Cell Systems

CellPress

Perspective

Engineering allosteric transcription factors guided by the LacI topology

Ashley N. Hersey,[1] Valerie E. Kay,[1] Sumin Lee,[1] Matthew J. Realff,[1] and Corey J. Wilson[1,*]
[1]Georgia Institute of Technology, School of Chemical & Biomolecular Engineering, Atlanta, GA, USA
*Correspondence: corey.wilson@chbe.gatech.edu
https://doi.org/10.1016/j.cels.2023.04.008

SUMMARY

Allosteric transcription factors (aTFs) are used in a myriad of processes throughout biology and biotechnology. aTFs have served as the workhorses for developments in synthetic biology, fundamental research, and protein manufacturing. One of the most utilized TFs is the lactose repressor (LacI). In addition to being an exceptional tool for gene regulation, LacI has also served as an outstanding model system for understanding allosteric communication. In this perspective, we will use the LacI TF as the principal exemplar for engineering alternate functions related to allostery—i.e., alternate protein DNA interactions, alternate protein-ligand interactions, and alternate phenotypic mechanisms. In addition, we will summarize the design rules and heuristics for each design goal and demonstrate how the resulting design rules and heuristics can be extrapolated to engineer other aTFs with a similar topology—i.e., from the broader LacI/GalR family of TFs.

INTRODUCTION

Allostery, as introduced by Jacob, Changeux, and Monod, describes a connection between the distal surfaces of a protein, which serves to couple the protein function based on environmental signals.[1] Allosteric communication between these functional surfaces commonly involves a conformational change of a protein on interaction with a specific small molecule input. Allostery can be observed in essential biological processes such as catalysis,[2,3] signal transduction,[4] and gene regulation.[5,6] Transcription factors (TFs) are essential to a myriad of processes involved in gene regulation. The term TF covers an expansive category of DNA-binding proteins with diverse mechanisms and structures. TFs can be broadly binned by mechanisms namely activators and repressors, where activators influence gene expression by enhancing the binding affinity of RNA polymerase and repressors disrupt RNA polymerase function.

A subset of activators and repressors, called allosteric TFs (aTFs), utilize allosteric communication to regulate interaction with specific elements of DNA. Objectively, aTFs are decision-making modules—i.e., aTFs can convert an environmental input into a gene expression (or non-coding) output. Their function is achieved when an allosteric network of residues transmits a signal from the ligand-sensing surface to the DNA-binding surface within the protein to modulate DNA affinity. aTF environmental cues include a myriad of inputs—e.g., small molecules,[1] protein interactions,[7] light,[8] and phosphorylation.[9] As evidenced by the broad range of inputs, aTFs have an ever-expanding role in fundamental research, biological engineering, and industrial applications.

The *Escherichia coli* (*E. coli*) lac repressor (LacI) is a canonical aTF (Figure 1A), which will be used as the principal exemplar of engineered functions throughout this perspective. LacI has been used in a broad range of technologies—i.e., synthetic biology systems, biomanufacturing, and medicine—in addition to serving as a model system for fundamental research of allostery and related functions. Despite the extensive use of LacI and similar aTFs in biotechnology applications, there remains a significant barrier to unlocking the full potential of engineered systems. Despite decades of extensive studies, we still do not know how to generally engineer aTFs or simple allostery *de novo*. In this perspective, we will review the design rules for aTFs through the lens of the LacI and paint a picture for the future of engineering allostery within this topology.

A MODEL SYSTEM FOR THE DESIGN OF aTFs

LacI is an excellent model system for developing workflows for designing, building, and testing bespoke aTFs, with similar topologies. Namely, the LacI is well studied structurally and functionally—i.e., the LacI TF has: (1) solved crystals structures with and without DNA and ligand,[10,11] (2) a known panoply of effector ligands,[12] (3) a defined DNA operator in terms of sequence and length,[13] and (4) been the subject of partial mutational scanning identifying phenotypic consequences of single mutations throughout the protein.[14]

LacI achieves genetic regulation through allosteric communication via two functional surfaces—i.e., (1) a DNA-sensing surface and (2) a ligand-sensing surface. LacI is represented in Figure 1A as a homodimer, with two major functional domains connected by a linker. The first approximately 41 N-terminal residues of each monomer contain the DNA-binding site and are commonly referred to as the DNA-binding domain (DBD), see Figure 1B. The regulatory core domain (RCD) is composed of

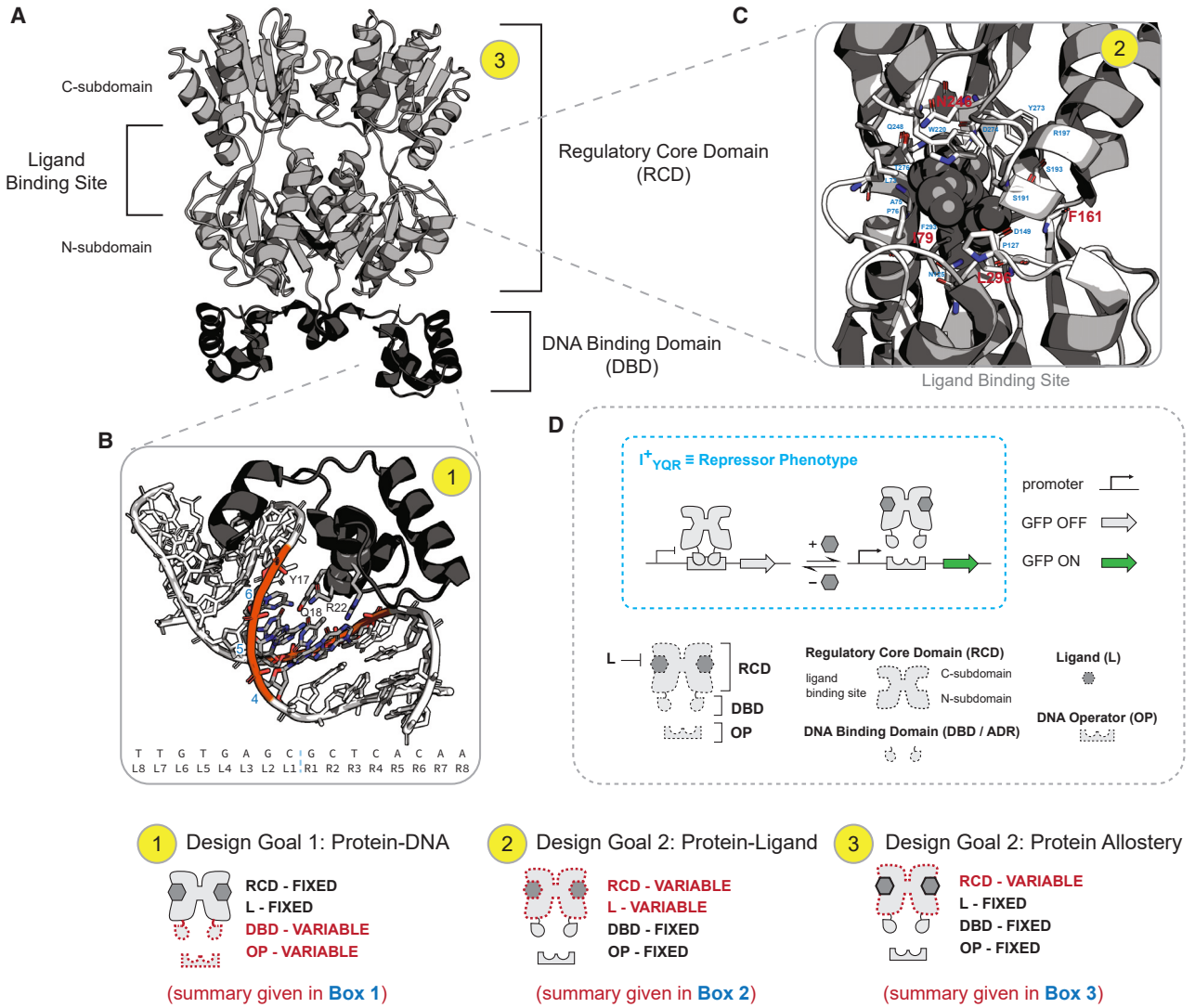

Figure 1. Structure of the lac repressor and design goals

(A) Crystal structure of the dimeric lac repressor (PDB: 1EFA). This aTF consists of two major domains: (1) the RCD that interacts with a cognate ligand (IPTG) at one functional surface and (2) a DBD which interacts with the DNA operator at a second distal functional surface. The RCD can be decomposed into C- and N-subdomains, which flank and comprise the ligand-binding site. The allosteric routes within this topology are hypothesized to exist between the two functional surfaces. Design goal 3 is focused on engineering alternate allosteric communication via a prescribed workflow that requires a super-repressor mutation—see Box 3.

(B) A model of a single LacI DBD bound to the cognate DNA operator. The DBD can be defined as amino acid positions 1–60 in the primary structure and naturally binds to the DNA operator given in the inset. Design goal 1 is to engineer alternate DNA binding by varying the DBD and cognate DNA operator at fixed positions concurrently—see Box 1. To confer alternate DNA recognition, amino acids at positions Y17, Q18, and R22 are modified to bind to synthetic DNA operators. The given residues bind to operator base pairs 4, 5, and 6, respectively.

(C) Model of the LacI ligand-binding site bound to cognate ligand (IPTG). Positions I79, F161, N246, and L296 are purported to confer ligand specificity. Design goal 2 is focused on engineering alternate ligand specificity by varying residues that contact IPTG in addition to residues in proximity to the binding pocket—see Box 2.

(D) Cartoon representation of the components of the lac repressor (abbreviated as I^{+}_{YQR}) in its genetic context. The regulatory core domain (RCD) interacts with ligand (L), while the DNA-binding domain (DBD) interacts with the DNA operator (OP). In this example, the LacI RCD (I) paired with its natural DBD (YQR) binds to the operator DNA in the absence of ligand (IPTG), blocking transcription and turning GFP expression off. In the presence of IPTG, I^{+}_{YQR} unbinds the operator, turning GFP expression on. Note the superscript "+" denotes the repressor phenotype.

residues 62–333 and contains the dimerization interface as well as one ligand-binding site per monomer. The ligand-binding site senses LacI's natural inducer allolactose or the non-hydrolyzable analog isopropyl-ß-D-1-thiogalactopyranoside (IPTG), see Figure 1C. DBD and RCD are connected by a linker region, also called the hinge or hinge helix, that becomes disordered on IPTG binding and leads to reduced DNA affinity.[15] Generically, the structure of LacI is representative of the LacI family of regulatory proteins.[5] The general function of LacI (abbreviated as I^{+}_{YQR}) regulating the production of green fluorescent protein (GFP) is given in Figure 1D. A clarifying note on LacI quaternary structure—the wild-type LacI is a homo-tetramer (or dimer of

dimers).[16] However, the minimal functional unit is a homodimer[17] and will represent the principal structure of our engineered systems throughout this perspective. The tetrameric form of LacI can bind two DNA sites in proximity to one another concurrently. The resulting DNA looping facilitates efficient regulation of the natural lac operon.[18] However, DNA looping is largely an unnecessary factor to consider when designing a synthetic circuit. Accordingly, from an engineering vantage point, we can regard the basic functional unit of LacI as a single dimer used to bind a single operator—i.e., regulating a single promoter. Therefore, the general structure of LacI illustrated in Figure 1A is represented as a dimer and does not include the C-terminal tetramerization domain as it is not required for the regulation of typical engineered systems.

In this perspective, we will use LacI as a model system to illustrate the successes and challenges in the space of aTF design (engineering) as we discuss efforts in the field that each aimed to engineer a single characteristic of LacI: (1) DNA specificity, (2) ligand specificity, and (3) functional phenotype—i.e., alternate allosteric communication. In turn, we will leverage the lessons learned from these exemplars toward the design, build, and test of new systems. In addition to specifying design rules and heuristics, we will discuss what will be needed for the *de novo* design of aTFs moving forward.

Engineering alternate LacI DNA binding

In a series of studies, Muller-Hill et al. were the first to alter the DNA specificity of LacI through selective mutations of residues in the DNA-binding domain. The authors observed that single mutations in the DNA recognition helix (amino acid residues 17–26) were not sufficient to alter specificity significantly.[19] However, combining at least two mutations at Y17, Q18, or R22 resulted in LacI repressors with affinity to single or double base pair-substituted (fully symmetric) lac operators. The authors suggested that residues at positions 17 and 18 interact with base pairs 4 and 5 of the operator, whereas 22 interacts with base pair 6 of the operator,[20,21] see Figure 1B.

In a later study, Milk et al. expanded on the observations made by Muller-Hill et al., with the goal of developing general design rules for alternate LacI DNA binding.[22] Milk et al. posited that DBD positions 17, 18, and 22 (noting that the wild-type residues Y17, Q18, and R22 are abbreviated as YQR) could be concurrently varied and paired with one or more putative symmetric operator DNA variant(s) with substitutions at positions 4, 5, and 6, whereas a putative DNA operator variant is defined as 5′-A ATT **XXX** AGC GCT **YYY** AAT T-3′ where "X" is any nucleotide and "Y" represents the nucleotides required to make the operator fully symmetric, see Figure 1B. The resulting combinatorial space for the DBD was defined by 20^3 (8,000) protein variants, whereas the operator combinatorial space resulted in 4^3 (64) DNA elements—for a total of $\sim 10^5$ putative protein DNA pairs. Milk et al. built and tested the given combinatorial space and determined that mutations at the three amino acid positions were sufficient to confer inducible DNA specificity to 26 of the 64 operators, the putative phenotype given in Figure 1D. Moreover, only 195 (of the 8,000 possible DBD protein variants) were found to bind one or more of the identified DNA operator variants. However, the authors reported few observable trends (rules) borne out of these results. Namely, there were no trends in amino acid frequencies at each position based on hydrophobicity, molecular weight, or isoelectric point. The repressor mutants (even among those that bound the same operator) differed in DNA affinity, dynamic range, leakiness, and promiscuity. The only universal rule observed was arginine was present at position 22 of the repressor when base pair 6 was guanine. Through these studies, the partial DNA-binding site of LacI was validated as positions 17, 18, and 22. The given positions were identified as important for DNA affinity because mutating one or more of the positions away from Y17, Q18, or R22 resulted in changes in expression levels of a reporter gene controlled by the symmetric lac operator regulating a given promoter compared with wild type. These positions appear to be important for DNA specificity in that mutations also changed expression levels of genes controlled by base pair-substituted DNA operators that regulate a given promoter. It was also concluded that DNA-binding residues can be isolated and engineered separately—i.e., DBD positions 17 and 18 bind base pairs 4 and 5, whereas position 22 binds base pair 6. Moreover, mutations not only change specificity but also change other protein characteristics like binding promiscuity and leakiness—although unpredictably.

The design heuristics and rules for LacI protein-DNA interactions largely end here. In summary, the data gathered from a randomized library of the partial DNA-binding domain and cognate operators provide minimal steadfast design rules—however, heuristics can be gleaned. From this study, we can conclude that protein-DNA specificity is a complex problem in the LacI aTF topology involving more than a pairwise ruleset—likely requiring other factors such as allosteric communication and related structural feedbacks. We have summarized the putative design rules and heuristics for engineering LacI protein-DNA interactions in Box 1.

Engineering alternate ligand specificity in the LacI topology

Barkley et al. demonstrated that wild-type LacI is responsive to several alternate ligands over four decades ago.[12] Accordingly, redesigning the binding pocket of LacI to respond to alternate ligands with a similar size and structure seems plausible. To test the extent to which LacI can be engineered to accommodate alternate ligands, Taylor et al. used a computational protein design to redesign the binding pocket of the LacI to be allosterically responsive to alternate ligands,[23] see Figure 1C. Taylor et al. posited that a canonical redesign of the binding pocket (and regions proximal thereof) would be sufficient to change the LacI ligand-binding function from the cognate ligand IPTG (an analog of the native ligand allolactose) to an alternate ligand with a similar structure and size—i.e., fucose, lactitol, or sucralose. Although the computational protein design approach resulted in variants of LacI that could be induced by the given alternate ligands, nearly all of the tested variants retained some sensitivity to IPTG. The most responsive LacI variants often contained four or more mutations; however, several single-mutant variants were observed. In addition, substitutions at positions 79 (a putative effector binding position) and 291 (proximal to the binding pocket) were frequently observed in variants that responded to new inducers, suggesting that these residues

Box 1. Engineering alternate LacI DNA binding—design rules and heuristics

Observation 1: DNA-binding domain amino-acid positions 17, 18, and 22 are critical to introducing variation in specificity, and are coupled to complementary changes to DNA operator positions 4, 5, and 6.
Observation 2: in addition to protein-DNA specificity, mutations also affect promiscuity, leakiness, and dynamic range—without any degree of predictability.
Supposition: this pairing constitutes a design heuristic opposed to a design rule in that precise interactions cannot be predicted from the minimal dataset—intimating that other factors contribute to this function, possibly allosteric communication.

might determine the ligand specificity of the binding pocket to this particular set of ligands.

In addition to computer-generated LacI variants, in a separate objective, Taylor et al. evaluated single-position variants to assess the capacity to change the ligand specificity of LacI to be inducible to gentiobiose—which has a similar structure and size to the native ligand allolactose. This study revealed that changes outside of the binding pocket were sufficient to alter ligand binding. In addition to a given point mutation to the binding pocket (i.e., Q291H), a point mutation to the dimerization interface (i.e., R255H) or DBD (i.e., V20A) resulted in responsiveness to gentiobiose. Congruent with the computational protein design variants the single-mutant variants were still responsive to IPTG. Taylor et al. posited that specificity to gentiobiose could be improved via the introduction of additional compensatory mutations—thus focused on improving the specificity of the Q291H variant as an exemplar. On analyzing the results, we noted that the authors utilized a super-repressor mutation[24] (also see Figure 2B), which mitigates the LacI response to IPTG. Although the authors tested the variant with compensatory mutations for reduced sensitivity to IPTG and induction via gentiobiose, Taylor et al. did not test the responsiveness to allolactose—which is closer to the structure of gentiobiose—thus, the result was not necessarily conclusive.

In a separate study, Wu et al. engineered a LacI variant to be responsive to lactulose.[25] Briefly, the authors exclusively targeted residues in the LacI-binding pocket (i.e., I79, F161, N246, and L296), also see Figure 1C. The resulting library consisted of all possible combinations of mutations (or 10^6 variants). Once the mutant library was constructed, the authors conducted (1) a positive screen in the presence of 10 mM lactulose and (2) a negative screen in the presence of 10 mM lactose or in the absence of an inducer. Note: the authors demonstrated that wild-type LacI was not responsive to lactulose. After several rounds of positive and negative screening, the authors identified 8 LacI variants that were responsive to lactulose. However, only one variant LacI-L5 (harboring one amino acid substitution, F161K) displayed reasonable performance metrics with a ~2.7-fold induction of GFP expression in response to 10 mM lactulose. In addition, the LacI-L5 variant had specificity to lactulose and was not responsive to other disaccharides tested (i.e., lactose, epilactose, maltose, sucrose, cellobiose, and melibiose). However, the LacI-L5 variant remained responsive to IPTG with a similar dynamic range to wild-type LacI.

These studies demonstrate that mutation of ligand-binding residues alone does not result in new ligand specificity; instead, the given engineering strategies result in aTFs with promiscuous binding to cognate ligands. However, a combination of randomly generated compensatory mutations along with an allosteric block (i.e., one or more super-repressor mutations—abbreviated as $X^S = I^S$, also see Figures 2A and 2B and inset (1)) can potentially result in binding specificity. We intimate that establishing an entirely new allosteric function in a given aTF (i.e., to support new ligand specificity) will likely require consideration of the allosteric network as a whole, not just the immediate functional surface. We summarize the observations and lessons learned from this section in Box 2.

Engineering alternate allosteric communication in the LacI topology

The antithetical phenotype to the LacI—i.e., anti-lac ($X^A_{ADR} = I^A_{YQR}$), also see Figure 2C—was initially observed over four decades ago achieved via simple mutagenesis.[14,26,27] In addition, more recent studies have reaffirmed the utility of mutagenesis (in many cases via multiple random mutations) to confer anti-repression in the LacI scaffold.[28–30] Although these studies clearly illustrate that the anti-lac phenotype can be achieved through a variety of mutations, no steadfast design rules can be directly credited to any one of the aforementioned.

In a set of studies, Wilson et al. developed an engineering workflow to confer alternate allosteric communication in the LacI scaffold[31,32]—moving one step closer to the rational design of allosteric communication in the given topology. Briefly, the authors posited that allosteric communication in the LacI scaffold is plastic, and on strategic mutation away from the functional surfaces, alternate allosteric routes (i.e., communication between the two functional surfaces) could be activated. This thesis was predicated (in part) on the discovery by Suckow et al. of 569 point mutations (124 of 328 positions) throughout the LacI scaffold that resulted in a super-repressor (I^s) phenotype[14]—see Figures 1A and 2B. This number represents only a portion of the super-repressor point mutations at each position as the authors performed a partial mutational scan by changing each position to 12 or 13 of the 20 amino acids. Using data from this study paired with structural information, Meyer et al.[31] postulated that super-repressor variants could be classified into two broad categories—(1) ligand-binding disrupting variants (class I) or (2) allosteric-blocking variants (class II). The authors intimated that blocking allosteric communication without disrupting the ligand-binding surface—i.e., via one or more class II I^s variant(s)—could facilitate the engineering of alternate allosteric routes in the LacI topology. In other words, the authors aimed to preserve both the ligand-binding function and DNA-binding function while short-circuiting communication between the two functional surfaces. Once the block in allosteric communication was achieved, compensatory mutations were introduced (via error-prone PCR [EP-PCR]) between residues 62 and 330, restricting allosteric recovery to variation within the RCD, see

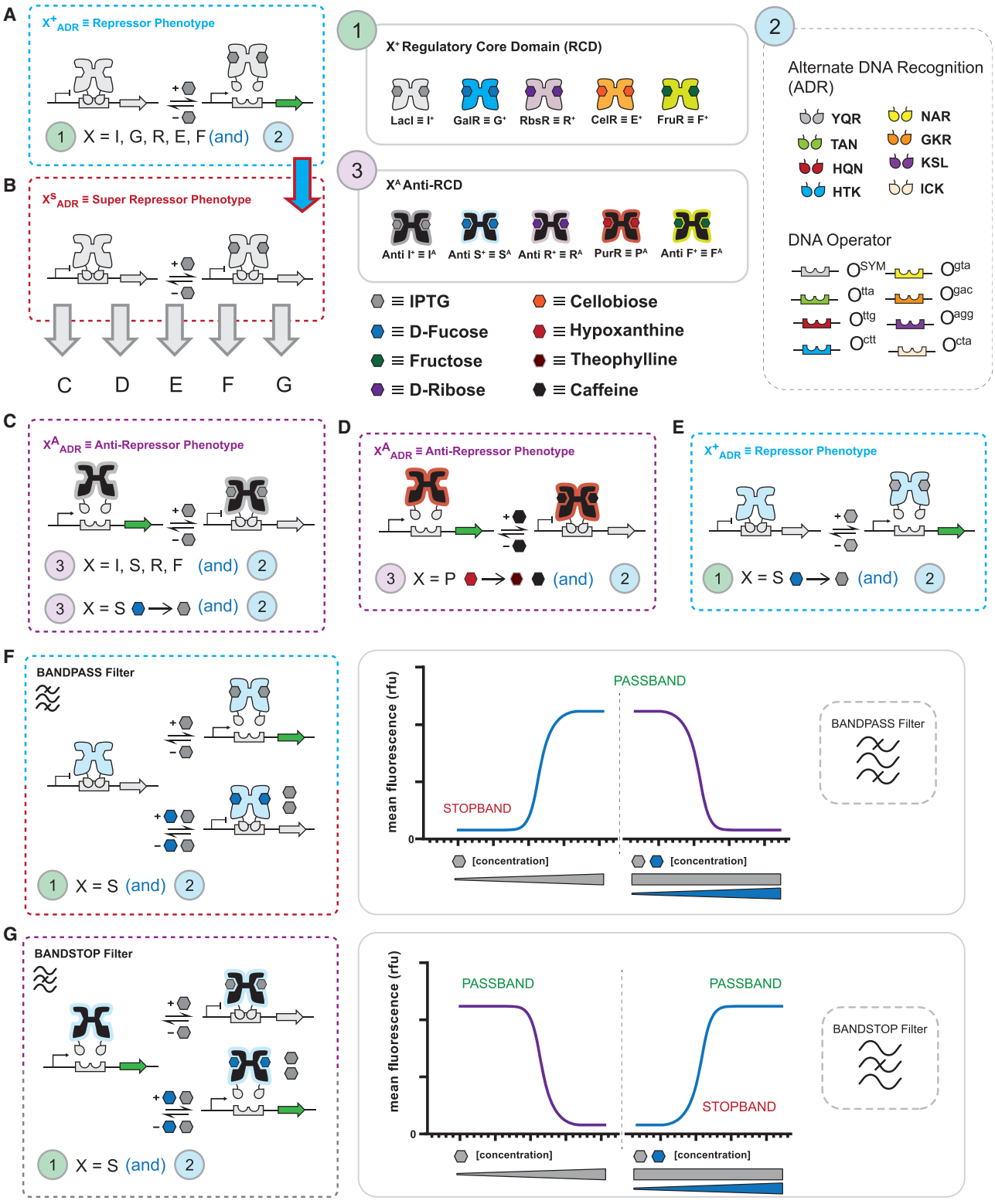

Figure 2. Workflows for engineering allosteric transcription factors

(A) General scheme for repressor phenotype and alternate DNA binding functions—abbreviated as X^{+}_{ADR}. X can be any of the RCD given in inset (1) and can be putatively paired with any of the alternate DNA recognition units given in inset (2).

(B) General mechanism of a class II super-repressor variant—abbreviated as X^{S}_{ADR}, where the superscript "S" denotes the phenotype. In this iteration of the super-repressor ligand binding and DNA binding functions are preserved; however, allosteric communication between the two functional surfaces is disrupted or blocked.

(legend continued on next page)

workflow given in Figures 2A–2C. In brief, random mutations beyond the super-repressor point mutation D88A resulted in restored LacI repression (I^{+}_{YQR}—see Figure 1D) and the identification of the anti-lac phenotype (see Figure 2C—where $X^{A}_{ADR} = I^{A}_{YQR}$) within a remarkably small search space of 10^{1}–10^{2} DNA sequences, with an average of 3 mutations per reading frame.

In a follow-up study, Richards et al. engineered anti-lacs by incorporating different super-repressor mutations from Suckow et al. into the workflow to demonstrate generalizability,[32] see Figures 2A–2C. The authors used four additional super-repressor variants to initially block allosteric communication—i.e., via class II point mutations K84A, V95A, V95F, and D275F (see Figures 2A and 2B). As with the previous study, compensatory mutations were introduced to this set of super-repressor variants via EP-PCR—with an average of 3 mutations per reading frame—and the resulting library was assessed by a black-white screen on solid media in the presence and absence of an inducer—validated via fluorescence-activated cell sorting (FACS) using GFP as the output modality. Eleven anti-lacs (see Figure 2C—where $X^{A}_{ADR} = I^{A}_{YQR}$) and three repressors were identified (again within a small search space of 10^{1}–10^{2} DNA sequences), each representing a unique solution to the design goal. Moreover, all resulting variants retained the initial super-repressor mutation after each round of evolution. Akin to the findings from Meyer et al., the additional substitutions varied widely in position and number, without changing the immediate ligand binding or DNA-binding sites. In most cases, several sets of compensatory mutations could confer alternate allosteric communication for a single super-repressor variant—however, no apparent mutational trends were observed within or between the sets of variants beyond the requirement of a super-repressor position.

We posited that using the workflows established by Wilson et al.[31,32] (illustration given in Figures 2A–2C), the search space for the identification of anti-lacs can be dramatically reduced. For example, Poelwijk et al. identified anti-repressor variants via the interrogation of a large library (10^{6}) of LacI variants generated via EP-PCR[28] with an average of 3.2 mutations per reading frame. In this study, the authors did not incorporate an initial super-repressor position but rather performed laboratory evolution on the wild-type scaffold. A postmortem assessment of the data by Richards et al.[32] revealed that a super-repressor variant was indeed observed in the Poelwijk et al. anti-lacs.

Reiterating this observation, in a more recent study, Tack et al.[29] generated 10^{5} LacI variants via EP-PCR with an average of 4.4 amino acid substitutions per variant. The results indicated that diverse substitutions could lead to the anti-lac phenotype; moreover, the authors estimated that approximately 0.35% of the measured library was composed of the given I^{A}_{YQR} phenotype. Notably, LacI amino acid substitutions associated with the anti-lac phenotype identified by Tack et al. contained super-repressor point mutations previously identified by Suckow et al.[14] Specifically, in the given collection of anti-lacs, 8 of the 10 positions analyzed by both groups are associated with the super-repressor phenotype via one or more point mutations that occur at positions 70, 84, 88, 96, 192, 200, 248, or 273. Noting that, Suckow et al. only conducted a partial mutational scan by changing each position to 12 or 13 of the 20 amino acids and restricted this analysis to positions 1–339, accordingly; the analysis is indeterminate for amino-acid positions 135 and 343.

In general, these observations demonstrate the power of the Wilson et al. workflow in reducing the search space from 10^{6} to 10^{2} variants via seeding the mutational process initially with a class II super-repressor mutation. The most significant design rule to be drawn from these studies is that blocking allostery appears to be critical to accessing evolutionary paths to anti-repression, echoing the importance of super-repressor mutations for switching ligand specificity—see Box 2. These exemplars show that there are many paths to alternate allostery in a single topology, given that allosteric blocks at many different locations throughout the scaffold can result in alternate allosteric communication on the addition of compensatory mutations via random mutagenesis. In addition to changing the phenotype, mutations that confer alternate allosteric communication can affect apparent ligand sensitivity and DNA affinity without changing residues in the ligand-binding site or the DBD. However, predicting how a given mutation (or set of mutations) will affect protein function is not clear from these studies. In Box 3, we articulate the putative design rules and heuristics from this section.

APPLYING THE LESSONS LEARNED FROM ENGINEERING LacI

In the following sections, we will demonstrate the application of the lessons learned from engineering individual functions of the LacI (illustrated in previous sections) and how we can expand and combine these putative workflows to engineer new TFs not observed in nature. Namely, we will:

(1) Describe studies and workflows in which the RCD and DBD are swappable, forming functional TF chimeras.
(2) Describe the process of engineering alternate allosteric communication in RCDs beyond the LacI TF.
(3) Describe how engineered RCDs can be combined with alternate DNA recognition (ADR) (i.e., alternate DNA-binding functions).
(4) Describe the process of engineering ligand binding, allosteric communication, and DNA-binding functions in a single scaffold.

(C) The terminal phenotype for engineered alternate allosteric communication—i.e., anti-repression abbreviated as X^{A}_{ADR}, where the superscript "A" denotes the phenotype. X can be any of the RCD given in inset (3) and can be putatively paired with any of the alternate DNA recognition units given in inset (2). In addition, the GalS RCD can be engineered to respond to an alternate ligand (i.e., IPTG).
(D) Summary of the workflow engineering PurR to responsive to alternate ligands caffeine and theophylline. This engineered aTF is also compatible with the ADR units given in inset (2).
(E) Illustration of the engineered GalS scaffold in which the ligand-binding function has been altered to be responsive to IPTG opposed to the cognate ligand D-fucose. The engineered transcription factor is also amenable to the ADR modular design, compatible with units given in inset (2).
(F) Illustration of the engineered BANDPASS function. The given BANDPASS functions are also compatible with the ADR modular design.
(G) Illustration of the antithetical operation to the BANDPASS—i.e., a BANDSTOP function. As with the BANDPASS operations, the BANDSTOP can be adapted with ADR functions.

Box 2. Engineering alternate ligand specificity in the LacI topology—design rules and heuristics

Observation 1: the computer-aided redesign of residues in the binding pocket of LacI are sufficient to confer changes to accommodate alternate ligands of similar size and structure to cognate ligands IPTG and allolactose. However, the redesign of the binding pocket to respond to alternate ligands often results in aTFs that retain sensitivity to native ligands.

Observation 2: mutations outside of the binding pocket can bestow alternate ligand binding—alone. In addition, compensatory mutations can potentially mitigate responsiveness to the cognate ligand IPTG; the given modifications appear to involve at least one super-repressor mutation, which is tantamount to a block in native allosteric communication. NOTE: super-repressor mutations are not restricted to the binding pocket.

Supposition: the redesign of a given ligand-binding pocket can involve more than modifications to the binding pocket—likely involving concurrent changes to allosteric communication. Although the concurrent modification of the binding pocket and allosteric route can result in alternate ligand binding, the specificity for alternate ligand binding alone is not conferred, although compensatory super-repressor mutations can potentially mitigate promiscuous ligand binding.

Furthermore, we will explore the impact of engineering the given functions concurrently and describe which of these characteristics can be engineered simultaneously as well as independently, and the impact on important functional properties—such as dynamic range, leakiness, ligand specificity, and sensitivity to the ligand.

Modularity of the regulatory core and DNA-binding functions

LacI is part of a larger family of proteins that share a similar topology and putative mechanism of action.[5] The LacI/GalR protein family is composed of over 1,000 homologs. Moreover, the LacI/GalR transcriptional regulatory proteins mediate responses to a wide range of environmental and metabolic changes. Structurally, the general LacI/GalR topology can be defined by two fundamental domains—i.e., (1) a RCD and (2) a DBD, see Figure 1D. Accordingly, we can regard this collection of paralogs as a putative design space—when carefully decomposed—positing that the given functional domains can be mixed and matched to form new aTFs.

To test this assertion, Meinhardt et al. posited that various RCDs—i.e., that bind different input ligands—could support the LacI DNA-binding function when paired properly.[33] In brief the authors selected nine RCDs—i.e., purine repressor (PurR), galactose repressor (GalR), galactose isorepressor (GalS), trehalose repressor (TreR), fructose repressor (FruR), ribose repressor (RbsR), cryptic asc operon repressor (AscG), cellobiose repressor (CelR), and cytidine repressor (CytR)—and replaced the native DBD with the LacI DNA-binding function (i.e., DBD = YQR). All but one showed inducible regulation of the lac promoter. The functional chimeric aTFs varied in affinity for DNA and dynamic range compared with each other—and relative to LacI. However, both characteristics were found to be tunable by mutating the non-conserved residues of the linker region—i.e., the region between the RCD and DBD. Other studies demonstrated that other native DBDs—beyond the LacI DBD—could function with other native RCD from the same paralog space.[34]

In turn, Rondon et al.[35] posited that a given RCD—from the collection tested by Meinhardt et al.—could support ADR using the design heuristics established by Milk et al. (summarized in Box 1). The authors selected five RCD—i.e., GalR, GalS, FruR, RbsR, and CelR—and systematically paired the given regulatory cores with seven ADR that were putatively orthogonal in terms of DNA-binding specificity. This resulted in 35 putative aTFs that were subsequently built and tested. From this design space, 27 (~80%) of the engineered aTFs displayed the correct qualitative repressor phenotype—i.e., were inducible by the appropriate cognate ligand, putative mechanism, and design space given in Figure 2A—also see Figure 2 insets (1) and (2).

In a separate study, Chure et al. introduced a series of point mutations in either the RCD or DBD of the LacI scaffold[36] (see Figures 1A and 1B, respectively). In turn, the authors characterized the dose response of the resulting variants and used simple models to predict how each mutation tunes the free energy at different inducer concentrations, repressor copy numbers, and DNA-binding strengths. The authors concluded that mutations in the DBD influence DNA-binding strength and that mutations within the RCD exclusively affect parameters that dictate the allosteric response, which is in good agreement with the apparent modularity between the RCD and DBD observed by Meinhardt et al. and Rondon et al.

These studies illustrate that the modular design of aTFs from the broader LacI/GalR family of proteins is possible—at least from a qualitative perspective. Unsurprisingly, features like the dynamic range cannot be predicted *a priori* from the aforementioned coarse-grain approaches; however, they are tunable—at least for engineered aTFs with mixed domains. Overall, these results successfully demonstrate design modularity. Although the majority of combinations of RCDs and DBDs are functional (~80%), a few chimeras do not result in functional aTFs. This observation implies that certain DBDs are only compatible with specific allosteric solutions.

Engineering alternate allosteric communication (beyond LacI) paired with alternate DNA binding

Groseclose et al.[37] posited that additional anti-repressors (beyond anti-lacs) could be engineered, using the workflows established in earlier studies.[31,32] Namely, the authors intimated that RbsR and FruR could be converted to anti-repressors using the design heuristics summarized in Box 3—workflow outlined in Figures 2A–2C. Thus, to engineer said anti-repressors, first, the DNA-binding function was normalized (fixed) to YQR—i.e., to the LacI DNA-binding function. Second, amino acids corresponding to super-repressor positions 84, 88, 95, and 96 equivalent to LacI were identified in FruR and RbsR via primary sequence alignment. The putative super-repressor positions were subjected to single-site saturation mutagenesis and

Box 3. Engineering alternate allosteric communication in the LacI topology—rules and heuristics

Observation 1: the LacI scaffold can support multiple alternate allosteric routes.
Observation 2: residues that are involved in the native allosteric route appear to have an important role in conferring alternate allosteric routes—specifically mutations that block or disrupt native allosteric communication.
Supposition: multiple allosteric routes can exist in a single topology. This observation intimates that allosteric communication design rules can be gleaned for systems that retain both (1) native ligand-binding and (2) DNA-binding functions. However, the putative design rules may only be sufficient to identify the phenotype. Quantitative properties such as dynamic range and sensitivity to ligand will likely require additional structural details. An emerging theme (heuristic) appears to be the role of positions that can block allosteric communication. In addition, modification to the allosteric route is inextricably link to other properties like dynamic range and sensitivity to the ligand.

phenotype screening via fluorescence measurements of a reporter gene controlled by the lac promoter. The variants with mitigated fluorescence independent of the inducer (i.e., variants that do not dissociate from DNA) were selected to serve as the super-repressor intermediates of the engineering workflow. Super-repressor mutations were found for both FruR and RbsR. Surprisingly, through screening for super-repressors, four single-point mutation RbsR anti-repressors emerged—in other words, putative positions in the allosteric pathway could be modified to confer single-mutation anti-repressors in the RbsR scaffold. However, FruR required additional compensatory mutations beyond the super-repressor point mutation—i.e., conforming to the canonical workflow. The discovery of single-mutation anti-repressors in the RbsR scaffold prompted the authors to reevaluate super-repressor positions in wild-type LacI to allow for variation of the amino acid. Indeed, two new single-mutation anti-lacs were identified—which required an elegant revision to the initial workflow to include the possibility of single-mutation anti-repressors within a given super-repressor position. All of the anti-repressors varied in dynamic range and leakiness both between scaffolds and within the same scaffold.

In turn, Wilson et al. demonstrated that in addition to engineering allosteric communication, alternate DNA binding could be concurrently engineered in a single scaffold in a series of studies.[37,38] This was first illustrated by Rondon and Wilson,[38] in which the authors tested eight alternate DNA-binding functions concurrently with nine disparate (engineered) anti-lac RCDs. Six of the eight putative ADRs resulted in orthogonal DNA binding, and 46 of the putative 54 engineered aTFs were functional. In addition, although all functional systems had the same phenotype and ligand specificity, there was significant variation in dynamic range and leakiness between the given TFs—intimating that allosteric communication can impact the apparent binding properties of an engineered system. Next, Groseclose et al. leveraged the design heuristics to confer alternate DNA binding in the anti-RbsR and anti-FruR scaffolds via the same ADR set with similar results.[37] These results successfully demonstrate that design heuristics and rules 1 (Box 1) and 3 (Box 3) can be combined, see Figure 2C.

Concurrently engineering alternate DNA binding, ligand binding, and allosteric communication

Rondon and Wilson posited that engineering alternate ligand binding could be generally applied to other members of the LacI/GalR family of TFs.[39] To demonstrate this supposition, Rondon and Wilson engineered an anti-repressor that responded to caffeine via the PurR scaffold—which is natively anti-induced by hypoxanthine, see Figure 2D. The workflow began with the identification of a class II super-repressor point mutation in the PurR scaffold to mitigate responsiveness to hypoxanthine—i.e., congruent with the lessons learned in Box 2, also see Figure 2B. Next, the authors introduced compensatory mutations via EP-PCR and screened for responsiveness to caffeine. Briefly, four variants were identified that responded to caffeine—with mitigated responsiveness to hypoxanthine; this result confirms the supposition summarized in Box 2. In addition, the authors repeated this workflow to identify PurR variants responsive to theophylline with mitigated response to hypoxanthine, see Figure 2D. Notably, an initial block in allosteric communication was required; screening of variants without an initial block did not yield any functional variants. In turn, Rondon and Wilson, posited that alternate ligand binding and alternate DNA binding could be concurrently engineered—i.e., in other words, design rules and heuristics 1 (Box 1) and 2 (Box 2) could be combined. Namely, the authors showed that the four engineered caffeine-responsive RCDs could be modularly paired with eight disparate ADR functions. In addition, two rounds of EP-PCR combined with site saturation of non-conserved linker region residues were used to improve the dynamic range of the given engineered aTFs.

In this final example, we will demonstrate the ability to concurrently engineer all three properties—i.e., alternate allosteric communication, alternate ligand binding, and alternate DNA binding—in a single scaffold.[40,41] The GalS has a native affinity for and response to D-fucose.[42] In work conducted by Groseclose et al., the GalS RCD paired with a LacI DNA-binding function was chosen as a scaffold to alter allosteric function, ligand response, and DNA specificity.[40,41] First, anti-repressor function was conferred in the GalS scaffold via the heuristics summarized in Box 3—workflow outlined in Figures 2A–2C. The workflow was adapted to screen for variants that were anti-induced by D-fucose (Figure 2C). Next, the authors aimed to demonstrate that the GalS repressor (S^{+}_{YQR}) and engineered anti-repressor ligand response could be changed from the cognate ligand (D-fucose) to IPTG (see Figures 2E and 2C, respectively) using design heuristics given in Box 2. Notably, each variant leveraged a block in allosteric communication (see Figures 2A and 2B) followed by one or more rounds of EP-PCR.

The authors concurrently evaluated the resulting repressor's (Figure 2F) and anti-repressor's (Figure 2G) responsiveness to D-fucose. In each case, the native ligand (D-fucose) did not induce or anti-induce the engineered systems, rather the native

ligand at high concentrations antagonized the response to IPTG—resulting in a BANDPASS operation (Figure 2F) or BANDSTOP operation (Figure 2G). Finally, the engineered aTFs were subjected to modifications to DNA-binding specificity through ADR via modular design, as described in Box 1. This final exemplar illustrates the ability to concurrently bring together design heuristics 1 (Box 1), 2 (Box 2), and 3 (Box 3), also see Figures 2A–2C and 2E–2G.

Alternatives to *a priori* design of aTFs

Although this perspective highlights many accomplishments—in addition to a means to organize and combine different workflows—it also highlights the gaps in our knowledge regarding the *a priori* design of aTFs. Although *in silico* tools like molecular dynamics (MD) simulations have been used to model putative allosteric routes in the LacI scaffold,[43] emerging machine-learning techniques and methods present an exciting path forward to better understand allosteric communication in the given scaffold. Namely, techniques used to find allosteric signatures have either been based on machine learning to extract the relevant features from experimental or computational data or on algorithms that propagate information along edges in a graph representation of the protein structure, Verkhivker et al. provide a comprehensive overview of recent approaches.[44]

Notably, machine-learning approaches paired with MD simulations can capture short-term fluctuations in protein side-chain configurations, and this approach has been used to find dynamic allostery using autoencoding machine-learning techniques.[45] This approach recognizes that in the bound and unbound state, the motion of the side chains on coupled residues will be different. An MD simulation time series is used to generate a 1,500-dimension vector of the distance between the center of mass of the side-chain atoms in each pair of residues in the two forms. An auto-encoder is trained on the unbound signatures for the residue pairs. The signatures from the bound pairs are processed by the network, and the signatures with the largest errors are found to correlate with known allosteric interactions. Zhou et al. demonstrate that the dynamic trajectories of distances between pairs of residue C-alpha atoms and the dihedral angles can be used as features of decision tree and neural network learning approaches.[46] This approach was able to identify specific residue pairs that were implicated in allostery by experimental observations, and the distances between the residues were well outside those assumed for contact, 29 Å. This work was extended to include additional features in Hayatshahi et al.[47]

Garruss et al. present the most relevant case study to this perspective via the convergence of deep mutational scanning with machine-learning approaches focused on LacI-mediated transcriptional repression.[48] Namely, the authors measured the transcriptional repression function of 43,669 LacI variants and develop a deep neural network to predict transcriptional repression mediated by LacI using experimental measurements of variant function. Using molecular simulation and alignment-based evolutionary methods, the authors could predict an unseen single-mutation effect; however, with only modest performance. Molecular simulation could distinguish some special cases that grossly affect protein stability but could not discriminate variant function generally. The authors noted that evolutionary analyses may have difficulty predicting the effect of deviations to a single protein's function given that many mutations arising from a synthetic library may not yet be part of the extant record—e.g., the aTF may have alternate ligand binding despite the overall sequence similarity. The authors intimate that a more effective prediction of mutational effect on basal repression will improve the understanding and design of bespoke aTFs.

CONCLUSIONS AND FUTURE DIRECTIONS

In this perspective, we summarize recent accomplishments with respect to engineering aTFs from the broader LacI/GalR family of TFs. In particular, we have identified three fundamental and complementary design heuristics for engineering DNA binding, ligand binding, and allosteric communication—individually and concurrently. This collection of design heuristics intimates the importance of allostery in all aspects of engineered systems. Highlighting the critical importance of knowing how and when to block allosteric communication will facilitate the engineering and discovery of new functions and functional systems.

Finally, we believe the lessons learned from engineering TFs that share the LacI topology can be extended to other aTFs beyond the immediate family of functional proteins. A critical requirement for engineering alternate allosteric communication between two or more functional surfaces in a particular scaffold will be the ability to systematically identify residues and positions that disrupt allosteric communication. Moreover, we posit that the ability to engineer allosteric communication will enable the bespoke design of additional functional properties including catalysis and iterations thereof. Although a discussion of how well the rules and heuristics developed in this perspective align with other aTFs outside of the LacI/GalR family is beyond the scope of this perspective, we have presented relevant discussions of how sequence-based statistical mapping can be used to glean such relationships elsewhere.[49]

ACKNOWLEDGMENTS

This work was supported by National Science Foundation grants GCR 1934836, MCB 2123855, MCB 1921061, CBET 1804639, MFB 2226663, MCB 2227598, and MCB 2225300 all awarded to C.J.W.

AUTHOR CONTRIBUTIONS

A.N.H. and C.J.W. wrote the manuscript with input from all the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Monod, J., Changeux, J.-P., and Jacob, F. (1963). Allosteric proteins and cellular control systems. J. Mol. Biol. *6*, 306–329. https://doi.org/10.1016/S0022-2836(63)80091-1.

2. Lisi, G.P., and Loria, J.P. (2017). Allostery in enzyme catalysis. Curr. Opin. Struct. Biol. *47*, 123–130. https://doi.org/10.1016/j.sbi.2017.08.002.

3. Gerhart, J.C., and Pardee, A.B. (1962). The enzymology of control by feedback inhibition. J. Biol. Chem. *237*, 891–896. https://doi.org/10.1016/S0021-9258(18)60389-8.

4. Nussinov, R., Tsai, C.-J., and Liu, J. (2014). Principles of allosteric interactions in cell signaling. J. Am. Chem. Soc. *136*, 17692–17701. https://doi.org/10.1021/ja510028c.

5. Swint-Kruse, L., and Matthews, K.S. (2009). Allostery in the LacI/GalR family: variations on a theme. Curr. Opin. Microbiol. *12*, 129–137. https://doi.org/10.1016/j.mib.2009.01.009.

6. Reichheld, S.E., Yu, Z., and Davidson, A.R. (2009). The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. Proc. Natl. Acad. Sci. USA *106*, 22263–22268. https://doi.org/10.1073/pnas.0911566106.

7. Göös, H., Kinnunen, M., Salokas, K., Tan, Z., Liu, X., Yadav, L., Zhang, Q., Wei, G.H., and Varjosalo, M. (2022). Human transcription factor protein interaction networks. Nat. Commun. *13*, 766. https://doi.org/10.1038/s41467-022-28341-5.

8. Wang, X., Chen, X., and Yang, Y. (2012). Spatiotemporal control of gene expression by a light-switchable transgene system. Nat. Methods *9*, 266–269. https://doi.org/10.1038/nmeth.1892.

9. Hunter, T., and Karin, M. (1992). The regulation of transcription by phosphorylation. Cell *70*, 375–387. https://doi.org/10.1016/0092-8674(92)90162-6.

10. Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science *271*, 1247–1254. https://doi.org/10.1126/science.271.5253.1247.

11. Friedman, A.M., Fischmann, T.O., and Steitz, T.A. (1995). Crystal Structure of lac Repressor Core tetramer and Its Implications for DNA Looping. Science *268*, 1721–1727. https://doi.org/10.1126/science.7792597.

12. Barkley, M.D., Riggs, A.D., Jobe, A., and Burgeois, S. (1975). Interaction of effecting ligands with lac repressor and repressor-operator complex. Biochemistry *14*, 1700–1712. https://doi.org/10.1021/bi00679a024.

13. Gilbert, W., and Maxam, A. (1973). The nucleotide sequence of the lac operator. Proc. Natl. Acad. Sci. USA *70*, 3581–3584. https://doi.org/10.1073/pnas.70.12.3581.

14. Suckow, J., Markiewicz, P., Kleina, L.G., Miller, J., Kisters-Woike, B., and 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*, 509–523. https://doi.org/10.1006/jmbi.1996.0479.

15. Spronk, C.A.E.M., Slijper, M., van Boom, J.H., Kaptein, R., and Boelens, R. (1996). Formation of the hinge helix in the lac represser is induced upon binding to the lac operator. Nat. Struct. Biol. *3*, 916–919. https://doi.org/10.1038/nsb1196-916.

16. Wilson, C.J., Zhan, H., Swint-Kruse, L., and Matthews, K.S. (2007). The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. Cell. Mol. Life Sci. *64*, 3–16. https://doi.org/10.1007/s00018-006-6296-z.

17. Daber, R., Sharp, K., and Lewis, M. (2009). One is not enough. J. Mol. Biol. *392*, 1133–1144. https://doi.org/10.1016/j.jmb.2009.07.050.

18. Krämer, H., Niemöller, M., Amouyal, M., Revet, B., von Wilcken-Bergmann, B., and Müller-Hill, B. (1987). lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. EMBO J. *6*, 1481–1491. https://doi.org/10.1002/j.1460-2075.1987.tb02390.x.

19. Lehming, N., Sartorius, J., Niemoller, M., Genenger, G., Wilcken-Bergmann, B., and Muller-Hill, B. (1987). The interaction of the recognition helix of lac repressor with lac operator. EMBO J. *6*, 3145–3153. https://doi.org/10.1002/j.1460-2075.1987.tb02625.x.

20. Sartorius, J., Lehming, N., Kisters, B., von Wilcken-Bergmann, B., and Müller-Hill, B. (1989). lac repressor mutants with double or triple exchanges in the recognition helix bind specifically to lac operator variants with multiple exchanges. EMBO J. *8*, 1265–1270. https://doi.org/10.1002/j.1460-2075.1989.tb03500.x.

21. Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B., and Müller-Hill, B. (1990). Mutant lac repressors with new specificities hint at rules for protein–DNA recognition. EMBO J. *9*, 615–621. https://doi.org/10.1002/j.1460-2075.1990.tb08153.x.

22. Milk, L., Daber, R., and Lewis, M. (2010). Functional rules for lac repressor-operator associations and implications for protein-DNA interactions. Protein Sci. *19*, 1162–1172. https://doi.org/10.1002/pro.389.

23. Taylor, N.D., Garruss, A.S., Moretti, R., Chan, S., Arbing, M.A., Cascio, D., Rogers, J.K., Isaacs, F.J., Kosuri, S., Baker, D., et al. (2016). Engineering an allosteric transcription factor to respond to new ligands. Nat. Methods *13*, 177–183. https://doi.org/10.1038/nmeth.3696.

24. Markiewicz, P., Kleina, L.G., Cruz, C., Ehret, S., and Miller, J.H. (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. https://doi.org/10.1006/jmbi.1994.1458.

25. Wu, J., Jiang, P., Chen, W., Xiong, D., Huang, L., Jia, J., Chen, Y., Jin, J.M., and Tang, S.Y. (2017). Design and application of a lactulose biosensor. Sci. Rep. *7*, 45994. https://doi.org/10.1038/srep45994.

26. Myers, G.L., and Sadler, J.R. (1971). Mutational inversion of control of the lactose operon of Escherichia coli. J. Mol. Biol. *58*, 1–28. https://doi.org/10.1016/0022-2836(71)90229-4.

27. Miller, J.H., and Schmeissner, U. (1979). Genetic studies of the lac repressor. X. Analysis of missense mutations in the lacI gene. J. Mol. Biol. *131*, 223–248. https://doi.org/10.1016/0022-2836(79)90074-3.

28. Poelwijk, F.J., de Vos, M.G.J., and Tans, S.J. (2011). Tradeoffs and optimality in the evolution of gene regulation. Cell *146*, 462–470. https://doi.org/10.1016/j.cell.2011.06.035.

29. Tack, D.S., Tonner, P.D., Pressman, A., Olson, N.D., Levy, S.F., Romantseva, E.F., Alperovich, N., Vasilyeva, O., and Ross, D. (2021). The genotype-phenotype landscape of an allosteric protein. Mol. Syst. Biol. *17*, e10179. https://doi.org/10.15252/msb.202010179.

30. Daber, R., Sochor, M.A., and Lewis, M. (2011). Thermodynamic analysis of mutant lac repressors. J. Mol. Biol. *409*, 76–87. https://doi.org/10.1016/j.jmb.2011.03.057.

31. Meyer, S., Ramot, R., Kishore Inampudi, K., Luo, B., Lin, C., Amere, S., and Wilson, C.J. (2013). Engineering alternate cooperative-communications in the lactose repressor protein scaffold. Protein Eng. Des. Sel. *26*, 433–443. https://doi.org/10.1093/protein/gzt013.

32. Richards, D.H., Meyer, S., and Wilson, C.J. (2017). Fourteen ways to reroute cooperative communication in the lactose repressor: engineering regulatory proteins with alternate repressive functions. ACS Synth. Biol. *6*, 6–12. https://doi.org/10.1021/acssynbio.6b00048.

33. Meinhardt, S., Manley, M.W., Jr., Becker, N.A., Hessman, J.A., Maher, L.J., 3rd, and Swint-Kruse, L. (2012). Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. Nucleic Acids Res. *40*, 11139–11154. https://doi.org/10.1093/nar/gks806.

34. Dimas, R.P., Jiang, X.-L., Alberto de la Paz, J., Morcos, F., and Chan, C.T.Y. (2019). Engineering repressors with coevolutionary cues facilitates toggle switches with a master reset. Nucleic Acids Res. *47*, 5449–5463. https://doi.org/10.1093/nar/gkz280.

35. Rondon, R.E., Groseclose, T.M., Short, A.E., and Wilson, C.J. (2019). Transcriptional programming using engineered systems of transcription factors and genetic architectures. Nat. Commun. *10*, 4784. https://doi.org/10.1038/s41467-019-12706-4.

36. Chure, G., Razo-Mejia, M., Belliveau, N.M., Einav, T., Kaczmarek, Z.A., Barnes, S.L., Lewis, M., and Phillips, R. (2019). Predictive shifts in free energy couple mutations to their phenotypic consequences. Proc. Natl. Acad. Sci. USA *116*, 18275–18284. https://doi.org/10.1073/pnas.1907869116.

37. Groseclose, T.M., Rondon, R.E., Herde, Z.D., Aldrete, C.A., and Wilson, C.J. (2020). Engineered systems of inducible anti-repressors for the next generation of biological programming. Nat. Commun. *11*, 4440. https://doi.org/10.1038/s41467-020-18302-1.

38. Rondon, R.E., and Wilson, C.J. (2019). Engineering a new class of anti-LacI transcription factors with alternate DNA recognition. ACS Synth. Biol. *8*, 307–317. https://doi.org/10.1021/acssynbio.8b00324.

39. Rondon, R., and Wilson, C.J. (2021). Engineering alternate ligand recognition in the PurR topology: A system of novel caffeine biosensing

transcriptional antirepressors. ACS Synth. Biol. *10*, 552–565. https://doi.org/10.1021/acssynbio.0c00582.

40. Groseclose, T.M., Rondon, R.E., Hersey, A.N., Milner, P.T., Kim, D., Zhang, F., Realff, M.J., and Wilson, C.J. (2021). Biomolecular systems engineering: unlocking the potential of engineered allostery via the lactose repressor topology. Annu. Rev. Biophys. *50*, 303–321. https://doi.org/10.1146/annurev-biophys-090820-101708.

41. Groseclose, T.M., Hersey, A.N., Huang, B.D., Realff, M.J., and Wilson, C.J. (2021). Biological signal processing filters via engineering allosteric transcription factors. Proc. Natl. Acad. Sci. USA *118*, e2111450118. https://doi.org/10.1073/pnas.2111450118.

42. Weickert, M.J., and Adhya, S. (1993). Control of transcription of gal repressor and isorepressor genes in Escherichia coli. J. Bacteriol. *175*, 251–258. https://doi.org/10.1128/jb.175.1.251-258.1993.

43. Flynn, T.C., Swint-Kruse, L., Kong, Y., Booth, C., Matthews, K.S., and Ma, J. (2003). Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. Protein Sci. *12*, 2523–2541. https://doi.org/10.1110/ps.03188303.

44. Verkhivker, G.M., Agajanian, S., Hu, G., and Tao, P. (2020). Allosteric regulation at the crossroads of new technologies: multiscale modeling, networks, and machine learning. Front. Mol. Biosci. *7*, 136. https://doi.org/10.3389/fmolb.2020.00136.

45. Tsuchiya, Y., Taneishi, K., and Yonezawa, Y. (2019). Autoencoder-based detection of dynamic allostery triggered by ligand binding based on Molecular Dynamics. J. Chem. Inf. Model. *59*, 4043–4051. https://doi.org/10.1021/acs.jcim.9b00426.

46. Zhou, H., Dong, Z., and Tao, P. (2018). Recognition of protein allosteric states and residues: machine learning approaches. J. Comp. Chem. *39*, 1481–1490. https://doi.org/10.1002/jcc.25218.

47. Hayatshahi, H.S., Ahuactzin, E., Tao, P., Wang, S., and Liu, J. (2019). Probing protein allostery as a residue-specific concept via residue response maps. J. Chem. Inf. Model. *59*, 4691–4705. https://doi.org/10.1021/acs.jcim.9b00447.

48. Garruss, A.S., Collins, K.M., and Church, G.M. (2021). Deep representation learning improves prediction of LacI-mediated transcriptional repression. Proc. Natl. Acad. Sci. USA *118*, e2022838118. https://doi.org/10.1073/pnas.2022838118.

49. Herde, Z.D., Short, A.E., Kay, V.E., Huang, B.D., Realff, M.J., and Wilson, C.J. (2020). Engineering allosteric communication. Curr. Opin. Struct. Biol. *63*, 115–122. https://doi.org/10.1016/j.sbi.2020.05.004.