



## Surface Sites for Engineering Allosteric Control in Proteins

Jeeyeon Lee, *et al.*  
*Science* **322**, 438 (2008);  
DOI: 10.1126/science.1159052

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of February 4, 2009):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/322/5900/438>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/cgi/content/full/322/5900/438/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/322/5900/438#related-content>

This article **cites 35 articles**, 10 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/322/5900/438#otherarticles>

This article appears in the following **subject collections**:

Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Our results test the hypothesis that variation in gene expression is dictated by regulatory regions, extending recent studies of expression by quantitative trait-loci mapping and comparative expression studies that have been confined to closely related species (26–30). The apparent absence of overt trans influences could be explained by the modest amount of human DNA provided by a single copy of human chromosome 21 when compared with the complete mouse genome, as well as the absence of liver-specific transcriptional regulators on chromosome 21. The extent to which protein coding and cis-regulatory mutations contribute to changes in morphology, physiology, and behavior is actively debated in evolutionary biology (3, 12, 13). Myriad points of control influence gene expression; however, it has also been an unresolved question as to which of these mechanisms has the most influence globally. Here, we show that each layer of transcriptional regulation within the adult hepatocyte, from the binding of liver master regulators and chromatin remodeling complexes to the output of the transcriptional machinery, is directed primarily by DNA sequence. Although conservation of motifs alone cannot predict transcription factor binding, we show that within the genetic sequence there must be embedded adequate instructions to direct species-specific transcription.

#### References and Notes

1. E. H. Davidson, D. H. Erwin, *Science* **311**, 796 (2006).
2. K. S. Zaret, *Mech. Dev.* **92**, 83 (2000).
3. G. A. Wray, *Nat. Rev. Genet.* **8**, 206 (2007).
4. B. Li, M. Carey, J. L. Workman, *Cell* **128**, 707 (2007).
5. E. Guccione *et al.*, *Nat. Cell Biol.* **8**, 764 (2006).
6. L. Elnitski, V. X. Jin, P. J. Farnham, S. J. Jones, *Genome Res.* **16**, 1455 (2006).
7. D. T. Odom *et al.*, *Nat. Genet.* **39**, 730 (2007).
8. A. M. Moses *et al.*, *PLOS Comput. Biol.* **2**, e130 (2006).
9. A. R. Borneman *et al.*, *Science* **317**, 815 (2007).
10. E. Birney *et al.*, *Nature* **447**, 799 (2007).
11. B. E. Bernstein *et al.*, *Cell* **120**, 169 (2005).
12. H. E. Hoekstra, J. A. Coyne, *Evolution* **61**, 995 (2007).
13. S. B. Carroll, *Cell* **134**, 25 (2008).
14. A. O'Doherty *et al.*, *Science* **309**, 2033 (2005).
15. Materials and methods are available as supporting material on *Science* Online.
16. D. Kampa *et al.*, *Genome Res.* **14**, 331 (2004).
17. J. S. Carroll *et al.*, *Cell* **122**, 33 (2005).
18. S. Cereghini, *FASEB J.* **10**, 267 (1996).
19. J. Eeckhoutte, B. Oxombre, P. Formstecher, P. Lefebvre, B. Laine, *Nucleic Acids Res.* **31**, 6640 (2003).
20. F. M. Sladek, M. D. Ruse Jr., L. Nepomuceno, S. M. Huang, M. R. Stallcup, *Mol. Cell Biol.* **19**, 6509 (1999).
21. A. Rada-Iglesias *et al.*, *Hum. Mol. Genet.* **14**, 3435 (2005).
22. M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, R. A. Young, *Cell* **130**, 77 (2007).
23. M. Vermeulen *et al.*, *Cell* **131**, 58 (2007).
24. R. J. Sims 3rd *et al.*, *Mol. Cell* **28**, 665 (2007).
25. A. Barski *et al.*, *Cell* **129**, 823 (2007).
26. Y. Gilad, A. Oshlack, G. K. Smyth, T. P. Speed, K. P. White, *Nature* **440**, 242 (2006).
27. P. Khaitovich *et al.*, *Science* **309**, 1850 (2005).
28. P. J. Wittkopp, B. K. Haerum, A. G. Clark, *Nat. Genet.* **40**, 346 (2008).
29. C. C. Park *et al.*, *Nat. Genet.* **40**, 421 (2008).
30. Y. Gilad, S. A. Rifkin, J. K. Pritchard, *Trends Genet.* **24**, 408 (2008).
31. We are grateful to E. Jacobsen, R. Stark, I. Spiteri, B. Liu, J. Marioni, A. Lynch, J. Hadfield, N. Matthews, the Cambridge Research Institute (CRI) Genomics Core, CRI Bioinformatics Core, and Camgrid for technical assistance, and B. Gottgens and J. Ferrer for insightful advice. Supported by the European Research Council (D.T.O.); Royal Society Wolfson Research Merit Award (S.T.); Hutchison Whampoa (D.T.O., ST); Medical Research Council (E.F., VT); Wellcome Trust (E.F., V.T.); University of Cambridge (D.T.O., D.S., N.B.M., S.T.); and Cancer Research U.K. (D.T.O., M.D.W., N.B.M., S.T., D.S.). Data deposited under ArrayExpress accession numbers E-TABM-473 and E-TABM-474. M.D.W., N.B.M., D.S., D.T.O., and C.M.C. designed and performed experiments; N.B.M., M.D.W., and E.F. analyzed the data; L.V., V.T., M.D.W., and E.F. created, prepared, and provided Tc1 mouse tissues; and M.D.W., N.B.M., D.T.O., and S.T. wrote the manuscript. D.T.O. oversaw the work. The authors declare no competing interests.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/1160930/DC1  
Materials and Methods  
SOM Text  
Figs. S1 to S12  
Table S1

27 May 2008; accepted 3 September 2008  
Published online 11 September 2008;  
10.1126/science.1160930  
Include this information when citing this paper.

## Surface Sites for Engineering Allosteric Control in Proteins

Jeeyeon Lee,<sup>1\*</sup> Madhusudan Natarajan,<sup>2\*</sup> Vishal C. Nashine,<sup>1</sup> Michael Socolich,<sup>2</sup> Tina Vo,<sup>2</sup> William P. Russ,<sup>2</sup> Stephen J. Benkovic,<sup>1</sup> Rama Ranganathan<sup>2†</sup>

Statistical analyses of protein families reveal networks of coevolving amino acids that functionally link distantly positioned functional surfaces. Such linkages suggest a concept for engineering allosteric control into proteins: The intramolecular networks of two proteins could be joined across their surface sites such that the activity of one protein might control the activity of the other. We tested this idea by creating PAS-DHFR, a designed chimeric protein that connects a light-sensing signaling domain from a plant member of the Per/Arnt/Sim (PAS) family of proteins with *Escherichia coli* dihydrofolate reductase (DHFR). With no optimization, PAS-DHFR exhibited light-dependent catalytic activity that depended on the site of connection and on known signaling mechanisms in both proteins. PAS-DHFR serves as a proof of concept for engineering regulatory activities into proteins through interface design at conserved allosteric sites.

Proteins typically adopt well-packed three-dimensional structures in which amino acids are engaged in a dense network of contacts (1, 2). This emphasizes the energetic importance of local interactions, but protein

function also depends on nonlocal, long-range communication between amino acids. For example, information transmission between distant functional surfaces on signaling proteins (3), the distributed dynamics of amino acids involved in enzyme catalysis (4–6), and allosteric regulation in various proteins (7) all represent manifestations of nonlocal interactions between residues. To the extent that these features contribute to defining biological properties of protein lineages, we expect that the underlying mechanisms represent conserved rather than idiosyncratic features in protein families.

On the basis of this conjecture, methods such as statistical coupling analysis (SCA) quantitatively examine the long-term correlated evolution of amino acids in a protein family—the statistical signature of functional constraints arising from conserved communication between positions (8, 9). This approach has identified sparse but physically connected networks of coevolving amino acids in the core of proteins (8–12). The connectivity of these networks is remarkable, given that a small fraction of total residues are involved and that no tertiary structural information is used in their identification. Empirical observation in several protein families shows that these networks connect the main functional site with distantly positioned secondary sites, enabling predictions of allosteric surfaces at which binding of regulatory molecules (or covalent modifications) might control protein function. Both literature studies and forward experimentation in specific model systems confirm these predictions (8–12). Thus, techniques such as SCA may provide a general tool for computational prediction of conserved allosteric surfaces.

The finding that certain surface sites might be statistical “hotspots” for functional interaction with active sites suggests an idea for engineering new regulatory mechanisms into proteins. What if two proteins were joined at surface sites such that their statistically correlated networks were juxtaposed and could form functional interactions (Fig. 1A)? If the connection sites are functionally linked to their respective active sites

<sup>1</sup>Department of Chemistry, Pennsylvania State University, University Park, PA 16802, USA. <sup>2</sup>Green Center for Systems Biology and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: rama.ranganathan@utsouthwestern.edu

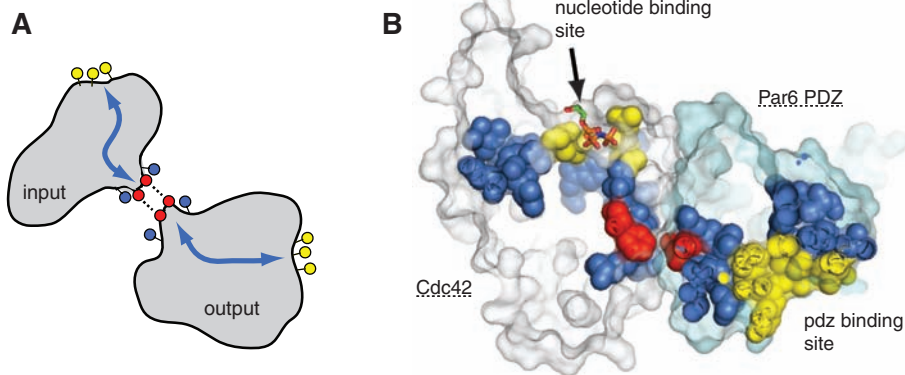
through allosteric mechanisms intrinsic to each protein, it should be possible to couple the activity of one protein to that of the other.

Evidence from natural systems supports this design concept (Fig. 1B). Ligand binding in a PDZ (PSD95/Dlg1/ZO1) protein interaction module in the human Par6 protein is allosterically regulated by interaction with the guanine nucleotide-binding protein (G protein) Cdc42 at a distant

surface site (13, 14). SCA for the PDZ and G protein families reveals a contiguous network of amino acids that connects the nucleotide-binding pocket of Cdc42 with the ligand-binding pocket of the PDZ domain through specific interactions across an allosteric interaction surface (Fig. 1B and figs. S1 to S3). Mutagenesis studies confirm that these networks in PDZ and G proteins contribute to allosteric signaling (11, 14). In both

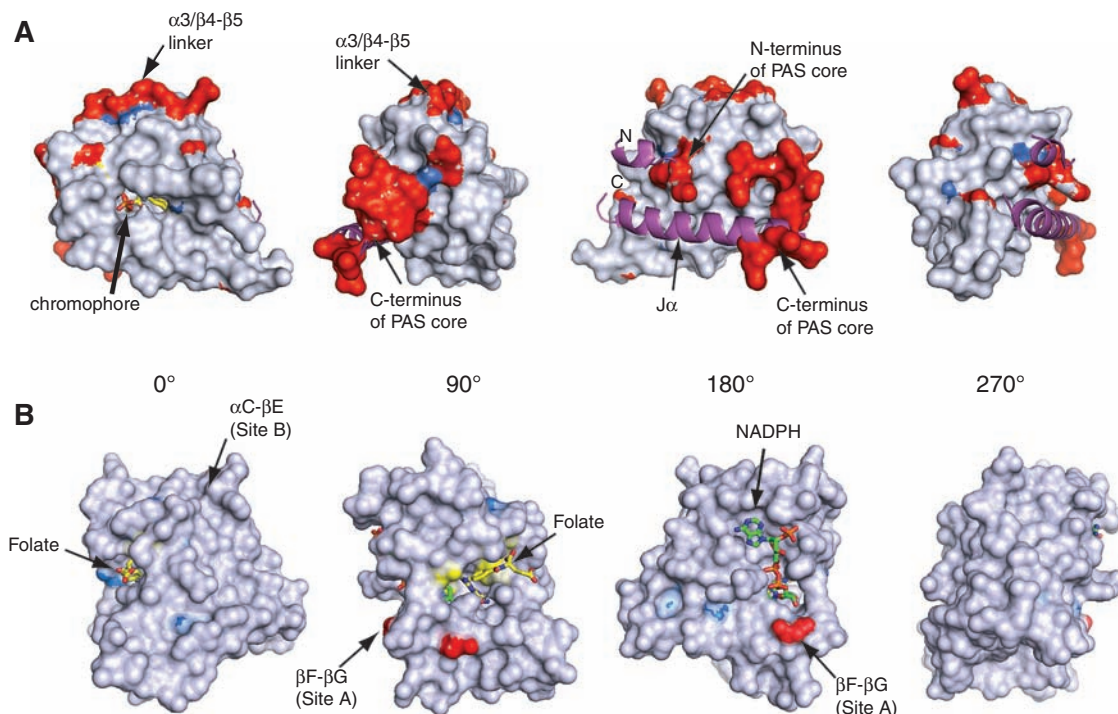
proteins, the allosteric site is uniquely identified as the surface-exposed residues of the SCA network that are nonetheless distant from the active site (fig. S3).

Can statistically correlated allosteric networks be joined to engineer functional communication between proteins? To test this idea, we selected two protein modules as components for creation of a non-natural allosteric two-domain protein in which a signal originating in one domain (the “input” module) is transmitted to influence the activity of the second domain (the “output” module) (Fig. 1A). For the input module, we chose a light-sensing domain from plant phototropin [*Avena sativa* LOV2 (15)], a member of the Per/Amt/Sim (PAS) family of signaling modules (16). PAS domains mediate biological responses to a diverse set of stimuli, including aromatic hydrocarbons, gases, redox potential, and light, and share a conserved core structure comprising a five-stranded antiparallel  $\beta$  sheet with two flanking  $\alpha$  helices (17) (fig. S4A). A deep ligand-binding pocket opens to one surface of the domain. In the phototropin LOV2 domain, signaling is initiated by light absorption by a flavin mononucleotide (FMN) chromophore bound within the binding pocket, which then transmits the signal through the structure to cause two large conformational changes at the opposite surface: destabilization and unbinding of helical extensions at both the C terminus (the  $J\alpha$  helix) and the N terminus of the core PAS domain (18, 19) (Fig. 2A and fig. S4A). The structural details of the N- and C-terminal extensions varies among members of the PAS domain family, but conformational change at these regions appears to be a



**Fig. 1.** A design concept for allosteric communication. **(A)** The SCA computational method identifies networks of statistically interacting amino acids in proteins (blue arrows). These networks often link primary functional sites (yellow) with distant surface positions (red) through the core (blue). This motivates the idea of functionally coupling the two proteins (denoted “input” and “output” modules) through linkage at the predicted allosteric sites. **(B)** A slice through the structure of a complex between the PDZ domain of the cell polarity protein Par6 (cyan surface) with its allosteric regulator, the Cdc42 G protein (white surface) shows an example of SCA network linkage in a natural protein-protein interaction. SCA residues [in CPK representation and colored as in (A)] constitute a contiguous network linking the site of nucleotide exchange on Cdc42 with the ligand-binding pocket of the PDZ domain (both in yellow) through specific residues at the allosteric interface (red).

**Fig. 2.** Design principle of the PAS-DHFR chimera. **(A)** and **(B)** Surface-exposed SCA sites shown on the LOV2 PAS domain [(A), PDB code 2VOU] and *E. coli* DHFR [(B), PDB code 1RX2] shown in four successive rotations of each molecule. As in Fig. 1A, SCA network positions are colored yellow (within 5 Å of substrate), red (surface-exposed), or blue (buried). A residue is considered buried if its fractional solvent-accessible surface area is  $<0.1$ . In the core PAS domain, this analysis reveals two surface-exposed nonsubstrate proximal SCA sites: (i) the N- and C-terminal regions that, in LOV2, mediate light-dependent interaction with the  $J\alpha$  and N-terminal helical extensions, and (ii) the  $\alpha3/\beta4$ - $\beta5$  region (see text and fig. S4). In DHFR, SCA reveals a network (fig. S5) that relates the enzyme active site to a specific distant surface loop ( $\beta$ F- $\beta$ G, site A, in red). The experiment is to insert the LOV2 domain into several DHFR positions at site A and at a control surface (site B,  $\alpha$ C- $\beta$ E loop).



conserved feature of allosteric signaling in this protein family (20, 21).

SCA for a multiple sequence alignment of 1104 core PAS domains mirrors the experimental findings. A spatially contiguous network of correlated amino acids links the ligand-binding pocket to surface-exposed residues at the N- and C-terminal regions of the core domain that, in LOV2, make direct interactions with the  $J\alpha$  and N-terminal helices (Fig. 2A and fig. S4, B and C). A second surface-exposed site is also evident (the  $\alpha 3$  helix and  $\beta 4$ - $\beta 5$  linker); this site undergoes light-dependent conformational change in another member of the PAS family, the photoactive yellow protein, but is close to the chromophore-binding pocket (22). These data uniquely identify the N- and C-terminal regions of the LOV2 domain as the logical connection sites in our design experiment.

As the output module, we chose *E. coli* dihydrofolate reductase (EcDHFR), an enzyme system in which extensive prior structural, biochemical, and theoretical work has established the basic catalytic mechanism and evidence for long-range control of activity. DHFR is an essential enzyme required for folate metabolism in all organisms; it catalyzes the stereospecific reduction of 7,8-dihydrofolate ( $H_2F$ ) to 5,6,7,8-tetrahydrofolate ( $H_4F$ ; Fig. 2B, yellow stick bonds), using nicotinamide adenine dinucleotide

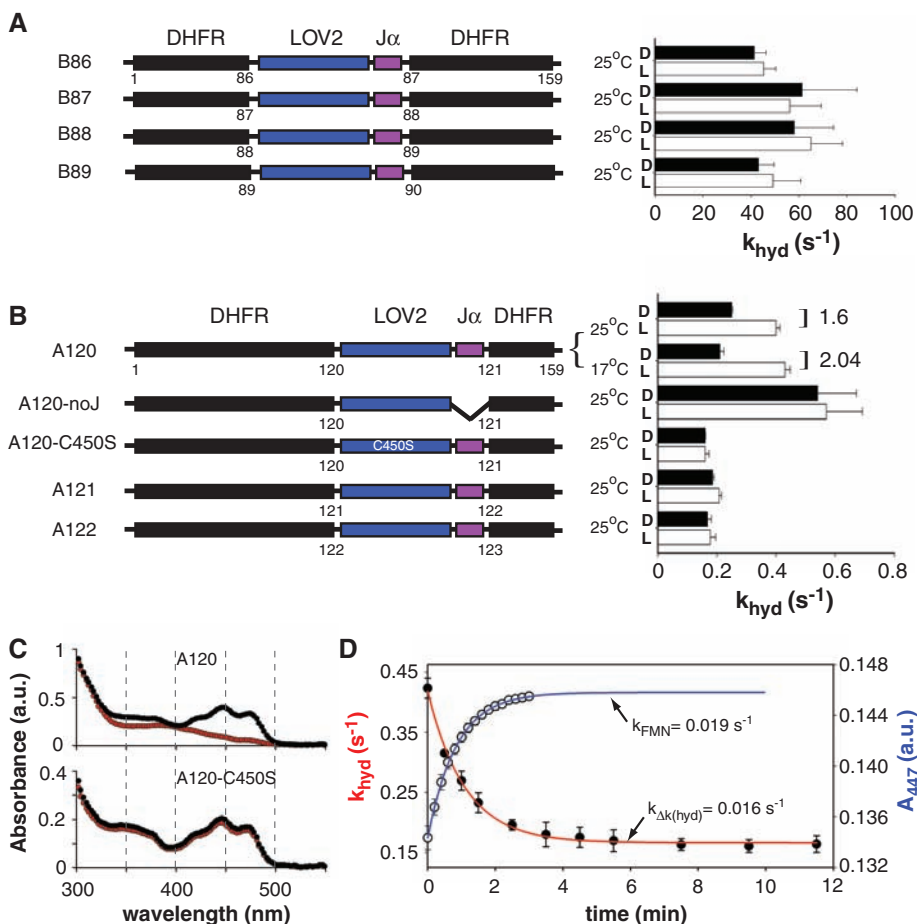
phosphate (NADPH; Fig. 2B, green stick bonds) as a cofactor (23). The structure of DHFR comprises a central eight-stranded  $\beta$  sheet ( $\beta$  strands A to H) and four flanking  $\alpha$  helices ( $\alpha B$ ,  $\alpha C$ ,  $\alpha E$ , and  $\alpha F$ ) that make up an active-site cleft that positions the substrate and cofactor for the catalytic step: transfer of the pro-R hydride from NADPH to position C-6 of  $H_2F$  (fig. S5A) (24). Changes in both the structure and dynamics of loops surrounding the active site are implicated in DHFR activity (25); in particular, the dynamics of the  $\beta F$ - $\beta G$  loop (residues 116 to 132) is thought to control transition-state stabilization and to enhance the rate of hydride transfer ( $k_{hyd}$ ). Consistent with these findings, SCA for an alignment of 418 members of the DHFR family uniquely identifies the  $\beta F$ - $\beta G$  loop as the most distant surface-exposed site showing strong correlated evolution with the enzyme active site (Fig. 2B). [See (26) and fig. S5B for details about SCA mapping and known allosteric mechanism in DHFR.]

To functionally couple the light-dependent allosteric mechanism in the LOV2 domain with the DHFR catalytic mechanism, we created sets of chimeric proteins in which the core LOV2 domain is inserted via its N- and C-terminal helical extensions into DHFR at two different surface sites (sites A and B; Fig. 2B and Fig. 3, A and B). Site A chimeras interrogate the computationally

and experimentally defined allosteric surface (A120 to A122,  $\beta F$ - $\beta G$  loop), and site B chimeras interrogate another surface site (B86 to B89,  $\alpha C$ - $\beta E$  loop) that is similarly distant but statistically uncorrelated with the active site; site B serves as a control for potential nonspecific coupling between the two domains.

We began by evaluating the independent activities of the DHFR and LOV2 domains in context of the chimeras. All LOV2-DHFR chimeras were well expressed and rescued growth in the DHFR auxotrophic *E. coli* strain (ER2566 $\Delta fol\Delta thy$ ) under minimal media conditions (27) (fig. S6), which indicated that insertion of the LOV2- $J\alpha$  domain did not abolish DHFR activity in any instance. Consistent with the known mutagenic sensitivity of the  $\beta F$ - $\beta G$  loop (25), site A chimeras showed a factor of  $\sim 1000$  reduction in hydride transfer rate (Fig. 3B, wild-type  $k_{hyd} = 220 \text{ s}^{-1}$ ), a value that translates to a factor of  $\sim 60$  change in the overall turnover rate  $k_{cat}$ . Site B chimeras showed a more modest decrease in catalytic activity (by a factor of 5 to 7), consistent with the prediction that this site is less influential for active-site function. With regard to LOV2 domain function, the absorbance spectra of dark-equilibrated proteins showed clear evidence of a 447-nm peak consistent with a noncovalently bound FMN chromophore (28), and light activation of all the chimeric enzymes triggered a char-

**Fig. 3.** Light-dependent catalysis in a LOV2-DHFR chimera. **(A and B)** Schematics of the chimeric constructs in sites A and B and associated hydride transfer rates ( $k_{hyd}$ ) carried out under single-turnover conditions. Data are measured either upon dark adaptation (black bars) or immediately after a 5-min exposure to intense white light (white bars) (26). The A120-noJ chimera lacks the C-terminal  $J\alpha$  helix (a major site of light-dependent conformational change in LOV2), and the A120-C450S chimera carries a point mutation that locks LOV2 in the dark state. No site B chimeras show light-dependent catalysis, but one site A chimera (A120) shows a modest but clear increase in  $k_{hyd}$  upon light activation. **(C)** Dark (black curves) and light-activated (red curves) absorbance spectra in the A120 and A120-C450S chimeras show the characteristic 447-nm peak of the FMN chromophore in the dark state of LOV2; in A120 alone, light activation shows the expected transition to the 390 nm-absorbing lit-state species. **(D)** The kinetics of dark recovery of the catalytic rate follows a single-exponential process (red curve), which closely mimics that kinetics of recovery of the 447 nm-absorbing dark state of LOV2 (blue curve). In all panels, error bars indicate SD.



acteristic spectral shift to a 390 nm-absorbing species due to formation of a covalent thiol-FMN adduct (Fig. 3C and fig. S7). This species showed single-exponential relaxation to the dark-state spectrum (Fig. 3D and table S4) with a rate constant ( $0.019 \text{ s}^{-1}$ ) nearly identical to that reported for the isolated domain (29). These findings show that the basic intrinsic features of the PAS domain and DHFR are structurally and functionally intact in the LOV2-DHFR chimeras.

We next examined the chimeras for light-dependent DHFR activity by comparing  $k_{\text{hyd}}$  for matched samples either maintained in the dark or immediately after exposure to light. Rates were measured under single-turnover conditions designed to minimize the effects of dark relaxation of the LOV2 domain. Consistent with the prediction that site B is uncoupled from active-site function, none of the site B chimeras showed light-dependent changes in enzyme activity (Fig. 3A). In contrast, one of the site A chimeras (A120) showed a modest but clear light-dependent DHFR activity (Fig. 3B and table S3). By varying temperature and pH conditions (fig. S8), we determined that light increased the catalytic rate of the enzyme by a factor of 2 at  $17^\circ\text{C}$  and by a factor of 1.6 at  $25^\circ\text{C}$  (Fig. 3B). The light-dependent effect depends on the known mechanism of signaling in the LOV2 domain; an A120 variant lacking the  $\alpha$  helix (19) (A120-noJ) or an A120 variant carrying a point mutation in LOV2 that is known to lock the molecule in the dark state (A120-C450S) (30, 31) showed no light dependence, although both constructs showed absolute DHFR activities similar to that of A120 (Fig. 3B). Shifting light-exposed A120 to the dark caused a decay of the light-dependent increase in  $k_{\text{hyd}}$  that followed a single-exponential relaxation ( $k_{\Delta t(\text{hyd})}$ ,  $0.016 \text{ s}^{-1}$ ) (Fig. 3D), observable through several cycles of excitation. The relaxation of enzyme activity nearly matched the rate of thermal relaxation of the LOV2 domain to the dark state ( $0.019 \text{ s}^{-1}$ , Fig. 3D) (29); this result implies that the light-dependent enzymatic activity in A120 is due to establishment of allosteric communication between LOV2 and DHFR.

The magnitude of the effect is small and hardly optimal, given the factor of  $\sim 600$  allosteric effect intrinsic to LOV2- $\alpha$  (32). In addition, it is

clear that not every chimera made at site A showed light-dependent activity (Fig. 3B). Nonetheless, these results show that site-specific connections at allosteric surfaces can, even without directed optimization or mechanism-based design, begin to produce coupled activities in designed chimeric proteins.

In addition to the hydride transfer step, the overall turnover cycle of DHFR involves several other reactions, including  $\text{NADP}^+/\text{NADPH}$  exchange after reduction of substrate and rate-limiting release of the product  $\text{H}_4\text{F}$  (26, 33). To examine potential allosteric control of these processes in the A120 chimera, we measured the light dependence of the  $\text{H}_4\text{F}$  off-rate and of the equilibrium dissociation constant for NADPH (Fig. 4). The  $\text{H}_4\text{F}$  off-rate, measured via methotrexate (MTX) competition under conditions saturating in  $\text{H}_4\text{F}$ , NADPH, and MTX (34), showed a factor of  $\sim 3$  decrease in the A120 chimera, a value comparable to a previously characterized point mutation in the  $\beta\text{F}-\beta\text{G}$  loop [G121V (25)] (Fig. 4A). We observed a small but statistically significant light dependence on  $\text{H}_4\text{F}$  off-rate that was abrogated in the background of the dark-locked LOV2 mutant (C450S) (table S4). In contrast, we observed no light dependence in the equilibrium dissociation constant for the NADPH cofactor, although the LOV2 insertion into the  $\beta\text{F}-\beta\text{G}$  loop showed an effect on NADPH binding similar to that of the G121V mutant (Fig. 4B). Previous work shows that, separate from the structural contributions of the  $\beta\text{F}-\beta\text{G}$  loop to cofactor or substrate binding, the dynamics of this loop specifically controls  $k_{\text{hyd}}$  and product release (35). Thus, the engineered interdomain allostery in the A120 chimera likely works mechanistically through light-dependent modulation of  $\beta\text{F}-\beta\text{G}$  loop dynamics.

Taken together, the data presented here are consistent with the notion that modular allosteric networks in each protein can be brought together to initiate the formation of new allosteric control. The installation of light-dependent enzymatic activity in the A120 LOV2-DHFR chimera occurred with minimal disruption of the internal biochemical features of each module. Few discernible alterations to the photodynamics of LOV2 were noted, and the changes in DHFR properties were no greater than the effect of point

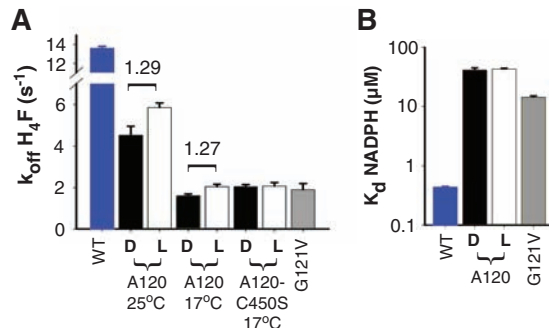
mutation at the surface site used for connection. Although the number of sites tested here is small, the emergence of interdomain allostery through insertion at the  $\beta\text{F}-\beta\text{G}$  loop but not at the  $\alpha\text{C}-\beta\text{E}$  loop is in agreement with the proposal that specific surface locations might act as evolutionarily conserved hotspots for allosteric control. Allosteric effects in proteins could also arise through idiosyncratic variation in individual family members (36), but these results suggest the notion that the linkage of conserved networks of amino acid interactions might represent a statistically preferential strategy for the evolution of allosteric signaling in multidomain proteins.

The engineering of light-dependent allosteric control in the LOV2-DHFR chimera represents an initial step toward a general scheme for the creation of allosteric multidomain systems. As methods become more refined, the computational prediction of potential allosteric surface sites should be combined with physics-based interface design and experimental screening to design high-performance allosteric systems.

#### References and Notes

1. F. M. Richards, W. A. Lim, *Q. Rev. Biophys.* **26**, 423 (1993).
2. J. Tsai, R. Taylor, C. Chothia, M. Gerstein, *J. Mol. Biol.* **290**, 253 (1999).
3. J. F. Swain, L. M. Gierasch, *Curr. Opin. Struct. Biol.* **16**, 102 (2006).
4. S. J. Benkovic, S. Hammes-Schiffer, *Science* **301**, 1196 (2003).
5. S. Hammes-Schiffer, S. J. Benkovic, *Annu. Rev. Biochem.* **75**, 519 (2006).
6. K. Henzler-Wildman, D. Kern, *Nature* **450**, 964 (2007).
7. I. Luque, S. A. Leavitt, E. Freire, *Annu. Rev. Biophys. Biomol. Struct.* **31**, 235 (2002).
8. S. W. Lockless, R. Ranganathan, *Science* **286**, 295 (1999).
9. G. M. Suel, S. W. Lockless, M. A. Wall, R. Ranganathan, *Nat. Struct. Biol.* **10**, 59 (2003).
10. A. D. Ferguson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 513 (2007).
11. M. E. Hatley, S. W. Lockless, S. K. Gibson, A. G. Gilman, R. Ranganathan, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14445 (2003).
12. A. I. Shulman, C. Larson, D. J. Mangelsdorf, R. Ranganathan, *Cell* **116**, 417 (2004).
13. S. M. Garrard *et al.*, *EMBO J.* **22**, 1125 (2003).
14. F. C. Peterson, R. R. Penkert, B. F. Volkman, K. E. Prehoda, *Mol. Cell* **13**, 665 (2004).
15. M. Salomon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12357 (2001).
16. B. L. Taylor, I. B. Zhulin, *Microbiol. Mol. Biol. Rev.* **63**, 479 (1999).
17. Y. Z. Gu, J. B. Hogenesch, C. A. Bradfield, *Annu. Rev. Pharmacol. Toxicol.* **40**, 519 (2000).
18. A. S. Halavaty, K. Moffat, *Biochemistry* **46**, 14001 (2007).
19. S. M. Harper, L. C. Neil, K. H. Gardner, *Science* **301**, 1541 (2003).
20. M. A. Cusanovich, T. E. Meyer, *Biochemistry* **42**, 4759 (2003).
21. P. J. Erbel, P. B. Card, O. Karakuzu, R. K. Bruick, K. H. Gardner, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15504 (2003).
22. R. Brudler *et al.*, *J. Mol. Biol.* **363**, 148 (2006).
23. P. T. Rajagopalan, S. J. Benkovic, *Chem. Rec.* **2**, 24 (2002).
24. M. R. Sawaya, J. Kraut, *Biochemistry* **36**, 586 (1997).
25. C. E. Cameron, S. J. Benkovic, *Biochemistry* **36**, 15792 (1997).
26. See supporting material on Science Online.

**Fig. 4.** Light dependence of product off-rate (A) and cofactor binding (B) in the A120 chimera. In (A), the product dissociation rate ( $k_{\text{off}, \text{H}_4\text{F}}$ ) shows a small light dependence in A120 (factor of  $\sim 1.3$ ) that is abrogated in the background of the dark-locked C450S mutation. (B), Cofactor binding ( $K_d$ , NADPH) shows no light dependence. In both assays, the overall effect of LOV2 domain insertion between positions 120 and 121 in the  $\beta\text{F}-\beta\text{G}$  loop is similar to that of the point mutation G121V. Error bars indicate SD.



27. M. C. Saraf, A. R. Horswill, S. J. Benkovic, C. D. Maranas, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4142 (2004).
28. J. M. Christie, M. Salomon, K. Nozue, M. Wada, W. R. Briggs, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8779 (1999).
29. M. Salomon, J. M. Christie, E. Knieb, U. Lempert, W. R. Briggs, *Biochemistry* **39**, 9401 (2000).
30. J. M. Christie, T. E. Swartz, R. A. Bogomolni, W. R. Briggs, *Plant J.* **32**, 205 (2002).
31. S. M. Harper, J. M. Christie, K. H. Gardner, *Biochemistry* **43**, 16184 (2004).
32. X. Yao, M. K. Rosen, K. H. Gardner, *Nat. Chem. Biol.* **4**, 491 (2008).
33. C. A. Fierke, K. A. Johnson, S. J. Benkovic, *Biochemistry* **26**, 4085 (1987).
34. J. T. Chen, K. Taira, C. P. Tu, S. J. Benkovic, *Biochemistry* **26**, 4093 (1987).
35. D. M. Epstein, S. J. Benkovic, P. E. Wright, *Biochemistry* **34**, 11037 (1995).
36. J. Kuriyan, D. Eisenberg, *Nature* **450**, 983 (2007).
37. We thank members of the Ranganathan and Benkovic labs for discussions, M. Rosen and K. H. Gardner for discussions and sharing of unpublished data and materials, and J. Wang for solvent accessibility calculations. Supported by the Robert A. Welch

foundation (R.R.) and a grant from the Defense Advanced Research Projects Agency (R.R. and S.J.B.).

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5900/438/DC1  
Materials and Methods

Figs. S1 to S8

Tables S1 to S4

References

14 April 2008; accepted 9 September 2008

10.1126/science.1159052

# A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell

Paul J. Choi,\* Long Cai,\*† Kirsten Frieda,‡ X. Sunney Xie§

By monitoring fluorescently labeled lactose permease with single-molecule sensitivity, we investigated the molecular mechanism of how an *Escherichia coli* cell with the *lac* operon switches from one phenotype to another. At intermediate inducer concentrations, a population of genetically identical cells exhibits two phenotypes: induced cells with highly fluorescent membranes and uninduced cells with a small number of membrane-bound permeases. We found that this basal-level expression results from partial dissociation of the tetrameric lactose repressor from one of its operators on looped DNA. In contrast, infrequent events of complete dissociation of the repressor from DNA result in large bursts of permease expression that trigger induction of the *lac* operon. Hence, a stochastic single-molecule event determines a cell's phenotype.

Genetically identical cells in the same environment can exhibit different phenotypes, and a single cell can switch between distinct phenotypes in a stochastic manner (1–4). In the classic example of lactose metabolism in *Escherichia coli*, the *lac* genes are fully expressed for every cell in a population under high extracellular concentrations of inducers, such as the lactose analog methyl- $\beta$ -D-thiogalactoside (TMG). However, at moderate inducer concentrations, the *lac* genes are highly expressed in only a fraction of a population, which may confer a fitness advantage for the entire population (5). We studied the molecular mechanism that controls the stochastic phenotype switching of a single cell.

Lactose metabolism is controlled by the *lac* operon (6, 7), which consists of the *lacZ*, *lacY*, and *lacA* genes encoding  $\beta$ -galactosidase, lactose permease, and transacetylase, respectively. Expression of the operon is regulated by a transcription factor, the *lac* repressor (8), which dissociates from its specific binding sequences of DNA, the *lac* operators, in the presence of an inducer to allow transcription (Fig. 1A). The production of

the permease increases inducer influx (9), which results in positive feedback on the permease expression level. Above a certain threshold of permease numbers, a cell will remain in a phenotype that is capable of lactose metabolism, and below this threshold, a cell will remain in a phenotype that is incapable of lactose metabolism (10, 11). The former has high fluorescence from the cell membrane when the permease is labeled with a yellow fluorescent protein (YFP), whereas the latter has low fluorescence. The image in Fig. 1B shows the coexistence of both phenotypes in a cell population at intermediate inducer concentrations, characterized by the bimodal distributions of fluorescence intensity in Fig. 1C.

Although it is known that the bistability in the *lac* operon arises from positive feedback (12, 13), the molecular mechanism underlying the initiation of switching between two phenotypes remains unclear. Novick and Weiner deduced that switching from the uninduced state to the induced state occurs through a single rate-limiting molecular process (12) rather than a multistep process. They further hypothesized that the random expression of one molecule of permease was enough to trigger induction. However, this has never been observed experimentally because of insufficient sensitivity.

Our group has previously shown that a single fluorescent protein molecule can be visualized in a living bacterial cell by using the method of detection by localization (14–16). Because a permease molecule is a membrane protein with slow

diffusion, its fluorescence label is highly localized as compared with a fluorescent protein in the cytoplasm, which allows for its detection above the background of cellular autofluorescence. Thus, we generated strain SX700, which possesses intact *lac* promoter elements and expresses a functional LacY-YFP fusion protein (fig. S1) from *lacY*'s native chromosomal position (fig. S2). Figure 1B shows a fluorescence image of SX700 cells that allows us to directly count the number of single LacY-YFP molecules present in a cell.

Figure 1D shows the histogram of permease copy numbers in the uninduced fraction of cells, free from the complication of autofluorescence background in the lower peak of Fig. 1C. Evidently, the uninduced cells have 0 to 10 LacY molecules with significant probability, independent of the inducer concentration. Had one permease been enough to trigger induction, all cells in the uninduced subpopulation would have contained zero LacY molecules. Thus, we conclude that a single copy of LacY is not sufficient to induce switching of the phenotype, and the threshold for induction must be much higher than several molecules per cell.

We then set out to determine the threshold of permease molecules for induction. It is experimentally difficult to capture the rare events of phenotype switching in real time. To overcome this difficulty, we prepared cells covering a broad range of permease copy numbers by first fully inducing the cells and subsequently washing out the inducer. We then allowed the cells to divide for one to six generations, during which the initial permeases were partitioned into daughter cells (17). Figure 2A shows fluorescence time traces of the cells, normalized by cell size, upon the reintroduction of 40  $\mu$ M TMG. Although the fluorescence in cells with low permease numbers continued to decay because of cell division and photobleaching, cells with more than 300 initial permease molecules induced again within 3 hours and showed increased fluorescence. Figure 2B shows the probability of induction as a function of initial permease number, which is well fit by a Hill equation with a Hill coefficient of 4.5 and a threshold of  $\sim$ 375 molecules. The large value of the threshold indicates that hundreds of permease molecules are necessary to switch the phenotype.

If the induction is due to a single rate-limiting event as Novick and Weiner argued, there must be a single large burst of permease expression to reach this high threshold. However, only small

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.

\*These authors contributed equally to this work.

†Present address: Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

‡Present address: Biophysics Program, Stanford University, Stanford, CA 94305, USA.

§To whom correspondence should be addressed. E-mail: xie@chemistry.harvard.edu