

Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins

Matthew Slattery,^{1,7,8} Todd Riley,^{2,3,7} Peng Liu,^{3,4,7} Namiko Abe,¹ Pilar Gomez-Alcala,^{2,5} Iris Dror,⁶ Tianyin Zhou,⁶ Remo Rohs,^{6,*} Barry Honig,^{3,4,*} Harmen J. Bussemaker,^{2,3,*} and Richard S. Mann^{1,*}

¹Department of Biochemistry and Molecular Biophysics, Columbia University, 701 West 168th Street, HHSC 1104, New York, NY 10032, USA

²Department of Biological Sciences, Columbia University, 1212 Amsterdam Avenue, New York, NY 10027, USA

³Center for Computational Biology and Bioinformatics, Columbia University, 1130 St. Nicholas Avenue, New York, NY 10032, USA

⁴Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, 1130 St. Nicholas Avenue, New York, NY 10032, USA

⁵Department of Electrical Engineering, Columbia University, 500 West 120th Street, New York, NY 10027, USA

⁶Molecular and Computational Biology Program, Department of Biological Sciences, University of Southern California, 1050 Childs Way, Los Angeles, CA 90089, USA

⁷These authors contributed equally to this work

⁸Present address: Institute for Genomics and Systems Biology, University of Chicago, 900 East 57th Street KCBD 10115, Chicago, IL 60637, USA

*Correspondence: rohs@usc.edu (R.R.), bh6@columbia.edu (B.H.), hjb2004@columbia.edu (H.J.B.), rsm10@columbia.edu (R.S.M.)

DOI 10.1016/j.cell.2011.10.053

SUMMARY

Members of transcription factor families typically have similar DNA binding specificities yet execute unique functions in vivo. Transcription factors often bind DNA as multiprotein complexes, raising the possibility that complex formation might modify their DNA binding specificities. To test this hypothesis, we developed an experimental and computational platform, SELEX-seq, that can be used to determine the relative affinities to any DNA sequence for any transcription factor complex. Applying this method to all eight *Drosophila* Hox proteins, we show that they obtain novel recognition properties when they bind DNA with the dimeric cofactor Extradenticle-Homothorax (Exd). Exd-Hox specificities group into three main classes that obey Hox gene collinearity rules and DNA structure predictions suggest that anterior and posterior Hox proteins prefer DNA sequences with distinct minor groove topographies. Together, these data suggest that emergent DNA recognition properties revealed by interactions with cofactors contribute to transcription factor specificities in vivo.

INTRODUCTION

Gene regulatory information is encoded in genomic DNA sequences and interpreted by transcription factors that bind to specific sequences. Although the in vitro binding properties of transcription factors have been studied for many years, it has proven notoriously difficult to predict in vivo genomic binding

from in vitro sequence specificity. Whether or not a predicted binding site is occupied in vivo depends strongly on sequence and chromatin context as well as cell type (Gaulton et al., 2010; Guertin and Lis, 2010; Kaplan et al., 2011). While the amount of genome-wide binding varies greatly between transcription factors, typically only a small fraction of a transcription factor's preferred DNA sequences are occupied in vivo.

What makes in vivo binding more specific than in vitro binding? One possible answer is that the organization of the chromatin—for example, the position of nucleosomes—limits access to transcription factor binding sites (Wunderlich and Mirny, 2009). A second explanation has its root in the combinatorial nature of gene regulation. Unlike individual transcription factors, complexes of interacting factors bind cooperatively to genomic regions that contain a favorable configuration of binding sites (Johnson, 1995). These mechanisms, however, are unlikely to be sufficient to account for the transcription factor specificities observed in vivo. In particular, confounding the issue of specificity is that most transcription factors are members of protein families that have very similar DNA binding domains with similar recognition properties. For example, in the mouse there are nineteen T-box factors that can bind to variations of the sequence TCACACC, 39 Hox family homeodomain proteins that bind to AT-rich binding sites, and nearly 60 basic helix-loop-helix (bHLH) factors, most of which bind to the DNA sequence CACGTG known as the “E-box” (Berger et al., 2008; Conlon et al., 2001; Jones, 2004; Noyes et al., 2008). Despite overlapping binding specificities, these factors carry out distinct functions in vivo (Alexander et al., 2009; Cao et al., 2010; Naiche et al., 2005; Pearson et al., 2005). Although some specificity is derived from the cell type specific expression of individual family members, the fundamental question of how they recognize distinct binding sites and regulate unique sets of target genes in vivo remains unsolved.

Although members of the same transcription factor family typically have very similar DNA binding domains these domains are rarely identical, raising the possibility that small differences in protein sequence could lead to significant differences in binding specificity. However, when assayed *in vitro*, using either classical or high-throughput methods, different members of the same protein family generally do not show large differences in binding specificity. For example, in *Drosophila* more than 50 homeodomain proteins bind to the six-base-pair sequences TAATTG and TAATTA, despite differences in their DNA binding domains (Berger et al., 2008; Noyes et al., 2008). On the other hand, subtle differences in homeodomain sequences, and transcription factor sequences in general, are often conserved across vast evolutionary distances, arguing that these differences are functionally important. The eight Hox paralogs in *Drosophila*, for instance, which execute distinct functions *in vivo*, each have recognizable orthologs in both vertebrates and other invertebrates. Hox orthologs can be recognized not only by their protein sequences but also from the order in which they are expressed along an animal's anteroposterior (AP) axis (Hueber et al., 2010). Moreover, orthologous Hox proteins often have conserved functions when expressed in a heterologous species (Lutz et al., 1996; McGinnis et al., 1990; Zhao et al., 1993). These observations suggest that sequence differences between related transcription factors, although evolutionarily conserved and functionally relevant, are not typically reflected in differences in their DNA binding preferences.

There are two plausible solutions to this paradox. One is that some of the sequence differences between related transcription factors do not play a role in DNA binding, but instead affect their ability to repress or activate their target genes. Several examples of this so-called "activity regulation" have been described, and suggest that the ability to recruit different coactivators or corepressors may be used to diversify transcription factor function (Gebelein et al., 2004; Joshi et al., 2010; Li and McGinnis, 1999; Taghli-Lamalle et al., 2007). An alternative mechanism, which we refer to here as "latent specificity," is that differences in the amino acid sequences of transcription factors within the same structural family may only impact DNA recognition when these factors bind with cofactors. This mechanism is distinct from conventional cooperativity, in which binding energetics are affected by the presence of a cofactor but nucleotide sequence specificity is not. By contrast, in latent specificity there is a cofactor-induced change in DNA recognition. For example, as shown by X-ray crystallography, the *Drosophila* Hox protein Sex combs reduced (Scr) has distinct DNA recognition properties when it binds as a heterodimer with its cofactor Extradenticle (Exd) (Joshi et al., 2007). By directly binding a Hox peptide known as the "YPWM" motif, Exd helps to position the N-terminal arm of Scr's homeodomain so that it can recognize a sequence-dependent narrow minor groove in its DNA binding site. The binding to narrow minor grooves, typically by Arg residues, is an example of the widely used mechanism of DNA shape recognition (Rohs et al., 2009). Although Exd and its mammalian orthologs Pbx1-3 can heterodimerize with all Hox family members, and differences in DNA sequence preferences for Exd-Hox complexes have been reported (Chan et al., 1994; Chang et al., 1996; Lu and Kamps, 1997; Mann and Chan,

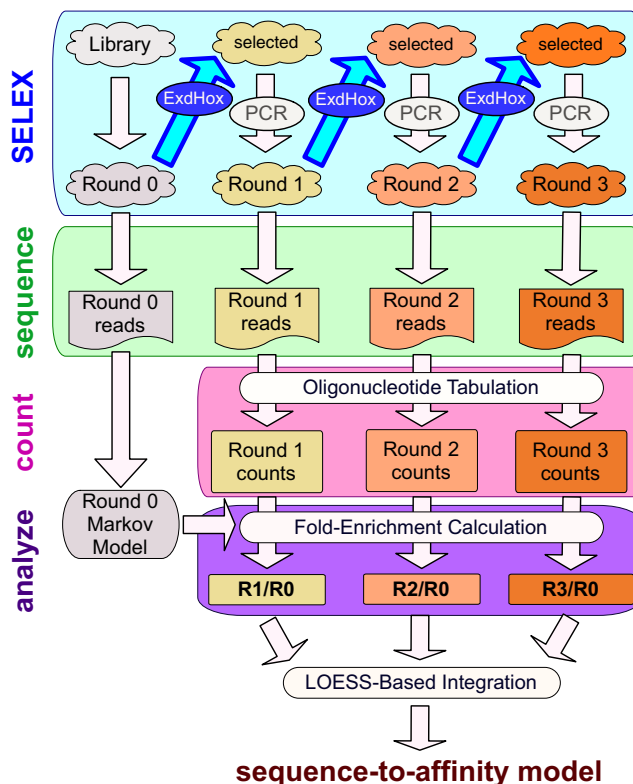


Figure 1. Overview of SELEX-seq

The starting point is a pool of synthesized DNA oligonucleotides containing a region of 16 random base pairs. This random pool is made double stranded and then sequenced using Illumina sequencing, resulting in a set of R0 reads. EMSAs are performed on the random pool and DNA molecules bound to Exd-Hox heterodimers are isolated and amplified by PCR. This enriched pool (R1) is sequenced. The affinity-based selection step is repeated multiple times. To accurately parameterize the sequence biases in R0, a Markov model is constructed. Relative fold-enrichments associated with the affinity-based selection step are calculated for all 12-mers. Information from earlier and later rounds of selection is combined using LOESS regression to estimate the relative binding affinity for each 12-mer with an optimal trade-off between accuracy and precision. See also Figure S1 and Tables S1 and S2.

1996), the degree to which the assembly of multi-protein complexes influences binding specificity has not been systematically analyzed for Hox proteins, or for any transcription factor family.

Here, we describe a high-throughput and systematic approach that demonstrates that complex formation between Hox factors and Exd uncovers latent DNA binding specificities that are only revealed upon heterodimerization. To do this, we combined Systematic Evolution of Ligands by Exponential Enrichment (Tuerk and Gold, 1990) with massively parallel sequencing (SELEX-seq) (Figure 1). The depth of the sequence information, combined with a biophysical model of the SELEX-seq data, allows us to calculate the relative affinity for any DNA sequence. We apply this method to all eight *Drosophila* Hox proteins in complex with the same cofactor, Exd. By analyzing the enrichment of oligonucleotides through several rounds of selection, we find that all Exd-Hox heterodimers prefer to bind the sequence GAYNNAY (where Y = T or C) and that the familiar

Download English Version:

<https://daneshyari.com/en/article/2036053>

Download Persian Version:

<https://daneshyari.com/article/2036053>

[Daneshyari.com](https://daneshyari.com)