Evolution of Eukaryotic Transcription Circuits

AAAS

In these cases, it is the speed of ligand binding and the speed of action of RNA polymerase that determine what metabolite concentration will regulate gene expression. This can be an advantage for riboswitches in their evolutionary competition with protein factors. Cells can change the concentration set point for a metabolite by mutating the aptamer or protein receptor to strengthen or weaken binding. With kinetically driven riboswitches, cells can tune the set point by orders of magnitude just by accruing mutations that alter the speed of riboswitch folding, or that alter the speed with which RNA polymerase completes riboswitch synthesis.

Riboswitch Discovery

Growing DNA sequence databases and improving bioinformatics tools are rapidly expanding the number of discovered riboswitch classes in bacteria (12, 29, 30), and similar efforts in eukaryotes might also begin to uncover more riboswitches in these organisms. Currently, few eukaryotic riboswitches have been discovered, and in the evolutionary race for dominance, perhaps proteins have put more distance between themselves and their ancient RNA counterparts in this branch of the evolutionary tree. Given that eukaryotes appear to express many noncoding

RNAs, researchers will need to keep open the possibility that riboswitches might also control the production and actions of these RNAs as well. As is the case with bacteria, if a protein factor cannot be found for a signaling compound, perhaps a complex sensor made of RNA is in action.

References and Notes

1. R. L. Coppins, K. B. Hall, E. A. Groisman, *Curr. Opin. Microbiol.* **10**, 176 (2007).
2. A. Nahvi *et al.*, *Chem. Biol.* **9**, 1043 (2002).
3. A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, M. S. Gelfand, *Trends Genet.* **20**, 44 (2004).
4. R. R. Breaker, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 3, 2006), pp. 89–107.
5. S. E. Osborne, A. D. Ellington, *Chem. Rev.* **97**, 349 (1997).
6. J. P. Gallivan, *Curr. Opin. Chem. Biol.* **11**, 612 (2007).
7. S. Thore, M. Leibundgut, N. Ban, *Science* **312**, 1208 (2006); published online 3 May 2006 (10.1126/science.1128451).
8. A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, *Nature* **441**, 1167 (2006).
9. D. J. Klein, A. R. Ferré-D'Amaré, *Science* **313**, 1752 (2006).
10. J. C. Cochrane, S. V. Lipchock, S. A. Strobel, *Chem. Biol.* **14**, 97 (2007).
11. T. E. Edwards, D. J. Klein, A. R. Ferré-D'Amaré, *Curr. Opin. Struct. Biol.* **17**, 273 (2007).
12. J. E. Barrick, R. R. Breaker, *Genome Biol.* **8**, R239 (2007).
13. R. T. Fuchs, F. J. Grundy, T. M. Henkin, *Nat. Struct. Mol. Biol.* **13**, 226 (2006).
14. S. D. Gilbert, R. P. Rambo, D. Van Tyne, R. T. Batey, *Nat. Struct. Mol. Biol.* **15**, 177 (2008).
15. N. Sudarsan, J. E. Barrick, R. R. Breaker, *RNA* **9**, 949 (2003).
16. T. Kubodera *et al.*, *FEBS Lett.* **555**, 516 (2003).
17. M. T. Cheah, A. Wachter, N. Sudarsan, R. R. Breaker, *Nature* **447**, 497 (2007).
18. S. Bocobza *et al.*, *Genes Dev.* **21**, 2874 (2007).
19. A. Wachter *et al.*, *Plant Cell* **19**, 3437 (2007).
20. M. T. Croft, M. Moulin, M. E. Webb, A. G. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20770 (2007).
21. R. Welz, R. R. Breaker, *RNA* **13**, 573 (2007).
22. N. Sudarsan *et al.*, *Science* **314**, 300 (2006).
23. A. Gutierrez-Preciado, R. A. Jensen, C. Yanofsky, E. Merino, *Trends Genet.* **21**, 432 (2005).
24. M. Mandal *et al.*, *Science* **306**, 275 (2004).
25. C. D. Stoddard, R. T. Batey, *ACS Chem. Biol.* **1**, 751 (2006).
26. R. Rieder, K. Lang, D. Graber, R. Micura, *ChemBioChem* **8**, 896 (2007).
27. J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, *Mol. Cell* **18**, 49 (2005).
28. W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside, S. M. Block, *Science* **319**, 630 (2008); published online 2 January 2008 (10.1126/science.1151298).
29. Z. Weinberg *et al.*, *Nucleic Acids Res.* **35**, 4809 (2007).
30. M. D. Kazanov, A. G. Vitreschak, M. S. Gelfand, *BMC Genomics* **8**, 347 (2007).
31. R. K. Montange, R. T. Batey, *Nature* **441**, 1172 (2006).
32. Riboswitch research in the Breaker laboratory is supported by grants from NIH and the Howard Hughes Medical Institute.

Evolution of Eukaryotic Transcription Circuits

Brian B. Tuch,[1,2] Hao Li,[1] Alexander D. Johnson[1,2]*

The gradual modification of transcription circuits over evolutionary time scales is an important source of the diversity of life. Over the past decade, studies in animals have shown how seemingly small molecular changes in gene regulation can have large effects on morphology and physiology and how selective pressures can act on these changes. More recently, genome-wide studies, particularly those in single-cell yeasts, have uncovered evidence of extensive transcriptional rewiring, indicating that even closely related organisms regulate their genes using markedly different circuitry.

Transcription of each gene in a eukaryotic organism is controlled by a collection of cis-regulatory sequences that are typically positioned in proximity to the coding sequence. The collection of cis-regulatory sequences associated with each gene specifies the time and place in the organism that the gene is to be transcribed. This information is read by sequence-specific DNA binding proteins [herein called transcription regulators (TRs)], which recognize these sequences and which themselves are typically expressed or active only at particular times and places in the life of the organism. It is the combination of active TRs present at a particular location and time that selects, through interactions with cis-regulatory sequences, those genes to be transcribed. Of course, there are many additional steps in transcription and in gene regulation; nevertheless, the cis-regulatory sequences and the TRs that recognize them form a critical layer of gene regulation.

Several properties of transcription regulation are especially important for considering its role in evolution (1–4). cis-regulatory sequences are

short (generally 5 to 10 nucleotides), degenerate (similar sequences confer equivalent TR binding), and their positions, relative to the gene whose transcription they control, can be variable. Different cis-regulatory sequences are often found close to one another, and TRs often bind cooperatively to these adjacent sites. This cooperative binding is a form of combinatorial control—the use of multiple, rather than single, TRs to control expression of a gene. cis-regulatory sequences often cluster into modules, each module acting independently to direct expression of the gene to a particular part of the organism at a specified time.

TRs are also modular and, in the laboratory, bits and pieces from different TRs can be recombined to produce novel types of regulation. Mutations can alter their DNA-binding specificity, their partner proteins, and their influence (activating or repressing) on transcription. Because many of the crucial protein-protein interactions made by TRs are relatively weak and nonspecific, even small changes to them can have large effects on gene regulation.

These properties make it easy to understand how new patterns of gene expression could arise through simple mutations. During the past decade, studies of single genes in animals have revealed many striking cases in which changes in cis-regulatory sequences likely underlie new morphological or physiological features (5, 6). These include the evolution of lactase persistence in humans (7), bone structures in fish (8), and trichomes (9) and pigmentation (10) in flies.

[1]Department of Biochemistry and Biophysics, University of California, San Francisco, CA, USA. [2]Department of Microbiology and Immunology, University of California, San Francisco, CA, USA.

*To whom correspondence should be addressed. E-mail: ajohnson@cgl.ucsf.edu

Gene Regulation
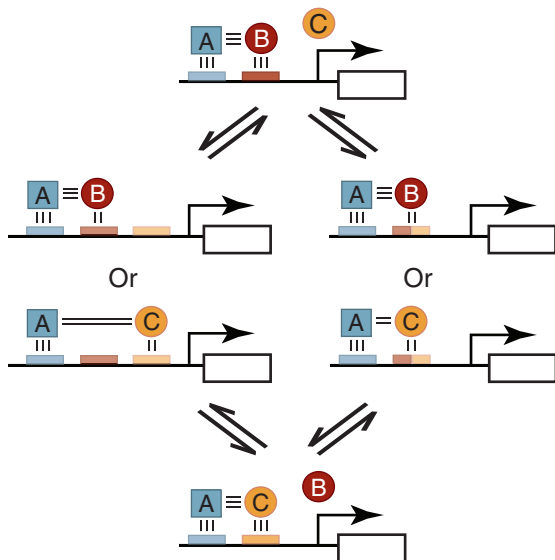
Fig. 1. Pathways to the rewiring of combinatorial circuitry. These two schemes can account for a handoff in the control of a gene (or a set of genes) from one TR to another. In both pathways an intermediate stage exists in which regulators B and C may act redundantly. Small black lines represent protein-protein and protein-DNA interactions, the number of these indicating the strength of the interaction. At any given time, each gene within a co-expressed set may have different control states (B only, C only, or B and C). The left pathway may be the route by which ribosomal genes and galactose-metabolizing genes were rewired in fungi (*14*, *19*). The right pathway is the likely route by which **a**-specific genes were rewired (*20*).

Although the emphasis is often placed on cis-regulatory sequences, changes to TRs can also underlie phenotypic change (*11*, *12*).

The animal studies outlined above typically start with an observable intra- or interspecies difference and trace its origins to changes in gene regulation at an individual locus. A complementary approach to studying evolutionary changes in transcriptional regulation (rewiring) begins with a molecular description of a transcription circuit, typically a large one consisting of several TRs and many target genes (i.e., genes they bind to and regulate). The circuit is then compared among two or more species. An advantage of this strategy is that it allows the entire landscape of circuit rewiring to be surveyed, without any bias as to the consequences of rewiring events to the organism. Of course, this is also its principal limitation; it is often difficult to discern whether the changes observed provide (or provided in the past) any benefit to the organism.

This genomic approach has been used to compare circuitry in closely related yeast, fly, and mammal species. Typically, bioinformatics, transcriptional profiling, and full-genome chromatin immunoprecipitation are used, often in combination. This approach has provided support for previous ideas and has also produced new insights.

First, high levels of transcriptional rewiring can occur over relatively short evolutionary time scales (*13*, *14*). Although the DNA-binding specificities of orthologous TRs (that is, TRs related by direct descent from the last common ancestor of a group of species) rarely differ substantially across species, the genes they directly regulate can differ considerably. For example, comparisons of the binding profiles of four liver-specific TRs across 4000 genes in mouse and human hepatocytes found that less than two-thirds of genes are conserved as targets of each TR (*15*). A genome-wide study of two TRs in three closely related yeast species (20 million years divergence) estimated that only a third of the TR–target gene connections seen in one species were preserved in the other two (*16*). Although some of these differences could be attributed to loss and gain of cis-regulatory sequences, others could not, and it remains to be seen what other types of molecular changes (e.g., changes in the activity level of TRs) contribute to this divergence.

A study that examined combinatorial circuitry involving the TR Mcm1 and its cofactors across three highly divergent yeasts (~300 million years divergence) also found evidence of massive rewiring (*17*). Only about 15% of the direct Mcm1–target gene interactions of *Saccharomyces cerevisiae* were preserved in two other yeast species. Mcm1 binds cooperatively to DNA with a set of cofactors to regulate many genes in each species, and the extensive rewiring observed was traced to high rates of gain and loss of cis-regulatory sequences as well as to the formation of new Mcm1-cofactor combinations and the breaking of old ones.
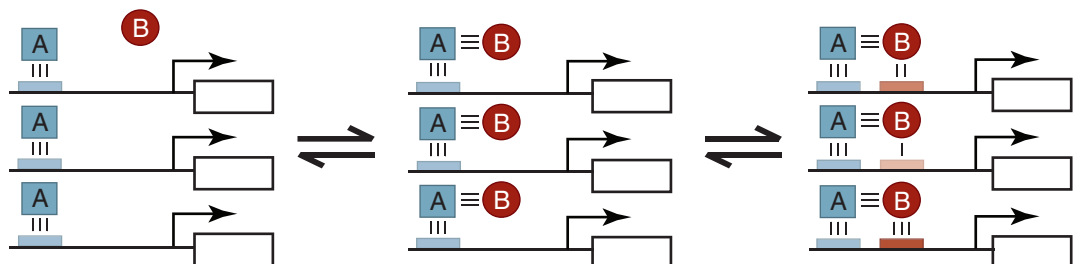
Second, genomic approaches have shown that the same set of coexpressed genes can be regulated by different mechanisms in different species. Earlier studies in flies showed that stabilizing selection can maintain the expression pattern of a single gene, while still allowing for considerable drift in the underlying regulatory mechanism [e.g., (*18*)]. Genome-wide studies in yeast have extended this idea, uncovering examples in which an entire group of genes remain coexpressed (that is, the genes respond as a group to changes in the environment or other perturbations) in different species, but the TR responsible for the regulation seems to have been swapped in one species relative to another. For example, in *S. cerevisiae* the presence of galactose induces transcription of genes that produce galactose-metabolizing enzymes via the TR Gal4. In another yeast, *Candida albicans*, the same enzymes are induced by galactose, but the Gal4 ortholog seems to have no role in this regulation; instead, these genes appear to be controlled by cis-regulatory sequences recognized by a different TR, and the Gal4 ortholog regulates glycolytic enzymes (*19*).

Another example occurs in mating-type regulation in fungi: In the lineage leading to *S. cerevisiae*, regulation of the coexpressed **a**-specific genes (transcribed in **a** cells and not in α cells) was "handed off" from a transcriptional activator to a transcriptional repressor (*20*). Because the activator and repressor are expressed in opposite cell types, the overall logic of the circuit is conserved. These replacements, of one TR with another, likely occurred through an intermediate state in which the target genes came under dual regulation, thus preserving coexpression throughout the transition (Fig. 1). Transition through a redundant intermediate has also been suggested for changes in the regulation of ribosomal genes in fungi (*14*).

It is not yet clear whether the rewiring of these coexpressed gene sets provides any advantage to the organism, as the overall expression pattern of the target genes seems, at least superficially, to have remained constant. It is possible that many examples of transcriptional rewiring are not adaptive at all but may simply reflect neutral evolution between alternative regulatory schemes (*21*).

Fig. 2. A plausible pathway to the concurrent rewiring of a large set of genes. In this scenario an interaction is acquired between TRs A and B, after which interactions between B and DNA are optimized gene-by-gene. Rewiring in this manner could avoid fitness barriers imposed by initially changing regulation one gene at a time.

Finally, genomic approaches provide evidence consistent with the idea that cooperative binding of TRs facilitates circuit changes. In its simplest form, the occupancy of two cooperatively binding TRs, A and B, on DNA depends on the concentration of each protein, the strength of each protein-DNA interaction, and the net favorable interaction between the two proteins. A decrease in any one of these parameters can be compensated by a gain in any other. This allows substantial shifts in the relative contribution of each component to the overall energetics without destroying the regulation; this flexibility, in turn, could catalyze regulatory change. For example, as shown in Fig. 1 (right path), the cis-regulatory sequence of B could drift away from consensus if the A-B interaction were sufficiently favorable. This drift could produce a weak cis-regulatory sequence for a third TR, C, whose expression might overlap that of B. If the A-C interaction were then strengthened by point mutation, the regulation of the gene would have changed from A-B to A-C through a series of small steps, none of which would destroy regulation of the gene. This scenario is but one of many that is made possible by cooperative binding. If the number of cooperative components is increased, the possibilities for "movement" in the system are multiplied.

A few studies provide direct support for these ideas. For example, the fungal mating circuit change already described roughly follows the scenario presented above (20). Further evidence comes from a whole-network analysis of the transcriptional circuitry of *S. cerevisiae* (22). Here, a strong correlation was observed between the number of TRs that regulate a gene and the fuzziness (departure from consensus) of the cis-regulatory sequences present at that gene. This fuzziness may indicate that the cooperative binding of multiple TRs to DNA relaxes the importance of any one TR-DNA interaction. It has also been shown, through the simulated evolution of systems of interacting components, that the existence of redundant intermediate states, such as those described above, greatly catalyzes change within these systems (23). Finally, Zuckerkandl has argued that the type of neutral changes permitted by cooperative assembly of TRs on DNA may have facilitated the formation of complex regulatory circuits (24). Although we have emphasized cooperative binding of TRs to DNA, other forms of combinatorial control (e.g., when two TRs bind DNA independently to control a target gene) could also facilitate circuit rewiring.

We propose that cooperativity may be especially important for coordinating changes in the regulation of entire sets of coexpressed genes. For example, the gain of a protein-protein interaction between two TRs could "jump-start" the rewiring of a set of genes at which one TR is already present (Fig. 2). Afterward, the new circuit could be improved, target gene by target gene, through the gradual formation of optimal cis-regulatory sequences. This idea may help to explain how regulatory changes could sweep through a complete set of coexpressed genes.

References and Notes

1. S. B. Carroll, J. K. Grenier, S. D. Weatherbee, *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell, Malden, MA, ed. 2, 2005).
2. E. H. Davidson, *Genomic Regulatory Systems: Development and Evolution* (Academic Press, San Diego, 2001).
3. G. A. Wray *et al.*, *Mol. Biol. Evol.* **20**, 1377 (2003).
4. M. Kirschner, J. Gerhart, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8420 (1998).
5. G. A. Wray, *Nat. Rev. Genet.* **8**, 206 (2007).
6. B. Prud'homme, N. Gompel, S. B. Carroll, *Proc. Natl. Acad. Sci. U.S.A.* **104** (suppl. 1), 8605 (2007).
7. S. A. Tishkoff *et al.*, *Nat. Genet.* **39**, 31 (2007).
8. M. D. Shapiro *et al.*, *Nature* **428**, 717 (2004).
9. A. P. McGregor *et al.*, *Nature* **448**, 587 (2007).
10. S. Jeong *et al.*, *Cell*, **132**, 783 (2008).
11. M. Ronshaugen, N. McGinnis, W. McGinnis, *Nature* **415**, 914 (2002).
12. R. Galant, S. B. Carroll, *Nature* **415**, 910 (2002).
13. A. P. Gasch *et al.*, *PLoS Biol.* **2**, e398 (2004).
14. A. Tanay, A. Regev, R. Shamir, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7203 (2005).
15. D. T. Odom *et al.*, *Nat. Genet.* **39**, 730 (2007).
16. A. R. Borneman *et al.*, *Science* **317**, 815 (2007).
17. B. B. Tuch, D. J. Galgoczy, A. D. Hernday, H. Li, A. D. Johnson, *PLoS Biol.* **6**, e38 (2008).
18. M. Z. Ludwig, C. Bergman, N. H. Patel, M. Kreitman, *Nature* **403**, 564 (2000).
19. M. Martchenko, A. Levitin, H. Hogues, A. Nantel, M. Whiteway, *Curr. Biol.* **17**, 1007 (2007).
20. A. E. Tsong, B. B. Tuch, H. Li, A. D. Johnson, *Nature* **443**, 415 (2006).
21. M. Lynch, *Nat. Rev. Genet.* **8**, 803 (2007).
22. Y. Bilu, N. Barkai, *Genome Biol.* **6**, R103 (2005).
23. E. S. Haag, M. N. Molla, *Evol. Int. J. Org. Evol.* **59**, 1620 (2005).
24. E. Zuckerkandl, *J. Mol. Evol.* **44** (suppl. 1), S2 (1997).
25. We thank B. Alberts, S. Carroll, A. Gann, S. Noble, and M. Ptashne for comments on the manuscript. We regret that due to space limitations we could not include original citations to much of the work in this field. This work was supported by grants from the NIH to A.D.J and H.L., and the Packard Fellowship in Science and Engineering to H.L. B.B.T. was supported by an NSF Predoctoral Fellowship.

10.1126/science.1152398