Ligand-induced changes in dynamics mediate long-range allostery in the lac repressor

Anum Glasgow\textsuperscript{1,2}, Helen T. Hobbs\textsuperscript{3}, Zion R. Perry\textsuperscript{4}, Susan Marqusee\textsuperscript{3,5,*}, and Tanja Kortemme\textsuperscript{1,*}

\textsuperscript{1} Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94158
\textsuperscript{2} Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032
\textsuperscript{3} Department of Chemistry, University of California, Berkeley, Berkeley CA 94720
\textsuperscript{4} Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511
\textsuperscript{5} Department of Molecular & Cell Biology, University of California, Berkeley, Berkeley CA 94720

* To whom correspondence may be addressed. Emails: marqusee@berkeley.edu, kortemme@cgl.ucsf.edu

Abstract

Allostery, broadly defined as a protein’s functional response to distal perturbations, is fundamental to biological regulation. In classical models, allosteric ligand binding produces a defined set of structural changes in the protein, resulting in a different low-energy conformation. Proteins that undergo ligand-induced allostery with few observable structural changes therefore frustrate interpretations by classical models. Here we used hydrogen-deuterium exchange with mass spectrometry (HDX/MS) to map the allostERIC effects in a paradigm ligand-responsive allosteric transcription factor, the lac repressor (LacI). X-ray crystal structures of the core domain of LacI bound to different small molecule ligands, or the DNA operator, show less than 1.5 Å difference in the protein all-atom root-mean-square-deviation (RMSD) between any two structures. Despite this high degree of similarity among static structures, our HDX/MS experiments reveal widespread and unexpected differences in the flexibility of secondary structures in the LacI core domain in each functional state. We propose a model in which ligand binding allosterically switches the functional response of the repressor by selectively changing the dynamics of particular secondary structure elements relative to each other, shifting the conformational ensemble of the protein between mutually incompatible DNA-bound and inducer-bound states. Our model also provides a mechanistic context for the altered functions of thousands of documented LacI mutants. Furthermore, our approach provides a platform for characterizing and engineering allosteric responses in proteins.
Introduction

Biological regulation ubiquitously depends on protein allostericity, but few allosteric mechanisms are understood in structural detail. In the absence of high-resolution structural information, the Monod-Wyman-Changeux (MWC) (Monod et al., 1965) and Koshland-Nemethy-Filmer (KNF) (Koshland et al., 1966) models for allostery have been used to explain the functional behavior of diverse allosteric systems. In a few cases, detailed structural information has enabled descriptions of allosteric mechanisms for proteins, both in terms of specific, sequential conformational changes resulting from perturbation (Daily and Gray, 2009; Lockless and Ranganathan, 1999; Yu and Koshland, 2001) and as shifts in the average conformational ensemble (Cooper and Dryden, 1984; McLeish et al., 2015; Pan et al., 2000). However, deducing an atomic-level mechanism for allosteric effects on protein function is challenging in several common scenarios: when the beginning- and end-state average structures are structurally indistinguishable; when the structural change occurs over a large distance with few structural differences in the allosteric site; or when the protein is disordered in one or both states.

A paradigm example for these challenging scenarios is the *E. coli* lac repressor (LacI), a model allosteric transcription factor that regulates expression of the genes of the lac operon. Wild-type (WT) LacI forms a dimer of dimers, and each dimer can bind to one of three DNA operator sequences in the *E. coli* genome. The MWC and KNF models for protein allostery were established in part from studies of LacI after observing bacterial responses to sugar ligands (Jacob et al., 1965). The ligands bind in a pocket in the core domain of LacI, 40 Å away from the DNA-binding surface of the DNA-binding domain (Fig. 1A), and allosterically modulate the binding affinity of LacI for its DNA operator. Interestingly, X-ray crystal structures of the core domain are essentially identical in different ligand-bound states as well as the apo state (0.5-1 Å heavy atom RMSD) (Fig. 1B, C). The high structural similarity of the core domains, combined with a lack of X-ray crystal structures for a non-operator-bound DNA-binding domain, complicates mechanistic understanding of ligand-regulated allostery in LacI despite more than five decades of study using structural (Bell and Lewis, 2000; Daber et al., 2007; Friedman et al., 1995; Lewis et al., 1996), genetic (Coulondre and Miller, 1977a, 1977b; Farabaugh et al., 1978; Nelson et al., 1997; Suckow et al., 1996; Swerdlov and Schaaper, 2014; Swint-Kruse et al., 2003), computational (Daily and Gray, 2009; Flynn et al., 2003; Parente and Swint-Kruse, 2013; Villa et al., 2005), and biophysical (Adler et al., 1972; Kalodimos et al., 2002; Lamerichs et al., 1990; Morgan et al., 2005; O’Gorman et al., 1980; Royer et al., 1990; Wick and Matthews, 1991; Wilson et al., 2007, 2005) methods.

Adding to this challenge, chemically similar molecules binding in the core domain can have opposite effects on LacI-regulated gene expression (Fig. 1D, E). When bound to the inducers allolactose or isopropyl-β-D-1-thiogalactopyranoside (IPTG), LacI binds approximately 1000-fold less efficiently to the DNA operator than apo-LacI, allowing for the transcription of downstream genes. However, when bound to the anti-inducer ortho-nitrophenyl-β-D-fucoside (ONPF), LacI is stabilized in the operator-bound state by 2- to 5-fold over apo-LacI. Other ligands such as ortho-nitrophenyl-β-galactoside (ONPG) bind LacI but produce no observable functional effects. Inducer potency has been hypothesized to depend on the molecule’s binding affinity for LacI, the presence of an O6 hydroxyl group on the sugar ring, an optimal size for the C1 substituent, and the flexibility of the substituent linkage to the sugar ring (Barkley et al., 1975; Chakerian et al., 1987).
Several models have been proposed to explain the inducer-specific effects on gene expression by LacI. Crystallography and molecular dynamics simulations support a model in which inducer binding leads to new intramolecular interactions in the N-terminal subdomain of the core and partial unfolding of the DNA-binding domain, with the C-terminal subdomain of the core remaining unchanged (Daber et al., 2007; Richards et al., 2017; Rondon and Wilson, 2019). Crystallographic studies of LacI bound to the inducer IPTG or to DNA revealed similar structures of the core domain, and a small shift (all-atom RMSD 1.5 Å) of the N-terminal subdomains of the core closer together in the IPTG-bound structure as compared to DNA-bound LacI (Fig. S1) (Daber et al., 2007). However, because of the structural similarity of the core domain in the apo, DNA-bound, inducer-bound, and anti-inducer-bound states of the protein, a detailed molecular mechanism for allosteric structural changes in the core domain upon ligand binding remains elusive.

Here, we map allosteric changes in the core domain of LacI in different functional states via hydrogen-deuterium exchange with mass spectrometry (HDX/MS). We define “functional states” by binding partner, where IPTG-LacI denotes the IPTG-bound protein; ONPF-LacI the ONPF-bound protein; ONPF-DNA-LacI the protein bound to ONPF and the lac operator at the same time; and

Figure 1. Ligands bind without dramatic structural changes in the core domain to control the structure of the DNA-binding domain allosterically. (A) X-ray crystal structure of dimeric LacI bound to the DNA operator. One subunit is shown in color (green: C-terminal subdomain; yellow: N-terminal subdomain; purple: DNA-binding domain) and the other subunit is shown in gray. PDB ID: 1EFA. (B) Overlaid structures of the core domains show similar backbone structures for apo-LacI (dark gray), IPTG-LacI (green, yellow, maroon), and ONPF-LacI (blue). PDB IDs: 1LBI, 2P9H, 2PAF. (C) Overlaid structures of the ligand binding site in IPTG-LacI and ONPF-LacI reveal similar sidechain-atom positions, colored as in (B). Thin sticks show amino acid sidechains. Thick sticks show ligands. (D) A cell culture fluorescence assay shows ligand-modulated gene expression by LacI. TMG- and IPTG-binding to LacI in cells results in GFP expression and fluorescence, but ONPF-binding results in increased repression. (E) Chemical structures of IPTG, TMG, and ONPF.
apo-LacI the unbound protein. We propose a model for mutually exclusive low-energy conformational ensembles of LacI in inducer- and DNA-bound states. Using a combined computational and experimental approach, we show that protein, ligand, and solvent atoms together play a central role in modulating the rigidity of secondary-structure elements in the ligand-binding core domain of LacI to control its functional state, without dramatic conformational changes. Our model for allosteric regulation in LacI can inform future efforts to re-engineer the protein to control gene expression in response to new ligands, and establishes a method for characterizing how allostery in other proteins is conserved, evolved, or affected in disease.

**Results**

We set out to determine the mechanism of allosteric changes in dimeric LacI by comparing structural and functional differences between DNA-bound states and inducer-bound states using HDX/MS. The DNA-bound and inducer-bound states are the key functionally relevant states of LacI. To study inducer-bound states, we complexed LacI with IPTG, a strong inducer molecule (LacI-IPTG $K_D \sim 5 \times 10^{-6} \text{ M}$), and thiomethyl β-D-galactoside (TMG), a weaker inducer molecule (LacI-TMG $K_D \sim 7 \times 10^{-4} \text{ M}$) (Barkley et al., 1975). To study DNA-bound states, we used the LacI-lac operator complex, with and without the anti-inducer molecule ONPF (LacI-operator $K_D < 10^{-12} \text{ M}$, LacI-ONPF $K_D \sim 7 \times 10^{-3} \text{ M}$) (Barkley et al., 1975). This strategy generated a set of LacI states with a range of binding affinities for the operator DNA (in order from unmeasurably weak to strongest): IPTG-LacI; TMG-LacI; DNA-LacI; and the ternary ONPF-DNA-LacI. We confirmed the effects on gene expression of IPTG, TMG and ONPF ligands binding to LacI using a plate-based cell culture assay, in which LacI regulates the expression of a green fluorescent protein (GFP) that is encoded downstream of the LacI DNA operator sequence on a plasmid (Fig. 1D, E).

For each state, we monitored amide hydrogen-deuterium exchange at the peptide level over a four hour time period using mass spectrometry (see Methods). From the full set of mass spectra representing more than 300 total peptides, we curated a set of 57 peptides from the core domain of LacI for which we could collect replicate data in six different states as well as a fully deuterated control HDX/MS experiment (in order to correct for back-exchange; see Methods). The curated dataset had 97% peptide coverage for the core domain, excluding the extreme C-terminal 22 amino acids (Fig. S2) for which we could not collect a complete dataset (data for each time point and each state with back-exchange).

**Comparison of inducer-bound and DNA-bound states.**

We first compared exchange patterns for core domain peptides in the IPTG-LacI and DNA-LacI states (Figs. 2, S3). We quantified the level of exchange in both states by calculating the deuterium uptake as a function of time for each peptide (Fig. S4), and fit these data to a nonlinear regression model to estimate the number of slow-, medium-, fast-, and non-exchanging protons in each peptide/state, and the resulting rates associated with each group (slow, medium, and fast) (Methods). In the deuterium-uptake plots in Figure 2A, increased rates of exchange for a peptide indicate increased structural flexibility associated with decreased participation in secondary structure hydrogen bonds by the peptide’s backbone amide groups. Figure 2B shows this behavior for the curated set of 57 peptides in
the core domain, and Figure 2C illustrates these differences in H-D exchange between IPTG-LacI and DNA-LacI mapped onto the structure of the core domain.

When comparing IPTG-LacI to DNA-LacI, we observed both regions of decreased deuterium uptake (less exchange in IPTG-LacI relative to DNA-LacI, teal, Fig. 2A,C, i-iii, Fig. S5A) and regions of increased deuterium uptake (more exchange in IPTG-LacI, purple, Fig. 2A, C, iv-vii). Regions with decreased deuterium uptake are located near the ligand-binding pocket (Fig. 2A, C, i, N-terminal end of helix 2) and the monomer-monomer interface in the N-terminal subdomain (Fig. 2A, C, ii, beta strand 3). We hypothesize that the decreased deuterium uptake for beta strand 3 reports on the formation of a hydrogen bond between the backbone atoms of position 96 across the beta sheets of both N-terminal subdomains in the LacI dimer (Fig. S6). We also observed decreased deuterium uptake in regions distal to the monomer-monomer interface (Fig. 2C, iii, N-terminal portion of helix 11). Other regions in this category include the loop connecting beta strand 8 (N-terminal subdomain) and helix 9 (C-terminal subdomain) and residues 159-162 (Fig. 2C).

Conversely, regions with increased deuterium uptake (purple, Fig. 2) in IPTG-LacI compared to DNA-LacI included alternating beta strands in the C-terminal subdomain beta sheet (strands 12, 14 and 18) and all of the N-terminal subdomain helices (C-terminal end of helix 2, and helices 4, 6 and 19), which provide an interacting surface for the DNA-binding domain. Peptides from N-terminal subdomain helices 2 and 4 are shown in Fig. 2C, panels iv and v, respectively. Other regions in this category included residues in the C-terminal subdomain at the monomer-monomer interface (helix 13, residues 221-235, and helix 17, residues 276-282, Fig. 2C, vi, Fig. S5B). Of the nine loops that flank each ligand binding site, five loops exchanged more in IPTG-LacI than DNA-LacI.

Taken together, these results suggest a model where DNA-bound and inducer-bound states of the lac repressor adopt mutually incompatible low-energy conformational ensembles (Fig. 3A). DNA-LacI is characterized by an overall rigid structure of both the N- and C-terminal core domains, where the N-terminal subdomain makes extensive interactions with a well-folded operator-bound DNA-binding domain in the DNA-LacI crystal structure. Upon binding of the inducer IPTG, the ligand-binding pocket periphery as well as several secondary structure elements at the monomer-monomer interface are further rigidified, resulting in increased interactions between the LacI core subdomains. Importantly, this rigidification upon ligand binding leads to a simultaneous increase in flexibility in all of the N-terminal subdomain helices at the interface with the DNA-binding domain, even in the absence of the operator DNA (Fig. S7, see ΔHX; IPTG-DNA plot). In addition, the C-terminal core subdomain helices 13 and 17,
which flank the central helix 15 at the monomer-monomer interface, become more flexible in the inducer-bound states, in concert with increased flexibility in every other beta strand in the C-terminal subdomain beta sheet.

To further characterize differences in the rigidity of secondary structure elements between inducer- and DNA-bound states, we also used an alternate inducer. With few exceptions localized to the ligand binding pocket, we found that H-D exchange in TMG-LacI was similar to IPTG-LacI (Fig. S7). Likewise, H-D exchange in ONPF-DNA-LacI was similar to DNA-LacI (Fig. S7).

Comparison to mutational data.

We next compared our model for inducer-triggered conformational changes in the LacI core domain (Fig. 3A) to the extensive mutational phenotype data for LacI, where “phenotype” is defined by the effect of point mutations on gene expression. In a landmark study, 4000 LacI point mutants were tested for effects on gene expression (Suckow et al., 1996). The authors binned each WT amino-acid residue position according to its tolerance to mutation, and mapped these classifications onto the first crystal structure of LacI (Friedman et al., 1995) (Table S1, Fig. S8). The main Figure 3. Structural differences correspond to functional differences in the two mutually incompatible low-energy states of LacI. (A) Differential H-D exchange between IPTG-LacI and DNA-LacI, visualized by core subdomains. Diagrams on the right are positioned and cross-sectioned according to the schematic of LacI on the left. Secondary structure elements are labeled according to the schematic in the top panel of Fig. 2C, with helices represented as cylinders or circles and sheets represented as arrows, rectangles, or twisted rectangles. Purple indicates less exchange in DNA-LacI, while teal indicates less exchange in IPTG-LacI. (B) LacI structural models colored according to the two main classes of observed mutational phenotypes: I\(\ddagger\) (constitutive repression) and I\(\ddagger\) (constitutive expression) (Suckow et al., 1996). The table summarizes the mutational phenotype data as percentages of residues in differentially rigidified regions observed in HDX/MS experiments. The final column shows the percentage of mutationally tolerant residues in the regions that are rigidified in IPTG-LacI or DNA-LacI but do not cause an I\(\ddagger\) or I\(\ddagger\) phenotype. PDB ID: 2P9H.
mutational phenotypes observed in the study were: similar inducer response to WT LacI; unresponsive to inducer binding (I\(^s\)); or constitutively inducing (I\(^-\)).

We find a striking correlation between the phenotypic effects of these point mutants and our H-D exchange data. We found that the regions rigidified in IPTG-LacI (teal in Figure 2C, see Methods) correlated with positions at which mutations commonly cause an I\(^s\) phenotype, whereas regions rigidified in DNA-LacI (purple in Figure 2C) correlated with positions at which mutations cause an I\(^-\) phenotype. Mutations at 71% of the core domain residues that we observed to be rigidified in IPTG-LacI cause the I\(^s\) phenotype (Fig. 3B), but mutations to only 8% of these rigidified residues cause the I\(^-\) phenotype. Conversely, mutations at 30% of the core domain residues in regions that we observed to be rigidified in DNA-LacI cause the I\(^-\) phenotype, while mutations to only 8% of these residues cause the I\(^s\) phenotype. There were also mutations in regions with altered rigidity that cause neither phenotype, but the majority of these positions are tolerant to substitutions (Fig. 3B, last two columns in the table). Overall, we find considerable agreement between Suckow's phenotype assignment groups and differences in exchange between DNA-LacI and IPTG-LacI. The HDX/MS and mutational phenotype data for LacI together support a model for allosteric response in which the differential rigidification of secondary structures relative to one another in each state modulates protein function. Mutations in regions that are uniquely rigidified in IPTG-LacI may destabilize the conformational ensemble of that state, leading to constitutive repression. Conversely, mutations in regions that are more rigid in DNA-LacI may destabilize the DNA-LacI conformational ensemble, leading to constitutive transcription. Notably, the regions of the protein that are rigidified in either state are not confined to the ligand binding pocket or the DNA-binding surface.

We also compared the HDX/MS data to functional data from decades of LacI biochemistry studies, focusing on the effects of mutations in regions of the core domain that respond to IPTG- and DNA-binding in different ways: the monomer-monomer interface and the binding-pocket periphery.

At the monomer-monomer interface in helix 2, mutations at A81 and K84 (Fig. S9A) changed the IPTG-LacI and DNA-LacI binding affinities, as well as the protein stability (Chakerian et al., 1985; Chang et al., 1993; Markiewicz et al., 1994; Nichols and Matthews, 1997; Swint-Kruse et al., 2005). One hypothesis informed by crystal structures is that an inter-subunit, inter-subdomain salt bridge between H74 in helix 2 and D278 in helix 17 of the other subunit (Fig. S9B) stabilizes the conformation of IPTG-LacI (Lewis et al., 1996) although mutations to these residues do not necessarily disrupt IPTG binding (Barry and Matthews, 1999). Our H-D exchange data alongside these mutational data suggest that several quaternary contacts at the monomer-monomer interface in helices 2, 13, and 17, and beta strand 3, facilitate the IPTG-specific shifts in the conformational ensemble (Figs. 2B-C, S5).

At the N-terminal subdomain monomer-monomer interface, mutations to residues 94-98 in beta strand 3 (Fig. S9C) have also previously been shown to disrupt LacI allostery (Zhan et al., 2010, 2009). This observation is consistent with our hypothesis that hydrogen bonds across the beta sheets in both subunits rigidify the inducer-bound states, which arises from the observation of decreased H-D exchange in beta strand 3 in IPTG-LacI as compared to DNA-LacI (Fig. 2A, ii, and see also peptide 94 -VVVSM in Fig. S7, ΔHX, DNA-TMG and ΔHX, IPTG-DNA plots). Crystal structures additionally show a decrease in the distance between the beta strands, which corresponds to the formation of direct
hydrogen bonds between V96 backbone atoms in the IPTG state, rather than water-mediated hydrogen bonds in the DNA state in which the sidechain of K84 interacts with the carbonyl oxygens of V94 and V96 (Fig. S6).

Finally, we considered residues in the binding pocket-peripheral loops that interact directly with IPTG, TMG, and ONPF. Mutagenesis to residues L148 and S151 (Fig. S9D) resulted in shifts in the binding affinities of LacI for both IPTG and the DNA operator (Swint-Kruse et al., 2003). In line with the shift in binding affinity for IPTG noted in these mutational studies, we observed attenuated H-D exchange in the pocket-peripheral loops for IPTG as compared to the weaker-affinity ligand TMG (Fig. S7, ΔHX, IPTG-TMG plot, peptides 98 - MVERSGVE, 158 - SIIFSH, 147 - FLDVSDQTPINS and 185 - LAGPLSSVSARL) (Barkley et al., 1975).

Effect of anti-inducer binding.

We next sought to identify differences in the conformational ensembles between IPTG-LacI and ONPF-LacI. The two ligands IPTG and ONPF have opposite functional effects, where IPTG destabilizes and ONPF stabilizes the DNA-bound state (Barkley et al., 1975; Daber et al., 2007). To facilitate the comparison and because IPTG-LacI does not show measurable binding to DNA, we compared H-D exchange for IPTG-LacI and ONPF-LacI without the DNA operator to the apo state (Fig. 4). We found that, compared to the apo state, rigidification (less exchange) of particular secondary structure elements near the ligand binding pocket and in the interface between the N- and C-terminal core subdomains (the N-terminal part of helix 2, strand 3, strand 8 and the loop C-terminal to strand 8, helix 11, helix 15 and the loop N-terminal to helix 15) was unique to IPTG-LacI and not observed in ONPF-LacI (Figs. 4A, B, iv-vi). In contrast, ONPF binding rigidified regions distal to the ligand binding site in the N-terminal subdomain (Fig. 4A, B, i-iii, the C-terminal end of helix 2 and the loop connecting it to strand 3, the loop between strand 3 and helix 4, the loop between helix 4 and strand 5, the loop between strand 5 and helix 6, the loop between helix 6 and strand 7, the loop connecting strands 7 and 8, and the C-terminal ends of helices 6 and 19). Many of these regions interface directly with the DNA-binding domain (although DNA is absent in these HDX/MS experiments). Taken together, these results suggest that IPTG and ONPF selectively stabilize different secondary structure elements in the core domain to shift the conformational ensemble of LacI to different, mutually exclusive low-energy states. Unlike IPTG binding, ONPF binding does not cause dramatic changes to the structure of LacI that are incompatible with the observed conformational ensemble of DNA-LacI (Fig. S7, compare ΔHX, ONPF-APO, ΔHX, IPTG-APO, and ΔHX, APO-DNA plots).

The distinctive rigidification of the N-terminal subdomain in ONPF-LacI in the absence of DNA suggests that ONPF binding could preorganize this subdomain and the DNA-binding domain for DNA-binding. To test this idea, we compared HDX/MS data for peptides in the DNA-binding domain in different functional states of LacI. The structure of the DNA-binding domain has only been solved by X-ray crystallography in the DNA-LacI state (Fig. S10A). In these crystal structures, the N-terminal subdomain beta strands and helices form electrostatic interactions with a well-folded DNA-binding domain. Interestingly, while HDX/MS confirmed that the DNA-binding domain is well-structured in DNA-LacI and ONPF-DNA-LacI, we observed a range of hydrogen exchange behavior in the DNA-binding domain for the non-DNA-bound ligand-bound states. The DNA-binding domain underwent the most H-D exchange across
A More rigid in ONPF-LacI

More rigid in IPTG-LacI

B Differential HDX/MS comparing IPTG-LacI and ONPF-LacI to the apo state

C Probability of a hydrogen bond with a computationally predicted structural water
Figure 4. Computationally predicted solvent-mediated hydrogen bonding networks in the ligand-binding pocket of LacI correspond with differences in H-D exchange between IPTG-Lacl and ONPF-Lacl. (A) Deuterium uptake plots for the peptides i-vi marked in (B). Peptides that are more rigid in ONPF-Lacl are shown in the top row (yellow line), and peptides that are more rigid in IPTG-Lacl are shown in the bottom row (blue line), with apo-Lacl (green line) shown for reference. The data points show averaged, back exchange-corrected HDX/MS data for $n = 2-5$ individual measurements and error bars show standard deviations. Lines are fit to averaged experimental data using a nonlinear regression model as described in the Methods. (B) Crystal structures of the core domain of ONPF-Lacl (left) and IPTG-Lacl (right), colored according to differences in H-D exchange compared to apo-Lacl. Green regions are rigidified in each ligand-bound state as compared to apo-Lacl. Dark gray regions indicate no H-D exchange data. (C) Overlay of models of the ligand-binding pockets with ONPF (left) and IPTG (right) bound, with rigidified residues colored green in the cartoon representation as in (B), and surface representations showing the predicted probabilities of residue interactions with structural water molecules by our semi-explicit solvation protocol. Sidechains that are predicted to interact with structural waters are shown as sticks and surfaces and colored showing the predicted probabilities calculated from 100 trials. Secondary structure elements that are rigidified in IPTG-Lacl but not ONPF-Lacl are highlighted in the dashed circles.

timepoints in IPTG-Lacl, particularly in the hinge helix (residues 50-57) that connects the DNA-binding domain to the core domain (Fig. S10B), and was therefore the least structured among the states. However, decreased H-D exchange in the DNA-binding domain for ONPF-Lacl when compared to IPTG-Lacl even in the absence of DNA suggests that ONPF binding preorganizes the DNA-binding domain, thereby improving the binding affinity of LacI for the DNA operator when compared to apo-Lacl (Fig. S10C).

Water-mediated interactions in the ligand binding site.

Several secondary structure elements near the ligand binding pocket that show decreased exchange only in IPTG-Lacl and/or TMG-Lacl, such as helix 2, strand 8 and helix 11 (Figs. S3, S7), do not make direct contacts with ligands in X-ray crystal structures (Fig. S11). We hypothesized that decreased exchange in these regions may result from the formation of additional solvent-mediated hydrogen bonds in the ligand binding pocket that form upon binding inducers, but not anti-inducers. Crystal structures of IPTG-Lacl show such solvent-mediated interactions (Daber et al., 2007). In addition, mutations at binding pocket residues D149 and N125, which form water-mediated hydrogen bonds to IPTG in the crystal structure, decrease the protein’s affinity for IPTG (Xu et al., 2011).

To explore the potential role of water-mediated hydrogen bonding networks in the ligand binding pocket, we applied a computational “semi-explicit” solvation protocol in the macromolecular modeling program Rosetta (Pavlovicz et al., 2020) to predict the positions of structural water molecules in IPTG-Lacl, ONPF-Lacl, and apo-Lacl. We defined “structural waters” as water molecules that make three or four hydrogen bonds with atoms in the protein or ligand. As inputs, we first relaxed the experimentally solved crystal structures of IPTG-Lacl and ONPF-Lacl (Supplemental computational methods). The computational protocol then places water molecules in hydrogen bonding geometries relative to protein polar groups. Highly coordinated water molecules in the protein structure are kept, with or without “solvated” sidechains allowed to change rotameric conformations in Monte Carlo simulations. Only one high-resolution structure of LacI that includes electron densities for water molecules in the ligand binding sites is available (protein databank (PDB) ID 2P9H, the core domain of IPTG-Lacl). Using our
computational strategy without sidechain repacking, we first confirmed that the positions of our computationally predicted structural water molecules in IPTG-LacI are within 0.5 Å of several water molecules observed in this structure (Fig. S12). We then placed structural water molecules in the ligand binding pockets of ONPF-LacI and apo-LacI from crystal structures using the same protocol (Figs. 4C, S13). Based on 100 models for each LacI state, we calculated the probability that an atom in the protein or ligand makes a hydrogen bond with a computationally placed structural water in the ligand binding pocket (Fig. S13).

We observed a distinctive correspondence between regions with ligand-induced rigidification of secondary structures as measured by HDX/MS (Figs. 4A, B) and the calculated probability for water-mediated hydrogen bonding among residues in the ligand binding pocket (Fig. 4C, circled regions correspond to regions iv, v, and vi highlighted in Fig. 4A, bottom row, and 4B, right panel, as well as helix 15). Most strikingly, we predict an extensive water-mediated hydrogen bonding network for IPTG-LacI (Fig. 4C, right). In contrast, this network is not observed in the ONPF-LacI simulations (Fig. 4C, left). This result supports a model in which structural water molecules play a role in differentially rigidifying LacI in different functional states, where water-mediated hydrogen bonding networks contribute to increased interactions in the ligand binding pocket between the N- and C-terminal core subdomains in IPTG-LacI but not ONPF-LacI.

**Discussion**

We used a computational and experimental approach to map the allosteric response of LacI to different binding partners in six different functional states, at secondary structure resolution. Because the X-ray crystal structures of the core domains of DNA-LacI and IPTG-LacI are almost identical, our approach addressed the long-standing question of how allosteric effects in LacI are transmitted across more than 40 Å in the absence of sizeable conformational changes between functional states. Our model for inducer-triggered relief of DNA operator binding centers on selective dynamics of particular secondary structure elements relative to each other, shifting the conformational ensemble of the protein between mutually incompatible DNA-bound and inducer-bound states. Our results provide a detailed molecular mechanism consistent with data from structural and biochemical studies, expanding on the original MWC and KWF phenomenological models. The differences in conformational ensembles between the inducer-bound states and DNA-bound states that we observed by HDX/MS and predicted using computational methods also provide a mechanistic context for the altered functions of thousands of documented LacI mutants.

Figure 5A illustrates the key differences in the mutually incompatible conformational ensembles of the core domains in the IPTG-LacI and DNA-LacI states. The main features of IPTG-LacI, compared to DNA-LacI, are: rigidification of specific secondary structure elements near the ligand binding pocket (teal); increased specific inter-subdomain and inter-subunit interactions in the periphery of the ligand binding site (orange lines), some of which are supported by hydrogen bonding of protein and ligand atoms with structural water molecules (red circles and lines); and increased flexibility of other secondary structures throughout the C-terminal subdomain and in the N-terminal subdomain adjacent to the DNA-binding domain (purple). In contrast, DNA-LacI is characterized by increased rigidification in both core subdomains away from the ligand-binding site and at the interface with the DNA binding...
domain. We propose that IPTG binding rigidifies the N-terminal end of helix 2 and the inter-subunit contacts at the expense of the rigidity of neighboring helices. These IPTG-triggered shifts in the conformational ensemble in the core domain lead to increased flexibility in the hinge helices that in turn increase flexibility in the DNA-binding domain and ultimately weaken the binding affinity of LacI for the DNA operator.

![Figure 5A](image1.png)

**A** IPTG-LacI vs DNA-LacI

![Figure 5B](image2.png)

**B** ONPF-LacI vs DNA-LacI

**Figure 5. Consolidated models for ligand-specific changes in the conformational ensemble in IPTG-LacI and ONPF-LacI, as compared to DNA-LacI.** Schematics summarizing the rigidity of LacI core domain secondary structure elements in IPTG-LacI and ONPF-LacI as compared to DNA-LacI, with rigidified regions in each ligand-bound state shown in teal, de-rigidified regions shown in purple, and regions for which there is no data shown with stripes. Red circles show the positions of predicted structural water molecules, and red lines show the interactions of the predicted structural water molecules with atoms in ligands or amino acid residues. Orange lines represent hydrogen bonds uniquely formed among protein atoms in each ligand-bound state as compared to DNA-LacI. Rigidified loops in each schematic are thick for emphasis. (A) In IPTG-LacI, hydrogen bond formation at the ligand-binding pocket periphery promotes interactions between the N- and C-terminal subdomains of the core, which causes small structural rearrangements in the N-terminal subdomains (arrows) to increase flexibility in the hinge helix (black X). (B) In ONPF-LacI, the C-terminal core subdomain is not extensively perturbed as compared to DNA-LacI. However, N-terminal core subdomain loops become more structured, which may stabilize the interface of the core domain with the DNA-binding domain to bind the operator.

**Figure 5B** shows the key differences in the conformational ensembles of the core domains of ONPF-LacI and DNA-LacI. The main feature of ONPF-LacI is increased rigidity (teal) in loops in the N-terminal subdomain of the core distal from the ligand binding site. We propose that because ONPF binding does not increase the flexibility of key secondary structure elements in the core domain to the same extent as inducer binding, ONPF binding is not incompatible with the DNA-bound conformational ensemble. The main functional consequence of ONPF binding may be that the binding interface of the core
domain with the DNA-binding domain is preorganized in ONPF-LacI as compared to apo-LacI, leading to an increase in the binding affinity of LacI for the DNA operator.

One interesting aspect of our model is that it includes allosteric differences in exchange not only in the N-terminal subdomain but also in the C-terminal subdomain of the core (Figs. 2B-C, 3A). This observation is in contrast to previous models because the C-terminal subdomain in this family of transcription factors has long been considered conformationally fixed (Lewis et al., 1996; Mowbray and Björkman, 1999; Schumacher et al., 2004, 1995; Swint-Kruse and Matthews, 2009), with “virtually no difference in the hydrogen bond patterns or the number of hydrogen bonds in the C-terminal subdomains” (Daber et al., 2007). Targeted molecular dynamics simulations have built on this assumption, focusing solely on dynamics in the N-terminal subdomain while anchoring the C-terminal subdomain in place (Flynn et al., 2003). However, differences in HDX on the C-terminal side of the binding pocket, at the C-terminal monomer-monomer interface, in the beta sheet spanning the C-terminal subdomain, and in C-terminal subdomain solvent-exposed loops far from the ligand binding pocket highlight conformational changes that uniquely result from inducer binding. These observations provide a structural explanation for why mutations to these residues, for example D274, disrupt the response of LacI to binding IPTG (Chang and Matthews, 1995; Markiewicz et al., 1994; Suckow et al., 1996).

Engineering allosteric regulation in proteins remains challenging because allosteric mechanisms are typically incompletely understood. Our study provides mechanistic understanding of differences in conformational ensembles that can inform efforts to reengineer natural allosteric proteins to respond to user-defined perturbations. Transcription factors regulate gene expression through two types of binding interactions: binding to a specific DNA sequence, and binding to a set of small molecule or protein partners that control DNA binding. Domain-swapping (Jiang et al., 2021; Tungtur et al., 2010), chimeric fusions (Juárez et al., 2018), directed evolution (Richards et al., 2017; Taylor et al., 2016), saturation mutagenesis (Suckow et al., 1996; Wu et al., 2017), and deep mutational scanning (Tack et al., 2021) have resulted in LacI variants that control gene transcription in unexpected and interesting ways that depart from WT behavior, or in response to other sugars. However, none of these approaches are predictive with respect to tuning ligand response or allostery. Our strategy for computationally predicting interactions between protein, ligand, and water molecules and observing changes in conformational flexibility can help pinpoint structural changes that govern allosteric behavior in transcription factors and other proteins towards predictively and precisely designing the genotype-phenotype landscape.

Materials and methods

Cell culture fluorescence assay.

The E. coli DH10B ΔlacI:220 strain was chemically transformed with a plasmid encoding superfolder GFP (sfGFP) under the control of the pLacO promoter and a plasmid encoding the WT LacI (pEZ22G-GFP-kan and pSc101-lacI-specR, gifts from the Raman lab at U. Wisconsin) (Taylor et al., 2016) and plated on Lysogeny broth (LB) agar with 50 µl/ml kanamycin and 50 µl/ml spectinomycin. A culture was grown overnight in LB medium with antibiotics at the same concentrations from a single colony. New 5
ml cultures were grown at 37 °C and shaking at 225 rpm for 16 hours in blocks by subculturing the overnight culture using a 1:100 dilution in fresh LB medium. IPTG, TMG, and ONPF (Sigma-Aldrich, St. Louis, MO) were added to separate cultures to final concentrations of 100 µM, 100 µM, and 3 mM, respectively, at the two hour mark. Fully grown cultures were spun down at 3,000 rpm for ten minutes and the supernatant was decanted. Pellets were washed two times in 5 ml 1X phosphate-buffered saline (PBS), resuspended in 5 ml 1X PBS, and diluted 1:20 in 200 µl PBS in a clear-bottom black 96 well microtiter plate (Costar). Fluorescence from sfGFP expression was read from 7 mm above the plate from all sample wells using an excitation wavelength of 485 nm and an emission wavelength of 528 nm. All fluorescence readings were normalized by culture growth, measured by absorbance at 600 nm.

Protein expression and purification.

A BL21(Al) ΔlacI-ZYA strain was generated by the method of Datsenko and Wanner (Datsenko and Wanner, 2000) and BL21(Al) ΔlacI-ZYA cells were transformed with the pET9a-6×His-LacI(1-331) vector. This construct encodes for dimeric LacI, without the tetramerization domain. All HDX/MS experiments were conducted with dimeric Lacl. Cells were grown at 37 °C while shaking at 225 rpm for 12-16 hours overnight in LB medium with 50 µg/ml kanamycin (LB-kan). The overnight culture was subcultured at a 1:50 dilution in LB-kan with 0.2% glucose and grown at 37 °C while shaking at 225 rpm for two hours, until the OD600 value of the culture was between 0.4 and 0.5. The temperature was then decreased to 16 °C, protein expression was induced by the addition of 0.2% arabinose, and cells were grown for an additional 16 hours. The culture was centrifuged at 6,000 × g for 20 minutes at 4 °C and the pellet was frozen at -80 °C for at least two hours. The pellet was then thawed on ice for 15 minutes and resuspended in 40 ml lysis buffer (50 mM Tris, 150 mM NaCl, pH 8.0, with 5 mM MgCl₂, 1 mM MnCl₂, 100 µM CaCl₂, a dissolved Pierce™ protease inhibitor cocktail tablet (Thermo Fisher Scientific, Waltham, MA), and 40 µg DNaseI (Thermo Fisher Scientific)) by gentle vortexing for 20 minutes. The resuspended cells were lysed by two passes of microfluidization. The lysed cells solution was centrifuged at 27,000 × g for 30 minutes at 4 °C. The soluble cell fraction was decanted, supplemented with 20 mM imidazole, stored on ice for one hour, centrifuged at 3,270 × g for 10 minutes at 4 °C, and decanted again. We added 2 ml Ni²⁺-NTA slurry (HisPur Ni-NTA resin, Thermo Fisher Scientific) in binding buffer (50 mM Tris, 150 mM NaCl, 20 mM imidazole, pH 8.0) to the soluble cell fraction and nutated the protein-resin solution at 4 °C for one hour. The protein-resin solution was then centrifuged at 500 × g for five minutes at 4 °C and the supernatant was removed. The pelleted resin was resuspended in 2 ml binding buffer and applied to a 20 ml chromatography column. The resin was washed in the column with 50 ml binding buffer, then the protein was eluted in 14 ml elution buffer (50 mM Tris, 150 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein solution was spin-concentrated to a volume of 1.5 ml using Amicon™ Ultra Centrifugal Filter Units (MilliporeSigma, Burlington, MA) with a 10,000 Da molecular weight cutoff (MWCO). A volume of 3 µl from a 2 mg/ml stock solution of tobacco etch virus (TEV) protease with an N-terminal 6× histidine tag and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) was added to the concentrated protein solution. The reaction was incubated at 30 °C for two hours. Then, we diluted the reaction to 3 ml in Lacl in dialysis buffer (50 mM MOPS, 200 mM NaCl, 0.5 mM TCEP, pH 7.0) and dialyzed the reaction in dialysis buffer at 4 °C while stirring, using a dialysis cassette with a 3,000 Da MWCO (Slide-A-Lyzer, Thermo Fisher Scientific). The dialysis buffer was changed in 1 L increments of approximately 2 hours, 16 hours, and 2 hours. The
A dialyzed protein solution was diluted to 17 ml in dialysis buffer with 20 mM imidazole and supplemented with 2 ml Ni²⁺-NTA slurry that had been pre-equilibrated in dialysis buffer, nutated at 4 °C for one hour, and then applied to a 20 ml chromatography column. The flowthrough containing purified LacI was collected. An additional 5 ml dialysis buffer was applied to the column and the flowthrough was also collected. The flowthrough fractions were combined and spin-concentrated to a final concentration of 10 µM. A small volume of the final LacI sample was run on a denaturing SDS-PAGE gel to confirm the protein size and purity (>95%). The oligomerization state was confirmed by size-exclusion chromatography and protein secondary structure assessed by circular dichroism.

Sample preparation for HDX/MS.

Exchange buffer (50 mM MOPS, 200 mM NaCl, 0.5 mM TCEP, pH 7.0) was prepared in deionized water. To make a D₂O exchange buffer, the H₂O exchange buffer was lyophilized and resuspended in the same volume of D₂O three times. LacI samples were prepared at 10 µM in 100 µl H₂O exchange buffer, with a small molecule ligand, DNA operator, or both added to separate protein samples at the following concentrations: IPTG, 10 µM; ONPF, 3 mM; TMG, 10 µM; DNA operator, 25 µM. These ligand concentrations were chosen to be saturating and to maintain solubility. Samples were equilibrated at 4 °C with binding partners for at least 30 minutes before the HDX/MS experiment. Fully deuterated samples were prepared at 10 µM in D₂O buffer with 8 M urea, diluted in D₂O exchange buffer 1:10, incubated at room temperature for 16 hours, and lyophilized two times for at least 16 hours, adding back an equal volume of D₂O each time. Then the fully deuterated samples were incubated at room temperature for an additional 24 hours.

HDX/MS experimental protocol.

All HDX labeling was performed in solution using a LEAP technologies HDX PAL robot (Carrboro, NC), with a temperature-controlled sample vials and buffer holder held at 4 °C and refrigerated compartments held at 1 °C. Samples were injected on a Thermo UltiMate 3000 LC that was connected in-line with an Q Exactive mass spectrometer equipped with an electrospray ionization source (Thermo Fisher Scientific). The samples were diluted 1:10 in the D₂O exchange buffer (50 mM MOPS, 200 mM NaCl, 0.5 mM TCEP, pH 7.0) to allow for hydrogen atoms in the protein backbone amide groups to exchange for deuterons in the buffer. Aliquots of the diluted samples were collected at nine timepoints (0 sec, 30 sec, 45 sec, 1 min, 5 min, 25 min, 1 hr, 2 hr, 4 hr), quenched 1:1 in an acidic quench buffer (6 M urea, 200 mM arginine, 100 mM TCEP, pH 2.0), and cooled to 1 °C to slow the HDX reaction. This range of timepoints was chosen to allow for the measurement of fast and slow exchange events in all functional states of LacI. Each sample was then enzymatically proteolyzed on a Waters Enzymate™ BEH Pepsin Column (30 mm length x 2.1 mm diameter, 5 µm particle size, Waters Corporation, Milford, MA), or a combination of pepsin and fungal protease type XIII (manually conjugated to POROS™ 20 AL Aldehyde Activated Resin, Thermo Fisher Scientific, and packed in a 2 mm ID x 2 cm IDEX C-130B column, Fisher Scientific, Hampton, NH), loaded on a C18 analytical column (Hypersil Gold, 10 mm length x 2.1 mm diameter, 3 µm particle size, Thermo Fisher Scientific) with a guard column attached to the inlet (Thermo Fisher Scientific), and injected in a mass spectrometer using a 100 µL sample loop (Thermo Q Exactive) in solvent A (defined below). Desalting in sample A occurred over 2 minutes at 300 µl/min. Five independently generated exchange reactions (technical replicates)
for all states were performed from four separately expressed and purified LacI samples (biological replicates) to ensure adequate technical and biological replicates, except for the TMG state, for which two biological replicates were performed. Back exchange was calculated on a peptide basis by conducting an additional HDX/MS experiment on an apo LacI “maximally labeled” or “fully deuterated” control sample (Masson et al., 2019; Zhang and Smith, 1993). The average deuteration for all fully deuterated peptides over the whole protein was 70% with a standard deviation of 15.7%, and 11.4% of fully deuterated peptides had 50% deuteration or less (44 out of 386 total peptides). In the curated set of 57 core domain peptides included in this study, the average deuteration for fully deuterated peptides was 70.5% with a standard deviation of 14.6% (Table S2).

Acetonitrile, formic acid (1 mL ampules, Fisher Optima grade, 99.9%), and MS-grade water (Thermo Fisher Scientific) were used to prepare mobile phase solvents. Solvent A was 99.9% water/0.1% formic acid and solvent B was 90% acetonitrile/10% water with 0.1% formic acid added (v/v). The elution program consisted of isocratic flow at 10% B for 4 min, a linear gradient to 45% B over 9 min, isocratic flow at 100% B for one min, isocratic flow at 10% B for one min, a linear gradient to 45% B over one min, isocratic flow at 100% B for one min, and then isocratic flow at 10% B for 3 min, at a flow rate of 20 mL/min. Full-scan mass spectra were acquired in the positive ion mode over the range m/z = 400 to 2000 using the Orbitrap mass analyzer, with a mass resolution setting of 70,000, AGC target of 1.00e6, and maximum IT set to 50 s. We also ran three tandem mass spectrometry (MS/MS) experiments for each sample with the same full MS settings as described above, and dd-MS2 settings as follows: resolution 17,500, AGC target 1e5, maximum IT 100 ms, loop count 6, isolation window 2.0 m/z, NCE 28, charge state 1 and ≥7 excluded, dynamic exclusion of 20 seconds. Data acquisition was controlled using Xcalibur software (version 4.1, Thermo Fisher Scientific).

**HDX/MS data analysis.**

For each sample, peptides were identified using combined MS/MS raw data by searching against the LacI amino acid sequence (UniProtKB, www.uniprot.org, accessed 6/18/2019) and the pepsin amino acid sequence using Proteome Discoverer software (version 1.3, SEQUEST algorithm, Thermo Fisher Scientific) or Byonic software (Protein Metrics, Cupertino, CA). The protein search configuration used in the search were: cleavage residues, FWYLI;FWYLI; digest cutter, N-terminal and C-terminal cutter; and spectrum-level FDR, autocut. In a typical sample MS/MS, 476 unique forward peptides were found; 1027 spectra were matched to forward peptides; coverage was 98.18% of the 330-amino acid dimeric LacI sequence; and estimated spectrum-level false-discovery rate (FDR) on true proteins was 0.0%. The following filters were applied when identifying peptides: minimum precursor mass, 350 Da; maximum precursor mass, 5000 Da; minimum peak count, 1; maximum collision energy, 1000; S/N threshold, 1.5; minimum peptide length, 4; maximum peptide length, 144; maximum number of peptides reported, 10; precursor mass tolerance, 10 ppm; and fragment mass tolerance, 0.02 Da. The exported peptide lists (sequence, charge state, and retention time) from separate samples were imported into HD Examiner 2.0 (Sierra Analytics, Modesto, CA) and combined. All peptide isotope distributions at each exchange timepoint were fit in HD Examiner and manually checked for correct assignments. For each peptide in each functional state, deuteration levels at each timepoint were calculated by subtracting mass centroids of undeuterated peptides (m₀) from mass centroids of deuterated peptides (mₖ), dividing by the difference between the mass centroid of the fully deuterated peptide (mₐ) and the
undeuterated peptide, and multiplying by the theoretical maximum number of exchangeable protons (\textit{max.protons}): 

\[ D(t) = \frac{m_r - m_0}{m_\infty - m_0} \]

For each peptide, the charge state with the highest quality data as determined by HD Examiner was used. We note that the fully deuterated sample was in 100% D exchange buffer, while the experimental samples at each timepoint were in 90% D exchange buffer after 1:10 dilution, which would result in an increase in the NE and C parameters (described below).

Uptake plots were constructed by graphing the calculated deuteration levels against time for each peptide, with the maximum number of exchangeable protons as the limit on the y axis and the error bars showing the standard deviation of \( n = 2 \) to 5 replicates. The uptake data were fit to a nonlinear regression model using the SciPy optimize.curve_fit function. All calculations for fractional exchange as shown in Fig. 2B and Fig. S7 were performed using averaged centroid data at each timepoint, not the fitting functions; the nonlinear regression function fits are shown as lines in uptake plots to guide the eye. The nonlinear regression function was:

\[ D(t) = \textit{max.protons} - A \cdot e^{-k_1 \cdot t} - B \cdot e^{-k_2 \cdot t} - C \cdot e^{-k_3 \cdot t} - NE \]

where

\[ \textit{max.protons} = A + B + C + NE - 2 - \text{num.prolines} \]
\[ k_1 > k_2 \]
\[ k_2 > k_3 \]
\[ k_1 > k_3 \]

\( D(t) \): number of deuterons exchanged for protons as a function of time
\( \textit{max.protons} \): theoretical maximum number of exchangeable protons for the peptide
\( A, B, C \): number of fast-, medium-, and slow-exchanging protons, respectively
\( k_1, k_2, k_3 \): rates of exchange for fast-, medium-, and slow- exchanging protons, respectively
\( NE \): number of non-exchanging protons
\( \text{num.prolines} \): number of prolines in the peptide

For slow-exchanging peptides in which the last timepoint had an experimentally observed value of less than 0.5 D, data were fit to a modified regression function with \( A = 0 \) and \( k_1 = 0 \).

The fractional difference in H-D exchange at all timepoints and the associated uncertainty were calculated as described previously using averaged centroid data (Weis, 2021). Appendix 1 contains the calculated parameters for H-D exchange fitting to experimental data points for all peptides, corresponding to each line for the uptake plots in Fig. S3. Appendix 2 contains the calculated
uncertainties for the differential exchange plots shown in Fig. S7. Appendix 3 contains the raw MS centroids for all replicates at each timepoint for each peptide and functional state.

**Computational prediction of structural water molecules.**

To predict the positions of structural water molecules, we applied a semi-explicit solvation method implemented in the macromolecular modeling program Rosetta (Pavlovicz et al., 2020) on each pre-relaxed LacI crystal structure (PDB 2PAF, 2P9H, 1LBI) using the beta_nov16 energy function. Predicted structural water molecules within a 6 Å shell of any atom of the ligand were retained. All hydrogen bonds within a heavy-atom distance of 3.2 Angstroms were calculated between water oxygen atoms and protein or ligand residues. The probability of hydrogen bonding was calculated as the number of times that a residue could form a hydrogen bond with a predicted structural water molecule out of 100 total output structures. See the [Supplemental computational methods](#) for the detailed protocol, scripts, and command lines.

**Acknowledgements**

We thank members of the Kortemme and Marqusee Labs at UC San Francisco and UC Berkeley for discussions, the Raman Lab at the University of Wisconsin, Madison for the *E. coli ΔlacI:220* strain and the pZEO22G-GFP-kan and pSC101-LacI expression vectors, Shawn Costello for suggestions on the manuscript, and Monita Muralidharan for advice on HDX/MS data analysis. Funding: This work was supported by grants from the National Institutes of Health (K99GM135529 to AG and R01GM050945 to SM), and fellowships from the UCSF IRACDA program (K12GM081266) and the UC Office of the President through the UCSF Chancellor’s Fellowship to AG. ZP was supported by the Rosetta Commons Research Experience for Undergraduates Summer Internship program (NSF Award 1950697). SM and TK are Chan Zuckerberg Biohub Investigators.

**Author contributions**

AG conceived the idea for the project, developed the core questions, and performed the majority of the experimental studies and analysis. HTH and AG carried out the HDX/MS experiments. ZRP performed the Rosetta simulations under the guidance of AG. SM and TK provided discussion, mentorship, and resources. AG wrote the manuscript with contributions from the other authors.
References

Adler K, Beyreuther K, Fanning E, Geisler N, Gronenborn B, Klemm A, Müller-Hill B, Pfahl M, Schmitz A. 1972. How lac repressor binds to DNA. Nature 237:322–327. doi:10.1038/237322a0

Barkley MD, Riggs AD, Jobe A, Bourgeois S. 1975. Interaction of effecting ligands with lac repressor and repressor-operator complex. Biochemistry 14:1700–1712. doi:10.1021/bi00679a024

Barry JK, Matthews KS. 1999. Thermodynamic analysis of unfolding and dissociation in lactose repressor protein. Biochemistry 38:6520–6528. doi:10.1021/bi9900727

Bell CE, Lewis M. 2000. A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol 7:209–214. doi:10.1038/73317

Chakerian AE, Olson JS, Matthews KS. 1987. Thermodynamic analysis of inducer binding to the lactose repressor protein: contributions of galactosyl hydroxyl groups and beta substituents. Biochemistry 26:7250–7255. doi:10.1021/bi00397a009

Chakerian AE, Pfahl M, Olson JS, Matthews KS. 1985. A mutant lactose repressor with altered inducer and operator binding parameters. J Mol Biol 183:43–51. doi:10.1016/0022-2836(85)90279-7

Chang WI, Matthews KS. 1995. Role of Asp274 in lac repressor: diminished sugar binding and altered conformational effects in mutants. Biochemistry 34:9227–9234. doi:10.1021/bi00028a036

Chang WI, Olson JS, Matthews KS. 1993. Lysine 84 is at the subunit interface of lac repressor protein. J Biol Chem 268:17613–17622.

Cooper A, Dryden DT. 1984. Allosteroy without conformational change. A plausible model. Eur Biophys J EBJ 11:103–109. doi:10.1007/BF00276625

Coulondre C, Miller JH. 1977a. Genetic studies of the lac repressor: III. Additional correlation of mutational sites with specific amino acid residues. J Mol Biol 117:525–567. doi:10.1016/0022-2836(77)90056-0

Coulondre C, Miller JH. 1977b. Genetic studies of the lac repressor: IV. Mutagenic specificity in the lacI gene of Escherichia coli. J Mol Biol 117:577–606. doi:10.1016/0022-2836(77)90059-6

Daber R, Stayrook S, Rosenberg A, Lewis M. 2007. Structural Analysis of Lac Repressor Bound to Allosteric Effectors. J Mol Biol 370:609–619. doi:10.1016/j.jmb.2007.04.028

Daily MD, Gray JJ. 2009. Allosteric Communication Occurs via Networks of Tertiary and Quaternary Motions in Proteins. PLOS Comput Biol 5:e1000293. doi:10.1371/journal.pcbi.1000293

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci 97:6640–6645. doi:10.1073/pnas.120163297

Farabaugh PJ, Schmeissner U, Hofer M, Miller JH. 1978. Genetic studies of the lac repressor: VII. On the molecular nature of spontaneous hotspots in the lacI gene of Escherichia coli. J Mol Biol 126:847–863. doi:10.1016/0022-2836(78)90023-2

Flynn TC, Swint-Kruse L, Kong Y, Booth C, Matthews KS, Ma J. 2003. Allosteric transition pathways in the lactose repressor protein core domains: Asymmetric motions in a homodimer. Protein Sci Publ Protein Soc 12:2523–2541.

Friedman AM, Fischmann TO, Steltz TA. 1995. Crystal structure of lac repressor core tetramer and its implications for DNA looping. Science 268:1721–1727. doi:10.1126/science.7792597

Jacob F, Ullmann A, Monod J. 1965. Déletions fusionnant l’opéron lactose et un opéron purine chez Escherichia coli. J Mol Biol 13:704–719. doi:10.1016/S0022-2836(65)80137-1

Jiang X-L, Dimas RP, Chan CTY, Morcos F. 2021. Coevolutionary methods enable robust design of modular repressors by reestablishing intra-protein interactions. Nat Commun 12:5592. doi:10.1038/s41467-021-25851-6

Juárez JF, Lecube-Azpeitia B, Brown SL, Johnston CD, Church GM. 2018. Biosensor libraries harness large classes of binding domains for construction of allosteric transcriptional regulators. Nat Commun 9:3101. doi:10.1038/s41467-018-05525-6

Kalogiannakis CG, Bonvin AMJJ, Salinas RK, Wechselberger R, Boelens R, Kaptein R. 2002. Plasticity in protein–DNA recognition: lac repressor interacts with its natural operator O1 through alternative
conformations of its DNA-binding domain. *EMBO J* **21**:2866–2876. doi:10.1093/emboj/cdf318

Koshland DE, Némethy G, Filmer D. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**:365–385. doi:10.1021/bi00865a047

Lamerichs RM, Boelens R, Van der Marel GA, Van Boom JH, Kaptein R. 1990. Assignment of the 1H-NMR spectrum of a lac repressor headpiece-operator complex in H2O and identification of NOEs. Consequences for protein-DNA interaction. *Eur J Biochem* **194**:629–637. doi:10.1111/j.1432-1033.1990.tb15662.x

Lewis M, Chang G, Horton NC, Kercher MA, Pace HC, Schumacher MA, Brennan RG, Lu P. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**:1247–1254. doi:10.1126/science.271.5253.1247

Lockless SW, Ranganathan R. 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**:295–299. doi:10.1126/science.286.5438.295

Markiewicz P, Kleina LG, Cruz C, Ehret S, Miller JH. 1994. Genetic studies of the lac repressor. XIV. Analysis of 4000 altered Escherichia coli lac repressors reveals essential and non-essential residues, as well as “spacers” which do not require a specific sequence. *J Mol Biol* **240**:421–433. doi:10.1006/jmbi.1994.1458

Masson GR, Burke JE, Ahn NG, Anand GS, Borchers C, Brier S, Bou-Assaf GM, Engen JR, Englander SW, Faber J, Garlish R, Griffin PR, Gross ML, Guttmann M, Hamuro Y, Heck AJR, Houde D, Iacob RE, Jørgensen TJD, Kaltashov IA, Klinman JP, Konermann L, Man P, Mayne L, Pascal BD, Reichmann D, Skehel M, Snijder J, Strutzenberg TS, Underbakke ES, Wagner C, Wales TE, Walters BT, Weis DD, Wilson DJ, Wintrode PL, Zhang Z, Zheng J, Schriemer DC, Rand KD. 2019. Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat Methods* **16**:595–602. doi:10.1038/s41592-019-0459-y

McLeish TCB, Cann MJ, Rodgers TL. 2015. Dynamic Transmission of Protein Allostery without Structural Change: Spatial Pathways or Global Modes? *Biophys J* **109**:1240–1250. doi:10.1016/j.bpj.2015.08.009

Monod J, Wyman J, Changeux J-P. 1965. On the nature of allosteric transitions: A plausible model. *J Mol Biol* **12**:88–118. doi:10.1016/S0022-2836(65)80285-6

Morgan MA, Okamoto K, Kahn JD, English DS. 2005. Single-Molecule Spectroscopic Determination of Lac Repressor-DNA Loop Conformation. *Biophys J* **89**:2588–2596. doi:10.1529/biophysj.105.067728

Mowbray SL, Björkman AJ. 1999. Conformational changes of ribose-binding protein and two related repressors are tailored to fit the functional need. *J Mol Biol* **294**:487–499. doi:10.1006/jmbi.1999.3271

Nelson BD, Manoil C, Traxler B. 1997. Insertion mutagenesis of the lac repressor and its implications for structure-function analysis. *J Bacteriol* **179**:3721–3728. doi:10.1128/jb.179.11.3721-3728.1997

Nichols JC, Matthews KS. 1997. Combinatorial mutations of lac repressor. Stability of monomer-monomer interface is increased by apolar substitution at position 84. *J Biol Chem* **272**:18550–18557. doi:10.1074/jbc.272.30.18550

O’Gorman RB, Rosenberg JM, Kallai OB, Dickerson RE, Itakura K, Riggs AD, Matthews KS. 1980. Equilibrium binding of inducer to lac repressor.operator DNA complex. *J Biol Chem* **255**:10107–10114.

Pan H, Lee JC, Hilser VJ. 2000. Binding sites in Escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc Natl Acad Sci U S A* **97**:12020–12025. doi:10.1073/pnas.220240297

Parente DJ, Swint-Kruse L. 2013. Multiple Co-Evolutionary Networks Are Supported by the Common Tertiary Scaffold of the LacI/GalR Proteins. *PLOS ONE* **8**:e84398. doi:10.1371/journal.pone.0084398

Pavlovicz RE, Park H, DiMaio F. 2020. Efficient consideration of coordinated water molecules improves
computational protein-protein and protein-ligand docking discrimination. *PLOS Comput Biol* **16**e1008103. doi:10.1371/journal.pcbi.1008103

Richards DH, Meyer S, Wilson CJ. 2017. Fourteen Ways to Reroute Cooperative Communication in the Lactose Repressor: Engineering Regulatory Proteins with Alternate Repressive Functions. *ACS Synth Biol* **6**:6–12. doi:10.1021/acssynbio.6b00048

Rondon RE, Wilson CJ. 2019. Engineering a New Class of Anti-LacI Transcription Factors with Alternate DNA Recognition. *ACS Synth Biol* **8**:307–317. doi:10.1021/acssynbio.8b00324

Royer CA, Chakerian AE, Matthews KS. 1990. Macromolecular binding equilibria in the lac repressor system: studies using high-pressure fluorescence spectroscopy. *Biochemistry* **29**:4959–4966. doi:10.1021/bi00472a028

Schumacher MA, Allen GS, Diel M, Seidel G, Hillen W, Brennan RG. 2004. Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. *Cell* **118**:731–741. doi:10.1016/j.cell.2004.08.027

Schumacher MA, Choi KY, Lu F, Zalik H, Brennan RG. 1995. Mechanism of corepressor-mediated specific DNA binding by the purine repressor. *Cell* **83**:147–155. doi:10.1016/0092-8674(95)90243-0

Suckow J, Markiewicz P, Kleina LG, Miller J, Kisters WR, Brennan RG, Hillen W, Schumacher MA, Allen GS, Diel M, Seidel G, Kisters WR, Brennan RG, Hillen W. 2003. Ligand interactions with lactose repressor intermediates. *Proc Natl Acad Sci* **100**:13713–13718. doi:10.1073/pnas.2232983100

Swint-Kruse L, Zhan H, Swint-Kruse L, Zhan H, Matthews KS. 2005. Integrated insights from simulation, experiment, and mutational analysis yield new details of LacI function. *Biochemistry* **44**:11201–11213. doi:10.1021/bi050404+

Swint-Kruse L, Zhan H, Matthews KS. 2007. Ligand interactions with lactose repressor DNA complex revealed by a multiscale simulation. *Proc Natl Acad Sci* **104**:13736–13741. doi:10.1073/pnas.0701412104

Swint-Kruse L, Zhan H, Matthews KS. 2009. Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol* **12**:129–137. doi:10.1016/j.mib.2009.01.009

Taylor ND, Garruss AS, Moretti R, Chan S, Arbing MA, Cascio D, Rogers JK, Isaacs FJ, Kosuri S, Baker D, Fields S, Church GM, Raman S. 2016. Engineering an allosteric transcription factor to respond to new ligands. *Nat Methods* **13**:177–183. doi:10.1038/nmeth.3696

Tungtur S, Meinhardt S, Swint-Kruse L. 2010. Comparing the Functional Roles of Nonconserved Sequence Positions in Homologous Transcription Repressors: Implications for Sequence/Function Analyses. *J Mol Biol* **395**:785–802. doi:10.1016/j.jmb.2009.10.001

Villa E, Balaeff A, Schulten K. 2005. Structural dynamics of the lac repressor–DNA complex revealed by a multiscale simulation. *Proc Natl Acad Sci* **102**:6783–6788. doi:10.1073/pnas.0409387102

Wick KL, Matthews KS. 1991. Interactions between lac repressor protein and site-specific bromodeoxyuridine-substituted operator DNA. Ultraviolet footprinting and protein-DNA cross-link formation. *J Biol Chem* **266**:6106–6112.

Wilson CJ, Das P, Clementi M, Matthews KS, Wittung-Stafshede P. 2005. The experimental folding landscape of monomeric lactose repressor, a large two-domain protein, involves two kinetic intermediates. *Proc Natl Acad Sci* **102**:14563–14568. doi:10.1073/pnas.0505808102

Wilson CJ, Zhan H, Swint-Kruse L, Matthews KS. 2007. Ligand interactions with lactose repressor protein and the repressor-operator complex: The effects of ionization and oligomerization on binding. *Biophys Chem*, Julian Sturtevant Memorial Issue **126**:94–105.
Wu J, Jiang P, Chen W, Xiong D, Huang L, Jia J, Chen Y, Jin J-M, Tang S-Y. 2017. Design and application of a lactulose biosensor. Sci Rep 7:45994. doi:10.1038/srep45994

Xu J, Liu S, Chen M, Ma J, Matthews KS. 2011. Altering Residues N125 and D149 Impacts Sugar Effector Binding and Allosteric Parameters in Escherichia coli Lactose Repressor. Biochemistry 50:9002–9013. doi:10.1021/bi200896t

Yu EW, Koshland DE. 2001. Propagating conformational changes over long (and short) distances in proteins. Proc Natl Acad Sci U S A 98:9517–9520. doi:10.1073/pnas.161239298

Zhan H, Camargo M, Matthews KS. 2010. Positions 94-98 of the lactose repressor N-subdomain monomer-monomer interface are critical for allosteric communication. Biochemistry 49:8636–8645. doi:10.1021/bi101106x

Zhan H, Sun Z, Matthews KS. 2009. Functional impact of polar and acidic substitutions in the lactose repressor hydrophobic monomer.monomer interface with a buried lysine. Biochemistry 48:1305–1314. doi:10.1021/bi801357f

Zhang Z, Smith DL. 1993. Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. Protein Sci 2:522–531. doi:10.1002/pro.5560020404