HHS Public Access

Author manuscript

Trends Genet. Author manuscript; available in PMC 2017 June 01.

Published in final edited form as:

Trends Genet. 2016 June ; 32(6): 334–347. doi:10.1016/j.tig.2016.03.004.

To be specific or not: the critical relationship between Hox and TALE proteins

Samir Merabet[1] and **Richard Mann**[2]

[1] Institut de génomique fonctionnelle de Lyon, Centre National de Recherche Scientifique, Lyon, France

[2] Columbia University College of Physicians and Surgeons, New York, NY, United States

Abstract

Hox proteins are key regulatory transcription factors that act in different tissues of the embryo to provide specific spatial and temporal coordinates to each cell. These patterning functions often depend on the presence of the TALE-homeodomain class cofactors, which form cooperative DNA-binding complexes with all Hox proteins. How this family of cofactors contributes to the highly diverse and specific functions of Hox proteins in vivo remains an important unsolved question.

Here we review the most recent advances in understanding the molecular mechanisms underlying Hox-TALE function. In particular, we discuss the role of DNA shape, DNA- binding affinity and protein-protein interaction flexibility in dictating Hox-TALE specificity. We propose several models to explain how these mechanisms are integrated with each other in the context of the many distinct functions Hox and TALE factors carry out in vivo.

Keywords

Hox proteins; homeodomains; TALE cofactors; SELEX-seq; SLIMs; DNA binding specificity

A widely used set of Hox cofactors: the TALE proteins

Embryonic development relies on the activity of a relatively small number of regulatory factors that provide spatial and temporal coordinates to each cell. Among these molecules are the Hox transcription factors (TFs), **homeodomain** (**HD**)-containing proteins that are required for patterning the anterior-posterior (AP) axis of all bilaterian animals [1]. Hox proteins are also implicated in organogenesis [2] as well as the specification of individual cell types [3], illustrating their wide range of activities during animal development.

Hox genes are organised in **paralog** groups that emerged from successive duplications over the course of animal evolution [4]. These duplication events allowed the diversification of

Corresponding Authors: Samir Merabet: ; Email: merabet@ens-lyon.fr, Richard Mann: ; Email: rsmann10@gmail.com.

Hox expression patterns and functions, which have been remarkably conserved across vast evolutionary distances. In total, Hox proteins are classified into anterior (paralogs 1-2), central (paralogs 3-8) and posterior (paralogs 9-13) paralog groups (Figure 1a). In both vertebrates and invertebrates, mutations that alter the normal order of Hox paralog expression along the AP axis lead to dramatic homeotic transformations where one body part develops in place of another, due to the deployment of Hox-controlled developmental programs at inappropriate positions in the embryo [1].

Hox proteins bind DNA via their HDs, a 60 amino acid domain that is highly similar even among distant Hox paralogs, particularly at residues known from structural studies to make direct contacts with the DNA [5]. Consistent with these observations, on their own Hox homeodomains bind to very similar DNA sequences *in vitro*: anterior and central paralogs typically bind sequences that contain a TAAT while posterior paralogs have a preference for TTAT [6]. These generic and overlapping DNA binding properties are difficult to reconcile with the paralog-specific functions Hox factors execute *in vivo*, leading to what has been referred to as the Hox specificity paradox [7]. One solution to this paradox is that Hox proteins bind DNA with the help of additional **cofactors** *in vivo*. Indeed, as first implied by genetic experiments [8] and then confirmed by *in vitro* studies [9], DNA binding cofactors modify Hox DNA-binding specificities. The main families of Hox cofactors are the **PBC**, Meis and Prep families, which all belong to the so-called **t**hree **a**mino acid **l**oop **e**xtension (**TALE, Box 1**) class of HD-containing TFs [10,11]. These cofactors have the capacity to bind DNA cooperatively with Hox proteins (Figure 1b), which is critical for their ability to modify Hox DNA binding specificity.

In this article, we focus on the role of TALE cofactors in Hox specificity, with the goal of integrating recent structure/function and high-throughput *in vitro* data with *in vivo* genetic and genome-wide binding data. We discuss the role of DNA-binding site affinity in dictating Hox-TALE specificity, and the existence of distinct interaction modes between Hox, TALE cofactors, and DNA. We speculate that Hox-TALE-DNA complexes have the flexibility to exist in multiple binding modes that are tailored to execute the many distinct functions these factors carry out *in vivo*.

Flavors of Hox-TALE target genes

Before discussing the nature of Hox-cofactor binding sites, it is important to take a step back and consider the types of target genes Hox and TALE proteins are thought to co-regulate. As discussed in an earlier review [7], not all interactions between Hox proteins and their target genes need to be exquisitely specific. There are two main reasons for this. For one, some target genes may be regulated by more than one Hox protein (Figure 2, Key Figure). For example, the *Drosophila* Hox target gene *Distalless* (*Dll*), which is expressed in the limb primoridia of thoracic segments, is repressed by several Hox proteins in the abdomen [12]. Similarly, in *C. elegans* the terminal differentiation gene *mec-3* is activated in two different touch receptor neurons by two different Hox proteins binding to the same Hox-TALE binding site [13]. Other so-called semi-paralog specific genes have been described in *Drosophila* [14] and mouse [15]. In these examples, the binding sites mediating this type of regulation do not need to be specific for a single Hox paralog. This regulation often, but not

always, depends on the TALE cofactors [14,16], illustrating the diversity of the underlying molecular mechanisms (see also below). Along the same lines, other target genes might be regulated by even more, or in principle all, Hox paralogs and are referred to as paralog non-specific 12,17] (Figure 2).

A second reason for why some Hox or Hox-TALE binding sites may not require paralog specificity is a consequence of Hox expression patterns and the combinatorial nature of transcriptional regulation. If there is only a single Hox protein expressed in a particular tissue or cell type, then it is not a problem if multiple Hox proteins have the potential to bind the same binding sites – in cells where the target gene is not regulated, the binding sites may be inaccessible due to chromatin structure or critical cofactors may be missing. Although the specific binding sites are not known, a possible example of this scenario is the regulation of genes in the tritotocerebrum by the *Drosophila* Hox protein Labial [18] where it has been shown that other Hox proteins have the capacity to take Labial's place. However, during normal development, other Hox proteins never have the opportunity to regulate genes in this part of the brain because they are not expressed there. In other words, in this case, Labial 'specificity' is provided to a large extent by the combination of transcription factors that create the 'tritocerebrum context' and apparently not by the Hox binding site itself (Figure 2).

These scenarios, however, contrast with those in which target gene specificity for a single Hox paralog is essential (Figure 2). Although it is not yet known how many target genes require paralog specificity, according to at least one set of measurements in which RNA-seq profiles were compared following the ubiquitous expression of individual Hox proteins in *Drosophila*, a large number of target genes are regulated in a paralog-specific manner [26]. However, it is unclear from these experiments how much of this regulation is due to direct DNA binding. Two different strategies have been described that result in paralog-specific gene regulation. In one strategy, specificity depends on paralog-specific DNA-binding sites. The best example is the regulation of the salivary gland promoting gene *forkhead* (*fkh*) by the Hox protein Sex combs reduced (Scr) in *Drosophila*. Scr is the only *Drosophila* Hox protein capable of initiating the salivary gland development program even though the progenitor cells are present in other segments where other Hox proteins are expressed. The solution in this case is the use of binding sites that are highly specific for Scr-TALE complexes and are much more poorly bound by most other Hox-TALE complexes [19,20]. The one exception is Deformed (Dfd), which also has the potential to bind these sites with Exd, albeit with lower affinity. However, in contrast to Scr, Dfd binding results in repression, not activation, of reporter genes made with these binding sites [21]. Thus, in this case specific target gene regulation relies on the combination of semi-paralog-specific binding sites and non-DNA binding domains N-terminal to the Hox homeodomain that control the sign of gene regulation. Two other Hox target genes that depend on highly specific binding sites have been described in *Drosophila* [22,23].

As illustrated by the regulatory differences between Dfd and Scr discussed above, target gene specificity can also emerge from paralog non-specific or semi-specific binding sites. In these cases, specificity is provided by Hox-specific regulatory activity that is independent of DNA binding (Figure 2). Another example is *Foxp1*, which is specifically repressed by

Hoxc9 in the mouse embryo despite the fact that Hox4-10 paralogs display comparable cooperative DNA-binding with Pbx3 to its *cis*-regulatory binding sites [24]. Paralog-specific repression is mediated by a region in the Hoxc9 N-terminus that has no appreciable impact on DNA-binding. Such specific regulatory output might be mediated by the direct recruitment of specific co-regulator(s) or by forming a structurally distinct complex with the TALE cofactors that recruits specific co-regulators (see below and [25]).

As the above discussion illustrates, Hox regulation of target gene expression *in vivo* is likely the result of a complex mixture of binding site specificity (ranging from paralog specific to paralog-non-specific), cellular context, and non-DNA binding mechanisms controlling their activity. Given these examples, we discriminate between five classes of Hox target genes that are regulated by binding sites with varying degrees of paralog specificity (Figure 2):

Class 1: Paralog specific target genes regulated by paralog specific binding sites.

Class 2: Paralog specific target genes regulated by paralog non-specific binding sites.

Class 3: Semi-paralog specific target genes regulated by binding sites with intermediate specificity.

Class 4: Paralog non-specific target genes that have multiple regulatory inputs, each of which depend on paralog specific binding sites or ones with intermediate specificity.

Class 5: Paralog non-specific target genes regulated by paralog non-specific binding sites.

The latent specificity of Hox-cofactor complexes

The existence of class 3, 4, and in particular class 1 target genes requires binding sites such as the one in *fkh* that can discriminate between different Hox paralogs to different degrees. Given the highly similar binding specificities of Hox homeodomains, binding site discrimination must depend on cofactors. Moreover, the observation that the same set of common cofactors, i.e. the TALE homeodomain proteins, can provide additional specificity for the entire set of Hox proteins suggests that specificity is built into each Hox protein and that cofactors are needed to uncover this specificity. This mechanism, termed **latent specificity**, was most strongly supported by *in vitro* experiments using a platform called **SELEX-seq (Box 2)**, in which a large portion of the binding site universe was sampled for both monomeric Hox proteins and for Hox-cofactor complexes [27,28]. Using this method, Hox proteins selected very similar binding sites to each other in the absence of cofactors (Figure 3a). But in the presence of cofactors, distinct binding site preferences were observed (Figure 3a). Thus, latent specificity information is revealed upon complex formation with TALE cofactors.

X-ray crystal structures of Hox-PBC-DNA complexes have provided some mechanistic insights into the latent specificity concept for one Hox-TALE-DNA complex. Most informative were a pair of complexes with the same proteins (the Hox protein Scr and cofactor Exd) bound to two different DNA sequences [20]. One sequence, *fkh250*, preferred to bind Scr-Exd over other Hox-Exd heterodimers, while the other sequence, $fkh250^{con}$, did not prefer a particular Hox-Exd heterodimer. There were three important differences

between these two structures: 1) the N-terminal arm and adjacent residues of the Scr homeodomain were more ordered in the *fkh250* structure, 2) careful measurements of minor groove width revealed that there were two local minima in the *fkh250* structure but only one minimum in the *fkh250*^{con} structure, and 3) the additionally ordered residues in the *fkh250* structure included an Arg and a His whose side chains inserted into the second minor groove minimum (Figure 3c); importantly, these side chains were necessary for Scr to activate an Scr-specific reporter gene, *fkh250-lacZ*, *in vivo*. Further, DNA structure predictions suggested that the differences in DNA structure were sequence dependent and not influenced by protein binding. Thus, at least for this Hox-cofactor complex, latent specificity was a consequence of a three-way collaboration between the Hox protein, the cofactor, and the DNA: the *fkh250* DNA sequence resulted in a minor groove binding site for basic side chains (Arg, His); Scr had the correct sequence of amino acids to read that structure; and due to a protein-protein interaction, the TALE cofactors positioned those side chains so they could insert into the minor groove width minima (Figure 3c).

DNA structure is a consequence of DNA sequence, making it difficult to tease these two modes of recognition apart. To determine if the recognition of DNA structure occurred independently of base recognition, SELEX-seq experiments were performed with Scr mutants in which the minor groove-inserting Arg and His residues were changed to alanines, thus abrogating their ability to recognize minor groove width minima [29]. The results from these and other experiments demonstrated that the minor groove width recognizing side chains were both necessary and sufficient for selecting binding sites with this particular structural feature. Further, conferring the ability to read Scr's preferred minor groove structure to a naive Hox protein endowed that factor with the ability to activate *fkh250-lacZ in vivo* [29]. Thus, the recognition of DNA structure is a direct mode of DNA recognition that is independent of other sequence-dependent modes of recognition such as forming hydrogen bonds with bases in the major groove, and this mechanism is important for specific DNA binding and gene regulation *in vivo*.

The latent specificity mechanism and the importance of DNA structure in protein-DNA recognition are not limited to Hox-cofactor-DNA complexes. Analogous phenomena have been described for several other transcription factors, as summarized in previous review articles [30,31]. Further, although the recognition of minor groove width is likely to be relevant for other protein-DNA complexes, we speculate that there are additional structure-based readout mechanisms, such as DNA flexibility, that are also used for specific binding. When available, additional high-resolution 3D structures of protein-DNA complexes will likely provide additional insights into these mechanisms. Together with DNA-induced allostery, where a DNA sequence induces changes in protein structure or binding [30–33] structure-based readout mechanisms reveal several additional layers of complexity beyond the simple recognition of DNA bases that protein complexes exploit to achieve specificity.

Non-consensus and low affinity DNA-binding sites underlie Hox specificity

The latent specificity mechanism goes a long way towards providing a plausible solution for the Hox specificity paradox because it reveals how new, more complex DNA binding specificities can be generated as a result of protein-protein interactions. However, there is

still a large disconnect between the *in vitro* specificities uncovered by SELEX-seq or other high throughput *in vitro* methods and the sequences recognized by transcription factors *in vivo* (Figures 3b and 4a). For example, although they are related to the SELEX-seq-derived DNA sequences, the exact binding sites recognized by Hox-cofactor complexes in the *Dll* and *fkh* genes were not discovered by SELEX-seq experiments. **Chromatin immunoprecipitation** (**ChIP**) experiments also highlight a general problem with consensus binding sites. Although ChIP experiments frequently reveal a statistically significant enrichment of consensus binding sites, it is often difficult to find exact matches to these sites in ChIPed DNA fragments [31,34–36]. While some of this may be due to higher order chromatin structure leading to indirect precipitation, recent results raise the possibility that binding *in vivo*, as identified by ChIP experiments, may frequently be due to low affinity binding sites that do not fit known consensus sequences.

For example, the *Drosophila* abdominal Hox proteins Ultrabithorax (Ubx) and Abdominal-A (AbdA) activate the gene *shavenbaby* (*svb*) through a set of enhancers positioned many kilobases upstream from the transcription start site [14]. According to ChIP experiments, Ubx is bound directly to these enhancers *in vivo* [37]. *In vitro*, the TALE cofactors Exd and Hth bind specifically and cooperatively with Ubx and AbdA to each enhancer via multiple Hox-cofactor binding sites. However, these binding sites do not fit either the consensus for Hox monomer or Hox-TALE composite sites (Figure 4a). Importantly, these sites are also low affinity: they have less than 5% of the affinity relative to the highest affinity Hox-cofactor binding site. *In vitro* experiments demonstrate that increasing the affinity of these sites results in a loss of paralog specificity and, consequently, *svb* enhancers that have been mutated to include higher affinity Hox-TALE binding sites drive expression in cells that normally do not express *svb*. Thus, at least for these *svb* enhancers, Hox specificity requires low affinity, non-consensus Hox-TALE binding sites. Previous examples of the importance of low affinity transcription factor binding sites have been reported in a variety of contexts, suggesting that this phenomenon is likely to be general [38–43]. For example, low affinity binding sites have been shown to be important for the activation of enhancers regulated by Hedgehog signalling in *Drosophila* [41] and for the correct timing of enhancer activity in *C. elegans* [40].

The purely *in vitro* SELEX-seq data for the *Drosophila* Hox-TALE complexes also support the idea that specific Hox-TALE binding sites must be low affinity [27]. Although comparisons between any two Hox-TALE complexes can identify binding sites in the medium to high affinity range (>0.3 relative affinity) that prefer one complex over the other, no binding site in this affinity range binds to only one of the eight Hox-cofactor complexes [27]. Even binding sites that prefer one of the three paralog groups (paralog groups 1-2, 3-8, or 9-13) over the other two have a relative affinity of ~0.1 or less. Thus, in this large and unbiased sampling of Hox-TALE binding sites, paralog specific sites were only found in the low affinity range (Figure 3b). Also significantly, the number of sequences in this affinity range is very large compared to the more stringent high affinity range, suggesting that nature has many sequences to choose from to achieve specificity. The predominance of low affinity binding sites that do not easily fit high affinity consensus sequences makes their computational identification in eukaryotic genomes extremely challenging.

Relying on low affinity binding sites is not the only method that enhancers employ to increase specificity. As recently demonstrated by an elegant high-throughput method for enhancer mutagenesis in *Ciona* embryos [44], enhancer specificity also depends on imperfect spacing between binding sites for multiple transcription factors. In these experiments, the activities of more than two million variants of a Fibroblast Growth Factor (FGF)-responsive enhancer were assayed *in vivo*. Replacing low affinity sites with ones that better fit consensus binding sites, or improving the spacing between sites, resulted in inappropriate and ectopic enhancer activity. Thus, the use of low affinity and suboptimal arrangements of binding sites may be a general feature of enhancers that must be active in specific cell types or tissues.

Flexibility in Hox-TALE interaction properties and the recognition of low affinity DNA-binding sites

Biochemical, mutagenesis, and structural studies demonstrated that the interaction between PBC proteins (Exd and Pbx) and Hox proteins is mediated by the **tryptophan containing W** (or **hexapeptide; HX**) **motif** in Hox proteins with the TALE loop of the PBC homeodomain [7]. Although these interactions have been primarily documented when bound to sites that fit the Hox-TALE consensus nnGAYnnAYnnn, they are also likely to occur on non-consensus, low affinity binding sites as well. For example, although the highest affinity binding sites discovered by Hox-TALE SELEX-seq experiments generally fit this consensus, many thousands of binding sites were identified that do not fit this simple consensus [27]. However, with the exception of Scr-Exd bound to the binding site in *fkh250*, there are currently no other structures of Hox-cofactor complexes bound to low affinity binding sites. We can therefore only speculate that other Hox proteins bound to low affinity sites also use latent specificity-based mechanisms that rely in part on the recognition of DNA shape (Figure 3b). Despite the limited amount of structural data, this idea is supported by the observation that residues surrounding the W motif are well conserved in a paralog-specific manner [5].

Importantly, the W motif is not the only PBC interaction motif in Hox proteins, and it is even dispensable for some PBC-dependent Hox functions (see [25] for review). The existence of alternative interaction modes between Hox and PBC proteins was definitively established with the identification of novel PBC interaction motifs in several *Drosophila* Hox proteins [45–47]. These motifs were shown to be important for multiple PBC-dependent functions and for Hox-PBC interactions *in vivo* and *in vitro*. Interestingly, these motifs are conserved to different evolutionary extents among Hox paralogs [46,48]. Further, they are used in a paralog- and/or species-specific manner for recruiting the TALE cofactors. Based on these properties we collectively refer to them as, “specific PBC interaction motifs “ (**SPIMs**) [25]. The use of different SPIMs in place of or in conjunction with the W motif (Figure 1a) could allow Hox proteins to adopt different 3D-conformations with TALE cofactors, thereby increasing the range of binding sites that are recognized. Consistent with this idea, a crystal structure of Exd bound to DNA with the Ubx homeodomain and one such SPIM, the so-called UbdA motif, suggests that this SPIM-mediated interaction with Exd might influence the positioning of the Ubx homeodomain within the DNA major groove

[49]. Such differences in DNA recognition mediated by SPIM-TALE interactions could be especially important when binding low affinity, non-consensus binding sites.

In sum, the emerging evidence suggests that due to the availability of multiple W motifs and SPIMs, Hox proteins utilize a striking degree of molecular and structural plasticity when interacting with their TALE partners. The detailed sequence of the DNA-binding site is also likely to play a critical role in specificity. High affinity consensus binding sites may be ideally suited to bind Hox-TALE complexes that only require the presence of a generic W-motif, explaining their ability to be recognised by many Hox-TALE complexes [19,47]. For low affinity binding sites that nevertheless fit the Hox-TALE consensus nnGAYnnATnn, such as Scr-Exd bound to *fkh250*, binding specificity would depend on paralog-specific residues revealed by latent specificity. Finally, the recognition of low affinity binding sites that do not fit this consensus may depend on additional SPIM-mediated conformation modes (Figure 4a), as observed for Ubx-Exd and AbdA-Exd complexes bound to the *Dll* repressor element [45,46]. According to this view, SPIMs would be particularly important for paralog specific binding to non-consensus binding sites and subsequent gene regulation [50]. Further tests of this idea may come from the structural analysis of additional paralog-specific binding sites such as those in *rho* [22] or *svb* [14].

Diversity in the composition and structure of Hox-TALE complexes

An additional layer of diversity in Hox-TALE complexes stems from the use of alternative protein isoforms of both the Hox and TALE factors. For example, *Drosophila Ubx* encodes several alternatively spliced isoforms that differ in the number of W motifs and SPIMs [51]. In several cases, different Ubx isoforms appear to have distinct transcriptional regulatory properties [52,53]. For the TALE factors, the DNA-binding of Meis/Hth has been suggested to affect W-independent interactions between several Hox proteins and PBC cofactors [47]. Interestingly, both *Meis* and *hth* encode HD-containing and HD-less isoforms (Figures 1b). These isoforms have similar embryonic expression profiles and could therefore be responsible for developmentally distinct W- or SPIM-dependent interaction modes. Consistent with this idea, the two Hth isoforms carry out distinct and largely separable functions *in vivo*, suggesting that their DNA targeting and functional properties differ [54].

In vertebrates, where the number of TALE family members is much larger compared to invertebrate species, the potential diversity of Hox-TALE complexes is greatly increased [11]. Based on several observations, these different complexes may carry out distinct functions *in vivo*. For example, Pbx1 but not Pbx2 was shown to be required with Prep1 for HoxB1-dependent expression of the human *α2(V) collagen* gene [55]. Analogously, posterior Hox proteins display preferential interaction affinities with Meis *in vitro* when compared to anterior and central paralogs [56]. Accordingly, Meis has been described to enhance the aberrant activities of posterior Hox proteins in several cancers and leukemia [57,58], while the Prep factors have either no function or opposite effects in these pathological contexts [59,60]. At least some of these functional differences could be a consequence of distinct DNA binding properties of different Hox-TALE complexes. Consistent with this idea, genome-wide ChIP studies showed that Prep-containing complexes tend to bind close to promoter regions, while Meis-containing complexes exhibit

a different genome-wide distribution [34]. In addition, Meis complexes are more frequently associated with canonical Hox binding sites compared to Prep binding sites, highlighting another potential difference in how these cofactors may contribute to both Hox-dependent and Hox-independent functions. Prep has also been described to negatively control the stability of Meis1 indirectly, by sequestering Pbx [61]. The general tumour-inhibiting activity of Prep was thus proposed to rely on Meisdestabilization.

Finally, Hox proteins can also regulate the transcription of TALE genes [16,62] and, interestingly, Hox and TALE mRNAs can be co-targeted by the same microRNA in a tissue-specific manner [63]. Together these observations illustrate that the relative abundance of TALE and Hox proteins is tightly controlled through both transcriptional and post-transcriptional cross-regulatory relationships. We speculate that subtle quantitative variations in relative TALE protein levels could have strong qualitative consequences on the choice of interaction modes and, as a result, the activity of Hox-TALE complexes *in vivo* (Figure 4b).

Concluding remarks

Most of our knowledge on the molecular mode of action of Hox and TALE proteins results from studies on high affinity DNA-binding sites. Although this important amount of work has provided key molecular insights, it could not account for all the molecular complexity underlying Hox/TALE function *in vivo*. Recent work showed that Hox/TALE specificity could in fact rely on DNA-shape recognition mechanisms and the binding to non-consensus and low affinity DNA-binding sites. In addition, Hox proteins could display a versatile usage of different PBC interaction motifs to eventually form complexes with distinct 3-dimentional conformations. Altogether these findings reveal the existence of an expected high degree of flexibility in protein-DNA and protein-protein interaction modes. How such interaction flexibility could account for the various degrees of specificity that Hox and TALE proteins carry out *in vivo* is a central issue for future research (see outstanding questions).

Glossary

Chromatin immunoprecipitation (ChIP)	Using antibodies against transcription factors or histone modifications to identify where they are bound in chromatin, usually by deep sequencing of the immunoprecipitated fragments.
Cofactor	A protein that functions together with a transcription factor. Some cofactors bind DNA cooperatively with transcription factors. For example, Pbx is a cofactor for the Hox transcription factors.
Homeodomain	A very common DNA binding domain in eukaryotes, comprised of three alpha helices and an unstructured N-terminal arm. Most Homeodomains are 60 amino acids; the exception is for the TALE homeodomains that are 63 amino acids long.

Hox genes	A set of genes encoding homeodomain-containing transcription factors that control cell fates along the anterior-posterior axis of all bilaterians. There are eight Hox genes in *Drosophila* and thirty-nine Hox genes in mammals.
Latent specificity	A mechanism in which a protein-protein interaction between a transcription factor and a cofactor uncovers novel DNA binding specificity.
Paralogs	Genes related by homology within the same organism. The eight Hox genes in *Drosophila* are paralogs.
PBC	A set of TALE family cofactors that include Extradenticle (Exd) from *Drosophila*, Pbx proteins from mammals, and Ceh-20 from *C. elegans*. See also Box 1.
SELEX-seq	A method to identify the DNA binding preferences for any transcription factor or transcription factor complex. See also Box 2.
SPIM	"Specific PBC Interaction Motif"; short PBC-interacting peptides that are present in some, but not all, Hox paralogs, and conserved at different evolutionary extends in animals.
TALE homeodomain	A subset of homeodomains that have a "three amino acid loop extension", three additional amino acids between helices one and two that provide a binding pocket for the W-motifs of Hox proteins. The PBC, Meis/Hth, and Prep proteins all have TALE homeodomains. See also Box 1.
W-motif/hexapeptide	Short tryptophan-containing peptides in Hox proteins N-terminal to the homeodomain that interacts with the TALE motif of PBC proteins. Although they are conserved in a paralog-specific manner, Hox proteins from anterior, central and posterior groups contain at least one W-motif.

References

1. Pick L, Heffer A. Hox gene evolution: Multiple mechanisms contributing to evolutionary novelties. Ann. N. Y. Acad. Sci. 2012; 1256:15–32. [PubMed: 22320178]
2. Hombría JC-G, Lovegrove B. Beyond homeosis--HOX function in morphogenesis and organogenesis. Differentiation. 2003; 71:461–76. [PubMed: 14641327]
3. Seifert A, et al. Role of Hox genes in stem cell differentiation. World J. Stem Cells. 2015; 7:583–95. [PubMed: 25914765]
4. Holland PWH. Evolution of homeobox genes. Wiley Interdisciplinary Reviews: Developmental Biology. 2013; 2:31–45. [PubMed: 23799629]
5. Merabet S, et al. Classification of sequence signatures: a guide to Hox protein function. Bioessays. 2009; 31:500–511. [PubMed: 19334006]
6. Noyes MB, et al. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell. 2008; 133:1277–1289. [PubMed: 18585360]

7. Mann RS, et al. Hox Specificity Unique Roles for Cofactors and Collaborators. Current Topics in Developmental Biology. 2009; 88:63–101. [PubMed: 19651302]
8. Rauskolb C, et al. Extradenticle, a Regulator of Homeotic Gene Activity, Is a Homolog of the Homeobox-Containing Human Proto-Oncogene Pbx1. Cell. 1993; 74:1101–1112. [PubMed: 8104703]
9. Mann RS, Chan SK. Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. Trends Genet. 1996; 12:258–262. [PubMed: 8763497]
10. Mukherjee K, Bürglin TR. Comprehensive analysis of animal TALE homeobox genes: New conserved motifs and cases of accelerated evolution. J. Mol. Evol. 2007; 65:137–153. [PubMed: 17665086]
11. Longobardi E, et al. Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. Dev. Dyn. 2014; 243:59–75. [PubMed: 23873833]
12. Gebelein B, et al. Specificity of Distalless Repression and Limb Primordia Development by Abdominal Hox Proteins. Dev. Cell. 2002; 3:487–498. [PubMed: 12408801]
13. Zheng C, Jin FQ, C.M. Hox Proteins Act as Transcriptional Guarantors to Ensure Terminal Differentiation. Cell Rep. Nov.2015 17:1343–1352. [PubMed: 26547238]
14. Crocker J, et al. Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. Cell. 2015; 160:191–203. [PubMed: 25557079]
15. Gould, a, et al. Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. Genes Dev. 1997; 11:900–913. [PubMed: 9106661]
16. Rivas ML, Espinosa-Vázquez JM, Sambrani N, Greig S, Merabet S, Graba Y, H.J. Antagonism versus cooperativity with TALE cofactors at the base of the functional diversification of Hox protein function. PLoS Genet. 2013; 9(2)
17. Sambrani N, et al. Distinct molecular strategies for Hox-mediated limb suppression in Drosophila: from cooperativity to dispensability/antagonism in TALE partnership. PLoS Genet. 2013; 9(3)
18. Hirth F, et al. Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of Drosophila. Development. 2001; 128:4781–4788. [PubMed: 11731458]
19. Ryoo HD, Mann RS. The control of trunk Hox specificity and activity by Extradenticle. Genes Dev. 1999; 13:1704–16. [PubMed: 10398683]
20. Joshi R, et al. Functional specificity of a Hox protein mediated by the recognition of minor groove structure. Cell. 2007; 131:530–43. [PubMed: 17981120]
21. Joshi R, et al. Dissecting the functional specificities of two Hox proteins. Genes Dev. 2010; 24:1533–1545. [PubMed: 20634319]
22. Li-kroeger D, et al. Hox and Senseless Antagonism Functions as a Molecular Switch to Regulate EGF Secretion in the Drosophila PNS. Dev. Cell. 2008; 15:298–308. [PubMed: 18694568]
23. Ebner A, et al. Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. Development. 2005; 132:1591–600. [PubMed: 15753213]
24. Jung H, et al. Evolving hox activity profiles govern diversity in locomotor systems. Developmental Cell. 2014; 29:171–187. [PubMed: 24746670]
25. Merabet S, Hudry B. Hox transcriptional specificity despite a single class of cofactors: are flexible interaction modes the key? Plasticity in Hox/PBC interaction modes as a common molecular strategy for shaping Hox transcriptional activities. Bioessays. 2013; 35:88–92. [PubMed: 23255231]
26. Hueber SD, et al. Comparative analysis of Hox downstream genes in Drosophila. Development. 2007; 392:381–392. [PubMed: 17166915]
27. Slattery M, et al. Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. Cell. 2011; 147:1270–1282. [PubMed: 22153072]
28. Riley TR, et al. SELEX-seq: A Method for Characterizing the Complete Repertoire of Binding Site Preferences for Transcription Factor Complexes. Methods Mol. Biol. 2014; 1196:255–78. [PubMed: 25151169]
29. Abe N, et al. Deconvolving the Recognition of DNA Shape from Article Deconvolving the Recognition of DNA Shape from Sequence. Cell. 2015; 161:1–12.

30. Siggers T, Gordân R. Protein-DNA binding: Complexities and multi-protein codes. Nucleic Acids Res. 2014; 42:2099–2111. [PubMed: 24243859]
31. Slattery M, et al. Absence of a simple code: how transcription factors read the genome. Trends Biochem. Sci. 2014; 39:381–399. [PubMed: 25129887]
32. Beh CY, El-Sharnouby S, Chatzipli A, Russell S, Choo SW, W.R. Roles of cofactors and chromatin accessibility in Hox protein target specificity. Epigenetics Chromatin. 2016; 8:1–17. [PubMed: 26753000]
33. Kim S, et al. Probing allostery through DNA. Science. 2013; 339:816–9. [PubMed: 23413354]
34. Penkov D, et al. Analysis of the DNA-Binding Profile and Function of TALE Homeoproteins Reveals Their Specialization and Specific Interactions with Hox Genes/Proteins. Cell Rep. 2013; 3:1321–1333. [PubMed: 23602564]
35. Donaldson IJ, et al. Genome-wide occupancy links Hoxa2 to Wnt-β-catenin signaling in mouse embryonic development. Nucleic Acids Res. 2012; 40:3990–4001. [PubMed: 22223247]
36. Li X, et al. Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol. 2008; 6(2)
37. Slattery M, et al. Genome-wide tissue-specific occupancy of the hox protein ultrabithorax and hox cofactor homothorax in Drosophila. PLoS One. 2011; 6(4)
38. Jiang J, Levine M. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell. 1993; 72:741–752. [PubMed: 8453668]
39. Scardigli R, et al. Direct and concentration-dependent regulation of the proneural gene Neurogenin2 by Pax6. Development. 2003; 130:3269–3281. [PubMed: 12783797]
40. Gaudet J, Mango SE. Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science. 2002; 295:821–825. [PubMed: 11823633]
41. Ramos AI, Barolo S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2013; 368:20130018. [PubMed: 24218631]
42. Rowan S, et al. Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity. Genes Dev. 2010; 24:980–985. [PubMed: 20413611]
43. Crocker J, Preger-Ben Noon E, S.D. The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. Curr. Top. Dev. Biol. 2016; 116
44. Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, L.M. Suboptimization of developmental enhancers. Science. 2015; 16:325–8. [PubMed: 26472909]
45. Merabet S, et al. A unique Extradenticle recruitment mode in the Drosophila Hox protein Ultrabithorax. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:16946–51. [PubMed: 17942685]
46. Lelli KM, et al. Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:21122–7. [PubMed: 22160705]
47. Hudry B, et al. Hox proteins display a common and ancestral ability to diversify their interaction mode with the PBC class cofactors. PLoS Biol. 2012; 10(6)
48. Merabet S, et al. Insights into Hox protein function from a large scale combinatorial analysis of protein domains. PLoS Genet. 2011; 7(10)
49. Foos N, et al. A Flexible Extension of the Drosophila Ultrabithorax Homeodomain Defines a Novel Hox/PBC Interaction Mode. Structure. 2015; 23:270–279. [PubMed: 25651060]
50. Merabet S, Lohmann I. Toward a New Twist in Hox and TALE DNA-Binding Specificity. Developmental Cell. 2015; 32:259–61. [PubMed: 25669880]
51. Lopez, a J.; Hogness, DS. Immunochemical dissection of the Ultrabithorax homeoprotein family in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 1991; 88:9924–8. [PubMed: 1719557]
52. Saadaoui M, et al. Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:2276–81. [PubMed: 21262810]
53. Reed HC, et al. Alternative splicing modulates Ubx protein function in Drosophila melanogaster. Genetics. 2010; 184:745–58. [PubMed: 20038634]
54. Noro B, et al. Distinct functions of homeodomain-containing and homeodomain-less isoforms encoded by homothorax. Genes Dev. 2006; 20:1636–1650. [PubMed: 16778079]

55. Penkov D, et al. Cooperative interactions between PBX, PREP, and HOX proteins modulate the activity of the alpha 2(V) collagen (COL5A2) promoter. J. Biol. Chem. 2000; 275:16681–9. [PubMed: 10748126]
56. Shen WF, et al. AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. Mol. Cell. Biol. 1997; 17:6448–58. [PubMed: 9343407]
57. Sitwala KV, et al. HOX proteins and leukemia. Int. J. Clin. Exp. Pathol. 2008; 1:461–474. [PubMed: 18787682]
58. Crijns, a P.G., et al. MEIS and PBX homeobox proteins in ovarian cancer. Eur. J. Cancer. 2007; 43:2495–505. [PubMed: 17949970]
59. Longobardi E, et al. Prep1 (pKnox1)-deficiency leads to spontaneous tumor development in mice and accelerates EmuMyc lymphomagenesis: a tumor suppressor role for Prep1. Mol Oncol. Apr. 2010 :126–34. [PubMed: 20106730]
60. Thorsteinsdottir U, et al. Defining Roles for HOX and MEIS1 Genes in Induction of Acute Myeloid Leukemia. Mol. Cell. Biol. 2001; 21:224–34. [PubMed: 11113197]
61. Dardaei L, et al. Prep1 and Meis1 competition for Pbx1 binding regulates protein stability and tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:E896–905. [PubMed: 24578510]
62. Zheng C, Diaz-Cuadros M, C.M. Hox Genes Promote Neuronal Subtype Diversification through Posterior Induction in Caenorhabditis elegans. Neuron. Nov.2015 4:514–27. [PubMed: 26539892]
63. Garaulet DL, et al. Homeotic Function of Drosophila Bithorax-Complex miRNAs Mediates Fertility by Restricting Multiple Hox Genes and TALE Cofactors in the CNS. Developmental Cell. 2014; 29:635–648. [PubMed: 24909902]
64. Bürglin TR. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res. 1997; 25:4173–4180. [PubMed: 9336443]
65. Mukherjee K, Bürglin TR. Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution. J. Mol. Evol. 2007; 65:137–53. [PubMed: 17665086]
66. Bürglin TR. The PBC domain contains a MEINOX domain: Coevolution of Hox and TALE homeobox genes? Dev. Genes Evol. 1998; 208:113–116. [PubMed: 9569353]
67. Williams TM, et al. Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction. Dev. Biol. 2005; 277:457–71. [PubMed: 15617687]
68. Rieckhof GE, et al. Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. Cell. 1997; 91:171–183. [PubMed: 9346235]
69. Mercader N, et al. Ectopic Meis1 expression in the mouse limb bud alters P-D patterning in a Pbx1-independent manner. Int. J. Dev. Biol. 2009; 53:1483–1494. [PubMed: 19247936]
70. Oliphant AR, et al. Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. Mol. Cell. Biol. 1989; 9:2944–2949. [PubMed: 2674675]
71. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science. 1990; 249:505–510. [PubMed: 2200121]
72. Ellington, a D.; Szostak, JW. In vitro selection of RNA molecules that bind specific ligands. Nature. 1990; 346:818–822. [PubMed: 1697402]
73. Zykovich A, et al. Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. Nucleic Acids Res. 2009; 37:e151. [PubMed: 19843614]
74. Jolma A, et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. 2010; 20:861–873. [PubMed: 20378718]
75. Jolma A, et al. DNA-binding specificities of human transcription factors. Cell. 2013; 152:327–339. [PubMed: 23332764]
76. Jolma A, et al. DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature. 2015; 527:384–388. [PubMed: 26550823]

Trends Box

- Hox proteins regulate a wide variety of target genes with different specificity requirements.
- Hox proteins achieve DNA binding specificity by binding with TALE family cofactors through a mechanism called latent specificity.
- The recognition of DNA shape is an independent mechanism of DNA binding site recognition used by Hox-TALE complexes.
- SELEX-seq and other high-throughput methods are useful approaches for characterizing the DNA binding preferences of DNA binding proteins and complexes of DNA binding proteins.
- Highly specific DNA binding sites tend to be low affinity, non-consensus sites, whereas high affinity sites tend to have lower specificity.
- Hox proteins have a variety of short peptide motifs that are used to interact with TALE cofactors.
- A variety of TALE cofactors, generated by gene duplication or from the expression of alternative protein isoforms, provides a large number of potential Hox-TALE complexes in vivo, particularly in vertebrates.
- The paralog specific regulation of Hox target genes can arise from both DNA-binding and non-DNA-binding based mechanisms.

Box 1 Three Amino acid Loop Extension Homeodomain Proteins

TALE (Three Amino acid Loop Extension) proteins belong to the superclass of homeodomain (HD)-containing transcription factors. They are characterized by the presence of three additional residues between the first and second helix of the HD when compared to the other classes of HDs. TALE family members are found from plants to fungi and animals [64]. In animals, they have been subdivided into five different subclasses: PBC, IRO, MKX, TGIF and MEIS [65]. The MEIS subclass is itself subdivided into the PREP and MEIS subfamilies, which derive from a common ancestral protein. PREP and MEIS are also referred to by the term of MEINOX, due to the similarity of the N-terminal Meis and Knox domains between animal and plant proteins [66].

PBC and MEIS/PREP are the only TALE subclasses described to form complexes with Hox proteins. In the case of PBC, interaction with Hox involves the three additional TALE-specific residues that create a hydrophobic pocket [7]. This pocket binds a tryptophan (W) that is part of the W-motif in Hox proteins (also called the YPWM motif or hexapeptide (HX) motif). In contrast, MEIS was shown to use carboxy-terminal sequences to interact with posterior Hox proteins [67]. The formation of Hox/PBC/MEIS or Hox/PBC/PREP complexes is thought to rely on Hox-PBC and PBC-MEIS or PBC-PREP interactions, but this has not yet been confirmed by structural studies.

The PBC subclass includes several proteins in vertebrates (Pbx 1-4, Pbx stands for pre-B cell leukemia homeobox) and nematodes (ceh-20, ceh-40, ceh-60; the last two being highly divergent), while only one representative is present in *Drosophila* (Extradenticle, Exd). Different MEIS (Myeloid Ecotropic Integration Site) and PREP (Pbx Regulatory Protein) members are also present in vertebrates while only one representative of each subfamily is present in nematodes (Figure 1a). *Drosophila* has also only one MEIS member (Homothorax, Hth) and no PREP representative. PREP proteins are however present in other insect species, suggesting that PREP was specifically lost in the *Drosophila* lineage. In *Drosophila*, the nuclear localization of Exd depends on the interaction with Hth [68]. The equivalent protein interaction domains are highly conserved in Pbx and Meis and in at least some cases their nuclear localization is also co-dependent as in *Drosophila* [69].

Box 2 SELEX-seq

SELEX, the Systematic Evolution of Ligands by EXponential enrichment, is an *in vitro* procedure that was originally used to identify RNA or DNA oligonucleotides that bind specific proteins or other ligands at high affinity [70–72]. The procedure usually begins with a large random library of oligonucleotides that is sequentially selected for binding to a specific ligand by multiple rounds of binding, purification, and amplification. In its original application, many rounds of selection resulted in the identification of a small number of high affinity oligonucleotides that were subsequently characterized by cloning and traditional sequencing.

SELEX-seq [27,28] uses next generation deep sequencing to analyze the bound oligonucleotides at each round of selection, including the initial unselected library (round 0). Relative affinities are calculated by comparing the enrichment of specific sequences (k-mers) in successive rounds of selection. Because of the sequencing depth afforded by deep sequencing, it is only necessary to carry out a few rounds of selection, which allows the identification of many thousands of binding sites that range in affinity by about two orders of magnitude. When used in conjunction with Electrophoretic Mobility Shift Assays (EMSAs), oligonucleotides that are selected to bind multiprotein complexes can be defined. Alternative SELEX-based procedures, such as Bind-n-Seq [73] and HT-SELEX [74], have also been described. Because these procedures are typically solution based, it is harder to distinguish between the binding of multiprotein complexes from monomer binding. However, they have been effectively used in a high throughput manner to define the specificities of thousands of human DNA binding proteins [75].

When applied to the *Drosophila* Hox-TALE complexes, SELEX-seq demonstrated that the TALE factors uncover latent DNA binding specificities that are present in Hox proteins [27]. These latent specificities are only revealed by complex formation with the TALE cofactors, presumably by inducing a conformational change in the Hox protein. In at least one case, this conformational change was shown to be the result of stabilizing the structure of the N-terminal arm of the Hox homeodomain, so that it can read the structure of a DNA binding site [20]. More recently, a high-throughput study using HT-SELEX and human transcription factors suggests the existence of many additional examples of the latent specificity mechanism [76].

Outstanding questions

- What are the different shapes of the DNA double helix that could be recognized by different Hox/TALE complexes? Does it constitute a general recognition mechanism for Hox/TALE specificity?

- What is the range of low affinity DNA-binding sites that could be recognized by Hox/TALE complexes and what is their relative contribution to Hox specific function in general?

- Is there a paralog-specific signature in the sequence and/or structure of low affinity DNA-binding sites?

- How low affinity and non-consensus DNA-binding sites could be identified from ChIP and/or predicted by bioinformatics approaches?

- What are the various SPIMs (**s**pecific **P**BC **i**nteraction **m**otifs) and how could they help Hox proteins recognize low affinity and non-consensus DNA-binding sites?

- How might Meis and Prep influence Hox-PBC activity at the protein-protein and protein-DNA interaction levels?

- Do Hox/TALE complexes require the presence of additional cofactors to recognize low affinity and non-consensus binding sites *in vivo*?

- How do the lessons learned from Hox-TALE studies apply to other transcription factor complexes?

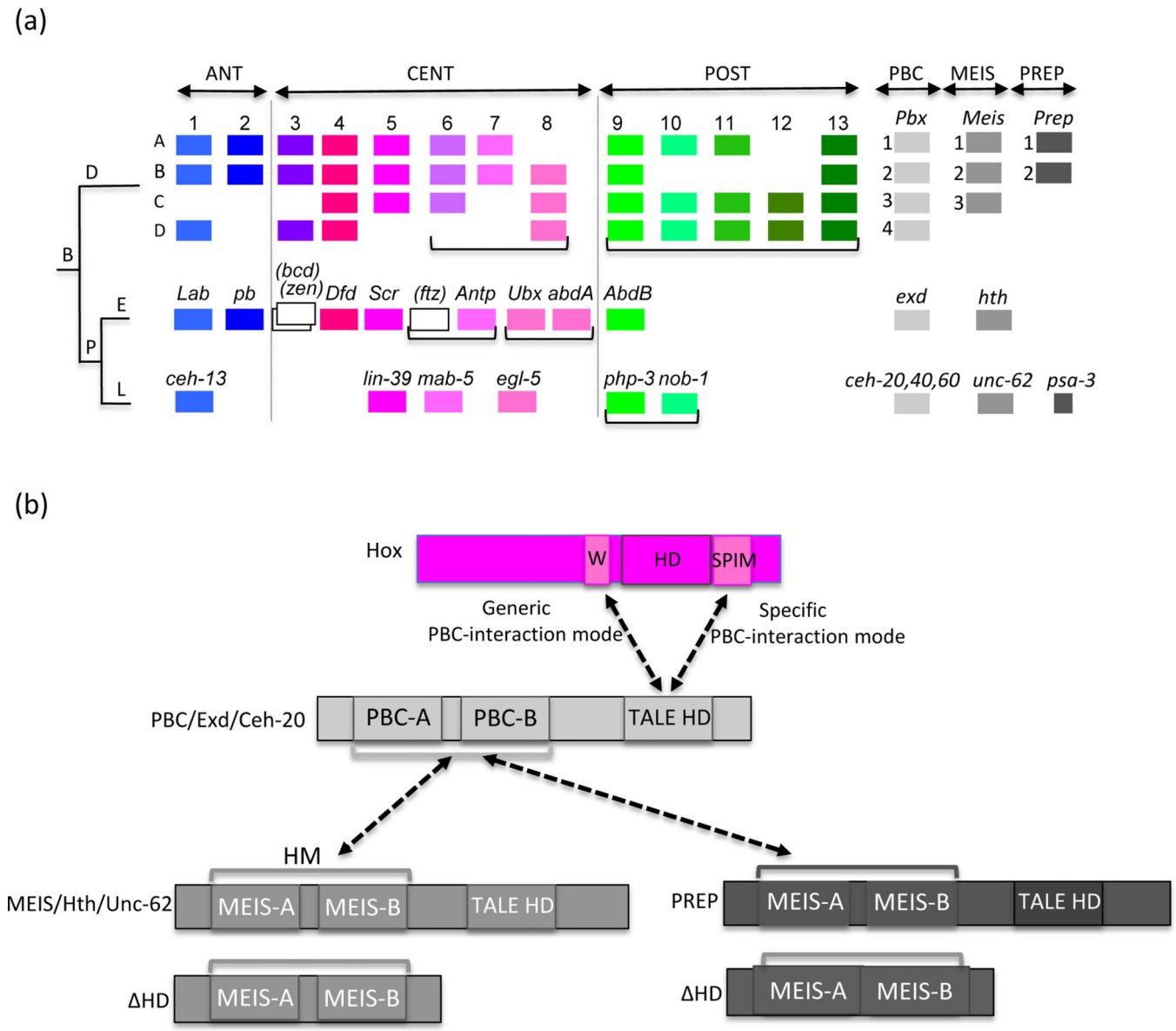

Figure 1. Hox and TALE members in vertebrates and invertebrates

(a). Evolution of Hox and TALE genes in Bilateria (B). Representative species for Deuterostomes (D) and for Ecdysozoa (E) and Lophotrochoza (L) branches from Protostomes (P) are *Mus musculus*, *Drosophila melanogaster* and *Caenorhabditis elegans*, respectively. Hox genes are organized into anterior (ant, blue-graded boxes), central (cent, pink-graded boxes) and posterior (post, green-graded boxes) paralog groups. This representation does not reflect the genomic cluster organisation. Independent duplications are indicated for central and posterior Hox genes (brackets). Boxes representing *bicoid* (*bcd*), *zerknült* (*zen*) and *fushi-tarazu* (*ftz*) are not colour-filled because these three genes strongly diverged in *Drosophila*. Note that PREP was specifically lost in *Drosophila* among Ecdysozoan species. **(b).** Schematic representation of motifs and domains involved in the Hox-TALE partnership. Hox proteins can use a generic (W-containing) motif and/or a specific PBC interaction motif (SPIM) to interact with the homeodomain (HD) of PBC.

PBC-A and PBC-B domains interact with the MEIS-A and MEIS-B domains of Meis or Prep. Note that these complexes could also form with HD-less isoforms of Hth and Meis.

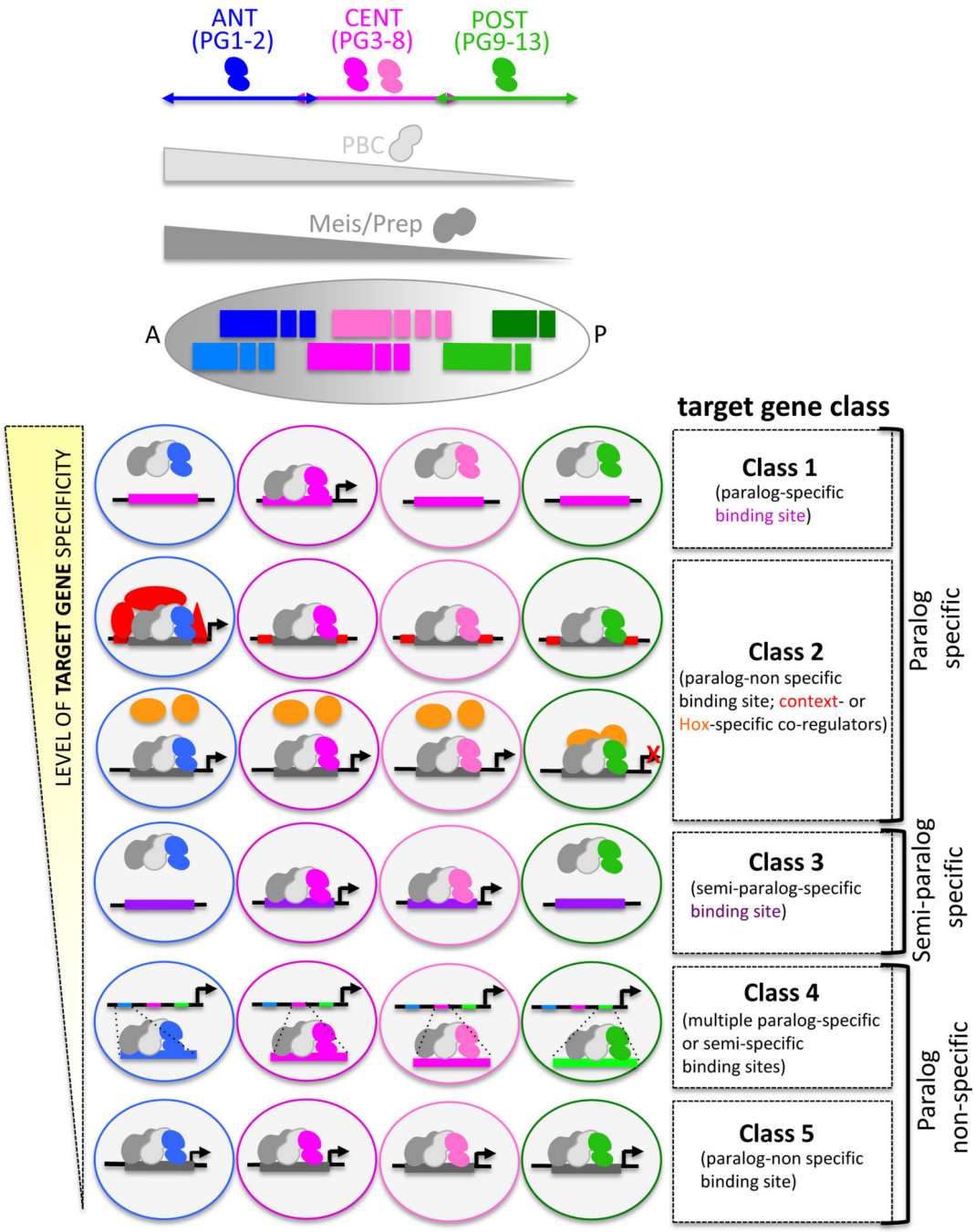

Figure 2. Key Figure

Different classes of target genes underlie Hox function *in vivo*. Target genes are classified (class 1 to 5) according to their level of specificity for Hox paralogs. Different classes of target genes contain Hox-TALE binding sites with various degrees of specificity, as indicated. Note that the level of binding site specificity does not necessarily correlate with the paralog specificity of the target gene. TALE-independent Hox target genes could also fall into those categories, as discussed elsewhere [7]. Known examples are described for class 1 (with specific activation:[19–23]), class 2 (lower row, with specific repression: [24]), class 3

(with activated or repressed target genes [12–15]) and class 5 (corresponding to artificial constructs: [12,19]). Target genes for class 2 (upper row, corresponding to the tritocerebrum context [18]) and class 4 are speculative (see also Figure 4a). Schematic expression profile of Hox and TALE proteins along the anterior-posterior (AP) axis of the *Drosophila* embryo is provided above the model. Differences in TALE expression levels might influence Hox function (see also Figure 4b). The color code for Hox paralogs and TALE cofactors is the same as in Figure 1. This color code is also used for binding sites that are paralog-specific or semi-paralog-specific. Non-paralog-specific binding sites are depicted in dark gray. Presence or absence of arrows, respectively, indicates regulation or not by the Hox-cofactor complex. Red and orange proteins in class 2 target genes highlight additional cofactors that provide cell-,(upper row) or Hox-specific (lower row) regulatory activity.

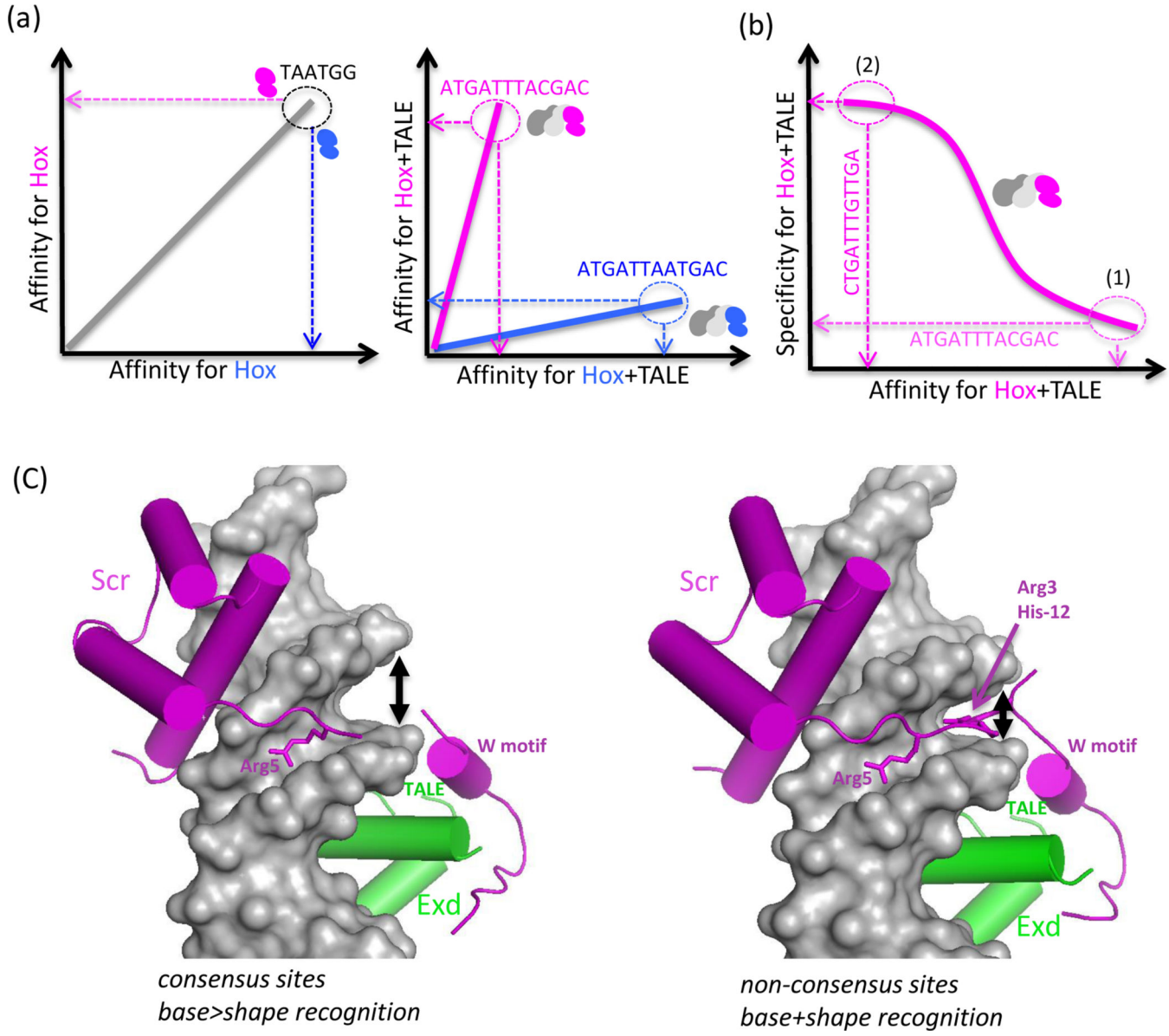

Figure 3. Specificity is inversely correlated to binding affinity

(a). Interaction with TALE cofactors allows Hox proteins from different paralog groups (highlighted in pink and blue) to preferentially bind to distinct sequences. In the absence of TALE factors, two Hox proteins bind to very similar sequences (left graph) while in the presence of TALE factors the same two Hox proteins exhibit distinct preferences (right graph). Representative nucleotide sequences recognized by monomers or Hox/TALE complexes are provided according to [6,27]. **(b).** The graph shows a hypothetical trade-off between specificity and affinity. Representative nucleotide sequences recognized by a Hox-TALE complex are provided according to [27]. High specificity and low affinity binding sites depend on both DNA shape and base recognition mechanisms as discussed in the main text. (c). Three dimensional structures illustrating the recognition of a paralog non-specific binding site (*fkhcon* [19], left, corresponding to case (1) in Figure 3b) and of a paralog-

specific binding site (*fkh* [19], right, corresponding to case (2) in Figure 3b) bound by the Hox protein Scr with its cofactor Exd. Note that complex formation on the paralog-specific site involves additional Scr-specific residues (Arg3 and His-12) and the recognition of a narrow minor grove (indicated by the double black arrow). The equivalent region of the paralog-non-specific complex has a wider minor groove. Adapted from [20] and the coordinates are from Protein Data Base (PDB) structures 2r5y and 2r5z, respectively.

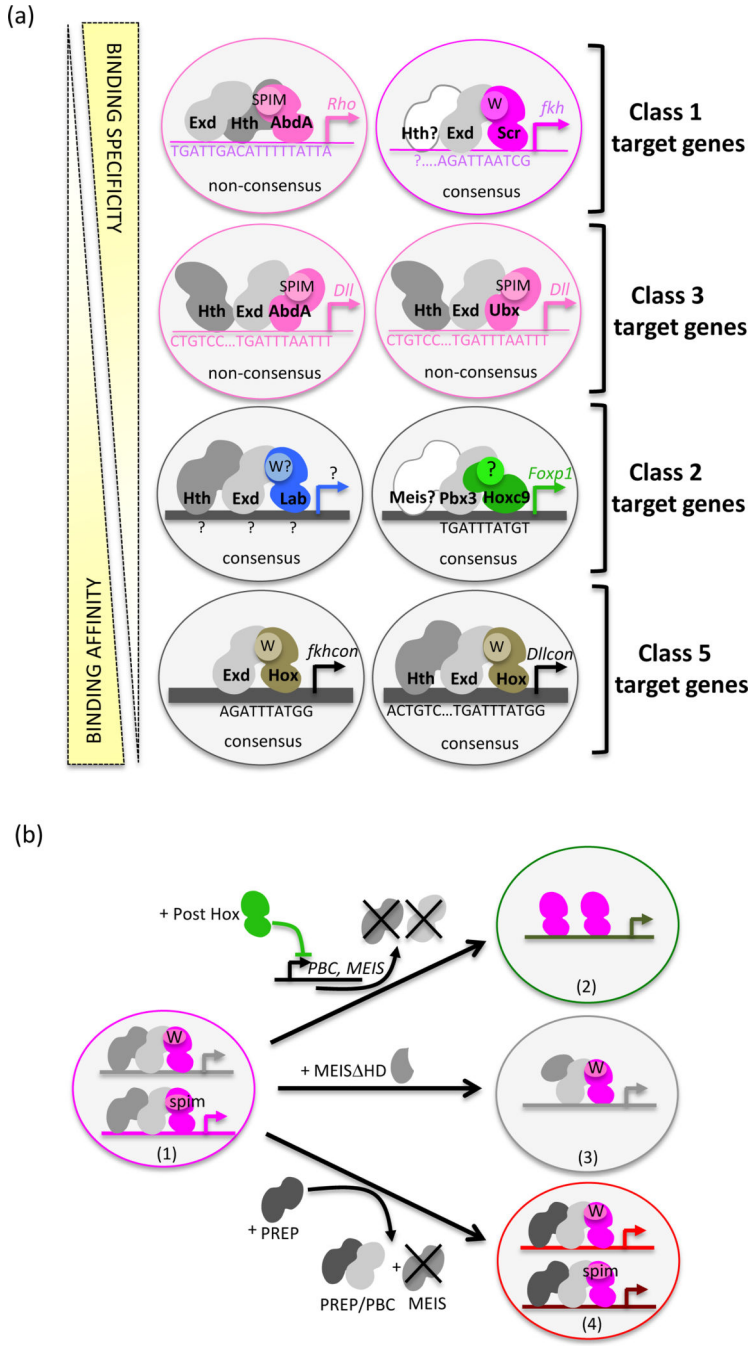

Figure 4. Various combinatorial relationships underlie the function of Hox-TALE complexes

(a) Examples of binding sites characterized in the *cis*-regulatory regions of Hox-TALE target genes. Hox-TALE binding sites that are specific or have intermediate specificity contain a consensus (e.g. *forkhead* (*fkh*, [19]) target gene) or non-consensus (e.g. *rhomboid* (*rho*, [22]), *Distalless* (*Dll*, [12]) and *shavenbaby* (*svb*, [14]) target genes) binding sites. Non-paralog specific Hox-TALE binding sites generally include a consensus binding site (e.g. *Foxp1* [24], *Dllcon* [12] and *fkhcon* [19]). Binding site specificity is illustrated by the color code (dark gray is non specific). Binding site affinity is illustrated by a variable line width.

The brown color of the Hox protein in class 5 target genes illustrates the ability of all paralogs to regulate a non-paralog specific enhancer. Motifs used for complex assembly on these different types of binding sites are indicated (W-containing or SPIM), as discussed in the main text. These motifs are speculative for class 2 target genes **(b).** Variation in the composition of Hox or TALE proteins can influence the interaction mode and, as a result, the activity of Hox-TALE complexes. An example is provided for a central Hox protein (pink) that could use either a W- motif or a SPIM to interact with the PBC and Meis cofactors in cell context (1). Three different additional cell contexts are proposed, based on published data. In cell context (2), the presence of a posterior Hox protein blocks the expression of TALE- encoding genes [16]. As a result, central Hox proteins will only regulate target genes that do not depend on TALE input. In cell context (3), the presence of a HD-less isoform in place of full length Meis does not allow the use of a SPIM for complex assembly [47]. As a consequence, the trimeric complex is not able to regulate paralog-specific target genes that depend on SPIM-mediated conformation modes. In cell context (4), the presence of Prep induces a degradation of Meis, indirectly through the titration of PBC [61], leading to the formation of Hox-PBC-Prep in place of Hox-PBC-Meis complexes. These complexes have the potential to assemble by using the W-containing motif and/or SPIMs but could potentially regulate distinct target genes compared to those regulated by Hox-PBC-Meis complexes.