not *FLO11*) requires *SIR3* and the yeast Ku genes, which are also required for TPE (Rusche et al., 2003). In contrast to TPE, however, silencing of *FLO10* and *FLO11* does not require *SIR2* and is promoter dependent since heterologous promoters placed at the *FLO10* and *FLO11* loci are not silenced. Thus, *FLO10* and *FLO11* epigenetic regulation is different from that seen at telomeres or at the silent mating loci; future work will shed light on the mechanistic details of this novel form of regulation.

How does the cell decide which *FLO* gene to express? In *C. glabrata*, a yeast pathogen closely related to *S. cerevisiae*, TPE governs expression of at least a subset of the *EPA* genes, a family of cell-surface adhesins (De Las Penas et al., 2003). Cells mutant in the silencing apparatus show expression of telomeric *EPA* genes that are normally silent. In *S. cerevisiae*, by contrast, strains mutant for *HST1/2* (or *SIR3*) do not express *FLO10*, suggesting that epigenetic chromatin modifications are not the primary regulators of which *FLO* gene is expressed. While chromatin modifications, as described above, play a role in determining what percent of cells in a population competent to express *FLO11* or *FLO10* actually do express the gene, the key regulatory decision of whether the cell population expresses *FLO10* at all depends on a different mechanism. *FLO10* is expressed only in mutants that were isolated at high frequency (10^{-3}) as colony morphology mutants. These mutants all contained inactivating mutations in the *IRA1* or *IRA2* genes, which encode GAPs (GTPase-activating proteins) for yeast Ras (Tanaka et al., 1990). Given the implication of protein kinase A signaling in induction of *FLO10* and *FLO11* expression (Halme et al., 2004 and references therein), a reasonable model is that increased Ras activity in the *ira1* or *ira2* mutants is responsible for derepression of *FLO10*.

Is it credible that mutation of *IRA1* or *IRA2* is the normal mechanism by which *FLO10* is switched on outside the laboratory environment? Presumably, if mutational inactivation is used as a regulatory mechanism, there must be a mechanism for reversion to re-repress *FL010*. In this regard, four out of the six identified *ira* mutations are frameshift insertions or deletions within long stretches of AT or TA base pairs. At least for these frameshift mutations, slipped strand mispairing within the AT/TA-rich tracts during DNA replication might also provide a mechanism for efficient reversion of the Flo10^{+} cells to a Flo10^{-} phenotype.

There is precedent for regulation of surface components by mutation. In *Neisseria meningitidis*, for example, slipped strand mispairing in the SiaD capsule biosynthesis gene governs phase variation between encapsulated and nonencapsulated meningococcal forms (Hammerschmidt et al., 1996). In a second key example, clinical *Pseudomonas aeruginosa* isolates from cystic fibrosis patients are mucoid, being gangbuster producers of alginate, an extracellular polysaccharide important in biofilm formation as well as bacterial survival in the lung. The majority of mucoid isolates from CF lungs have mutations (primarily frameshifts) in *MucA*, an anti-sigma factor that normally antagonizes the alternative sigma factor *AlgU*, which drives alginate expression (Boucher et al., 1997). Thus, in these pathogens, mutation of a key regulator drives adaptation to the environment. In *S. cerevisiae*, mutational disruption of Ras homeostasis may be the critical regulatory step in altering surface expression of *FLO* gene family members.

Candida experts reading this paper will be struck by the similarity in colony morphology of the *ira1* and *ira2* mutants to certain colony morphology phenotypes in *C. albicans*. *C. albicans* can flip between metastable colony morphology types (called switch phenotypes) at rates of around 10^{-3} to 10^{-4} (Soll, 1992). It will be interesting to determine if high-frequency mutational disruption of Ras signaling might underlie some of the colony switching behavior of *C. albicans*, for which the molecular basis is still a mystery.

Brendan Cormack
Department of Molecular Biology and Genetics
Johns Hopkins University School of Medicine
Baltimore, Maryland 21210

Selected Reading

Barry, J.D., Ginger, M.L., Burton, P., and McCulloch, R. (2003). Int. J. Parasitol. *33*, 29–45.

Boucher, J.C., Yu, H., Mudd, M.H., and Deretic, V. (1997). Infect. Immun. *65*, 3838–3846.

De Las Penas, A., Pan, S.J., Castano, I., Alder, J., Cregg, R., and Cormack, B.P. (2003). Genes Dev. *17*, 2245–2258.

Hagblom, P., Segal, E., Billyard, E., and So, M. (1985). Nature *315*, 156–158.

Halme, A., Bumgarner, S., Styles, C., and Finu, G.R. Cell *116*, this issue, 405–415.

Hammerschmidt, S., Muller, A., Sillmann, H., Muhlenhoff, M., Borrow, R., Fox, A., van Putten, J., Zollinger, W.D., Gerardy-Schahn, R., and Frosch, M. (1996). Mol. Microbiol. *20*, 1211–1220.

Reynolds, T.B., and Fink, G.R. (2001). Science *291*, 878–881.

Rusche, L.N., Kirchmaier, A.L., and Rine, J. (2003). Annu. Rev. Biochem.*72*, 481–516.

Soll, D.R. (1992). Clin. Microbiol. Rev. *5*, 183–203.

Sundstrom, P. (2002). Cell. Microbiol. *4*, 461–469.

Tanaka, K., Nakafuku, M., Satoh, T., Marshall, M.S., Gibbs, J.B., Matsumoto, K., Kaziro, Y., and Toh-e, A. (1990). Cell *60*, 803–807.

Allostery and Coupled Sequence Variation in Nuclear Hormone Receptors

The analysis of correlated sequence variation in evolutionarily related proteins is beginning to provide useful information regarding allosteric coupling between different functional sites. Such an analysis has been carried out for the nuclear hormone receptors, and the conclusions tested by making mutations that switch the allosteric response to ligands of RXR heterodimers.

In our quest to understand how proteins work, things would be much easier if protein domains were rigid and simply did whatever it is that they are supposed to do without jiggling around too much. Unfortunately, the fluctuating nature of proteins has been apparent since

the first protein structure was determined, when it was soon realized that the static structure of the oxygen binding protein myoglobin contains no path from the outside to the internal site where oxygen is bound to the heme iron. Thus, even in this monomeric and apparently non-allosteric cousin of hemoglobin, protein dynamics is an essential element of proper function. To make things worse, recent studies are pointing to a level of complexity in myoglobin that was not previously appreciated (Frauenfelder et al., 2003). Instead of hopping in and out of the ligand binding pocket by using a proximal and readily available histidine-gated door, ligands appear instead to explore several conserved interior cavities in myoglobin, jumping from the proximal to the distal side of the heme group. These cavities are now thought to be important for an emerging function of myoglobin in nitric-oxide metabolism, one in which allostery may also be a component (Frauenfelder et al., 2003). Thus, instead of being the simpleton of the oxygen binding family, myoglobin is revealing itself as a surprisingly sophisticated molecular machine.

Myoglobin provides one illustration of the general concept that proteins often have features to their design that go beyond what might be expected for the chemical or binding events that represent the central or best understood function of the protein. These nuances in design provide nature with opportunities for the development of highly complex allosteric proteins, which evolve through the interlocking of many such responsive protein elements. These elements are not always self-evident in static views of crystal structures, and defining the pathways of allosteric communication in proteins remains a major challenge in structural biology. One potential route forward is to take advantage of the fact that aspects of design that are important for function are likely to be preserved through evolution, leaving a trace in the pattern of sequence conservation that is characteristic of members of the protein family.

Comparative sequence analyses provide extremely powerful probes of protein function because of the sheer breadth of sequence information that is now available for variant forms of most proteins. Comparative alignments, particularly those that are leveraged by three dimensional structural information, often lead to the discovery of unexpected interactions or modulatory elements that were not anticipated in the original structural and functional characterization of the protein. To pick just one of many recent examples, an evolutionary analysis of substitution patterns in the hormone leptin, combined with knowledge of the structure, suggests that the human leptin protein has acquired an additional interaction surface that is distinct from the receptor binding site (Gaucher et al., 2003). This observation might help guide studies aimed at explaining why the human protein behaves differently in terms of its effects on body weight when compared to leptin in other mammals that lack the second interaction site. This kind of "first-order" analysis, which treats each position in the protein chain independently, is powerful at identifying functional elements within proteins, but stops short of providing direct information regarding allosteric coupling between such functional sites. This shortcoming of comparative sequence analysis was the issue addressed by a stimulating paper published in 1999 by Lockless and Ranganathan, who realized that the dramatic growth in sequence databases now enables statistically significant "second-order" analysis of patterns of sequence conservation in proteins (Lockless and Ranganathan, 1999).

Lockless and Ranganathan did something that is pretty obvious, but which is difficult to do reliably unless one has an alignment of a large number of sequences of sufficiently divergent forms of a protein. After first obtaining a multiple sequence alignment for 274 PDZ domains, they analyzed the statistical coupling between the amino acid distributions at different positions along the protein chain. This analysis identifies pairs of positions in the chain that exhibit some mutual interdependence in sequence, in at least a subset of the protein sequences. By clustering these linkages together, they identified "networks" of residues that are statistically coupled and which appear to link different regions of the protein together. Thermodynamic mutant cycle analysis, a method pioneered by Alan Fersht for studying energetic coupling between residues (Carter et al., 1984), was used to verify experimentally that the statistical coupling in "evolutionary space" corresponded to energetic coupling in "real space." This correspondence suggests that the network of statistically coupled residues may be relevant for understanding how conformational changes are transmitted through the protein during allosteric modulation (Figure 1).

This statistical coupling analysis has since been applied by Ranganathan and coworkers to the G protein-coupled receptors, serine proteases, and even the hemoglobin family (although the authors do not comment on myoglobin's new found notoriety) (Suel et al., 2003). In each case, networks of coupled residues are discovered that do not appear to be trivial reflections of the architectural elements of the structure. In this issue of *Cell*, Mangelsdorf, Ranganathan, and coworkers (Shulman et al., 2004) present a detailed dissection of ligand-mediated allostery in nuclear hormone receptors, in which the statistical coupling analysis is put to the test. RXR heterodimers are nuclear receptors in which the retinoid X receptor (RXR) is coupled with another nuclear hormone receptor. The ligand binding domain of each of the two hormone receptors can be activated by distinct ligands, and the combinatorial response of the heterodimer depends on the nature of the RXR partner. Allosteric communication between the two ligand binding domains and the DNA binding domain leads to different responses, depending on the RXR partner. In permissive heterodimers, ligand binding to either partner evokes a response, and an increased response is obtained when both ligands are bound. In conditional heterodimers, the RXR domain is inert unless the partner protein has its ligand binding site occupied.

The new paper presents a statistical coupling analysis for 250 nuclear hormone receptor ligand binding domains, which leads to the identification of a coupled network of 27 residues that link the heterodimerization interface, the ligand binding domain, and two regions that are important for transmission of the signal. The key result of the paper is that the mutation of residues that are within this network can result in fundamental alteration in the properties of the heterodimer, converting permissive heterodimers into conditional ones. Residues that are located outside the network, but

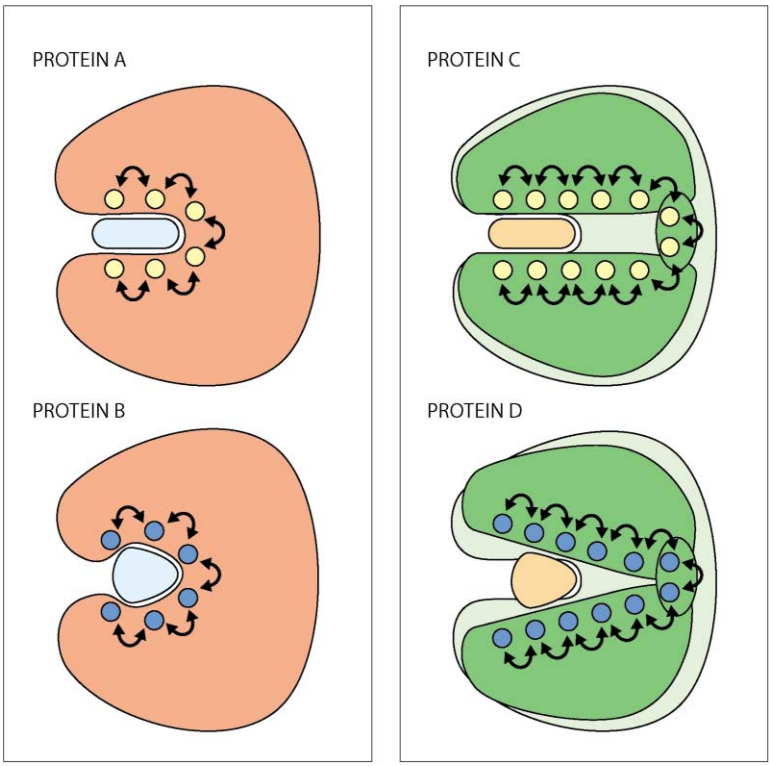

Figure 1. Schematic Diagram Illustrating Correlated Sequence Variation in Protein Families

The left panel shows two proteins, A and B, which belong to a family in which specificity toward different ligands has evolved as a consequence of local sculpting of the active site. Correlated changes in amino acid sequence across the family are most obvious in regions immediately adjacent to the active site (yellow and blue circles, with correlated changes indicated by curved arrows). The panel on the right shows two proteins, C and D, that are members of family in which a hinge bending motion underlies differential specificity toward ligands. In this case, correlations in sequence variation extend through a network of residues (yellow and blue circles) that extend away from the active site and into the hinge region. Proteins in this family could exhibit allosteric behavior if the hinge region is coupled to some other function, such as a second ligand binding site. Illustration by Lore Leighton.

which are otherwise spatially close to the important regions of the protein, do not yield such dramatic effects when mutated. The statistical coupling analysis thus appears to be a useful tool for identifying the subset of residues that are critically important for allosteric function.

An upcoming issue of *Molecular Cell* (Nettles et al., 2004) contains another analysis of allostery in nuclear hormone receptors, one that is complementary in its approach to the RXR heterodimer study. While the RXR paper takes a more global approach to the family of nuclear hormone receptors, Greene and coworkers present the results of a more focused study of two closely related estrogen receptor subtypes, ERα and ERβ, that share high sequence similarity yet differ in their response to ligands. Taking advantage of the remarkable ability of some compounds to stimulate one of these two receptors but inhibit the other, Greene and coworkers probe the origins of the differential specificity of the receptors. Based on the analysis of a number of chimeric receptors, in which various portions of the two receptors are swapped with corresponding segments of the other receptor, as well as structure-guided mutagenesis, they reach the conclusion that the recognition of various ligands depend on a spatially distributed but coupled set of structural features. This principle of ligand recognition by proteins has been appreciated for some time, but the importance of allosteric effects for inhibitor design in clinically relevant targets such as the estrogen receptor is now coming to the forefront.

In contrast to the RXR paper, which relies primarily on the sequence correlation analysis to define the allosteric network, the paper on the estrogen receptors shows us how the accumulation of structural information regarding various states of activation of nuclear receptors is now leading to insights into the allosteric mechanism of these proteins. Fascinating as both papers are, it is still the case that neither study provides a satisfying physical mechanism for how the various sites on these proteins actually couple to each other energetically. It is hoped that, eventually, computer simulations of protein dynamics (Karplus and McCammon, 2002) will enable accurate mapping of the physics of interatomic collisions into the biology of allosteric communication.

John Kuriyan
Howard Hughes Medical Institute
Departments of Molecular and Cell Biology and of Chemistry
University of California
Berkeley, California 94707

Selected Reading

Carter, P.J., Winter, G., Wilkinson, A.J., and Fersht, A.R. (1984). Cell *38*, 835–840.

Frauenfelder, H., McMahon, B.H., and Fenimore, P.W. (2003). Proc. Natl. Acad. Sci. USA *100*, 8615–8617.

Gaucher, E.A., Miyamoto, M.M., and Benner, S.A. (2003). Genetics *163*, 1549–1553.

Karplus, M., and McCammon, J.A. (2002). Nat. Struct. Biol. *9*, 646–652.

Lockless, S.W., and Ranganathan, R. (1999). Science *286*, 295–299.

Nettles, K.W., Sun, J. Radek, J.T., Sheng, S., Rodriguez, A.L., Katzenellenbogen, J.A., Katzenellenbogen, B.S., and Greene, G. (2004). Mol. Cell *13*, in press.

Shulman, A.I., Larson, C., Mangelsdorf, D.J., and Ranganathan, R. (2004). Cell *116*, this issue, 417–429.

Suel, G.M., Lockless, S.W., Wall, M.A., and Ranganathan, R. (2003). Nat. Struct. Biol. *10*, 59–69.