

Substitutions at Nonconserved Rheostat Positions Modulate Function by Rewiring Long-Range, Dynamic Interactions

Paul Campitelli,¹ Liskin Swint-Kruse,² and S. Banu Ozkan^{*,1}

¹Department of Physics, Center for Biological Physics, Arizona State University, Tempe, AZ

²Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, Kansas City, KS

*Corresponding author: E-mail: banu.ozkan@asu.edu.

Associate editor: Claus Wilke

Abstract

Amino acid substitutions at nonconserved protein positions can have noncanonical and “long-distance” outcomes on protein function. Such outcomes might arise from changes in the internal protein communication network, which is often accompanied by changes in structural flexibility. To test this, we calculated flexibilities and dynamic coupling for positions in the linker region of the lactose repressor protein. This region contains nonconserved positions for which substitutions alter DNA-binding affinity. We first chose to study 11 substitutions at position 52. In computations, substitutions showed long-range effects on flexibilities of DNA-binding positions, and the degree of flexibility change correlated with experimentally measured changes in DNA binding. Substitutions also altered dynamic coupling to DNA-binding positions in a manner that captured other experimentally determined functional changes. Next, we broadened calculations to consider the dynamic coupling between 17 linker positions and the DNA-binding domain. Experimentally, these linker positions exhibited a wide range of substitution outcomes: Four conserved positions tolerated hardly any substitutions (“toggle”), ten nonconserved positions showed progressive changes from a range of substitutions (“rheostat”), and three nonconserved positions tolerated almost all substitutions (“neutral”). In computations with wild-type lactose repressor protein, the dynamic couplings between the DNA-binding domain and these linker positions showed varied degrees of asymmetry that correlated with the observed toggle/rheostat/neutral substitution outcomes. Thus, we propose that long-range and noncanonical substitution outcomes at nonconserved positions arise from rewiring long-range communication among functionally important positions. Such calculations might enable predictions for substitution outcomes at a range of nonconserved positions.

Key words: Allostery, protein dynamics, protein network, flexibility, *Lacl*.

Introduction

Historically, protein substitution experiments have been biased toward amino acid positions that are conserved during evolution (Kumar et al. 2009). Changes at these positions are often disease-causing and much effort has been spent toward understanding and predicting their outcomes. However, advancements in genome sequencing have also shown thousands of variations at nonconserved protein positions. By analogy to paralog evolution, changes at these positions can also alter protein function. Predicting the outcome of mutations at nonconserved positions, particularly for cases that do not cause large-scale structural differences from the wild-type structure, remains a considerable challenge (Swint-Kruse 2016; Miller et al. 2017).

We previously used the lactose repressor protein (“*Lacl*,” fig. 1) as a model system for understanding the functional roles of nonconserved positions. In this work, we focused on nonconserved positions in the linker region that forms interfaces with the two major *Lacl* functional domains (Bell and Lewis 2000; Swint-Kruse et al. 2002). Many linker positions are

nonconserved (supplementary table S1, Supplementary Material online) and their substitutions often showed non-canonical outcomes (Meinhardt et al. 2013): When a range of amino acids were substituted into each of the positions, the resulting changes in transcription repression and/or DNA-binding affinity ranged several orders of magnitude (supplementary fig. S1, Supplementary Material online and table 1). Notably, when the functional data for each position’s substitutions were arranged by their rank order, the resulting plots did not correlate with side chain chemistries nor could be explained by side chain characteristics (including their effects on helical propensity, hydrophobicity, side chain size, etc.) or evolutionary frequency (Zhan et al. 2006; Meinhardt et al. 2013). Since each variant was capable of binding DNA (Meinhardt et al. 2013), we reasoned that structural rearrangements must be minimal. However, previous experimental and computational studies indicated that the *Lacl* linker is very flexible and that changes in its folding (and thus flexibility) are integral to *Lacl* function (Wade-Jardetzky et al. 1979; Ha et al. 1989; Spolar and Record 1994; Lewis et al. 1996;

Spronk et al. 1996; Kalodimos et al. 2001, 2004; Kalodimos, Bonvin, et al. 2002; Swint-Kruse et al. 2002; Taraban et al. 2008). Furthermore, the LacI family exhibits allosteric behavior, where coevolution and evolutionary analysis has been conducted highlighting the complex interactions between domains (Dimas et al. 2019). Thus, we hypothesized that

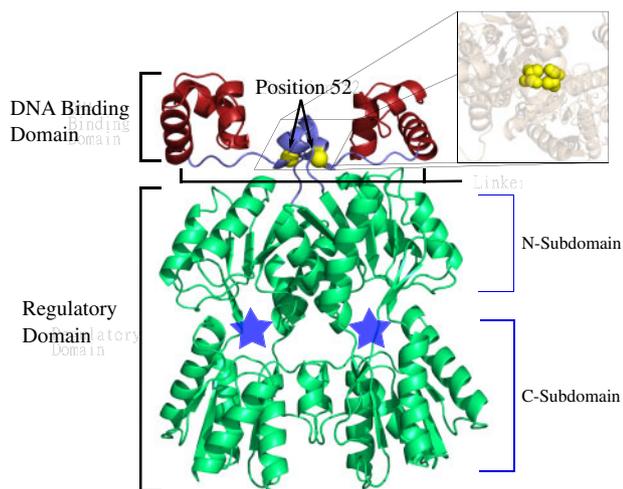


Fig. 1. LacI structure (pdb ID 1efa [Bell and Lewis 2000]) colored and labeled by domain. The minimal LacI structure for high affinity DNA binding is a homodimer (reviewed in Swint-Kruse and Matthews [2009]). Each monomer comprises two major domains—a DNA-binding domain (red) and a γ -regulatory domain (green)—that are connected by a flexible linker region (blue). DNA-binding involves positions on the top surface of the DNA-binding domains as well as several linker positions. The $C\alpha$'s of linker position 52 are highlighted with yellow spheres; the inset shows a top-down view of both V52 side chains using space-filling spheres. In the regulatory domains, locations of the allosteric effector-binding sites are indicated with blue stars. Small molecule “inducer” binding to this site diminishes DNA-binding affinities by up to 1,000-fold (Swint-Kruse and Matthews 2009). The structure shown and used in this work was extracted from a crystal structure (1efa, Bell and Lewis [2000]) that was determined in the presence of bound DNA and ortho-nitrophenyl-beta-D-fucopyranoside (“ONPF,” an anti-inducer that binds in the allosteric site and enhances DNA binding). This figure was rendered with pyMOL (Schrödinger 2010).

altered dynamics contribute to the noncanonical outcomes that arise from amino acid variation in the LacI linker.

To test this hypothesis, we performed in-depth analyses for substitutions at LacI linker position 52. We chose this position because, among all linker positions, 52 had the widest range of substitution outcomes and the most extensive biochemical data available (Zhan et al. 2006; Sousa et al. 2016; Hodges et al. 2018). Experimentally, substitution outcomes for this position spanned more than 4 orders of magnitude (Zhan et al. 2006; Meinhardt et al. 2013) (table 1).

In vitro DNA-binding affinities (K_d values) were available for wild-type LacI and 11 variants, of which six overlapped with the set of variants characterized by quantitative in vivo repression assays. Repression showed a strong correlation with K_d (supplementary fig. S1, Supplementary Material online), which suggests that repression changes were due to altered DNA-binding affinity. From these two experimental data sets, we used 11 variants (including wild-type LacI), which we designate as the “V52X” set throughout this text.

Here, we explored whether V52X substitutions affected flexibility at mechanistically important regions of the LacI structure and/or altered dynamic coupling between the DNA-binding domain and the linker. To that end, we first calculated site-specific flexibility of wild-type LacI and V52X variants using the dynamic flexibility index (“DFI”). The DFI parameter measures each position’s sensitivity to perturbations within a network of interactions and represents a given amino acid position’s ability to explore its local conformational space (flexibility). For LacI, the flexibilities of the DNA-binding domain and linker region differed significantly among the V52X variants, and the flexibility changes correlated strongly with changes in DNA-binding affinities.

Linker position 52 does not directly contact DNA and, indeed, is distal to many DNA-binding domain positions. Thus, we further reasoned that the flexibility changes observed in the DNA-binding domain must be propagated from the V52X variants via altered long-range dynamic coupling. To that end, we calculated the dynamic coupling index (“DCI”) for these variants. DCI measures the displacement response of an individual position to the perturbation of a second position or group of positions, relative to the average

Table 1. Experimentally Measured Functional Parameters for LacI V52X Variants (Zhan et al. 2006; Meinhardt et al. 2013).

Variant	$O_1 K_d (M \times 10^{-11})$	$O_1 + IPTG^a K_d (M \times 10^{-11})$	$[IPTG]_{mid} (M \times 10^6)$ Operator Release	$\Delta \ln$ Repression
WT	1.5 ± 0.4	>10,000	2.9 ± 0.6	0.00 ± 1.40
V52A	0.2 ± 0.03	6.2 ± 3.5	2.0 ± 0.3	-1.42 ± 3.12
V52I	0.94^b	n.d.	n.d.	0.27 ± 1.20
V52T	18.8 ± 8.0	>10,000	2.6 ± 0.5	4.42 ± 1.01
V52W	2.1 ± 1.0	400 ± 140	4.2 ± 0.7	n.d.
V52F	4.1 ± 2.0	>10,000	1.7 ± 0.2	n.d.
V52G	9 ± 2.0	>10,000	2.6 ± 0.3	2.90 ± 1.03
V52H	0.3 ± 0.2	18 ± 5	6.9 ± 2.0	n.d.
V52L	2.9 ± 0.2	>10,000	3.0 ± 0.1	n.d.
V52Q	7 ± 0.2	>10,000	1.9 ± 0.1	2.19 ± 0.99
V52S	0.9 ± 0.2	76 ± 8	4.5 ± 0.2	0.86 ± 1.50

NOTE.—n.d., Not determined.

^aIsopropyl β -D-1-thiogalactopyranoside.

^bExtrapolated, as shown in supplementary fig. S1, Supplementary Material online.

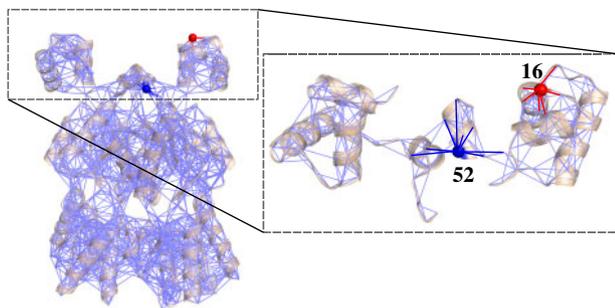


FIG. 2. Protein structure interactions can be represented as inhomogeneous networks. In this example using *Lacl*, each amino acid is represented as a node in the network, and covalent and noncovalent interactions are represented as edges. The expanded area in the dashed box shows a close-up view of the DNA-binding domains and linkers, with two positions highlighted (blue and red) to exemplify asymmetric coupling. Due to the inhomogeneous interaction networks between positions 52 and 16 and their partner positions, a perturbation at the red site may induce a response at the blue site that differs in magnitude from the response arising at the red site with the blue site is perturbed.

response to any perturbation of all possible positions. For each of the V52X variants, coupling to positions in the DNA-binding domain did indeed show different profiles, suggesting that changes in long-range dynamic coupling may underlie the altered flexibilities of the DNA-binding domains.

Given the compelling results for this representative nonconserved position, we next compared DFI and DCI calculations for all 13 nonconserved positions in the wild-type *Lacl* linker to four conserved linker positions. Within the linker region (positions 46–62), 13 positions were varied among homologs in the broader *Lacl*/*GalR* family (Tungtur et al. 2011) and exhibited a range of sequence entropies (Parente and Swint-Kruse 2013). In contrast, the other four linker positions were very strongly conserved across the whole family (Tungtur et al. 2011; Parente and Swint-Kruse 2013) with low sequence entropies (supplementary table S1, Supplementary Material online). Most of the nonconserved positions had substitution outcomes similar to position 52, with intermediate changes in biochemical function (Meinhardt et al. 2013), but the range of change varied from <1 to >4 orders of magnitude. In contrast, the four conserved linker positions showed canonical substitution outcomes, with most substitutions abolishing repression (Markiewicz et al. 1994; Suckow et al. 1996; Miller et al. 2017). Neither DFI nor DCI values discriminated these patterns of substitution outcomes. However, DCI values measure coupling that is inherently asymmetric: Every amino acid position has a unique network of direct, local interactions that gives rise to a unique network of highly coupled partner positions (fig. 2). Across the protein structure, this gives rise to an inhomogeneous, overall 3D network, and the asymmetry provides directionality to long-distance coupling. Thus, for a particular pair of coupled amino acids (i and j), their unique network constraints differentiate the coupling of i to j from the coupling of j to i . When DCI analyses were used to assess asymmetric coupling between linker positions and those in

the DNA-binding domain, distinct patterns were identified that correlated with different patterns of substitution outcomes.

Taken together, the current analyses provide a rationale for the substitution outcomes at nonconserved linker positions for which no previous relationships could reconcile and suggest that these positions are nodes in the dynamic communication network that exists between the functional domains of *Lacl*.

Results

The DFI calculation combines Perturbation Response Scanning and Linear Response Theory to evaluate each position's displacement response to random force perturbations at other locations in the protein (Gerek and Ozkan 2011; Nevin Gerek et al. 2013). For each position i , the DFI value quantifies its total displacement response relative to the total displacement response for all positions in the protein. To identify the most flexible positions, ranked %DFI values are used: For each position i , its percentile rank in the observed DFI range is calculated from the ratio of the number of positions with DFI values \leq DFI $_i$ to the total number of positions. %DFI values range from 0 to 1. The DFI parameter can be considered a measure of a given position's ability to explore its local conformational space; that is, DFI is a measure of an amino acid's flexibility. This parameter has been used to identify important structural elements of a protein, such as hinges, and to show the flexibility of functionally critical areas, such as binding domains and catalytic regions (Gerek and Ozkan 2010, 2011; Bolia et al. 2012; Nevin Gerek et al. 2013; Kumar et al. 2015; Bolia and Ozkan 2016; Campitelli et al. 2018; Modi et al. 2018; Modi and Ozkan 2018).

Figure 3 presents the %DFI profile for wild-type *Lacl* mapped onto its structure, using a color spectrum of red (highest flexibility) to blue (lowest flexibility). Calculations were performed using the protein structure extracted from the DNA-bound crystal structure 1efa (Bell and Lewis 2000) (i.e. no DNA was present in the calculation). The calculated DFI profile indicates that the wild-type structure exhibits a relatively flexible linker region and DNA-binding domain and a comparably stiff regulatory domain. The two subunits of the homodimer show symmetry in their flexibilities. These results agree with a variety of structural and biochemical evidence: As mentioned in the Introduction, a range of studies (Wade-Jardetzky et al. 1979; Ha et al. 1989; Spolar and Record 1994; Lewis et al. 1996; Spronk et al. 1996; Kalodimos et al. 2001, 2004; Kalodimos, Bonvin, et al. 2002; Swint-Kruse et al. 2002; Taraban et al. 2008) indicate flexibility of the DNA-binding domains and linker regions. In addition, comparisons of crystal structures for different ligand-bound forms of *Lacl* show that the C-subdomains of the regulatory domains remain fixed, whereas the two N-subdomains rotate relative to each other and to the C-subdomains (Lewis et al. 1996; Mowbray and Björkman 1999; Flynn et al. 2009).

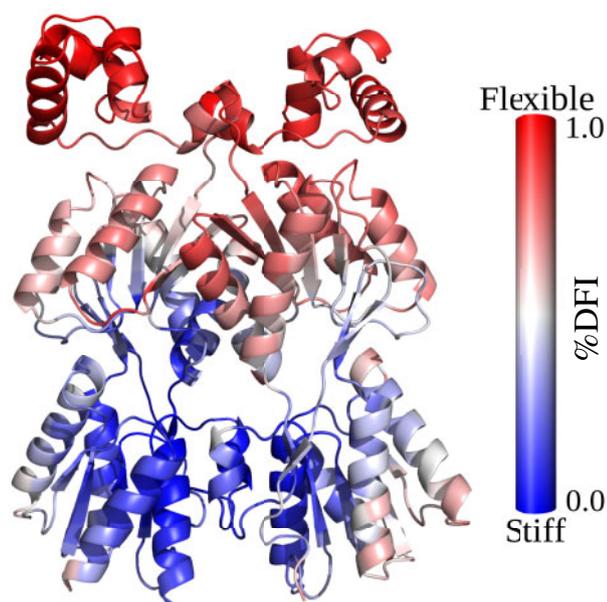


FIG. 3. WT LacI (pdb ID 1efa [Bell and Lewis 2000]) structure colored by %DFI. The profile was calculated using each position on both chains from simulations of the DNA-bound structure in the absence of DNA or anti-inducer ONPF. Note that the DNA-binding domains and the linker regions are more flexible than the distal parts of the regulatory domains, in agreement with known conformational changes.

LacI V52X Variants Induce Local and Nonlocal Flexibility Changes in the DNA-Binding Domain and Linker Region

Next, we determined whether amino acid substitutions in the LacI linker changed the flexibility profile of the DNA-binding domain and whether the altered flexibility correlated with altered DNA-binding affinities (K_d values). To that end, we first modeled the amino acid changes for 11 variants at position 52 and performed 250 ns molecular dynamics (MD) simulations with the AMBER ff14SB force field (Case et al. 2017). Relative to wild-type, the overall fold of the modeled structures remained unchanged for the DNA-binding and -regulatory domains (supplementary fig. S2A, Supplementary Material online). However, the substitutions gave rise to differences in local interactions, as expected for a position that makes linker–linker interactions (supplementary fig. S2B and C, Supplementary Material online) (Swint-Kruse et al. 2002), and subsequently we expected that these changes in local interactions would lead to changes in flexibility amongst the variants. Thus, to evaluate any flexibility changes, we calculated %DFI for positions in each of the V52X variants by using covariance matrices obtained from MD simulations (see Materials and Methods section).

Figure 4A shows the DNA-binding domain and linker regions of wild-type LacI and three V52X variants colored by %DFI. Figures for the other variants are in supplementary figure S3, Supplementary Material online and plots of %DFI values per position are shown in supplementary figures S4 and S5, Supplementary Material online. V52I had a similar %DFI profile in its DNA-binding domain and linker region as

wild-type LacI (fig. 4A), whereas seven variants with diminished DNA-binding affinities exhibited stiffening throughout the linker region as well as varied and distinct changes in the DNA-binding domain (V52T in fig. 4A and other variants in supplementary fig. S3, Supplementary Material online). Finally, V52A—which exhibited the highest repression and enhanced DNA binding of the V52X variants in this study (table 1 and supplementary fig. S1, Supplementary Material online)—showed much more stiffening in both the DNA-binding domain and linker.

Next, we reasoned that DNA-binding affinities would emerge as a property of structural flexibilities in the DNA-binding domain and possibly the linker region. The latter also contains residues that directly contact DNA (Bell and Lewis 2000; Kalodimos et al. 2001, 2004; Kalodimos, Bonvin, et al. 2002; Swint-Kruse et al. 2002). Therefore, we 1) summed the raw DFI values (“ Σ DFI”) for various groupings of these positions and 2) calculated the change in Σ DFI relative to that of wild-type LacI (“ $\Delta\Sigma$ DFI”). Finally, we correlated the changes in Σ DFI for various groups of positions with changes in measured DNA-binding affinities (fig. 4B and C and supplementary fig. S6, Supplementary Material online). Since DFI is a measure of local conformational entropy, it is appropriate to correlate this parameter with $\ln K_d$ values, which is proportional to free energy. Strikingly, for the DNA-binding domain positions that directly contact DNA, enhanced flexibility correlated with diminished DNA-binding affinity (fig. 4B). In contrast, no meaningful correlation was observed for linker positions that directly contact DNA (fig. 4C). Similar results were observed when all positions in the DNA-binding domain and linker region were included in Σ DFI (supplementary fig. S6A–C). Thus, we conclude that scores for DNA-binding domain positions with direct DNA contacts appeared contribute the most signal to the correlation.

V52X Substitutions Alter Dynamic Coupling between the Linker- and DNA-Binding Domains

The DFI results suggest that V52X substitutions have long-range repercussions on the flexibilities of the DNA-binding domains. Thus, we used the DCI to quantify long-range dynamic coupling between position 52 and individual positions in the DNA-binding domain (fig. 5A). The DCI parameter captures the strength of a displacement response for position i upon perturbation of position j , relative to the average fluctuation response of position i to all other positions in the protein. As such, this parameter represents a measure of the dynamic coupling between i and j . As in the case of %DFI, the DCI results are presented as %DCI, a percentile rank of the DCI range observed with values ranging from 0 to 1. Note that these two metrics are distinct: DFI is a measure of flexibility and DCI is a measure of coupling. Furthermore, DCI specifically quantifies the coupling between individual positions and, as such, DCI values depend explicitly upon the positions selected for analysis.

For each V52X variant, we calculated %DCI values between position 52 of chain A and all positions in the two DNA-binding domains (fig. 5B and supplementary fig. S7, Supplementary Material online). Results show that each

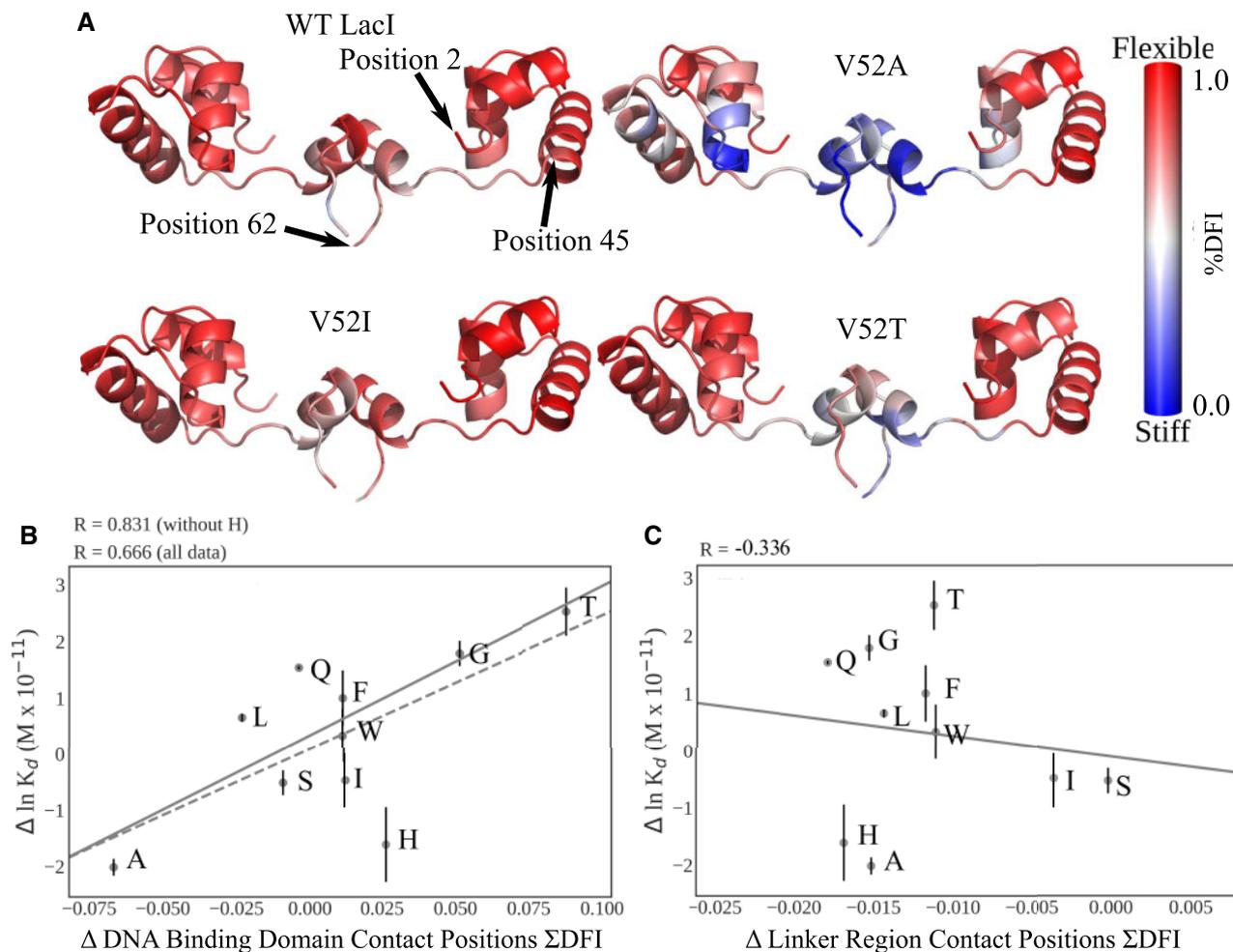


Fig. 4. (A) V52X variants showed altered flexibility. The DNA-binding domains (positions 2–45) and linker regions (positions 46–62) of LacI chains A and B are colored by %DFI for WT and three variants. All results are shown on the WT LacI crystal structure (pdb ID 1efa [Bell and Lewis 2000]). Note that V52A and V52T alter the flexibility profiles of DNA-binding domain and linker region, whereas V52I preserves the WT flexibility. For variants not shown here, see [supplementary figure S3, Supplementary Material](#) online. Plots of %DFI per position are shown in [supplementary figures S4 and S5, Supplementary Material](#) online. (B, C) Change in DNA-binding affinity versus change in ΣDFI for LacI positions that directly contact DNA. In the DNA-binding domain, the direct contact positions include 3–7, 14–19, 21, 22, 24, 25, 28–32, and 34. In the linker region, the direct contact positions include 50, 53, 54, 56, 57, and 59. Calculations included positions on both monomers of a dimer. Black lines indicate the best fit from linear regression of the data. The DNA-binding positions (B) show a strong correlation whereas no meaningful correlation exists between linker positions that contact DNA (C). Thus, we propose that V52X mutations lead to the long-distance changes in flexibility of DNA-binding domain that modulate-binding affinity. The V52H variant is a notable outlier on panel (B). One possibility arises from the fact that calculations were performed for a structure that is very similar to the DNA-bound structure used to initiate MD simulations. In contrast, changes in $\ln K_d$ can arise from changes in either the DNA-bound or -unbound states. The V52X substitutions may also induce flexibility changes in the apo structure. This could be a contributing factor to both the V52H outlier as well as the imperfect correlation in (B).

amino acid substitution led to dramatic dynamic coupling changes with many positions in the DNA-binding domains. Furthermore, the changes were amino acid specific. For example, even though wild-type, V52I, and V52W exhibited similar binding affinities (table 1) and similar flexibilities (fig. 4A and [supplementary fig. S3, Supplementary Material](#) online), the three variants have significantly different coupling patterns with positions in the DNA-binding domains (fig. 5B).

To assess whether any similarities exist among the coupling patterns of the V52X variants, we performed principle component analysis of the %DCI profiles. The singular value decomposition of each variant's %DCI profile was first assessed by plotting the first two principle components

(fig. 6A). Surprisingly, this analysis suggested a correlation with an additional aspect of LacI function: V52W, V52A, V52H, and V52S all cluster on the positive side of the first principle component; and DNA binding by these variants are *less* sensitive to allosteric regulation that occurs when inducer binds to the regulatory domains (fig. 6B) (Note that biochemically measured values for V52I inducibility are not available and thus V52I is not represented in fig. 6B.). To compare the top three principle components, we created a dendrogram (fig. 6C). In this analysis, observed clusters reflect groups with similar DNA-binding K_d values (table 1). For example, four variants with K_d near or lower than wild-type LacI (V52A, V52S, V52H, and V52I) are closely clustered (red oval).

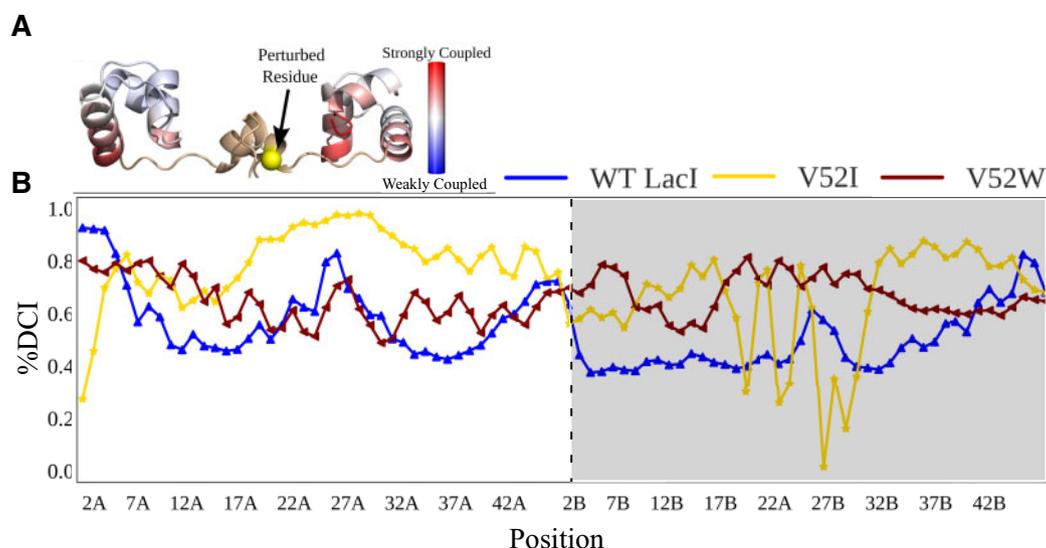


Fig. 5. (A) Example DCI calculations mapped on the Lacl DNA-binding domain/linker structure. The DCI values for all DNA binding and linker positions that result from perturbing position 52 (yellow sphere) are shown with the red to blue color gradient. (B) Percentile-ranked dynamic coupling between position 52 of chain A and the DNA-binding domain positions of WT Lacl, V52W, and V52I. The “A” and “B” designations on the position numbers and the white/gray shading indicate the two monomers of a dimer. Connecting lines are to aid visual inspection of the data. Note the significant differences in the score profiles, even though both variants have K_d values that are similar to WT. See [supplementary figure S7, Supplementary Material](#) online for %DCI scores for other V52X variants.

Dynamic Coupling Asymmetries Can Be Used to Distinguish among Rheostat, Toggle, and Neutral Positions

Since the different V52X variants altered binding affinities across a wide range of values, we previously defined position 52 as a “rheostat” position ([supplementary fig. S8, left panel, Supplementary Material](#) online) ([Meinhardt et al. 2013](#)). Rheostatic substitution behavior contrasts with the behavior often observed for conserved positions: For conserved positions in the Lacl linkers, most substitutions are catastrophically deleterious ([Markiewicz et al. 1994](#); [Suckow et al. 1996](#)) and we have referred to such positions as “toggle” positions ([supplementary fig. S8, middle panel, Supplementary Material](#) online) ([Meinhardt et al. 2013](#); [Miller et al. 2017](#); [Hodges et al. 2018](#)). The Lacl linker contains both toggle and rheostat positions ([supplementary fig. S8, Supplementary Material](#) online) ([Markiewicz et al. 1994](#); [Suckow et al. 1996](#); [Zhan et al. 2006](#); [Meinhardt et al. 2013](#)). Furthermore, the rheostat positions exhibited varied degrees of rheostat character, with some skewed more toward toggle behavior and others skewed more to “neutral” behavior ([Hodges et al. 2018](#)). At neutral positions, most substitutions have little to no effect; for example, Lacl nonconserved position 62 is nearly perfectly neutral ([supplementary fig. S8, right panel, Supplementary Material](#) online).

Since the DFI and DCI calculations predicted the functional outcomes for the individual substitutions at rheostat position 52, we were curious whether the wild-type structure alone could be used to predict the overall substitution outcome of each position. To that end, we used wild-type Lacl to investigate 1) whether other positions with strong rheostatic characteristics showed similar coupling characteristics as position 52, and 2) whether their characteristics of strongly

rheostatic positions were distinct from positions that were dominated by toggle and neutral substitution outcomes. First, we performed two separate %DCI calculations: DNA-binding domain responses to linker position perturbations and linker position responses to DNA-binding domain perturbations. Neither of these methods distinguished the linker positions’ degree of rheostatic, toggle, or neutral substitution behavior. However, we and others previously identified an intriguing characteristic in other model proteins ([Schreiber 2000](#); [Guo et al. 2015](#); [Hacisuleyman and Erman 2017](#); [Campitelli et al. 2018](#)). In these systems, ligand-binding altered the fluctuations at their local-binding sites and modulated dynamics at distal sites via unidirectional communication. We reasoned that unidirectional communication might also be altered by amino acid substitutions and that analyzing the asymmetry in dynamic coupling could discriminate rheostatic or toggle substitution behavior.

As described in the Introduction, asymmetry can be captured in the DCI values because dynamic coupling is inherently asymmetric: For a particular pair of amino acids (i and j), their unique interactions constrain and differentiate the coupling of i to j from the coupling of j to i ([fig. 2](#)). Thus, we used the wild-type structure of Lacl to calculate 1) %DCI _{i} , how strongly each linker position is coupled to each DNA-binding domain position, 2) %DCI _{j} , how strongly each DNA-binding domain position is coupled to each linker position. From these, we calculated 3) “%DCI_{asym}” from (%DCI _{i} – %DCI _{j}) to assess the asymmetry in coupling. [Figure 7A](#) shows %DCI_{asym} values obtained for positions representing the most rheostatic (V52), toggle-like (A53), and neutral (L62) positions; score distributions are presented as histograms in [figure 7B and D](#). Results showed intriguing patterns, particularly when considering their coupling to positions 2–14 in

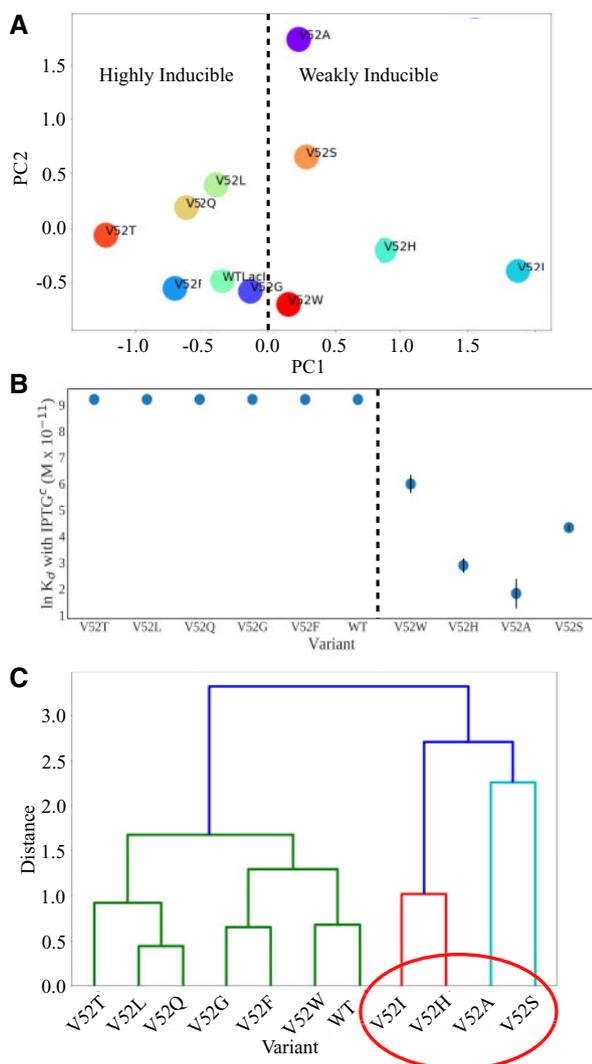


Fig. 6. Principle component analyses of V52X %DCI scores. (A) Scatter plots of the first (PC1) versus second (PC2) principle components from the V52X to DNA-binding domain %DCI scores suggest a separation in the first principle component correlates with altered inducibility (table 1), where the dashed line separates positive from negative values of the first principle component. Strikingly, this also appears to identify variants with altered inducibility (table 1 and panel [B]). (B) Point plot of $\ln K_d$ of DNA binding in the presence of IPTG. The dashed line marks the separation of highly inducible variants (K_d with IPTG $> 10,000 M \times 10^{-11}$) from weakly inducible variants and coincides with the dashed line in (A). The exception to this trend in PC1 may be V52I, which may comprise a third group as indicated by its position to the far right of panel (A). Biochemically measured values for V52I inducibility are not available, but this variant showed WT inducibility in repression assays (Meinhardt et al. 2013). (C) Hierarchical clustering dendrogram using Ward's method (Ward 1963) over first three principle components. Clusters associate with DNA-binding K_d values both in the presence of IPTG^c and without inducer (table 1). The colors represent clusters of variants within a given Ward distance cutoff used to define subgroups.

the DNA-binding domain. Values for the toggle position 53 were mostly negative, showing high unidirectionality, (the DNA-binding domain was more strongly coupled to the linker region). However, values for the rheostat position 52

were less negative, whereas the values for neutral position 62 were near zero or slightly positive.

To assess whether these trends were broadly observed, we performed a principle component analysis of the %DCI_{asym} values for all of the linker rheostat, toggle, and neutral positions. Figure 8A shows a plot of the first two principle components from this decomposition. The indicated groups were empirically determined in the absence of detailed knowledge about functional outcomes for each position. When functional data (fig. 8B and supplementary fig. S9, Supplementary Material online) were compared with computational results, PC2 appeared to separate the toggles from nontoggle positions (supplementary fig. S10, Supplementary Material online). When additional analyses were conducted for nontoggle positions using PC1, PC3, PC4, and PC5, the resulting dendrogram (supplementary fig. S11A, Supplementary Material online) as well as a direct comparison of PC1 and PC3 (supplementary fig. S11B, Supplementary Material online) suggest that the nontoggle positions form clusters that are similar to groups shown in figure 8A.

When the experimental substitution behaviors were compared with the principle component analyses (fig. 8B and supplementary fig. S9, Supplementary Material online), the groups from the %DCI_{asym} analysis matched reasonably well with the groups derived from experimental data (Hodges et al. 2018): For the three positions in group A, no catastrophic (“dead”) substitutions were identified in either of two experimental studies; instead most substitutions had neutral or modest outcomes. Positions with strong rheostat scores—which indicate that many intermediate outcomes were observed—fell into group B (see also supplementary fig. S9, Supplementary Material online). For the two positions in group C, most substitutions were catastrophic, but a few substitutions showed modest or neutral effects. Finally, for positions in group D, neither study identified modest nor neutral substitutions and the majority of substitutions were catastrophic (as indicated by high toggle scores and neutral scores of zero). Figure 8C shows the individual positions of each group from 8A and 8B color-coded on the LacI structure. Interestingly, the positions in groups A and D generally exhibited %DCI_{asym} distributions that were similar to those shown for example toggle and neutral positions in figure 7D and B.

In summary, %DCI_{asym} analysis was able to largely discriminate the different types of substitution outcomes experimentally observed for the LacI linker positions. The toggle-like and the neutral-like positions each comprised one group, and the rheostat positions were divided into two groups with different maximal severities from substitution.

Discussion

For missense variants, the prediction accuracy of current diagnostic tools is astonishingly low. In addition to their participation in multifactorial diseases, one source of the low accuracy arises from variants at nonconserved positions (the fastest evolving positions) (Subramanian and Kumar 2006; Miller et al. 2017). These positions are often *not* in close proximity to functionally critical sites (Jiménez-Osés et al.

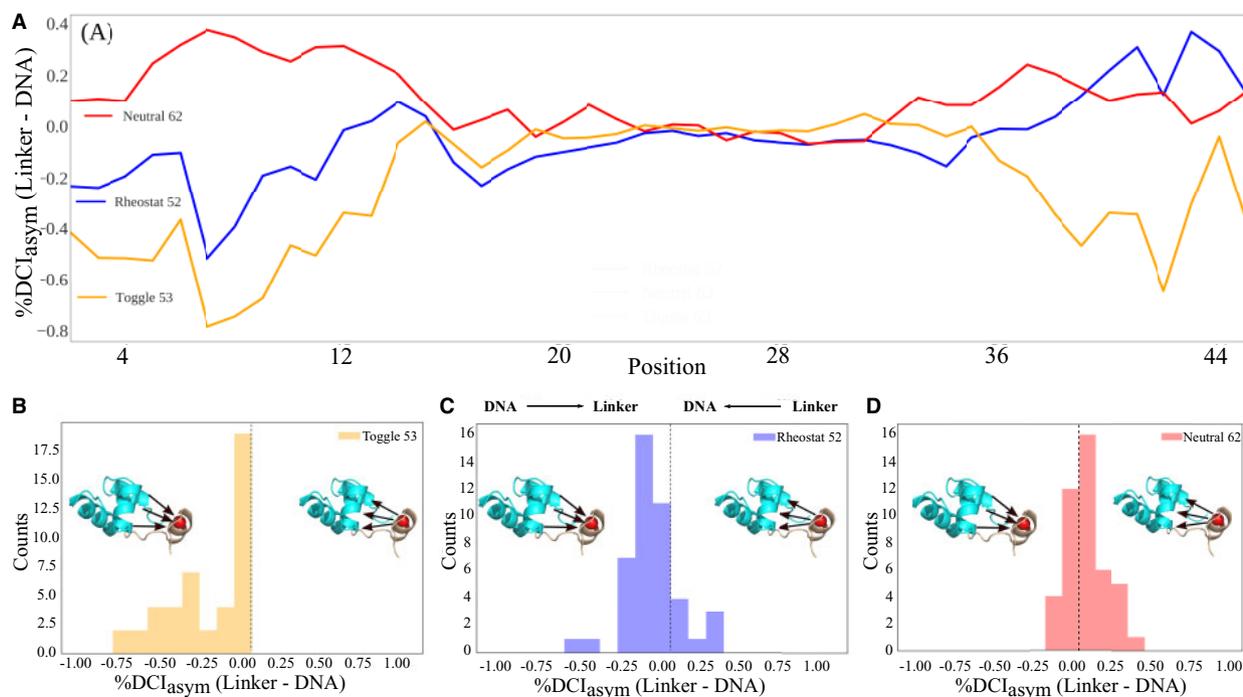


Fig. 7. (A) %DCI_{asy m} (linker to DNA-binding domain–DNA-binding domain to linker) scores for representative rheostat (52), neutral (62), and toggle (53) positions of chain A in WT Lacl. (B–D) Distributions of %DCI_{asy m} values for each position.

2014; Townsend et al. 2015; Modi and Ozkan 2018; Saavedra et al. 2018; Curado-Carballada et al. 2019), which has made it difficult to rationalize how their substitutions alter function. Similar observations have been made in directed evolution studies, which frequently identify substitutions distal from active, conserved sites that are required to evolve functional variation (Swint-Kruse 2016; Wilding et al. 2019).

In attempts to understand the effects of missense variants in the Lacl linker, extensive correlations with analyses of multiple sequence alignments were previously employed (Tungtur et al. 2011; Meinhardt et al. 2013; Parente and Swint-Kruse 2013; Parente et al. 2015). The nonconserved Lacl linker positions do not have strong coevolution scores, but they have reasonably high phylogeny scores. Since the current data are limited to the linker region of Lacl, the number of positions is too small to determine a definitive correlation. Work is ongoing to determine whether there is a general relationship among the toggle, neutral, and rheostat positions and scores from various sequence analyses (including sequence entropy). In addition, both phylogeny and/or coevolution scores have been assessed for other types of structural or functional correlations in various Lacl/GalR homologs, such as ligand specificity or direct structural contacts; to date, no relationship has been found (Meinhardt and Swint-Kruse 2008; Taraban et al. 2008; Tungtur et al. 2010, 2019; Parente and Swint-Kruse 2013).

Here, we have analyzed whether protein structural dynamics can provide insights how substitutions at nonconserved sites modulate protein function. Using Lacl as a model system, with specific focus on the relationship between the linker region and DNA-binding domain, our calculations showed correlations with experiment for individual substitutions at

one nonconserved linker position and could be used to discriminate the overall substitution responses for linker positions with a range of substitution profiles. Results provide mechanistic understanding to Lacl function, suggest avenues for future computations to further understand the substitution/function relationship in Lacl, and suggest trends that may be generally applicable to other proteins.

DFI and DCI Calculations Inform Several Aspects Lacl Function

Experimentally, when apo-Lacl binds to DNA, the protein undergoes a large conformational change (Ha et al. 1989; Kalodimos, Boelens, et al. 2002; Swint-Kruse et al. 2002; Taraban et al. 2008). The DFI calculations (fig. 4 and supplementary fig. S3, Supplementary Material online) suggest that V52X substitutions, and perhaps substitutions for other linker rheostat positions, alter the equilibrium dynamics of the DNA-bound conformation. In turn, these entropic changes would alter the free energy of DNA binding that was captured in experimental measures of DNA-binding affinity. Although not considered in the current computations, the V52X substitutions might also alter the equilibrium dynamics of the unbound state or induced states. Differential changes in each of the conformations may help explain why the correlations between binding affinity and DFI were not perfect (fig. 4B). Changes in flexibility were calculated on DNA-bound ensembles only, whereas the experimentally determined free energy of binding is defined by the difference between bound and unbound states. In future studies, the predicted entropic changes could be tested using thermodynamic experiments designed to detect the entropic component of DNA binding.

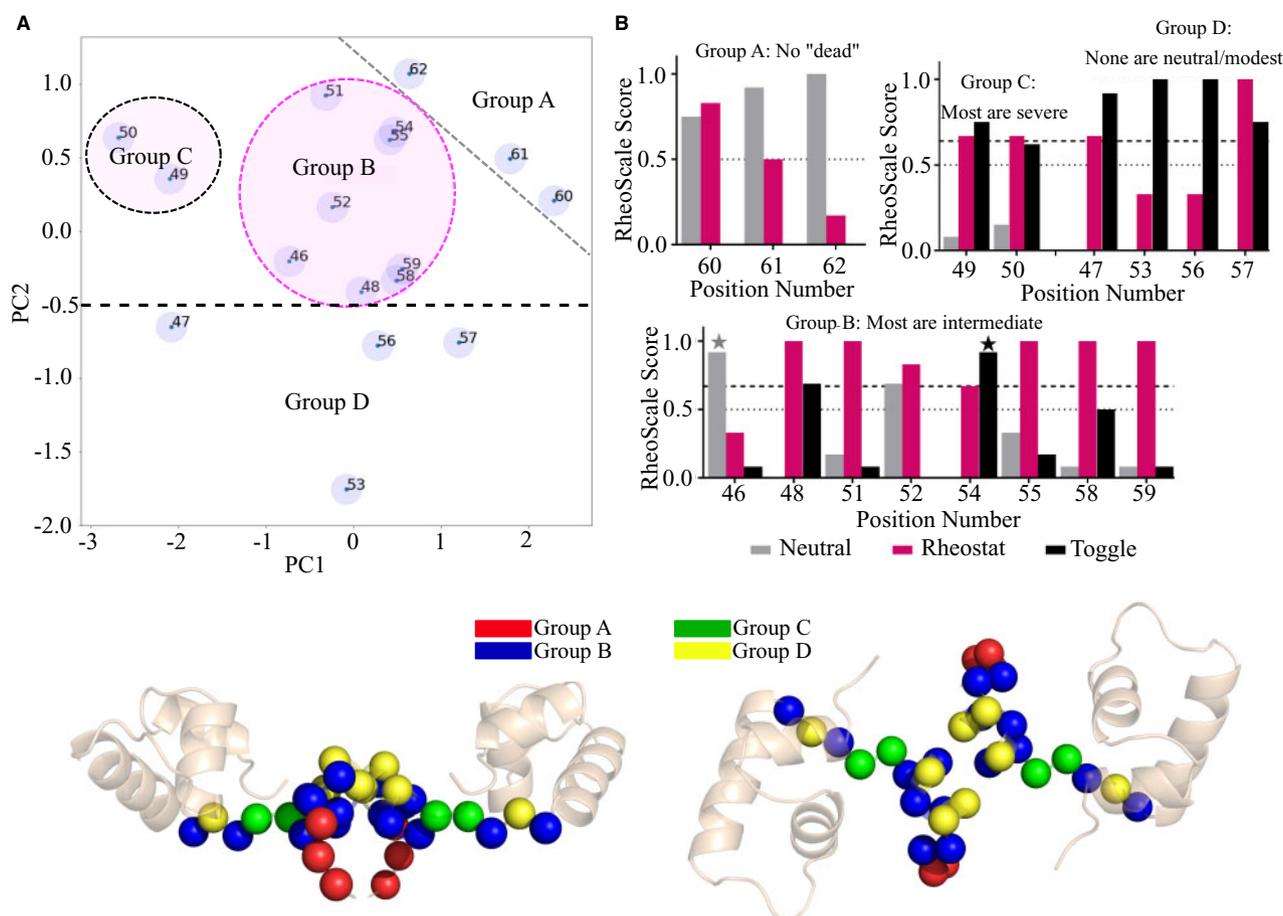


Fig. 8. Comparison of %DCI_{asy} analyses with functional attributes of each linker position. (A) 2D plot of the first two principle components from singular value decomposition of %DCI_{asy} profiles. Four groups were identified and indicated on this plot. (B) Aggregate of the experimental outcomes for all substitutions at each Lacl linker position. For each position, the outcomes from all available substitutions were analyzed with the “RheoScale” calculator (Hodges et al. 2018) to describe each position’s overall substitution behavior. “Neutral” scores (gray bars) reflect the fraction of substitutions that have little effect on function. “Rheostat” scores (magenta bars) reflect the fraction of the functional range that was accessed by at least one substitution; for the high-resolution data set, we previously defined rheostat scores ≥ 0.5 (dotted line) as significant (Hodges et al. 2018). “Toggle” scores (black bars) reflect the fraction of substitutions that are greatly damaging to function (“dead”); we previously defined toggle scores ≥ 0.67 (dashed line) as significant (Miller et al. 2017). This panel shows calculations for the low-resolution experimental data, which are available for all linker positions; calculations for the high-resolution data, which are available for 11 linker positions, are shown in [supplementary figure S9, Supplementary Material](#) online; for further details about these two data sets, see Materials and Methods section. Note that the neutral and toggle scores determined from low-resolution assay data are overestimates; many of these substitutions have intermediate outcomes. Scores for two positions affected by this overestimate are indicated with gray and black stars. (C) Linker position C α carbons colored by cluster for face-forward (left) and top-down (right) views.

In addition to DNA binding, Lacl binds inducer molecules via its regulatory domain (fig. 1), and the DNA-binding and -regulatory domains exhibit allosteric communication (reviewed in Swint-Kruse and Matthews [2009]). The latter function was also altered by substitutions at position 52 (Zhan et al. 2006). Interestingly, clustering the DCI of the V52X positions captured both the changes in the DNA binding and the changed response to inducer (Zhan et al. 2006) (fig. 6). Analogous to the DNA-binding positions that showed strong correlation in figure 4B, we speculate that a group of “sensor” positions in the inducer-binding site will show long-range flexibility changes in response to amino acid changes at position 52. In addition to the DNA-bound conformation used herein, such computations should be pursued using the inducer-bound conformation of Lacl. It will also be interesting to

determine whether DCI/DFI calculations can be used to predict the locations of other nonconserved positions for which substitutions modulate Lacl-inducer binding or allosteric regulation. More generally, this observation suggests a mechanism by which substitutions change in long-distance dynamic coupling to modulate allosterically regulated proteins.

In addition to alternative protein conformations, alternative DNA sequences might contribute to computational outcomes. For this work, the Lacl structure was extracted from a structure that was bound to DNA. During the simulations, Lacl did not transform all the way into the apo-form: the hinge helices in the linker stayed folded and the DNA-binding domains maintained similar conformations ([supplementary fig. S1C, Supplementary Material](#) online). Thus, some

of the information from the DNA sequence may have persisted during the simulations and DFI/DCI calculations. In the crystal structure used for this work (pdb 1efa), LacI was bound to the symmetric operator “Osym” (Bell and Lewis 2000). However, the experiments referenced herein used the natural operator “O1” (Zhan et al. 2006; Meinhardt et al. 2013). This did not concern us, because the effects of most V52X variants are comparable for these two DNAs (Tungtur et al. 2019) (When a substitution enhanced Osym binding, O1 binding was enhanced by a comparable amount.). However, this relationship between Osym and O1 binding is not known for V52I, which is an outlier in figure 6A. Thus, one prediction could be that V52I-Osym binding in the presence of inducer may differ from the trend followed by the rest of the V52X variants.

DCI_{asym} Distributions Capture Substitution Outcomes Trends at Rheostat, Toggle, and Neutral Positions

When we expanded our calculations to consider the dynamic coupling between 17 linker positions with the DNA-binding domain in the wild-type protein, we observed that degree of asymmetry in long-range dynamic coupling was in agreement with the functional outcomes of substitution. As discussed above, the dynamic coupling between linker position *i* and the DNA-binding position *j* is not necessarily symmetric, due to the asymmetry inherent to each positions' local network of interactions. This asymmetry gives rise to unidirectional information transfer, a concept which has previously been used to identify asymmetry in allosteric systems and to predict mutational impact (Kamberaj and van der Vaart 2009; LeVine and Weinstein 2014; Hacısuleyman and Erman 2017).

Historically, we have considered the “obvious” explanations for the origins of toggle positions. For example, LacI toggle position A53 directly contacts DNA; most other substituted side chains might make a steric clash; the A53G substitution (also catastrophic) might lack a critical interaction. The current work provides another explanation—not only does position 53 participate in the DNA interaction, it is a key node in the LacI coupling network. Indeed, all four toggle-like positions in the linker may function as key nodes in the coupling network. These positions exhibited large coupling asymmetry, with greater dynamic coupling of the DNA-binding domain to the linker region than the reverse. In other words, all the DNA-binding domain positions dominated the long-range communications to these linker positions, suggesting that substitution at these positions could compromise these critical communication channels with the DNA-binding domain. In contrast, the three neutral-like positions in the linker exhibited largely symmetric distributions, with a balance between DNA-dominant and linker-dominant coupling.

Intriguingly, the distributions of %DCI_{asym} values for rheostat linker positions were intermediate to the two extremes (toggle and neutral). This may allow (or be requisite for) broad modulation of function: Substitutions at rheostat positions may need only to slightly alter the dynamic coupling/communication to increase or decrease dynamic coupling among distal positions (i.e. rewire a new dynamic coupling

network). This would explain the long-range effects of substitutions and gives rise to another intriguing possibility for the underlying basis of nonadditivity often observed among two or more substitutions. Finally, this dynamic coupling network provides a satisfactory framework for understanding the biophysical mechanisms that underlie protein evolution and could explain why many of the rheostat positions identified to date occur at nonconserved positions.

Conclusion

In summary, these calculations suggest a number of promising avenues for understanding the functional outcomes of amino acid substitutions, particularly at nonconserved sites. These results add to the growing body of evidence that substitutions at nonconserved sites modulate function by fine-tuning flexibilities of distal functional sites. Modulation of flexibilities was also observed in studies that compared the equilibrium dynamics of resurrected ancestral proteins to modern-day homologs (Kim et al. 2015; Zou et al. 2015; Modi and Ozkan 2018). Similarly, long-range dynamics regulation and altered DCI profiles were observed in evolutionary variants of thioredoxin (Modi et al. 2018) and disease-associated human missense variants (Kumar et al. 2015). Thus, one mechanism by which proteins can adapt a new environment or acquire a new function (i.e. degrade a new substrate) is via substitutions that change the flexibilities of functionally important distal sites (Hilser 2010; Jiménez-Osés et al. 2014; Butler et al. 2015; Townsend et al. 2015; Modi and Ozkan 2018; Saavedra et al. 2018; Curado-Carballada et al. 2019). If the change is too large, the extreme change manifests as disease.

Studying the changes in equilibrium dynamics and, particularly, analyzing how substitutions modulate conformational states could provide mechanistic insights as to how substitutions at nonconserved sites affect function. In the future, this type of analysis can be expanded to other regions of LacI and to other types of proteins. In combination with dynamic spectroscopic analysis (single molecule or ensemble), computational mapping of the full interaction networks, particularly those existing between functional domains, will allow clear identification of critical positions in these networks, such as multipathway nodes. Such analyses could provide a means to predict the severity of substitution outcomes at nonconserved positions.

Materials and Methods

Functional Data for the LacI V52X Variants

As described in figure 1, a dimer is the minimal LacI functional unit for high affinity DNA binding. However, full-length wild-type LacI is a dimer-of-dimers, with tetramerization facilitated by an additional C terminal domain that forms a coiled coil. This domain can be cut off, leaving dimer with essentially identical intrinsic-binding affinity for a single DNA-binding site (Chen and Matthews 1994). DNA-binding affinities reported in table 1 were measured for the tetrameric versions of LacI (Zhan et al. 2006). High-resolution repression values (table 1) (Meinhardt et al. 2013) were measured using a

dimeric version of Lacl to compare high-resolution in vivo repression to that other dimeric paralogs (repression by tetrameric Lacl is enhanced by binding and two looping two operator sequences) (Swint-Kruse and Matthews 2009). In all studies carried out to date, substitution outcomes on DNA-binding affinities are the same in tetrameric and dimeric Lacl (Swint-Kruse et al. 2005; Tungtur et al. 2019). This is supported by the strong correlation shown in [supplementary figure S2, Supplementary Material](#) online. In this manuscript, we use the term “wild-type” interchangeably for the dimeric and tetrameric forms. A third data set comprising low-resolution in vivo repression data for variants of tetrameric Lacl (Markiewicz et al. 1994; Suckow et al. 1996) was used for RheoScale analyses and is further described below.

In both biochemical and high-resolution in vivo studies, all substitution outcomes appeared to manifest in functional changes, rather than as changes in structure or stability: In the in vitro study, the variant proteins purified and behaved very similar to wild-type Lacl (Zhan et al. 2006). In the in vivo studies, control experiments showed that all protein variants were expressed and capable of binding DNA (Meinhardt et al. 2013). Together, these observations indicate that the gross structural features of the Lacl V52X variants were not distorted.

Modeling of V52X Variant Structures

Bell and Lewis 2000) (PDB ID 1efa); the models for DNA operator and anti-inducer ONPF were removed prior to calculations. Mutagenesis was performed using the PyMol Mutagenesis Wizard on valine at position 52 of both chains A and B to create 10 additional variants: V52A, V52F, V52G, V52H, V52I, V52L, V52Q, V52S, V52T, and V52W. Topology files for all structures were prepared using the AMBER LEaP program with the ff14SB force field (Case et al. 2017), where hydrogen atoms were added and each structure was surrounded by an 18.0 Å cubic box of water molecules using the TIP3P (MacKerell et al. 1998) water model. Na⁺ and Cl⁻ atoms were added for neutralization. To remove any unfavorable torsional angles or steric clashes and ensure that the system reached a local energetic minimum, each system was energy-minimized using the AMBER SANDER package (Case et al. 2017). First, the protein was kept fixed with harmonic restraints to allow surrounding water molecules and ions to relax, followed by a second minimization step in which the restraints were removed and the protein-resolution was further minimized. Both minimization steps employed the method of steepest descent followed by conjugate gradient.

The systems were then heated from 0 to 300 K over 250 ps, at which point long-range electrostatic interactions were calculated using the particle mesh Ewald method (Darden et al. 1993). Direct-sum, nonbonded interactions were cut off at distances of 9.0 Å or greater. The systems were then simulated using MD at constant temperature and pressure with 2 fs time steps for 250 ns. Covariance matrix data were calculated over the final 40 ns of each simulation, using a 20 ns moving window that overlap by 5 ns. Additional references used to validate these models were (Swint-Kruse et al. 1998, 2001;

Swint-Kruse 2004; Swint-Kruse and Brown 2005), as further detailed in the supplement.

Dynamic Flexibility and Dynamic Coupling

The DFI utilizes a Perturbation Response Scanning technique that combines the Elastic Network Model and Linear Response Theory (“LRT”) (Gerek and Ozkan 2011; Nevin Gerek et al. 2013).

$$[\Delta\mathbf{R}]_{3N \times 1} = [\mathbf{H}]_{3N \times 3N}^{-1} [\mathbf{F}]_{3N \times 1}. \quad (1)$$

Using LRT, $\Delta\mathbf{R}$ is calculated as the fluctuation response vector of residue j as a result of unit force's \mathbf{F} perturbation on residue i , averaged over multiple unit force directions to simulate an isotropic perturbation (2). \mathbf{H} is the Hessian, a $3N \times 3N$ matrix which can be constructed from 3D atomic coordinate information and is composed of the second derivatives of the harmonic potential with respect to the components of the position's vectors of length N . For this work, the Hessian matrix was extracted directly from MD simulations as the inverse of the covariance matrix. This method allows one to implicitly capture-specific physiochemical properties and more accurate residue-residue interactions via atomistic force fields and subsequent all-atom simulation data.

Each position in the structure was perturbed sequentially to generate a Perturbation Response Matrix \mathbf{A}

$$\mathbf{A}_{N \times N} = \begin{bmatrix} \Delta R^1_1 & \cdots & \Delta R^N_1 \\ \vdots & \ddots & \vdots \\ \Delta R^1_N & \cdots & \Delta R^N_N \end{bmatrix}, \quad (2)$$

where $\Delta R^j_i = \sqrt{(\Delta R)^2}$ is the magnitude of fluctuation response at position i due to perturbations at position j . The DFI value of position i is then treated as the displacement response of position i relative to the net displacement response of the entire protein, which is calculated by sequentially perturbing each position in the structure (3).

$$\text{DFI}_i = \frac{\sum_{j=1}^N \Delta R^j_i}{\sum_{i=1}^N \sum_{j=1}^N \Delta R^j_i}. \quad (3)$$

It is also often useful to quantify position flexibility relative to the flexibility ranges unique to individual structures. To that end, DFI can be presented as a percentile rank

$$\% \text{DFI}_i = \frac{n_{\leq i}}{N}, \quad (4)$$

where $n_{\leq i}$ is the number of positions with a DFI value $\leq \text{DFI}_i$. The denominator is the total displacement of all residues, used as a normalizing factor. All %DFI calculations present in this work used the DFI value of every residue of the full Lacl structure for ranking. The DFI parameter can be considered a measure of a given amino acid position's ability to explore its local conformational space.

Dynamic Coupling Index

Similar to DFI, the DCI also utilizes Perturbation Response Scanning with the Elastic Network Model and Linear Response Theory. The DCI captures the strength of displacement response of a given position i upon perturbation to a single functionally important position (or subset of positions) j , relative to the average fluctuation response of position i when all of the positions within a structure are perturbed.

$$\text{DCI}_i = \frac{\sum_j^{N_{\text{functional}}} \Delta R_i^j / N_{\text{functional}}}{\sum_{j=1}^N \Delta R_i^j / N}. \quad (5)$$

As above, DCI can also be presented as a percentile rank as

$$\% \text{DCI}_i = \frac{m_{\leq i}}{N}, \quad (6)$$

where $m_{\leq i}$ is the number of positions with a DCI value $\leq \text{DCI}_i$. As such, this parameter represents a measure of the dynamic coupling between i and j upon a perturbation to j . Additional information such as preferential information transfer through asymmetric dynamic coupling can also be determined by analyzing the coupling asymmetry between positions i and j , which can be calculated as

$$\text{DCI}_{\text{asym}} = \text{DCI}_i - \text{DCI}_j, \quad (7)$$

$$\% \text{DCI}_{\text{asym}} = \% \text{DCI}_i - \% \text{DCI}_j, \quad (8)$$

where a positive DCI_{asym} value indicates communication from position i to position j . For this work, DCI analysis was performed using force perturbations to chain A positions.

RheoScale Calculations Aggregate Experimental Functional Data for Multiple Variants

Rather than thinking about the role of individual amino acid residues within a protein, we have found it useful to consider the evolutionary role of each protein position: Conserved positions often require a specific amino acid to carry out an evolutionarily conserved function (or structure). Most amino acid substitutions at these locations are catastrophic to the protein's structure or function. As such, the outcomes for multiple substitutions resemble a toggle switch, with function either "on" or "off." Alternatively, some nonconserved positions can be varied to modulate function, an evolutionarily advantageous method to help organisms adapt to new niches. We have shown that substitutions at these positions can lead to a wide range of intermediate outcomes (Meinhardt et al. 2013; Hodges et al. 2018). Finally, some positions can tolerate any amino acid substitution; functional outcomes are "neutral" (Martin et al. 2020; Miller et al. 2019).

These three categories represent idealized behaviors. Substitution data for real protein positions often fall on the continuum. Thus, we developed a scale to quantify how toggle-like, rheostatic, and neutral the aggregate functional data are for a given position (Hodges et al. 2018). "Neutral" scores reflect the fraction of substitutions that have little effect on function. "Rheostat" scores reflect the fraction of the functional range that was accessed by at least one

substitution; for the high-resolution data set, we previously defined rheostat scores ≥ 0.5 (dotted line) as significant (Hodges et al. 2018). "Toggle" scores reflect the fraction of substitutions that are greatly damaging to function; we previously defined toggle scores ≥ 0.67 (dashed line) as significant (Miller et al. 2017).

For the Lacl linker positions, two data sets are available for RheoScale calculations, both comprised in vivo repression measurements: 1) The Swint-Kruse lab used a high-resolution functional assay to assess outcomes for 7–14 substitutions at each of 11 nonconserved linker positions in dimeric Lacl (Meinhardt et al. 2013). High-resolution data for position 52 are used in table 1 and RheoScale calculations for position 52 and other nonconserved positions (Hodges et al. 2018) are shown in supplementary figure S9, Supplementary Material online. 2) The Miller lab used a low-resolution assay to assess outcomes for 12–13 substitutions at all linker positions (and most other positions) in tetrameric Lacl (Markiewicz et al. 1994; Suckow et al. 1996); calculations from these data are shown in supplementary figure S8, Supplementary Material online and more details are given below. In all Lacl studies to date, substitutions have the same outcome on DNA binding for tetrameric and dimeric Lacl (Swint-Kruse et al. 2005). The two functional studies have also some differences in their sets of substituted amino acids used, but where overlapping data is available, the RheoScale scores were in good agreement (Hodges et al. 2018).

RheoScale calculations for the low-resolution data were as follows: For each substitution, repression was scored as either 1) "+" (reporter gene activity, relative to "no repression," was more than 200-fold diminished); 2) "+/–" (diminished from 200- to 20-fold); 3) "–/+" (diminished from 20- to 4-fold); or 4) "–" (no more than 4-fold diminished) (Markiewicz et al. 1994; Suckow et al. 1996). Since the RheoScale calculator requires numerical data to perform histogram analyses, we reassigned these categories as numerical values, with 1 being the best repression and 4 being the worst. Histogram analyses were carried out using 4 bins. Note that the "+" data encompasses some substitutions with modest changes in repression that would be detected in the low-resolution assay; the "+/" data also encompass any substitutions that enhance repression. Likewise, the "–" substitutions encompass some positions that retain some weak ability to repress. Thus, both the neutral and the toggle scores were overestimated for these data, whereas the rheostat score was underestimated.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the National Science Foundation Division of Molecular and Cellular Biosciences (award 1715591), the National Institute of Health (grant number GM118589), the Gordan and Betty Moore Foundation and the W.M. Keck Foundation.

References

- Bell CE, Lewis M. 2000. A closer view of the conformation of the Lac repressor bound to operator. *Nat Struct Biol.* 7(9):209–214.
- Bolia A, Gerek ZN, Keskin O, Banu Ozkan S, Dev KK. 2012. The binding affinities of proteins interacting with the PDZ domain of PICK1. *Proteins* 80(5):1393–1408.
- Bolia A, Ozkan SB. 2016. Adaptive BP-Dock: an induced fit docking approach for full receptor flexibility. *J Chem Inf Model.* 56(4):734–746.
- Butler BM, Gerek ZN, Kumar S, Ozkan SB. 2015. Conformational dynamics of nonsynonymous variants at protein interfaces reveals disease association. *Proteins* 83(3):428–435.
- Campitelli P, Guo J, Zhou H-X, Ozkan SB. 2018. Hinge-shift mechanism modulates allosteric regulations in human Pin1. *J Phys Chem B* 122(21):5623–5629.
- Case DA, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Greene D, Homeyer N, et al. 2017. Amber 2017. San Francisco: University of California.
- Chen J, Matthews KS. 1994. Subunit dissociation affects DNA binding in a dimeric Lac repressor produced by C-terminal deletion. *Biochemistry* 33(29):8728–8735.
- Curado-Carballada C, Feixas F, Osuna S. 2019. Molecular dynamics simulations on *Aspergillus niger* monoamine oxidase: conformational dynamics and inter-monomer communication essential for its efficient catalysis. *Adv Synth Catal.* 27:30391.
- Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: an N ·log(N) method for Ewald sums in large systems. *EMBO J.* 98(12):10089–10092.
- Dimas RP, Jiang X-L, La Alberto de Paz J, Morcos F, Chan CTY. 2019. Engineering repressors with coevolutionary cues facilitates toggle switches with a master reset. *Nucleic Acids Res.* 47(10):5449–5463.
- Flynn TC, Swint-Kruse L, Kong Y, Booth C, Matthews KS, Ma J. 2009. Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. *Protein Sci.* 12(11):2523–2541.
- Gerek ZN, Ozkan SB. 2010. A flexible docking scheme to explore the binding selectivity of PDZ domains. *Protein Sci.* 19(5):914–928.
- Gerek ZN, Ozkan SB. 2011. Change in allosteric network affects binding affinities of PDZ domains. Analysis through perturbation response scanning. *PLoS Comput Biol.* 7(10):e1002154.
- Guo J, Pang X, Zhou H-X. 2015. Two pathways mediate inter-domain allosteric regulation in Pin1. *Structure* 23(1):237–247.
- Ha J-H, Spolar RS, Record MT. 1989. Role of the hydrophobic effect in stability of site-specific protein–DNA complexes. *J Mol Biol.* 209(4):801–816.
- Hacisuleyman A, Erman B. 2017. Entropy transfer between residue pairs and allostery in proteins: quantifying allosteric communication in ubiquitin. *PLoS Comput Biol.* 13(1):e1005319.
- Hilsen VJ. 2010. An ensemble view of allostery. *Science* 327(5966):653–654.
- Hodges AM, Fenton AW, Dougherty LL, Overholt AC, Swint-Kruse L. 2018. RheoScale: a tool to aggregate and quantify experimentally determined substitution outcomes for multiple variants at individual protein positions. *Hum Mutat.* 39(12):1814–1826.
- Jiménez-Osés G, Osuna S, Gao X, Sawaya MR, Gilson L, Collier SJ, Huisman GW, Yeates TO, Tang Y, Houk KN. 2014. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat Chem Biol.* 10(6):431–436.
- Kalodimos CG, Biris N, Bonvin AMJJ, Levandoski MM, Guennegues M, Boelens R, Kaptein R. 2004. Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science* 305(5682):386–389.
- Kalodimos CG, Boelens R, Kaptein R. 2002. A residue-specific view of the association and dissociation pathway in protein–DNA recognition. *Nat Struct Biol.* 9(3):193–197.
- Kalodimos 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(12):2866–2876.
- Kalodimos CG, Folkers GE, Boelens R, Kaptein R. 2001. Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. *Proc Natl Acad Sci U S A.* 98(11):6039–6044.
- Kamberaj H, van der Vaart A. 2009. Extracting the causality of correlated motions from molecular dynamics simulations. *Biophys J.* 97(6):1747–1755.
- Kim H, Zou T, Modi C, Dörner K, Grunkemeyer TJ, Chen L, Fromme R, Matz MV, Ozkan SB, Wachter RM. 2015. A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins. *Structure* 23(1):34–43.
- Kumar A, Glembo TJ, Ozkan SB. 2015. The Role of conformational dynamics and allostery in the disease development of human ferritin. *Biophys J.* 109(6):1273–1281.
- Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipowski AJ. 2009. Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. *Genome Res.* 19(9):1562–1569.
- LeVine MV, Weinstein H. 2014. NBIT—a new information theory-based analysis of allosteric mechanisms reveals residues that underlie function in the leucine transporter LeuT. *PLoS Comput Biol.* 10(5):e1003603.
- 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(5253):1247–1254.
- MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616.
- 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(5):421–433.
- Martin TA, Wu T, Tang Q, Dougherty LL, Parente DJ, Swint-Kruse L, Fenton AW. 2020. Identification of biochemically neutral positions in liver pyruvate kinase. *Proteins* 88(10):1340–1350.
- Meinhardt S, Manley MW, Parente DJ, Swint-Kruse L. 2013. Rheostats and toggle switches for modulating protein function. *PLoS One* 8(12):e83502.
- Meinhardt S, Swint-Kruse L. 2008. Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: bioinformatics-based predictions generate true positives and false negatives. *Proteins* 73(4):941–957.
- Miller M, Bromberg Y, Swint-Kruse L. 2017. Computational predictors fail to identify amino acid substitution effects at rheostat positions. *Sci Rep.* 7(1):41329.
- Miller M, Vitale D, Kahn PC, Rost B, Bromberg Y. 2019. funtrp: identifying protein positions for variation driven functional tuning. *Nucleic Acids Res.* 47(21):e142.
- Modi T, Huihui J, Ghosh K, Ozkan SB. 2018. Ancient thioredoxins evolved to modern-day stability–function requirement by altering native state ensemble. *Philos Trans R Soc Lond B Biol Sci.* 373(1749):20170184.
- Modi T, Ozkan SB. 2018. Mutations utilize dynamic allostery to confer resistance in TEM-1 β -lactamase. *Int J Mol Sci.* 19(12):3808.
- 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(2):487–499.
- Nevin Gerek Z, Kumar S, Banu Ozkan S. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evol Appl.* 6(3):423–433.
- Parente DJ, Ray JC, Swint-Kruse L. 2015. Amino acid positions subject to multiple co-evolutionary constraints can be robustly identified by their eigenvector network centrality scores. *Proteins* 83(12):2293–2306.

- 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(12):e84398.
- Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ. 2018. Dynamic allostery can drive cold adaptation in enzymes. *Nature* 558(7709):324–328.
- Schreiber. 2000. Measuring information transfer. *Phys Rev Lett*. 85(2):461–464.
- Schrödinger LLC. 2010. The PyMOL molecular graphics system, version 2.0.
- Sousa FL, Parente DJ, Shis DL, Hessman JA, Chazelle A, Bennett MR, Teichmann SA, Swint-Kruse L. 2016. AlloRep: a repository of sequence, structural and mutagenesis data for the LacI/GalR transcription regulators. *J Mol Biol*. 428(4):671–678.
- Spolar RS, Record MT. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263(5148):777–784.
- Spronk CAEM, Slijper M, van Boom JH, Kaptein R, Boelens R. 1996. Formation of the hinge helix in the Lac repressor is induced upon binding to the lac operator. *Nat Struct Mol Biol*. 3(11):916–919.
- Subramanian S, Kumar S. 2006. Evolutionary anatomies of positions and types of disease-associated and neutral amino acid mutations in the human genome. *BMC Genomics* 7(1):306.
- Suckow J, Markiewicz P, Kleina LG, Miller J, Kisters-Woike B, Müller-Hill B. 1996. Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *J Mol Biol*. 261(4):509–523.
- Swint-Kruse L. 2004. Using networks to identify fine structural differences between functionally distinct protein states. *Biochemistry* 43(34):10886–10895.
- Swint-Kruse L. 2016. Using evolution to guide protein engineering: the devil is in the details. *Biophys J*. 111(1):10–18.
- Swint-Kruse L, Brown CS. 2005. Resmap: automated representation of macromolecular interfaces as two-dimensional networks. *Bioinformatics* 21(15):3327–3328.
- Swint-Kruse L, Elam CR, Lin JW, Wycuff DR, Shive Matthews K. 2001. Plasticity of quaternary structure: twenty-two ways to form a LacI dimer. *Protein Sci*. 10(2):262–276.
- Swint-Kruse L, Larson C, Pettitt BM, Matthews KS. 2002. Fine-tuning function: correlation of hinge domain interactions with functional distinctions between LacI and PurR. *Protein Sci*. 11(4):778–794.
- Swint-Kruse L, Matthews KS. 2009. Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol*. 12(2):129–137.
- Swint-Kruse L, Matthews KS, Smith PE, Pettitt BM. 1998. Comparison of simulated and experimentally determined dynamics for a variant of the LacI DNA-binding domain, Nlac-P. *Biophys J*. 74(1):413–421.
- Swint-Kruse L, Zhan H, Matthews KS. 2005. Integrated insights from simulation, experiment, and mutational analysis yield new details of LacI function. *Biochemistry* 44(33):11201–11213.
- Taraban M, Zhan H, Whitten AE, Langley DB, Matthews KS, Swint-Kruse L, Trewheella J. 2008. Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. *J Mol Biol*. 376(2):466–481.
- Townsend PD, Rodgers TL, Glover LC, Korhonen HJ, Richards SA, Colwell LJ, Pohl E, Wilson MR, Hodgson DRW, McLeish TCB, et al. 2015. The role of protein–ligand contacts in allosteric regulation of the *Escherichia coli* catabolite activator protein. *J Biol Chem*. 290(36):22225–22235.
- 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(4):785–802.
- Tungtur S, Parente DJ, Swint-Kruse L. 2011. Functionally important positions can comprise the majority of a protein’s architecture. *Proteins* 79(5):1589–1608.
- Tungtur S, Schwingen KM, Riepe JJ, Weeramange CJ, Swint-Kruse L. 2019. Homolog comparisons further reconcile in vitro and in vivo correlations of protein activities by revealing over-looked physiological factors. *Protein Sci*. 28(10):1806–1818.
- Wade-Jardetzky N, Bray RP, Conover WW, Jardetzky O, Geisler N, Weber K. 1979. Differential mobility of the N-terminal headpiece in the lac-repressor protein. *J Mol Biol*. 128(2):259–264.
- Ward JH. 1963. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*. 58(301):236–244.
- Wilding M, Hong N, Spence M, Buckle AM, Jackson CJ. 2019. Protein engineering: the potential of remote mutations. *Biochem Soc Trans*. 47(2):701–711.
- Zhan H, Swint-Kruse L, Matthews KS. 2006. Extrinsic interactions dominate helical propensity in coupled binding and folding of the lactose repressor protein hinge helix. *Biochemistry* 45(18):5896–5906.
- Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. 2015. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol Biol Evol*. 32(1):132–143.