Engineering allostery

Srivatsan Raman^{1,2*}, Noah Taylor^{1,2*}, Naomi Genuth^{1,2}, Stanley Fields³, and George M. Church^{1,2}

¹Wyss Institute for Biologically-Inspired Engineering, Harvard University, 3 Blackfan Circle, Boston, MA 02115, USA

² Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

³ Howard Hughes Medical Institute, Departments of Genome Sciences and Medicine, University of Washington, Seattle, WA 98195, USA

Allosteric proteins have great potential in synthetic biology, but our limited understanding of the molecular underpinnings of allostery has hindered the development of designer molecules, including transcription factors with new DNA-binding or ligand-binding specificities that respond appropriately to inducers. Such allosteric proteins could function as novel switches in complex circuits, metabolite sensors, or as orthogonal regulators for independent, inducible control of multiple genes. Advances in DNA synthesis and next-generation sequencing technologies have enabled the assessment of millions of mutants in a single experiment, providing new opportunities to study allostery. Using the classic Lacl protein as an example, we describe a genetic selection system using a bidirectional reporter to capture mutants in both allosteric states, allowing the positions most crucial for allostery to be identified. This approach is not limited to bacterial transcription factors, and could reveal new mechanistic insights and facilitate engineering of other major classes of allosteric proteins such as nuclear receptors, two-component systems, G proteincoupled receptors, and protein kinases.

Unlocking the power of allostery in synthetic biology

Allosteric regulation mediates virtually every biological process, including transcription, signal transduction, and enzyme activity and transport. Allostery can be broadly defined as activity at one site in a protein regulating function at a spatially distant site. Allosteric regulation occurs through an allosteric effector, generally a small molecule, which binds at one active site and triggers a conformational change that affects function at a distant site. Because of their ability to respond to small molecules by a change of state, allosteric proteins play an important role in synthetic biology. Nevertheless, our ability to engineer allosteric proteins is highly constrained by our limited understanding of the molecular details of allostery, and thus we have barely scratched the surface of how allosteric proteins can be applied in this emerging field.

Allosteric proteins are used as switches in synthetic circuits. Although synthetic biologists would like to build more complex circuits, a major limitation is the lack of

Corresponding author: Raman, S. (sraman@genetics.med.harvard.edu).

^{*}These authors contributed equally to this work.

0168-9525/

© 2014 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.tig.2014.09.004

orthogonal switches (allosteric proteins that bind to different inducers and different DNA sequences with little crosstalk). A suite of well-characterized orthogonal switches would vastly enhance our ability to build higher-order synthetic circuits with real-world applicability [1]. For example, such switches could serve as analog-to-digital converters that convert a continuous chemical gradient into a digital output. Bacteria possessing synthetic circuits combining many such analog-to-digital converters could then be used as whole-cell biosensors of the gut [1].

Allosteric proteins can also be used as in vivo metabolite sensors for engineering biosynthetic pathways [2]. These sensors detect and respond to the level of some sought-after metabolite, enabling genetic selections in which the best producers are identified from a large number of variant organisms. Despite an increasing demand for allosteric sensors detecting industrially useful chemicals, we are limited to the ligand-binding domains of known transcription factors; however, this bottleneck could be removed by designing new allosteric proteins. For instance, new small-molecule sensors could be generated from chimeras of a well-characterized DNA-binding domain with ligand-binding domains identified in the sequences of metagenomic samples. Alternatively, we might be able to mutate binding-site residues in an existing sensor to create new ligand specificities without affecting allosteric communication [3].

Besides their biotechnological applications, designer allosteric proteins can provide independent, temporal regulation of multiple genes, a useful tool for developmental biology. The Tet-on/off activator system, based on the *Escherichia coli* tetracycline resistance (Tet) repressor, is widely used for mammalian gene regulation, but it does not allow the independent control of multiple genes. With multiple orthogonal regulators similar to the Tet-on/off elements we could, for instance, gain exquisite control over stem cell differentiation pathways by modulating each differentiation factor independently.

Finally, redesigning allosteric proteins to respond to molecules that cross the blood-brain barrier would enable the activation of specific neural circuits in the brains of live animals simply by incorporating the inducers in the diet. To engineer allosteric proteins, however, we need to take a closer look at how allostery works at the molecular level.

Efforts to understand allostery have largely focused on biophysical models to explain the conformational transition between two states, corresponding to the presence and

Opinion



Figure 1. Engineering novel allosteric proteins. (A) Designing novel ligand specificities: a known allosteric protein can be made to recognize different small-molecule inducers by mutating the amino acids in the ligand-binding domain. (B) Designing novel DNA specificities: allosteric transcription factors can be targeted to new DNA sequences by altering the DNA-binding domain. (C) Novel allosteric chimeras: allosteric protein domains that are not capable of binding to DNA, such as periplasmic-binding proteins, can be attached to a DNA-binding domain to create novel. chimeric transcription factors.

absence of the effector [4]. Protein dynamics shows that allosteric transitions occur as a consequence of local conformational changes such as disorder or unfolding that are propagated to distal regions, shifting the overall conformational equilibrium [1,5-7]. Whereas these models focus on the thermodynamic drivers of allostery, little is understood about the underlying molecular basis. Alternatively, a genetics approach to understanding allostery could be taken by making large numbers of single mutations and combinations of double mutations in an allosteric protein and then determining which mutant proteins maintain allosteric coupling. This approach, in conjunction with biophysical measurements derived from nuclear magnetic resonance (NMR) or molecular dynamics simulations, may tell us how nature designs allosteric control, and might let us develop rules for engineering allosteric proteins to use in synthetic biology applications. The grand goal would be to engineer *de novo* an allosteric transition in a protein that is not normally allosteric. For instance, could we convert an immunoglobulin domain to actuate an allosteric response when it binds to an antigen? While this may appear a daunting challenge, a near-term achievable goal might combine domains from two different allosteric proteins to result in a functional chimera (Figure 1).

Our ability to engineer proteins to bind to DNA, small molecules, or other proteins draws from a wealth of research studying these interactions at molecular resolution. However, allostery has proved to be recalcitrant to engineering because we do not fully understand the molecular connectivity involved in allosteric communication. Previous studies indicate that a network of structurally contiguous residues act in concert to transmit the allosteric signal [1,8–11]. We review here a classic allosteric regulator and describe how new technologies - based on proteinwide mutational perturbation - could build a molecular 'wiring diagram' of allostery by deconstructing the role of each amino acid in the transmission of the allosteric signal. Knowledge of the wiring diagram should allow us to preserve allosteric connections as we design new functions into an allosteric protein. We conclude with how similar strategies might be applied to other broad classes of allosteric proteins.

Lessons from Lacl

Allosteric transcription factors in bacteria, one of the largest annotated families of proteins, regulate adaptive responses to environmental cues. The best- and longest-studied allosteric protein is the *E. coli* repressor LacI which regulates the *lac* (lactose-utilization) operon [2,12]. LacI is composed of ligand-binding and DNA-binding domains. In the absence of the ligand, LacI has high affinity for DNA; when bound to inducer, the protein undergoes a conformational change that causes the DNA-binding domain to lose affinity for DNA, thereby dissociating from the operator site and unblocking the path of RNA polymerase to transcribe downstream genes.

The structure of LacI [3,12,13] can be divided into three sections: the N-terminal 60 residues form a helix-turnhelix motif that binds to DNA; the core of the protein (residues 61–330) is made up of N- and C-terminal subdomains where the ligand binds; and the C-terminal 30 residues are involved in tetramerization (Figure 2). LacI is a functional dimer that makes extensive monomermonomer contacts across the N- and C-terminal ligandbinding subdomains [4,12,13]. The ligand binds in the cleft between the core subdomains, inducing a Venus flytraplike allosteric motion. Upon induction, the allosteric signal is communicated by the relative motion between the N- and C-terminal ligand-binding subdomains, causing the DNA-binding domain to undergo a helix-to-coil structural transition, and hence lose its DNA-binding activity.



Figure 2. Lacl protein architecture and long-range allosteric connections. Structure of a Lacl dimer bound to DNA is shown on the left side; key regions are highlighted in different colors. The broken lines on the right side show long-range connection between four mutations (M42, A133, D149, and S151) that rescue allostery in Y282D.

The molecular basis of allosteric communication in LacI has been the focus of extensive genetic and biochemical studies. Jeffrey Miller's laboratory made 4000 point mutants of LacI [14–16] and used a sensitive genetic screen to classify each mutant into one of three phenotypes: wild type-like activity (WT); unable to bind to DNA (I⁻); or unable to bind to inducer or broken allostery (I^s). These studies showed how structural integrity, dimerization, allosteric signal transduction, ligand-binding, and DNA-binding are tightly interlinked and act in concert for protein function.

The I⁻ phenotype was generally caused by mutations in the DNA-binding domain, or in the C-terminal ligandbinding subdomain which acts as a scaffold for the dimerization required for DNA-binding. Long-range allosteric communication was evident when over 20 second-site mutations scattered across the structure reverted the I⁻ phenotype of Y282D, a change that disrupted dimerization [17,18] (Figure 2). These second-site mutations were clustered in the core-pivot region.

The I^s phenotype resulted from mutations that affected the extensive interactions between the DNA-binding domain and the N-terminal ligand-binding subdomain. Mutation at a single position, A110, in the dimerization interface of N-terminal ligand-binding subdomain dimerization interface caused opposite phenotypes depending on the substitution: higher inducer affinity and lower DNA affinity, or lower inducer affinity and higher DNA affinity, highlighting the key role of the interface in allostery [19].

Co-evolution is a molecular signature of functionally coupled residues undergoing coordinated sequence changes across multiple members of a protein family. To preserve protein function, co-evolving pairs undergo mutually compensatory changes during evolution, which provide insight into residue connectivity [20]. The LacI family of bacterial transcription factors despite sharing the same protein architecture but have a sequence similarity of less than 30% [21]. Although residues in the ligand-binding pocket and at the DNA-binding interface vary depending on inducer and operator sequence, residues involved in allostery have lower sequence entropy. Indeed, most coevolving residues involved in allostery within a subfamily were found in spatially disconnected regions, showing that complex epistatic networks participate in allostery [22,23]. In LacI, these positions are scattered across the N- and C-terminal dimerization interfaces, core-pivot and hinge regions, and DNA-binding domain, consistent with genetic and biochemical studies of the mutants [21].

Within a close subfamily, structural residues in the ligand-binding pocket are more conserved and ligand specificity is achieved through mutations at less-conserved residues. Comparison across more distant families shows that approximately 19% of the residues are conserved, suggesting that these positions may be indispensable or 'hardwired' for structural integrity and function for that protein-fold family [21]. As the scaffold evolved to acquire specific functions, it may have adapted to a particular niche

through minor alterations in mechanism. For instance, although LacI and PurR share highly similar structures and sequences, their allosteric wiring is opposite: upon ligand-binding, LacI dislodges from DNA, whereas PurR binds to DNA.

Deep mutational scanning using a toggled selection system for allostery

Biochemical and evolutionary studies of LacI suggest that amino acids involved in allostery are intricately coupled with amino acids recognizing ligand and DNA. Thus, a first step toward unraveling this allosteric network would be to determine the role of all the amino acids in LacI function, classifying each as participating in one of the following categories: binds to DNA or to ligand, is required for structural stability, is involved in allostery, or none of the above. As with conventional protein biochemistry, the functional importance of each amino acid can be determined by mutational analysis. Because allostery is a systemic effect, however, we need to scale-up mutagenesis by orders of magnitude to interrogate the effect of every single mutation and every combination of double mutations. The double mutations are particularly important because they can reveal long-range allosteric interactions, as seen with the Y282D mutation in LacI. Although allostery most likely depends on higher-order network interactions beyond pairwise couplings, sufficient pairs of long-range double mutants should identify linchpin positions in this network.

How can we obtain the phenotype associated with every single and double mutation in an allosteric protein? An approach called 'deep mutational scanning' [24] can generate a comprehensive mutation map which shows the effect of all neutral and loss-of-function, and any gain-offunction or hyper-activating changes, as well as double mutations that restore function lost in a single mutant or that exacerbate the combined effect of single mutations more than predicted. Deep mutational scanning has successfully been used to analyze antibody affinity maturation [25,26], protein-peptide interactions [27,28], proteinsmall-molecule affinity [29], ubiquitination [30], protein stability [31], splicing [32], and antibiotic resistance [33], among other protein activities.

Deep mutational scanning leverages two major advances: first, the generation of targeted DNA libraries with over a million unique sequences through doped oligonucleotide assembly or pre-specified microarray-generated oligonucleotides [34]; and, second, enormous read numbers from next-generation sequencing that allow thousands of clones to be assayed simultaneously by linking activity to readfrequency of each unique sequence [31]. Genotype and phenotype are linked through an in vitro or in vivo selection that amplifies mutants with the desired function. Because the total sequence read-count is orders of magnitude greater (e.g., 400 million reads with the Illumina HiSeq) than the number of selected genotypes, the activity of a protein variant, to a first approximation, is reflected in the relative sequence abundance of its encoding DNA. Enthalpies derived from sequencing statistics has been shown to linearly correlate with experimental measurements and computational structure-based ΔG predictions [26,35]. For LacI, the library of all single mutations is a little over 6000,

and even the library of all double mutations, at nearly 2.5×10^7 , is well within the transformation efficiency (10^9) of commercially competent cells.

Once the mutants are constructed, they need to be assayed in a suitable selection system. Selection systems typically enrich for only one function; for instance, with protein-peptide interactions, phage display enriches for stronger binders. However, a selection system for allosteric proteins should enrich for either of two states: the induced or the uninduced. Such selections can be accomplished using a dual selectable marker as the reporter gene. For example, the *E. coli tolC* gene [36,37] and the *Saccharomyces cerevisiae URA3* gene allow enrichment of cells expressing (positive selection) or not expressing (negative selection) the reporter using different selection agents. Dual selection can thus identify mutants with the three phenotypes found in Miller's study: WT, I⁻, and I^s.

First, subjecting the library to positive selection in the absence of the inducer reveals mutants that are not capable of binding to DNA (I⁻ phenotype) because they constitutively express the reporter and are enriched (Figure 3, left arrow). Second, subjecting the library to negative selection in the absence of the inducer identifies mutants that bind to DNA; these repress the reporter and are insensitive to negative selection (Figure 3, center arrow). Subjecting this DNA-bound population to positive selection in the presence of the inducer enriches for mutants with WT phenotype because they respond to the inducer to activate the reporter (Figure 3, center arrow). Finally, exposing the DNA-bound population to negative selection in the presence of the inducer enriches for mutants that do not respond to the inducer (I^s phenotype), while those exhibiting WT-like activation perish (Figure 3, right arrow).

The next step is to identify specific positions from the I⁻ and I^s set that play a role in allostery. To delineate this subset, we need to reclassify I⁻ and I^s into subgroups. The I⁻ set comprises mutants that do not recognize DNA, structurally unstable mutants, and mutants that have broken allostery. The I^s set comprises mutants that do not recognize the inducer and those that fail to transmit the allosteric signal despite inducer binding. To narrowdown the allosteric subset from the I⁻ and I^s sets, we can carry out computational structure-based ΔG calculations of folding, ligand-binding, and DNA-binding of each mutant. Because the allosteric subset is deduced by the process of elimination, we need to prune this set further to isolate a minimal subset of positions that maximally contribute to allostery. NMR chemical shift perturbation measurement detects the local change in chemical environment for each amino acid upon conformational change. This approach has been used to identify the allosteric connectivity in protein kinase A [11]. Because amino acids in the vicinity of the allosteric positions are likely to undergo conformational change, chemical shift perturbations alone cannot identify the allosteric network. However, via the intersection of mutational data from genetic screens, structure-based enthalpy calculations, and biophysical measurements of local conformational change by NMR, we can pinpoint the minimal set of amino acids that constitute the allosteric network. Molecular simulations



Figure 3. Toggled selection scheme for high-throughput functional evaluation of Lacl mutants. Lacl mutants ('mut') are shown in yellow; red circles represent mutations. Positive selection in the absence of the inducer enriches for I⁻ mutants (left arrow). Negative selection without inducer followed by positive selection with inducer enriches for wild type (WT)-like mutants (center arrow). Negative selection with inducer enriches for I^s mutants (right arrow).

can independently validate the allosteric network by computing the conformational change of individual amino acids across the two allosteric states [8,38].

Thus, in one multiplexed assay, millions of LacI variants can be functionally characterized in a manner that is facile and might be extendable to other proteins. Extending this selection system for other members of the LacI family might be accomplished by swapping the promoter regulating the reporter with a promoter that is bound by a different allosteric protein.

Engineering allosteric proteins

New ligand and DNA specificities engineered into existing proteins

To engineer an orthogonal transcription factor, we can begin by redesigning the ligand- and DNA-binding specificities of natural transcription factors. Once the allosteric connections have been identified, and the residues responsible for allostery have been distinguished from those involved in binding ligand or DNA, we can incorporate this information into the design protocol. For example, in the case of LacI, mutating residues that cause an I^s phenotype is more likely to result in altered specificity because a subset of I^s mutants retain allostery despite losing binding affinity for the inducer, isopropyl β -D-1-thiogalactopyranoside (IPTG, a lactose analog). By contrast, to change the DNA specificity, we target I⁻ mutants to restore allostery and to achieve new specificity because both are intimately linked in the DNA-binding domain of LacI.

Despite different ligand specificities, the overall binding-pocket architectures of LacI-like proteins are similar. This similarity suggests that an altered specificity for a new ligand can be achieved while still preserving allostery. The binding pocket of each allosteric protein family can, in principle, be redesigned to access chemical diversity around its cognate ligand. For instance, the LacI family of proteins may be able to accommodate many derivatives of sugars. With respect to DNA recognition, the helix-turn-helix domain is commonly used across several bacterial transcription factor families. Thus, by analyzing the binding-site preferences across helix-turnhelix family members, we might gain the knowledge to engineer new DNA specificity without compromising allostery.

Chimeric allosteric proteins

An alternative to redesigning existing proteins is to engineer new allosteric proteins by mixing and matching protein domains. The LacI family is thought to have evolutionarily diverged from the structurally similar periplasmic-binding protein family by acquiring a DNA-binding domain [39]. Periplasmic-binding proteins respond to a large repertoire of small molecules by regulating ABC transporters. A chimera of a periplasmic-binding protein and a DNA-binding domain could be engineered such that an allosteric change in the periplasmic-binding protein is communicated to the DNA-binding domain to change DNA affinity. Another DNA-binding domain family that might be amenable to engineering is the zinc-finger protein class, which is comparable in size to the helix-turn-helix domain class. Because the specificity of these proteins can be programmed, we can build completely orthogonal switches. A further approach is to engineer a chimeric protein that, upon ligand binding, exposes a binding surface that recruits a sigma factor, resulting in the activation of transcription. Engineering functional chimeras will involve extensive optimization of the linker between the domains, the inter-domain interface, and the dimerization interface.

Although thousands of bacterial allosteric transcription factors have been annotated in the sequences of metagenomic samples, their ligand and DNA specificities remain unknown. However, a clue to the type of ligand that is recognized might come from characterization of the nearby operon that is regulated by a transcription factor. This set of transcription factors is a treasure trove for the design of biosensors for many industrially valuable molecules, allowing their biosynthetic pathways to be optimized for higher production. If the rules of allostery are known, we can build a functional chimera with a known DNA-binding domain and identify its inducer using a reporter-gene assay against a panel of ligands.

In general, knowledge of allosteric connections can help favorably bias the search for new function by avoiding changes that break or weaken allostery. Computational and directed evolution design protocols could incorporate deep mutational scanning data when choosing residues that can be mutated.

Application to the wider world of allostery

As DNA synthesis and sequencing become increasingly cheaper and higher in throughput, the rate-limiting step for the analysis of other classes of allosteric proteins is the development of functional assays. The bacterial transcription factor family allows one of the easiest functional assays because allostery is directly coupled to transcription. In this section we briefly describe functional assays for other major classes of allosteric proteins.

Direct transcriptional readout (nuclear receptors)

The toggled selection scheme for LacI can be adapted for other allosteric transcription factors such as nuclear receptors (Figure 4A). Binding of the ligand to the nuclear receptor induces a reorientation of the ligand-binding domain and, in the case of steroid receptors, directly leads to transcriptional activation. Other nuclear receptors, such as the receptors for retinoic acid and vitamin D, may also be successfully interrogated with this strategy, although upon ligand binding they have a more complex set of binding interactions involving additional transcriptional activators and repressors [40,41].



Figure 4. Functional assays for other allosteric protein classes. (A) Direct transcriptional readout: applicable to allosteric transcription factors such as nuclear receptors. The presence of the receptor ligand is necessary to activate expression of a selectable marker. (B) Indirect transcriptional readout: applicable to two-component systems. Allosteric activation of a histidine kinase causes phosphorylation (P) of a response regulator, leading to transcription of a selectable marker. (C) Split reporter assay: applicable to G protein-coupled receptors (GPCRs) and heterodimeric receptor tyrosine kinases (RTKs). Complementary halves of GFP are attached to the receptor and to the factor recruited upon allosteric activation. A functional GFP is formed only when the activated allosteric protein binds to its partner and brings the two GFP halves into sufficient proximity. (D) Environment-sensitive fluorophores: applicable to protein kinases. A fluorophore is attached to a peptide substrate of the kinase in close proximity to the phosphorylated residue, such that the fluorophore will fluoresce only when the phosphate group is present. (E) Domain-inserted reporter: applicable to cytosolic allosteric protein is inserted into GFP such that a functional GFP is formed only when the allosteric protein changes to a ligand-bound conformation.

Indirect transcriptional readout (two-component system)

Many allosteric proteins alter gene expression indirectly by regulating downstream transcription factors. A transcriptional readout may be employed to evaluate such allosteric proteins provided that target gene expression robustly reflects protein activity (Figure 4B). An example of such pathways is the bacterial two-component system containing a membrane-bound histidine kinase, that serves as an environmental sensor, and a cytoplasmic response regulator, which is phosphorylated by the activated kinase and often acts directly as a transcription factor [42]. Systematic mutagenesis could identify the amino acids needed for the sensor histidine kinase to change conformation upon ligand activation [43] by employing a toggled selection marker at the locus that is bound by the response regulator as a readout of pathway activation.

Split reporter assay [G protein-coupled receptors (GPCRs) and receptor tyrosine kinases (RTKs)]

For allosteric receptors far upstream of the transcription factors in their pathway, a different readout is needed. Membrane-bound receptors such as GPCRs and RTKs, upon allosteric activation, colocalize with a partner: GPCRs associate with β -arrestin [44] and EGFR family RTKs can form heterodimers [45]. Activity of these proteins can be measured with a split reporter assay in which the receptor is fused to one domain of a reporter protein and the recruited partner is fused to the other domain; when brought into close proximity the two domains can combine into a functional whole (Figure 4C) [46–49]. This assay can be performed using a divided GFP and fluorescence-activated cell sorting (FACS) to sort cells based on fluorescence intensity.

Environment-sensitive fluorophores (protein kinases)

Allosteric protein kinases can be difficult to study owing to the transitory nature of their interactions with their substrates. Kinases that do not directly regulate transcription factors may be interrogated using fluorescent sensors dependent on target phosphorylation. For tyrosine kinases whose phosphorylated targets are recognized by SH2 (Src homology 2) domains, the kinase target can be fused to SH2, together with two fluorescence resonance energy transfer (FRET)-compatible fluorophores at either end of the construct, such that SH2 binding to the phosphotyrosine brings the fluorophores together. Alternatively, as a more general tyrosine/serine phosphorylation sensor, a peptide substrate of the kinase is attached to a fluorophore with a separation of only several angstroms; when the peptide is phosphorylated, the change in the microenvironment activates the fluorophore (Figure 4D) [50-52]. FACS can then be used to sort cells based on fluorescence intensity.

Domain-inserted reporter

As a general methodology for cytoplasmic allosteric proteins, various domain-inserted reporter systems have been developed. The allosteric protein is inserted into a reporter protein in such a way that reporter activity is dependent on the conformation of the allosteric protein (Figure 4E). β -Lactamase has been used as such a reporter with maltose-binding protein, leading to maltose-inducible ampicillin resistance [53]; similarly, a calcium-sensitive GFP reporter was made through insertion of calmodulin [54]. In this way, a wide array of allosteric proteins, such as enzymes and ion sensors that do not regulate transcription, can be tied directly to fluorescent and antibioticresistance readouts.

Concluding remarks

The ability to engineer the above allosteric protein classes paves the way for new synthetic biology applications: designer GPCRs that can respond to a drug overdose, two-component system proteins that enable bacterial chemotaxis toward a specific molecule, dynamic rewiring of kinase signaling, and controlling the composition and function of engineered microbiota with quorum-sensing switches.

The approach outlined here shows how the power of deep sequencing can be harnessed to address a longstanding question in biology: how protein sequence affects allostery. We envisage that this rich mutational dataset will motivate new studies in kinetics of allostery through molecular dynamics and NMR experiments.

Acknowledgments

We thank Lea Starita and James Carothers for comments on the manuscript. This work has been supported by the Wyss Technology Development Fellowship to S.R, the US Department of Energy (DE-FG02-02ER63445 to G.M.C.), and the US National Institutes of Health (P41 GM103533 to S.F.). S.F. is an investigator of the Howard Hughes Medical Institute.

References

- 1 Lu, T.K. et al. (2009) Next-generation synthetic gene networks. Nat. Biotechnol. 27, 1139–1150
- 2 Dietrich, J.A. *et al.* (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth. Biol.* 2, 47–58
- 3 Rodriguez, G.J. et al. (2010) Evolution-guided discovery and recoding of allosteric pathway specificity determinants in psychoactive bioamine receptors. Proc. Natl. Acad. Sci. U.S.A. 107, 7787–7792
- 4 Hilser, V.J. et al. (2012) Structural and energetic basis of allostery. Annu. Rev. Biophys. 41, 585–609
- 5 Bai, F. et al. (2010) Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch. Science 327, 685-689
- 6 Hilser, V.J. (2010) An ensemble view of allostery. Science 327, 653–654
- 7 Goodey, N.M. and Benkovic, S.J. (2008) Allosteric regulation and catalysis emerge via a common route. Nat. Chem. Biol. 4, 474-482
- 8 Amaro, R.E. et al. (2007) A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry* 46, 2156–2173
- 9 Gandhi, P.S. et al. (2008) Structural identification of the pathway of long-range communication in an allosteric enzyme. Proc. Natl. Acad. Sci. U.S.A. 105, 1832–1837
- 10 Ricketson, D. et al. (2007) A conformational switch in the ligandbinding domain regulates the dependence of the glucocorticoid receptor on Hsp90. J. Mol. Biol. 368, 729–741
- 11 Masterson, L.R. et al. (2008) Allosteric cooperativity in protein kinase A. Proc. Natl. Acad. Sci. U.S.A. 105, 506–511
- 12 Lewis, M. (2005) The lac repressor. C. R. Biol. 328, 521-548
- 13 Bell, C.E. and Lewis, M. (2000) A closer view of the conformation of the Lac repressor bound to operator. Nat. Struct. Biol. 7, 209–214
- 14 Kleina, L.G. and Miller, J.H. (1990) Genetic studies of the lac repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. J. Mol. Biol. 212, 295–318
- 15 Markiewicz, P. et al. (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

Opinion

- 16 Suckow, J. et al. (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
- 17 Xu, J. and Matthews, K.S. (2009) Flexibility in the inducer binding region is crucial for allostery in the *Escherichia coli* lactose repressor. *Biochemistry* 48, 4988–4998
- 18 Swint-Kruse, L. et al. (2003) Perturbation from a distance: mutations that alter LacI function through long-range effects. Biochemistry 42, 14004–14016
- 19 Müller-Hartmann, H. and Müller-Hill, B. (1996) The side-chain of the amino acid residue in position 110 of the Lac repressor influences its allosteric equilibrium. J. Mol. Biol. 257, 473–478
- 20 de Juan, D. et al. (2013) Emerging methods in protein co-evolution. Nat. Rev. Genet. 14, 249–261
- 21 Parente, D.J. and Swint-Kruse, L. (2013) Multiple co-evolutionary networks are supported by the common tertiary scaffold of the LacI/ GalR proteins. *PLoS ONE* 8, e84398
- 22 Süel, G.M. et al. (2002) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat. Struct. Biol. 10, 59–69
- 23 Shulman, A.I. et al. (2004) Structural determinants of allosteric ligand activation in RXR heterodimers. Cell 116, 417–429
- 24 Fowler, D.M. and Fields, S. (2014) Deep mutational scanning: a new style of protein science. Nat. Methods 11, 801–807
- 25 Fujino, Y. et al. (2012) Robust in vitro affinity maturation strategy based on interface-focused high-throughput mutational scanning. Biochem. Biophys. Res. Commun. 428, 395–400
- 26 Whitehead, T.A. et al. (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat. Biotechnol. 30, 543–548
- 27 Ernst, A. et al. (2010) Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. Mol. Biosyst. 6, 1782
- 28 Araya, C.L. et al. (2012) A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc. Natl. Acad. Sci. U.S.A. 109, 16858–16863
- 29 Tinberg, C.E. *et al.* (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 501, 212–216
- 30 Starita, L.M. et al. (2013) Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. Proc. Natl. Acad. Sci. U.S.A. 110, E1263-E1272
- 31 Fowler, D.M. et al. (2010) High-resolution mapping of protein sequence–function relationships. Nat. Methods 7, 741–746
- 32 Findlay, G.M. et al. (2014) Saturation editing of genomic regions by multiplex homology-directed repair. Nature 513, 120–123
- 33 Cheng, A.A. et al. (2014) Enhanced killing of antibiotic-resistant bacteria enabled by massively parallel combinatorial genetics. Proc. Natl. Acad. Sci. U.S.A. 111, 12462–12467
- 34 Kosuri, S. et al. (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nat. Biotechnol. 28, 1295–1299
- 35 McLaughlin, R.N., Jr et al. (2013) The spatial architecture of protein function and adaptation. Nature 490, 138-142

- 36 DeVito, J.A. (2008) Recombineering with tolC as a selectable/counterselectable marker: remodeling the rRNA operons of *Escherichia coli*. *Nucleic Acids Res.* 36, e4
- 37 Isaacs, F.J. et al. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333, 348–353
- 38 Daily, M.D. et al. (2008) Contact rearrangements form coupled networks from local motions in allosteric proteins. Proteins 71, 455–466
- 39 Fukami-Kobayashi, K. et al. (2003) Parallel evolution of ligand specificity between LacI/GalR family repressors and periplasmic sugar-binding proteins. Mol. Biol. Evol. 20, 267–277
- 40 Gronemeyer, H. et al. (2004) Principles for modulation of the nuclear receptor superfamily. Nat. Rev. Drug Discov. 3, 950–964
- 41 Wurtz, J.M. et al. (1996) A canonical structure for the ligand-binding domain of nuclear receptors. Nat. Struct. Biol. 3, 87–94
- 42 Galperin, M.Y. (2006) Structural classification of bacterial response regulators: diversity of output domains and domain combinations. J. Bacteriol. 188, 4169–4182
- 43 Capra, E.J. et al. (2010) Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of twocomponent signaling pathways. PLoS Genet. 6, e1001220
- 44 Ferguson, S.S. et al. (1996) Role of beta-arrestin in mediating agonistpromoted G protein-coupled receptor internalization. Science 271, 363–366
- 45 Schlessinger, J. (2000) Cell signaling by receptor tyrosine kinases. Cell 103, 211–225
- 46 Stynen, B. et al. (2012) Diversity in genetic in vivo methods for proteinprotein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. *Microbiol. Mol. Biol. Rev.* 76, 331–382
- 47 Petschnigg, J. et al. (2014) The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. Nat. Methods 11, 585–592
- 48 Kittanakom, S. et al. (2014) CHIP-MYTH: a novel interactive proteomics method for the assessment of agonist-dependent interactions of the human β 2-adrenergic receptor. Biochem. Biophys. Res. Commun. 445, 746–756
- 49 Yan, Y-X. et al. (2002) Cell-based high-throughput screening assay system for monitoring G protein-coupled receptor activation using beta-galactosidase enzyme complementation technology. J. Biomol. Screen. 7, 451–459
- 50 González-Vera, J.A. (2012) Probing the kinome in real time with fluorescent peptides. *Chem. Soc. Rev.* 41, 1652–1664
- 51 Morris, M.C. (2013) Fluorescent biosensors probing protein kinase function in cancer and drug discovery. *Biochim. Biophys. Acta* 1834, 1387–1395
- 52 Chen, C-A. et al. (2004) Biosensors of protein kinase action: from in vitro assays to living cells. Biochim. Biophys. Acta 1697, 39–51
- 53 Guntas, G. et al. (2005) Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proc. Natl. Acad. Sci. U.S.A. 102, 11224–11229
- 54 Baird, G.S. et al. (1999) Circular permutation and receptor insertion within green fluorescent proteins. Proc. Natl. Acad. Sci. U.S.A. 96, 11241–11246