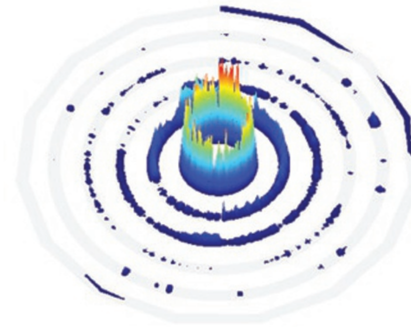
GENE REGULATION

# The DNA-binding landscape

**Visualizing binding data in the form of specificity landscapes yields new insight into the behavior of DNA-binding molecules.**

Proteins or small molecules designed to bind certain DNA sequences and to regulate target genes have much promise in both basic and applied research. But it is important that these tools be specific. As part of their efforts to build such synthetic transcription factors, Aseem Ansari and colleagues at the University of Wisconsin, Madison, therefore tested the specificity of factor binding to DNA. They used custom DNA arrays displaying every possible 10-mer sequence—almost a half a million of them—to examine the binding profiles of engineered hairpin polyamides and protein transcription factors, as well as of several natural DNA-binding proteins, across sequence space.

They quickly ran into a problem. The data were just too complex to understand intuitively. "We had this comprehensive binding dataset," says Ansari, "and the first question was, how do you look at these data? If you use colors or tables or graphs to represent it, things begin to get very complicated very quickly." They saw that protein transcription factors, in particular, bind fairly broadly across sequence space and that, although it is possible to extract a consensus binding motif, this often did not tell the whole story. What is more, the consequences of changing particular residues in a binding motif were not always simple or additive. So they had to find a new way of visualizing the data.



The sequence specificity landscape of the transcription factor Nkx2.5. Image courtesy of Clayton Carlson and Aseem Ansari.

Inspired by several sources, from the energy funnels used to represent protein folding to the painting of twentieth-century master of shape and color Wassily Kandinsky, Ansari and colleagues finally settled on what they call sequence specificity landscapes (SSLs) to represent the binding data. These are plots shown as several concentric circles, in which the innermost circle represents the perfect match to a particular seed motif and each subsequent circle, moving outward, denotes increasing sequence mismatches from the motif. Thus, sequences in the first circle have a single mismatch, those in the second circle have two mismatches, and so on. The intensity of the peaks in each circle represent the binding intensity on the array. The landscapes can also be represented in linear format, but it is the circular arrangement that allows the entire dataset to be visualized in a simple and intuitively interpretable form.

© 2010 Nature America, Inc. All rights reserved.

CHROMATIN TECHNIQUES

## FISSION YEAST DEFIES THE CODE

**A genome-wide map of nucleosome positions in *Schizosaccharomyces pombe* points to a positioning mechanism different from that of other organisms.**

Philipp Korber from the University of Munich considers the view that there is a universal genomic code that determines nucleosome positioning a prominent hypothesis; one that he, together with colleagues and collaborators Karl Ekwall from the Karolinska Institute and Guo-Cheng Yuan from Harvard University, recently helped call into question.

Genomic DNA does not exist in random coils in the nucleus but is wrapped around octamers of histone proteins, the nucleosomes, which in turn can form higher-order structures. The positioning of the nucleosomes determines the accessibility of DNA to other proteins such as polymerases and thereby influences transcription. Knowing where along a DNA molecule the nucleosomes are positioned allows one to better understand how genes are turned on or off.

Histones are among the most conserved proteins across all eukaryotes, and their higher affinity for certain DNA sequences has led scientists to postulate a positioning code. Korber calls this idea that eukaryotic DNA has evolved information that is sufficient to determine the positioning of the nucleosomes 'intellectually pleasing', but he says: "we and others had indications from *in vitro* reconstruction work (in [Saccharomyces] *cerevisiae*) that DNA sequence alone does not sufficiently determine positioning."

Among researchers working in the chromatin field, *S. pombe* is emerging as a preferred model organism because it has many features in common with multicellular organisms. Yet when Alexandra Lantermann, a doctorate student in Korber's group, set out to determine genome-wide nucleosome positioning *in vivo*, she found published nucleosome maps for only 3 loci—a negligible number given the approximately 5,000 genes in the *S. pombe* genome—and no reliable information on transcription start and termination sites (TSSs and TTSs). So Lantermann first used a laborious method to determine nucleosome positioning with locus-specific probes to obtain a gold-standard set of positions. Then, in what Korber calls a herculean effort, Lantermann annotated TSSs and TTSs of all *S. pombe* genes by hand.

To get a genome-wide positioning map, Lantermann digested chromatin with micrococcal nuclease and hybridized the DNA protected by the nucleosomes to a tiling array at 20-base-pair resolution. When she aligned the hybridization signals of all genes at their TSS, several interesting findings emerged. Downstream of the TSS Lantermann saw regular arrays, that is, similar average spacing between nucleosomes across all genes. To Korber, this suggests that the passage of the polymerase sets up the array. Upstream of the TSS, the researchers saw a nucleosome-depleted region (NDR). Similar NDRs have been

It became quickly apparent that the specificity landscapes of several binders, including natural transcription factors, are rugged. Notably, this is true even of the innermost circle, which represents the binding motif. As the points on this circle represent the motif in the context of different flanking sequences, what this observation points to is the importance of contextual information in the binding of proteins to DNA. "When you look at a landscape like this, you see immediately that context matters," Ansari says, adding: "That sort of information gives us a much better way of annotating the genome than consensus motifs."

The specificity landscapes also allow an almost at-a-glance evaluation of the specificity of binding of a given factor because peaks in outer circles immediately indicate that the binder targets sequences other than the consensus motif. The researchers found that synthetic polyamide binders are in fact highly specific for their designed target sites, often more so than natural DNA-binding proteins. Gratifyingly, the *in vitro* binding data for one polyamide could explain discrepancies that have been observed in its regulation of target genes in living cells. Further study of how the *in vitro* data relate to binding within the cell will undoubtedly provide a more nuanced picture.

Ansari and colleagues hope to have a user interface available soon so that others can use this visualization tool. Exploration of the peaks and valleys of a specificity landscape will then be just a click and a drag away.

**Natalie de Souza**

**RESEARCH PAPERS**
Carlson, C.D. *et al*. Specificity landscapes of DNA binding molecules elucidate biological function. *Proc. Natl. Acad. Sci. USA* **107**, 4544–4549 (2010).

© 2010 Nature America, Inc. All rights reserved.

reported in *S. cerevisiae*, and there, poly(A+T) stretches had been shown to be crucial in keeping the DNA nucleosome free. However, in *S. pombe* such sequences are not enriched in NDRs. Apparently, *S. pombe* uses a different, yet to be determined, mechanism to keep NDRs free of nucleosomes. In general, the correlation between DNA sequence and nucleosome positioning was very different between both yeast, arguing against universally conserved DNA sequence rules.

Interestingly, promoter regions enriched for the histone variant H2A.Z, which is considered an epigenetic mark for silent chromatin, also had arrays upstream of the TSS, indicating that histone H2A.Z has a role in positioning as well, again in contrast to *S. cerevisiae*.

Although much remains to be done to elucidate mechanisms, the value of this map is obvious. As Korber describes it, "everyone can turn to this map and find out where in their gene of interest is an NDR or a regular array and how high the nucleosome occupancy is." And in addition every gene is annotated with a TSS and a TTS.

Elucidating the mechanism of positioning will give important clues as to the intricate interplay between chromatin packaging and all DNA-related processes such as transcription, replication or DNA repair.

**Nicole Rusk**

**RESEARCH PAPERS**
Lantermann, A.B. *et al. Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of *Saccaromyces cerevisiae*. *Nat. Struct. Mol. Biol.* **17**, 251–257 (2010).

## NEWS IN BRIEF

STEM CELLS

### From fibroblasts to neurons

No biologist would be surprised these days if one cell type was converted to another by going through a pluripotent stage, nor if this reprogramming was done directly between two closely related cell lineages. But Vierbuchen *et al.* now demonstrate a way to directly convert fibroblasts into distantly related excitatory neurons. By introducing a specific combination of genes encoding transcription factors (*Ascl1, Brn2* and *Myt1l*) into mouse embryonic and postnatal fibroblasts, they made induced neuronal cells with characteristics and functional properties of mature neurons.

Vierbuchen, T. *et al. Nature* **463**, 1035–1041 (2010).

SYSTEMS BIOLOGY

### Validating functional CRMs

*Cis*-regulatory modules (CRMs) control gene expression as parts of complex regulatory networks. To experimentally validate CRMs, which are often identified based on computational predictions, Nam *et al*. recently developed a high-throughput assay. The researchers barcoded each CRM that drove expression of a reporter gene, then injected the constructs into a sea urchin embryo and subsequently analyzed the isolated mRNA by quantitative PCR. This approach will be a valuable tool for regulatory systems biology.

Nam, J. *et al. Proc. Natl. Acad. Sci. USA* **107**, 3930–3935 (2010).

PROTEIN BIOCHEMISTRY

### Routine NMR structures of large proteins

It is challenging to solve the structures of large proteins by nuclear magnetic resonance (NMR) spectroscopy without resorting to deuterium labeling. Raman *et al.* show that protein structures of up to 25 kilodaltons can be solved via NMR spectroscopy by using sparse information about backbone chemical shifts and residual dipolar couplings. The researchers used these data to guide Rosetta-based structural modeling of several proteins with known and novel structures.

Raman, S. *et al. Science* **327**, 1014–1018 (2010).

NEUROSCIENCE

### Recording the flying brain

An ideal way to understand how the nervous system transforms sensory information into locomotor actions is by directly recording the activity of relevant neurons in animals that are receiving the stimulus while they are free to move. Maimon *et al.* developed a way to perform whole-cell patch-clamp recordings from genetically identified neurons in *Drosophila melanogaster* in flight, allowing them to study the activity of a class of visual neurons and monitor optomotor responses in the tethered, flying flies.

Maimon, G. *et al. Nat. Neurosci.* **13**, 393–399 (2010).

SYNTHETIC BIOLOGY

### Designer promoters

The ability to tune the amounts of genes expressed in different cell types would permit more precise reverse genetic experiments. Schlabach *et al.* isolated strong synthetic enhancers by screening a library of 100-mer sequences consisting of tandem repeats of all possible 10-mers. Reporter expression varied in different cell lines and could be additionally modulated by screening for synergistic effects between enhancers.

Schlabach, M.R. *et al. Proc. Natl. Acad. Sci. USA* **107**, 2538–2543 (2010).