model of water flux, cell volume and turgor pressure. This allows them to close the feedback loop used by the cell to adapt to its new high-osmolarity environment (**Fig. 1**).

The resulting model essentially contains three distinct feedback loops: (i) loss of turgor pressure closes the Fps1 glycerol efflux channel, resulting in accumulating glycerol and restoration of turgor; (ii) signaling through the HOG pathway causes phosphorylated Hog1 to upregulate glycerol synthesis; and (iii) Hog1 also upregulates phosphatases that downregulate the HOG pathway.

The model provides interesting new insights, some of which seem to go against the grain of conventional wisdom. It had been suggested that upregulation of phosphatases might be the primary factor in shutting down the HOG pathway and ending the osmotic shock response. However, the model of Klipp *et al.* indicates that this mechanism is far too weak and slow to play such a role. Instead, the HOG pathway is shut down by the restoration of turgor pressure itself. A key initial player in this restoration seems to be the Fps1 efflux channel, whose almost instantaneous closure upon osmotic shock allows the buildup of intracellular glycerol well before HOG-induced regulation of glycolysis enzymes cranks up the production of glycerol from its initial basal synthesis levels. Moreover, both experiments and simulations show that without stopping the outflow through Fps1, most of the newly synthesized glycerol would simply escape into the extracellular environment.

Some intriguing questions remain. For example, why do we see such a strong and transient spike in HOG signaling and concomitant gene expression—at its peak up to ten times the levels after adaptation to the high-osmolarity environment? One answer may be that this strong spike allows for a fast protective response of the system at a time when other stress response mechanisms are rapidly shutting down mRNA transcription and ribosome synthesis.

The model of Klipp *et al.* is still open to improvement, for example, by using experimentally measured mRNA decay rates[6], and especially by optimizing the parameters to fit all available data, not just the standard osmotic shock experiment. The current model could even be used to design experiments that would optimally test specific parameters, for example, using pulses of high osmolarity with different durations.

Perhaps more importantly, the model could be expanded by including additional features or integrating it with other modules. Alternative models for the omitted and poorly understood Sho1 branch of the HOG pathway could be tested by integrating them into the larger model. Another obvious expansion would be to include the *hypo*-osmotic shock response, which partly mirrors the hyperosmotic response, but also acts through the PKC MAP kinase pathway to protect the cell wall from bursting.

Expanding and building upon the current model would be greatly simplified by the use of a common interchange format for biochemical models, such as the Systems Biology Markup Language (SBML; http://sbml.org/)[7]. *Nature* journals and *Molecular Systems Biology* now accept and encourage submissions in this format, and a repository of curated SBML models is gradually accumulating (http://www.ebi.ac.uk/biomodels/). Perhaps one day, building a model such as the one presented by Klipp *et al.* will be as simple as downloading a phospho-relay module, a model of the HOG pathway and glycolysis, linking them together with some generic gene expression circuitry and a custom turgor pressure formula, and *presto*… Systems Biology!

1. Klipp, E. *et al. Nat. Biotechnol.* **23,** 975–982 (2005).
2. Dutrochet, H. L'agent Immédiat du Mouvement Vital Dévoilé Dans La Nature et Dans Son Mode D'action Chez Les Végétaux et Les Animaux (JB Ballière, Paris, 1826).
3. De Vries, H. *Arch Néer Science Exacte Naturelles* **6**, 117–126 (1871).
4. Chamberlin, M.E. & Strange, K. *Am. J. Physiol.* **257**, 159–173 (1989).
5. Gasch, A.P. *et al. Mol. Biol. Cell* **11**, 4241–4257 (2000).
6. Wang, Y. *et al. Proc. Natl. Acad. Sci. USA* **99**, 5860–5865 (2002).
7. Hucka, M *et al. Bioinformatics* **19**, 524–531 (2003).

Discovering DNA regulatory elements with bacteria

Martha L Bulyk

A bacterial one-hybrid selection system offers a low-tech alternative for determining the DNA-binding specificities of transcription factors.

Following nearly a decade's worth of genome-scale gene expression profiling and the more recent sequencing of multiple higher eukaryotic genomes, attention is now shifting towards determining the regulatory mechanisms underlying gene expression patterns. A major obstacle in understanding transcriptional regulatory networks has been a lack of data on the DNA-binding specificities of most transcription factors. In this issue, Meng *et al.*[1] describe a modified bacterial one-hybrid system[2] that will help address this challenge. Using this system, they identified the DNA-binding-site motifs of eight metazoan transcription factors, including one *Drosophila melanogaster* protein (Odd-skipped (Odd)) whose DNA-binding specificity was previously unknown. The newly discovered binding sites then allowed the authors to predict and experimentally validate two new Odd target genes.

DNA-binding proteins are important in both lower organisms and more complex metazoans in numerous cellular processes such as transcription regulation, DNA repair and replication. The largest class of these proteins is regulatory transcription factors, which, by binding in a sequence-specific fashion to DNA-binding sites in the genome, modulate the expression of target genes. These interactions mediate normal progression through the cell cycle and the response to environmental stimuli, and, in higher organisms, frequently regulate gene expression in a cell type– and developmental stage–specific manner.

Despite the importance of transcription factors, relatively few of their DNA-binding specificities have been characterized in depth. Without binding site data, it is difficult to identify the target genes directly regulated by a given transcription factor and to identify the *cis* regulatory elements through which this regulation occurs. Currently, prediction of such *cis* regulatory elements requires experimental data on the DNA-binding

Martha L. Bulyk is in the Division of Genetics, Department of Medicine, Department of Pathology and the Harvard-MIT Division of Health Sciences & Technology, Brigham & Women's Hospital and Harvard Medical School, Harvard Medical School New Research Building, Room 466D, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.
e-mail: mlbulyk@receptor.med.harvard.edu

specificities of transcription factors. Some methods for high-throughput binding-site determination, such as microarray-based readout of chromatin immunoprecipitation ('ChIP-chip')[3–5], rely on specific antibodies, which may not always be available, whereas other methods, such as *in vitro* selection[6] and protein-binding microarrays[7], require purified protein. In contrast, the bacterial one-hybrid system not only uses *in vivo* selection, but also offers a low-tech alternative to microarray-based technologies.

In their study, Meng *et al.* expressed the DNA-binding domain of a given transcription factor as a fusion to the alpha subunit of RNA polymerase. A library of randomized oligonucleotides was cloned into a vector containing the selectable genes *HIS3* and *URA3*. If the given DNA-binding domain (the 'bait') binds a potential DNA target site (the 'prey') in the bacterium, it will recruit RNA polymerase to the promoter and activate transcription of the reporter genes (**Fig. 1a**). The two reporter genes *HIS3* and *URA3* are yeast genes that allow for positive and negative selection, respectively, when propagated in a bacterial strain in which the bacterial homologs of these genes have been deleted. Specifically, growth of cells on minimal medium containing 3-amino-triazole (3-AT), which is a competitive inhibitor of HIS3, provides positive selection, whereas growth on medium containing 5-fluoro-orotic acid (5-FOA), which is converted into a toxic compound by the uracil biosynthesis pathway, provides negative selection. Positive clones are then sequenced, and the sequences of the selected clones are examined with preexisting motif-finding tools (MEME, BioProspector) to identify the recognition binding site motif of the query transcription factor (**Fig. 1b**). The incorporation of a negative selectable marker to reduce background and the use of randomized candidate DNA-binding sites have both been used in yeast one-hybrid selection[8] and represent two important advantages of Meng *et al.*'s approach over the existing bacterial one-hybrid selection system[9].

To demonstrate that their system works, Meng *et al.* first used it to identify the known binding specificities of two mammalian Cys_2His_2 zinc finger proteins, Zif268 (also known as Egr1) and PLAG1, whose DNA-binding specificities were previously known. Before proceeding with analysis of additional proteins, the authors grew the original prey library alone in the presence of 5-FOA in an attempt to eliminate self-activating prey and thus reduce the false-positive rate. The authors then used this 'purified' prey library to determine the DNA-binding specificities of four individual transcription factors from *C. elegans* (LAG-1) and *D. melanogaster* (Dorsal, Paired, Odd), one of which (Odd) had not been characterized previously. Importantly, they also identified the binding specificities of the *Drosophila* proteins Runt and Big-brother (Bgb), which bind DNA with high affinity only as a heterodimer, thereby showing that their method works not only for monomeric proteins but also for proteins that bind DNA as complexes. In

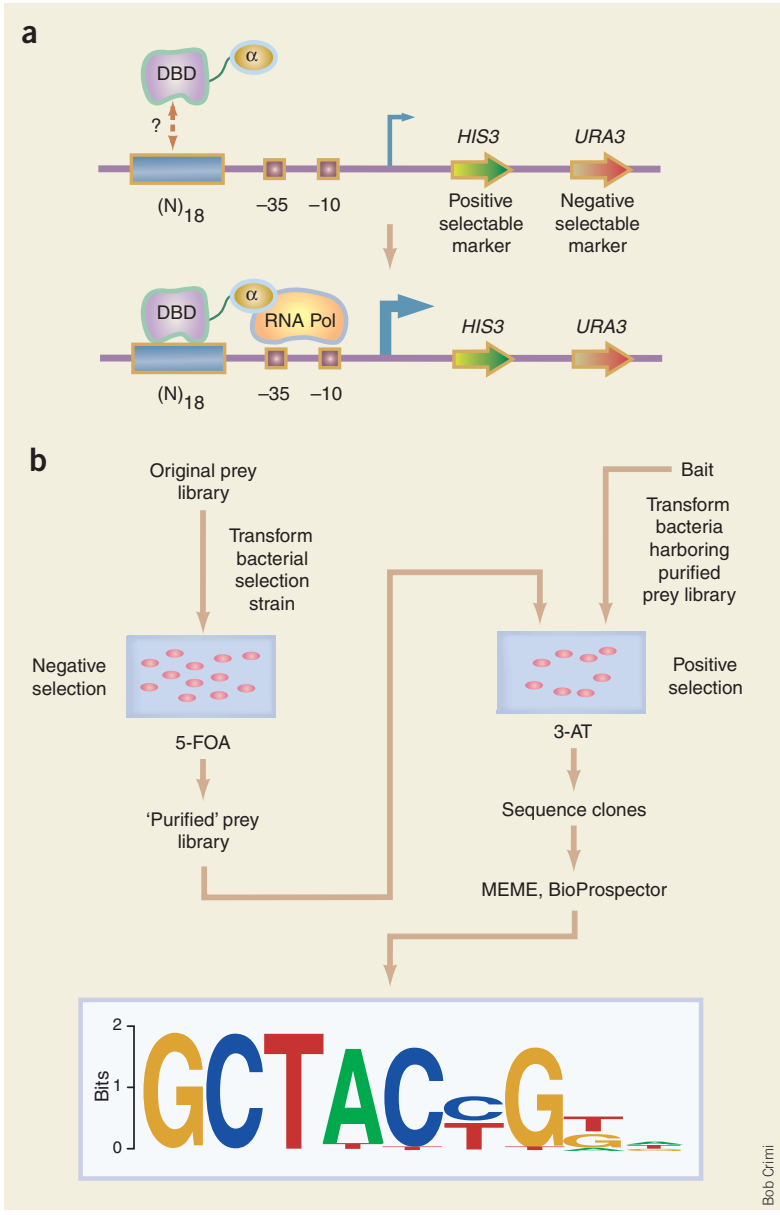

Figure 1 The bacterial one-hybrid system of Meng *et al.* (**a**) A library of randomized 18-bp oligonucleotides (the 'prey') is cloned upstream of the *HIS3* (positive) and *URA3* (negative) selectable markers in a bacterial strain lacking the bacterial *HIS3* and *URA3* homologs (*hisB* and *pyrF*, respectively). A plasmid containing the DNA-binding domain (DBD) of a query transcription factor fused to the alpha subunit of RNA polymerase (the 'bait') is transformed into bacteria harboring the prey library. If the query DNA-binding domain interacts with a prey DNA sequence, RNA polymerase is recruited, resulting in expression of the *HIS3* and *URA3* selectable marker genes. (**b**) The original prey library of candidate DNA-binding sites undergoes a round of negative selection on plates containing 5-FOA to reduce the proportion of self-activating sequences. The 'purified' prey library is transformed with the bait plasmid, and the bacteria undergo positive selection at a range of stringencies by growing the cells on a series of plates spanning a range of 3-AT concentrations. Prey from individual colonies are isolated and sequenced. Finally, the prey sequences are examined with motif-finding tools (MEME, BioProspector) to identify DNA-binding sites.

addition, these six proteins represent several structural classes of DNA-binding domains (Rel homology region, CSL-type DNA-binding domain, CBFα/β, paired domain and homeodomain) in addition to the Cys_2His_2 zinc finger domain, thus demonstrating the generalizability of the approach.

To explore the biological relevance of the DNA-binding-site motif that they identified for Odd, Meng *et al.* searched the *D. melanogaster* and *D. pseudoobscura* genomes for conserved, syntenic regions that contained at least two Odd binding sites. This type of search is used typically because *cis* regulatory modules frequently contain multiple copies of a given motif, and searches for individual binding sites can result in many false-positive target-gene predictions. Although the most appropriate way to measure functional conservation of binding sites is currently unclear, phylogenetic conservation within syntenic regions likely enriches for regulatory regions. A number of the regions that the authors found in their search were adjacent to genes with similar biological functions as that of Odd, including two genes (gooseberry (*gsb*) and Goosecoid (*Gsc*)) that had not been previously identified as direct targets of Odd regulation. *In situ* hybridizations indicated diminished expression of *gsb* and *Gsc* upon induction of ectopically expressed Odd, thus validating that Odd is regulating these genes.

Although the original prey library, consisting of 2×10^7 unique clones, contained only a very small fraction of all possible 18-bp sequences ($\sim 7 \times 10^{10}$), this subset of clones still covers enough sequence to allow a sufficiently large subset of binding sites to be sampled for most proteins. Nevertheless, without a more complex prey library, it may prove difficult to determine the binding specificities of transcription factors with lengthy binding sites (that is, much longer than 12 bp). Because self-activating sequences are removed in generating the purified prey library, a query transcription factor that has a close homolog in *Escherichia coli* that is active in the selection strain would also likely fail to be characterized by this approach.

Meng *et al.* used multiple stringencies (that is, concentrations of 3-AT) in their positive selection step to identify positive clones. Nevertheless, because the background in their Runt/Bgb selections was unacceptably high at even the highest 3-AT concentration, an additional negative selection step was required—here, the same concentration of 5-FOA was used as in selection of the purified prey library, which retained some self-activating sequences. Thus, to apply the bacterial one-hybrid approach generally, one would likely need to perform selections at a range of 3-AT concentrations, with presumably higher-affinity binding sites being selected at higher concentrations of 3-AT, as well as a range of 5-FOA concentrations, to keep the proportion of false-positive colonies to a minimum. It is encouraging that for eight of the nine transcription factors examined in this study, excluding the one protein that resulted in toxicity, the authors were able to successfully identify their DNA-binding-site motifs, despite the fact that they used only three different 3-AT concentrations over no more than a fivefold range.

An important point to keep in mind is that the degeneracy of the discovered binding site motifs will be reflective not only of the number of positive clones that are sequenced but also of the stringency of the selections. If only a small number of clones from a more stringent selection are sequenced, then the motifs will likely represent only the higher-affinity binding sites, even though weaker sites may also be biologically significant. Therefore, this system would be improved by the incorporation of a high-throughput sequencing step, such as by concatemerization of positive clones before sequencing, as in serial analysis of gene expression[10]. This would permit the discovery of more accurate motifs by sequencing a greater number of clones, including those from less stringent selection conditions.

One advantage that a bacterial one-hybrid system offers over a yeast one-hybrid system[8] is that the higher bacterial transformation efficiency allows more complex libraries to be examined more readily. In the present study, only a single large plate was required at each selection stringency, with multiple stringencies used for each transcription factor. Although expression in *E. coli* of proteins from higher eukaryotes will be problematic for some proteins, the authors were able to resolve this problem for one protein (Odd) by substituting rare codons with preferred synonymous codons; on the other hand, expression of another attempted protein was toxic. Still, the effects of any post-translational modifications that are important for DNA-binding specificity would be missed, as would any conformational changes of the DNA-binding domain caused by the rest of the protein, as only DNA-binding domains were examined in this study. Nevertheless, given the lack of binding site data for most transcription factors in both model organisms and humans, even imperfect binding site data would be extremely valuable. For example, recent analysis suggests that there are ~1,960 transcription factors, corresponding to ~8% of genes, in the human genome[11], and the sequence specificities and functions of most of these proteins have not yet been determined.

Meng *et al.*'s bacterial one-hybrid system provides another tool in our arsenal for identifying the DNA-binding specificities of transcription factors and thus predicting their target genes and genomic DNA regulatory elements. Since coregulation in higher eukaryotes frequently occurs through binding by a combination of transcription factors, analysis of transcription factor binding in these genomes will require further studies of homotypic and heterotypic binding site clustering, along with more sophisticated algorithms for the consideration of phylogenetic conservation.

The method of Meng *et al.* should also allow examination of the effects of protein-protein interactions on DNA binding, which may further guide the prediction of *cis* regulatory modules based on binding-site clustering. As suggested by the authors' studies on Odd, results from these analyses also could be used to predict the regulatory roles of uncharacterized transcription factors. The integration of data from such studies will certainly help to achieve our goals of delineating the regulatory networks that govern cellular gene expression.

1. Meng, X., Brodsky, M.H. & Wolfe, S.A. *Nat. Biotechnol.* **23**, 995–1001 (2005).
2. Dove, S.L., Joung, J.K. & Hochschild, A. *Nature* **386**, 627–630 (1997).
3. Iyer, V.R. *et al. Nature* **409**, 533–538 (2001).
4. Lieb, J.D., Liu, X., Botstein, D. & Brown, P.O. *Nat. Genet.* **28**, 327–334 (2001).
5. Ren, B. *et al. Science* **290**, 2306–2309 (2000).
6. Oliphant, A.R., Brandl, C.J. & Struhl, K. *Mol. Cell. Biol.* **9**, 2944–2949 (1989).
7. Mukherjee, S. *et al. Nat. Genet.* **36**, 1331–1339 (2004).
8. Wilson, T.E., Fahrner, T.J., Johnston, M. & Milbrandt, J. *Science* **252**, 1296–1300 (1991).
9. Joung, J.K., Ramm, E.I. & Pabo, C.O. *Proc. Natl. Acad. Sci. USA* **97**, 7382–7387 (2000).
10. Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, K.W. *Science* **270**, 484–487 (1995).
11. Messina, D.N., Glasscock, J., Gish, W. & Lovett, M. *Genome Res.* **14**, 2041–2047 (2004).