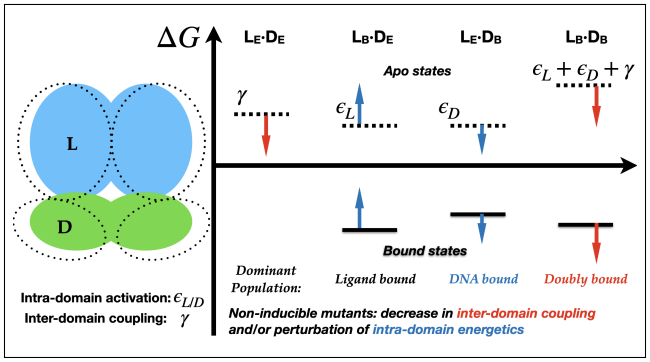
(71) Bonati, L.; Trizio, E.; Rizzi, A.; Parrinello, M. A unified framework for machine learning collective variables for enhanced sampling simulations: mlcolvar. *J. Chem. Phys.* **2023**, *159*, 014801.

(72) Mehdi, S.; Smith, Z.; Herron, L.; Zou, Z.; Tiwary, P. Enhanced Sampling with Machine Learning: A Review, arXiv (2023), https://arxiv.org/abs/2306.09111. **accessed 27 Aug 2023**,

(73) Ward, M. D.; Zimmerman, M. I.; Meller, A.; Chung, M.; Swamidass, S. J.; Bowman, G. R. Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets. *Nat. Commun.* **2021**, *12*, 3023.

(74) Mendels, D.; Bylehn, F.; Sirk, T.; de Pablo, J. J. Systematic modification of functionality in disordered elastic networks through free energy surface tailoring. *Sci. Adv.* **2023**, *9*, eadf7541.

TOC Graphic


ΔG
L_E·D_E
L_B·D_E
L_E·D_B
L_B·D_B
Apo states
γ
ϵ_L
ϵ_D
$\epsilon_L + \epsilon_D + \gamma$
L
D
Bound states
Dominant Population:
Ligand bound
DNA bound
Doubly bound
Intra-domain activation: $\epsilon_{L/D}$
Inter-domain coupling: γ
Non-inducible mutants: decrease in inter-domain coupling and/or perturbation of intra-domain energetics