A Partner Evokes Latent Differences between Hox Proteins

Aseem Z. Ansari[1,*] and Kimberly J. Peterson-Kaufman[1]
[1]Department of Biochemistry and The Genome Center, University of Wisconsin, Madison, Madison, WI 53706, USA
*Correspondence: ansari@biochem.wisc.edu
DOI 10.1016/j.cell.2011.11.046

Hox transcription factors bind highly related DNA sequences in vitro, yet they regulate different genes and play distinct roles in anterior-posterior patterning in animals. Slattery et al. report that a common cofactor, Exd, accentuates latent sequence specificities of all eight Hox proteins and directs binding to relevant sites across the genome.

In eukaryotes, transcription factors face acute challenges. They must identify functional binding sites within a crowded chromatin context to regulate gene expression. In many cases, regulation of fate-determining gene networks depends on recognition of short DNA binding sites, typically 6–10 base pairs, and each factor tolerates a surprising degree of variability in binding site sequences. Thus, for any given transcription factor, thousands of potential binding sites occur across a genome, yet only a fraction of those sites are used to regulate target genes. This problem is further compounded by the ability of members within a closely related family of transcription factors, such as the Hox proteins, to bind nearly identical sites in vitro. How do such factors regulate distinct gene networks and confer distinct fate decisions on cells? In this issue of *Cell*, Mann and colleagues report that a shared Hox protein cofactor, Extradenticle (Exd), accentuates differences between eight *Drosophila* Hox proteins (Slattery et al., 2011), providing insight that bridges what we understand about Hox specificity in vitro and their regulatory capacity in vivo.

Eight Hox orthologs (Lab, Pb, Dfd, Scr, Antp, Ubx, AbdA, and AbdB) play a central role in anterior-posterior patterning of *Drosophila melanogaster*. Genes encoding these eight proteins are collinearly placed on the chromosome and are expressed in a colinear manner. Ectopic expression of posterior Hox proteins in anterior regions leads to dramatic homeotic transformations. For example, anterior expression of Antennaepedia (Antp) replaces antennae with legs.

To explore the in vivo specificity of Hox-Exd heterodimers, Slattery et al. purified recombinant Hox family members and examined the spectrum of sequences bound by the Hox-Exd complexes using recently developed high-throughput SELEX and sequencing methods (Stormo and Zhao, 2010). The binding motifs that emerged clustered into three main classes that followed the Hox colinearity patterns. Lab and Pb complexes preferentially bound nTGATTGATnnn; Dfd and Scr preferred nTGATTAATnnn; and Antp-AbdB proteins preferred nTGATT TATnnn. Seven other composite sites with varying degrees of preference for different subsets of Hox-Exd complexes were also identified. In essence, differences from Hox monomer binding sites were evident in the composite 12-mer site bound by Hox-Exd complexes.

In the 12 base pair sites, nTGAT is bound by Exd, and the remaining site is targeted by the Hox partner. In each composite motif, retention of the core sequences preferred by Exd (nTGAT) and Hox (TAAT or TTAT) is evident, and most of the latent specificity emerges at the interface of the two binding sites (Figure 1). Similar studies with synthetic Hox analogs revealed that Exd only needs the nGAn core to bind, thus the latent specificity of natural Hox proteins evolved to finely balance cooperative assembly with the enhanced specificity of Hox-Exd complexes (Warren et al., 2006). The results are reminiscent of the altered specificity of Ets1 when bound to Pax5. For that complex, the crystal structure of the ternary complex revealed that a key tyrosine of Ets1 released its sequence-specific contacts with DNA and docked into Pax5, thereby altering the sequence recognition by Ets1 (Garvie et al., 2001). In the case of Scr-Exd, the crystal structure revealed that residues of the Scr "linker" that connected the homeodomain to the Exd-binding YPWM module were making sequence-specific contacts via the minor groove at the Exd-Scr interface (Joshi et al., 2007). Similar linker-dependent sequence readout at the interface of Hox-Exd sites might be at the heart of the distinct latent specificities of other members of the family.

In addition to the Hox YPWM module binding to Exd, there exist other modes of interaction that stabilize cooperative complexes formed by Hox proteins. A conserved "UbdA" octapeptide found at the C terminus of the Ubx and AbdA homeodomains stabilizes cooperative complexes with Exd (Merabet et al., 2007). Whether the UbdA stabilized Ubx-Exd or AbdA-Exd complexes target unique sites and regulate a distinct set of genes remains an exciting and important question. Moreover, additive cooperative binding of Ubx multimers (see A-A complexes in Figure 1) might be particularly relevant for animal development, as Ubx also functions in regions that do not express Exd (Gallant et al., 2002). This is somewhat incongruent with Slattery et al., which reports that Ubx-Exd-Hth sites are more commonly observed within genomic loci identified by ChIP-seq analysis in specific leg and haltere imaginal discs in vivo. Whether different types of cooperative complexes (Hox-Exd, Hox-Hox, Hox-other factors) unmask additional latent specificity determinants to

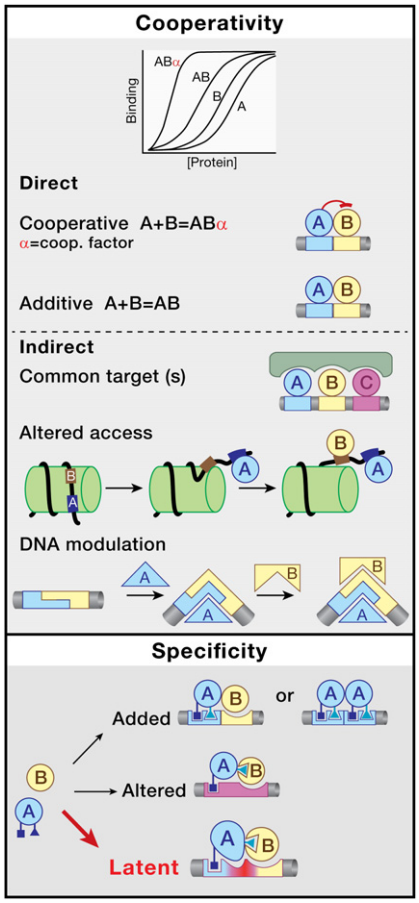

Figure 1. Specificity and Cooperativity Determine Transcription Factor Binding

"Latent specificity" is placed in the context of current models of cooperativity and specificity that guide the assembly of transcription factor complexes at target loci across the genome. Cooperativity, through direct or indirect means, plays an important role in the assembly of transcription factors at specific composite sites. The direct cooperativity of the DNA binding event can either be cooperative, whereby the binding of one factor increases the affinity of the other for DNA, or additive, whereby assembly of the complex is a function of coupled binding equilibria of protein-protein and protein-DNA interactions. Indirect cooperativity in the assembly of multiple protein-DNA interactions is often observed due to independent interactions with a common target in the transcriptional machinery or the alteration of chromatin or DNA structure to permit sequential protein binding. The DNA binding specificity of transcription factors A and B together can be the sum of the two individual target sequences (added), a completely new sequence (altered), or a combination of the two that becomes apparent upon interaction between all three components (latent). Such latent specificity determinants within each Hox protein play a significant role in sequence discrimination by Hox-Exd complexes.

regulate ortholog-specific gene expression would be an extremely important issue to resolve.

Looking deeper into the recognition patterns for the Hox-Exd complexes, Slattery et al. offer an interesting structural interpretation of the observed specificities. They model the predicted minor groove dimensions across the 12-mer binding site and suggest that a narrow minor groove between residues 4 and 5 and residues 8 and 9 of the motif ($nTGA_4$ $TNNA_8Ynnn$) favors binding by the anterior set of Hox orthologs, whereas the posterior orthologs prefer a wider minor groove at positions 8 and 9. Though intriguing, this interpretation will require validation via high-resolution structural analyses and targeted perturbation of minor groove geometry. An earlier study using sequence-specific minor groove binding small molecules showed that perturbations of the groove geometry improved Exd binding to its target TGAT sequence by 10-fold (Moretti et al., 2008). Thus, microstructural perturbations of the binding site that favor protein-DNA interactions likely contribute measurably to the overall assembly at one site over another.

The question of how transcription factors identify functionally relevant binding sequences is not a new one. Many mechanisms have been proposed, including variations on the theme of cooperative assembly of multiple transcription factors, like the Hox-Exd complexes. Cooperative assemblage provides several benefits. It limits the number of targets in the genome to those where the desired combination of DNA binding sites regulates different genes in a combinatorial manner with a small set of transcription factors. Cooperative interactions also offer mechanistic variability, as association with other proteins can convert a transcription factor from an activator of gene expression at one promoter to a repressor at another. These kinds of interactions also provide the basis for stable regulatory switches that maintain their regulatory state despite background fluctuations in signals and levels of proteins (Ptashne and Gann, 2002).

However, Biggin and coworkers have challenged the notion that direct cooperative assembly of transcription factors is important for targeting unique genomic loci. Using genome-wide location analyses (ChIP-seq), they reported overlapping binding patterns for several transcription factors at open chromatin regions across the genome (Li et al., 2011). Their analyses provided little support for the widespread role for cooperative assemblage at specific loci and, rather, suggest that "indirect" cooperativity of multiple factors independently assisting in opening chromatin regions may be more relevant (Figure 1).

The patterns of selectivity observed for Hox binding (Slattery et al., 2011) resonate most strongly with the concept of combinatorial recognition and provide a new and potentially broadly applicable mechanism for distinguishing family member function. Upon association with a common cofactor, latent specificity determinants within each Hox protein play a role in sequence discrimination. However, full understanding of the specificity determinants for Hox factor-dependent gene regulation is far from solved, and much remains to be learned about how distinct Hox factors identify and regulate their unique set of genes.

REFERENCES

Gallant, R., Walsh, C.M., and Carroll, S.B. (2002). Development *129*, 3115–3126.

Garvie, C.W., Hagman, J., and Wolberger, C. (2001). Mol. Cell *8*, 1267–1276.

Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., and Mann, R.S. (2007). Cell *131*, 530–543.

Li, X.Y., Thomas, S., Sabo, P.J., Eisen, M.B., Stamatoyannopoulos, J.A., and Biggin, M.D. (2011). Genome Biol. *12*, R34.

Merabet, S., Saadaoui, M., Sambrani, N., Hudry, B., Pradel, J., Affolter, M., and Graba, Y. (2007). Proc. Natl. Acad. Sci. USA *104*, 16946–16951.

Moretti, R., Donato, L.J., Brezinski, M.L., Stafford, R.L., Hoff, H., Thorson, J.S., Dervan, P.B., and Ansari, A.Z. (2008). ACS Chem. Biol. *3*, 220–229.

Ptashne, M., and Gann, A. (2002). Genes & Signals (Cold Spring Harbor, N.Y., USA: Cold Spring Harbor Laboratory Press).

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., and Mann, R.S. (2011). Cell *147*, this issue, 1270–1282.

Stormo, G.D., and Zhao, Y. (2010). Nat. Rev. Genet. *11*, 751–760.

Warren, C.L., Kratochvil, N.C., Hauschild, K.E., Foister, S., Brezinski, M.L., Dervan, P.B., Phillips, G.N., Jr., and Ansari, A.Z. (2006). Proc. Natl. Acad. Sci. USA *103*, 867–872.