

Fattening the perspective of Hox protein specificity through SLiMming

LUCREZIA RINALDI, ANDREW J. SAURIN and YACINE GRABA*

Aix Marseille Univ, CNRS, IBDM, Marseille, France

ABSTRACT The functional identification and dissection of protein domains has been a successful approach towards the understanding of Hox protein specificity. However, only a few functional protein domains have been identified; this has been a major limitation in deciphering the molecular modalities of Hox protein action. We explore here, by *in silico* survey of short linear motifs (SLiMs) in Hox proteins, the contribution of SLiMs to Hox proteins, focusing on the mouse, chick and *Drosophila* Hox complement. Our findings reveal a widespread and uniform distribution of SLiMs along Hox protein sequences and identify the most apparent features of Hox associated SLiMs. While few motifs have been associated with Hox proteins so far, this work suggests that many more contribute to Hox protein functions. The potential and difficulties to apprehend the full contribution of SLiMs in controlling Hox protein functions are discussed.

KEY WORDS: *Hox, SLiM, ELM, peptide motif*

Introduction

Transcriptional regulation relies on the assembly of diverse and multi-protein complexes, which ultimately recruit the general transcription machinery. How transcription regulatory complexes assemble and are appropriately targeted in the genome is a central question in biology, still poorly understood (Badis *et al.*, 2009; Lambert *et al.*, 2018; Romero *et al.*, 2012; Smith and Matthews, 2016). Gaining insight into differential gene expression requires grasping the logic and molecular mechanisms underlying the assembly and function of transcription regulatory complexes. Strong protein interactions involving pre-folded and well-structured globular protein domains have largely contributed to our current understanding of how such regulatory complexes assemble, and ultimately provide specificity in gene regulation.

In parallel to this view, the recognition of the functional importance of intrinsically disordered protein region (IDR) argues that protein function also relies on non-globular and intrinsically disordered regions (Tompa, 2012; Tompa *et al.*, 2015). IDRs are at least 30 residues long, often lack any 3D structure, are present in more than 30% of eukaryotic proteins and impact on a wide spectrum of biological functions (Arai, 2018). One of the properties of IDRs is their capacity to transition from disorder to order upon interaction with globular domains or other IDRs (Dunker *et al.*, 2001). Another attribute of IDRs is the presence of Short Linear Motifs (SLiMs, sometimes also referred to as “peptide motifs” or “Eukaryotic Linear

Motifs” (ELM; (Dinkel *et al.*, 2012; Gouw *et al.*, 2018; Tompa *et al.*, 2015)). SLiMs are between 3 and 11 residues on average, and mediate interactions that are usually weak, transient, and dynamic (Davey *et al.*, 2012; Tompa, 2012; Van Roey *et al.*, 2014), features that contrast with globular domain mediated interactions. In principle, SLiM mediated interactions offer a much higher interaction potential than globular domain mediated interactions: while the number of classical globular protein domains is relatively limited (below 100, likely involved in approximately 30 000 interactions in the human proteome), that of peptide motifs is much higher (likely in the range of several thousands, involved in up to 1 000 000 interactions in the human proteome; (Tompa *et al.*, 2014)). SLiMs evolve very easily due the low evolutionary pressure, thus defining a mean for adding/modifying functions to proteins, conferring a high degree of evolvability (Tokuriki and Tawfik, 2009). SLiMs are however also found in conserved and ordered regions, including globular domains, where they mainly localize at the surface of the domain. The large interaction possibilities of SLiM mediated

Abbreviations used in this paper: AbdA, abdominal A; AbdB, abdominal B; Antp, antennapedia; CLV, cleavage site; DEG, degenon; DOC, docking site; ELM, eukaryotic linear motif; HD, homeodomain; HX, hexapeptide; IDR, intrinsically disordered region protein; Lab, labial; LIG, ligand site; MOD, post translational modification site; Pb, proboscipedia; PG, paralog group; RegExp, regular expression; Scr, sex comb reduced; SLiM, short linear motif; TALE, three amino acid loop extension; TRG, subcellular targeting sequence; Ubx, ultrabithorax.

*Address correspondence to: Yacine Graba. Campus de Luminy Case 907, 13288 Marseille Cedex 9, France. Tel: + 33 491 26 96 13. E-mail: yacine.graba@univ-amu.fr
web: <http://www.ibdm.univ-mrs.fr/equipe/mechanisms-of-gene-regulation-by-transcription-factors>  <https://orcid.org/0000-0003-1324-1308>

Submitted: 13 September, 2018. Accepted: 17 September, 2018.

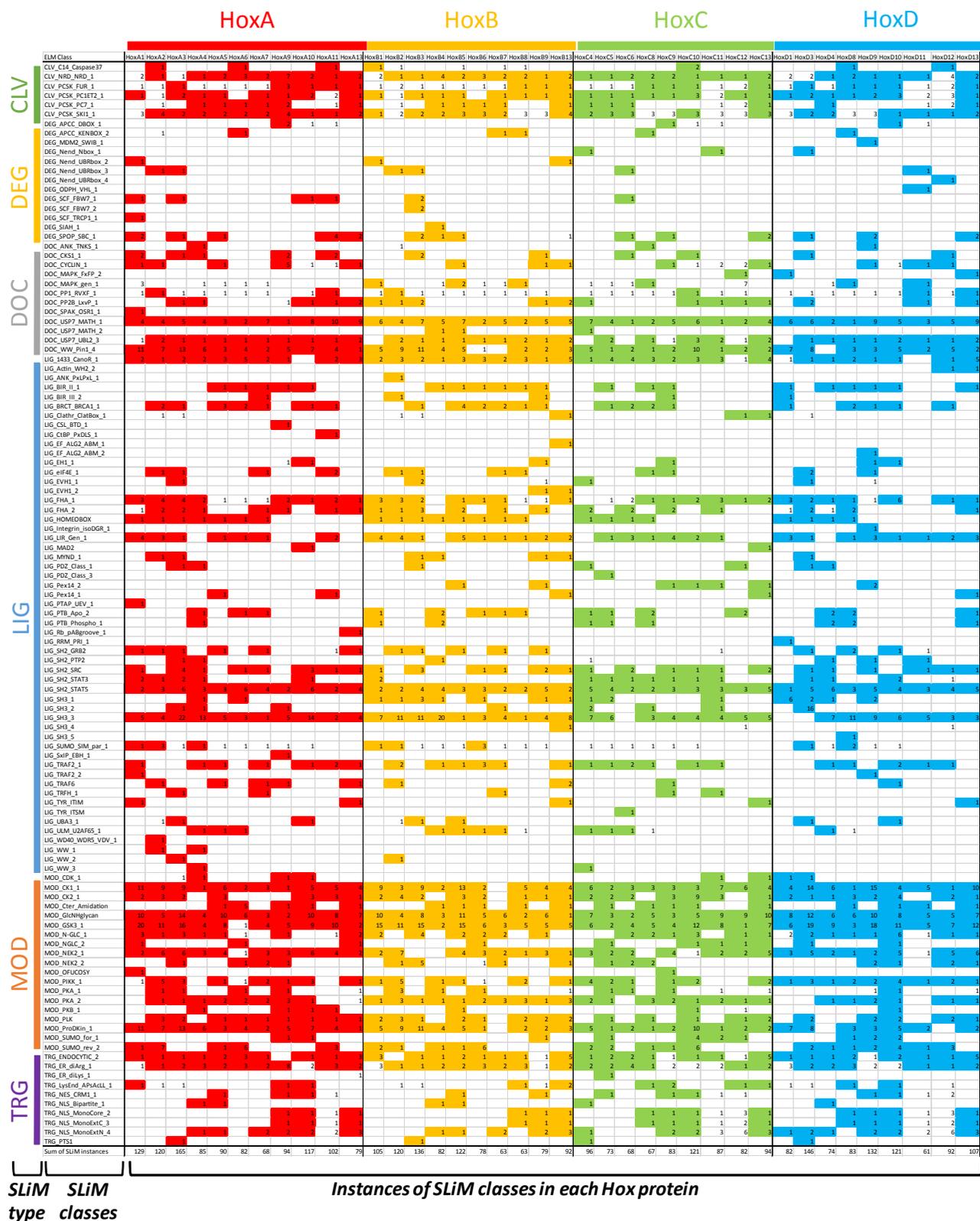


Fig. 1. Short linear motifs (SLiMs) in mouse Hox proteins. SLiMs are grouped in 6 different motif types according to their functions: cleavage or processing site (CLV), degradation site (DEG), enzyme recruitment site (DOC), ligand site (LIG), a post-translational modification (MOD) or a subcellular targeting sites (TRG). The first column displays the SLiM classes. Numbers represents the SLiM instances for each SLiM class, with either a colored background or not. When the background is white, SLiMs were filtered out by the ELM structural filter. Hox A, B, C, and D cluster encoded proteins are respectively in red, orange, green and blue. The bottom row of the table shows the total number of SLiMs instances per protein.

TABLE 1

COMPARISON OF SLiMs IDENTIFIED THROUGH ELM AND SLiMPred

Protein		Number of SLiM predicted with ELM	Number of <i>de novo</i> predicted SLiMs with SLiMPred
Anterior	Labial	168	7
	Proboscipedia	234	8
	Deformed	118	4
	cHoxb4	104	1
	mHoxb4	82	2
Central	Sex comb reduced	121	1
	Antennapedia	80	2
	Ultrabithorax	83	6
	Abdominal A	70	5
	cHoxb8	69	5
	mHoxb8	63	4
Posterior	Abdominal B	119	4
	cHoxb9	78	0
	mHoxb9	79	2

contacts are prone to adapt to the specific environment of many different protein complexes. The intrinsic feature of these protein interactions, weak and transient, render their identification and their functional contribution to biological processes difficult. Yet, a few examples of SLiM mediated interactions have been recognized as essential for the activity of a few protein complexes involved in a number of cellular processes, including protein homeostasis, sub cellular targeting, cellular signaling pathways and regulation of CDK/cyclin activity (Tompa *et al.*, 2014).

Hox proteins are key evolutionary conserved transcription factors controlling diversified morphogenesis in development and evolution (McGinnis and Krumlauf, 1992; Merabet and Galliot, 2015; Merabet and Mann, 2016; Rezsóhazy *et al.*, 2015). Two salient features were early recognized in Hox proteins. The first one is a typical globular domain, the homeodomain (HD), found in all Hox proteins, which serves as the unique DNA binding motif (Burglin and Affolter, 2016; Gehring *et al.*, 1994; Qian *et al.*, 1989; Scott and Weiner, 1984). The second one is a short Hexapeptide (HX) motif, shared by most Hox proteins (with the exception of a few posterior class proteins) mediating interaction with PBC class homeodomain-containing proteins. The HX, despite being pre-folded (Slupsky *et al.*, 2001), has SLiM type features (Baeza *et al.*, 2015) and directly contacts PBC class protein, Pbx and Exd in vertebrates and *Drosophila*, respectively (Mann and Chan, 1996). On the PBC side, the interaction involves a three amino acid loop extension specific to TALE class homeodomains, which defines a hydrophobic pocket that hosts the HX motif (Joshi *et al.*, 2007; LaRonde-LeBlanc and Wolberger, 2003; Longobardi *et al.*, 2014; Passner *et al.*, 1999; Piper *et al.*, 1999). This Hox-PBC interaction was shown to mediate cooperative DNA binding and therefore improved Hox functional specificity (Merabet and Lohmann, 2015; Pearson *et al.*, 2005); (Mann *et al.*, 2009).

Early work on the *Drosophila* Hox Ultrabithorax (Ubx) protein showed that it is mostly composed of IDRs (Liu *et al.*, 2008). It was further established that disordered regions of Ubx modify Ubx monomeric DNA binding, as well as cooperative DNA binding with Exd (Bondos *et al.*,

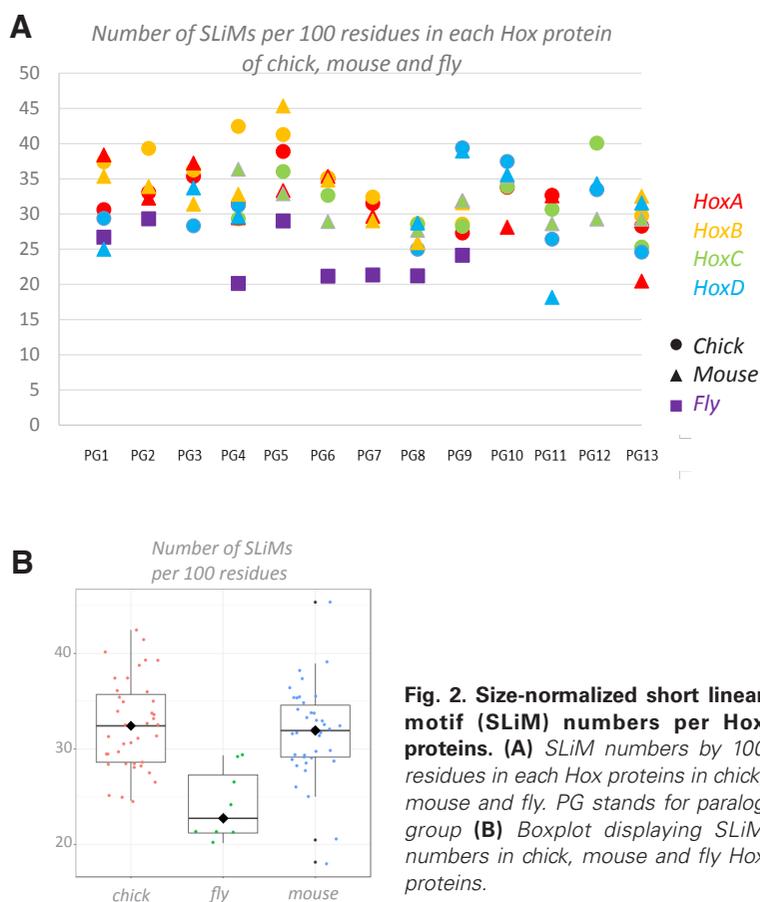
2015; Liu *et al.*, 2008, 2009). Additionally, Ubx IDRs were shown necessary for partnering with structurally folded proteins (Hsiao *et al.*, 2014), and for self-assembly as fibers and films of sheets (Howell *et al.*, 2016; Patterson *et al.*, 2015). How Ubx IDRs mediate such functions however remains elusive, and in particular whether these IDR-mediated activities rely on particular SLiMs has not been defined. More recent work showed that a few SLiMs functionally contribute to the activity of the *Drosophila* Hox proteins Sex combs reduced (Scr), Antennapedia (Antp), Ubx, Abdominal A (AbdA) and AbdB (Baeza *et al.*, 2015; Foos *et al.*, 2015; Hudry *et al.*, 2014; Lelli *et al.*, 2011; Merabet *et al.*, 2011), as well as the human HoxB3, HoxA7 and HoxC8 proteins (Dard *et al.*, 2018). These data collectively identify a few situations where SLiMs contribute to the activity of Hox proteins.

Here we aimed at exploring the extent to which SLiMs are intrinsic to Hox protein sequences. This was initiated using the full Hox complement of three widely used model animals, mouse, chick and *Drosophila*, allowing for comparisons over short and longer evolutionary distances. While not extensive with regard to Hox proteins sequences available, this study clearly highlights the wide contribution of SLiMs to Hox proteins, and delineates the most apparent features of Hox associated SLiMs.

Results

Mapping SLiMs in Hox proteins

To explore the presence of SLiMs in Hox proteins, we ran the 39 *Gallus gallus*, 39 *Mus musculus* and 8 *Drosophila melanogaster* Hox protein sequences through the ELM database. Results of this



analysis are summarized in Fig. 1 for the mouse Hox proteins (see also Supplementary Fig. 1 for the whole dataset). The SLiMs are grouped into 6 different motif types according to their functions: they correspond to a cleavage or processing site (CLV), induce a degradation of the protein (DEG), an enzyme recruitment (DOC), a ligand site (LIG), a post-translational modification (MOD) or a subcellular targeting sites (TRG). The sequences of the SLiMs are normalized as a regular expression (RegExp) (Davey et al., 2012), each defining a “consensus” to which all SLiMs from the same class will obey. Over the 248 SLiM classes registered in the ELM database, 129 were found to be present in Hox proteins. Fig. 1 displays SLiM classes grouped according to the different motif type, for each of the mouse Hox proteins. The instances of each SLiM classes, defined as the number of times the sequence of a SLiM, defined by its RegExp, is present in each Hox protein is variable, ranging from 0 to 31 (see numbers in Fig. 1). Some SLiMs are filtered out by the ELM structural filter, taking into account the predicted structure of the protein (numbers in a white background in Fig. 1). The total instances of SLiMs per protein ranges from 62 to 234 (numbers in the last row in Fig. 1), with a mean of 98. These numbers are necessarily influenced by the size of the protein that displays important variability (from 211 amino acids for chicken HoxD1 to over 800 amino acids for *Drosophila* Proboscipedia (Pb)). To circumvent the issue of protein size, we normalized SLiM numbers by defining the total number of SLiMs for each protein per 100 amino acids. Results displayed in Fig. 2A show that the number of SLiMs per protein seems relatively constant, with the median number of SLiMs per 100 residues being 32 for the vertebrate species and 23 for the fly (Fig. 2B). The mapping of SLiMs in Hox proteins thus reveals a high SLiM content, providing an unprecedented potential to identify novel functional domains in Hox proteins.

We next questioned whether mapping SLiMs in Hox proteins using the ELM resource provides a reasonably complete view of SLiMs in Hox proteins. For that, we turned to the SLiMPred resource (Mooney et al., 2012), that differs from the ELM resource in allowing *de novo* SLiM discovery. We focused this pilot approach on the full fly Hox complement, and a single representative of anterior, central and posterior class Hox proteins from mouse and chick. We found that SLiMPred identified only a few SLiMs not previously identified by the ELM resource (Table 1). This indicates that using SLiMs identified through the ELM resource should provide a reasonably representative idea of SLiMs in Hox proteins. We thus decided to conduct our analyses using ELM-identified SLiMs.

General features of Hox-associated SLiMs

We next analyzed the relative contribution of different SLiM types to the overall set of SLiM present in Hox proteins. Fig. 3A shows the relative contribution of each motif type taking into account the overall set of RegExp motifs from the entire ELM data set: the LIG type of SLiM is the one most represented (56%), followed sequentially by the MOD (13%), DEG (10%), DOC (9%), TRG (8%) and CLV (4%) classes. We found that this distribution is largely conserved for Hox-associated SLiMs, taking collectively Hox sequences from mouse, chick and *Drosophila*, with the exception of the LIG motif type, that display a slight negative enrichment (Fig. 3B, 49% instead of 56%). Overall this suggests that qualitatively, Hox proteins do not seem to display any particular bias toward a given SLiM motif type. This conclusion also holds true when the same analysis is performed using the mouse, chick or *Drosophila* datasets in isolation (Supplementary Fig. 2).

Next, we checked the distribution/conservation of the SLiM classes over the three Hox complements under survey, without taking into account the conservation of the position within the protein, or how often a SLiM is present (Fig. 3C). Of the 129 classes present in our Hox dataset, 66 are found in all three species tested (51%), 32 are found in mouse/chicken (25%), whereas only 1 (1%) was found in both mouse/*Drosophila* and chick/*Drosophila*. SLiMs present in only one species are as follows: 15 SLiMs (12%) for the mouse, 11 (8%) for the chicken and 3 (2%) for the *Drosophila*. This indicates that the majority of SLiM classes contribute equally to the SLiM qualitative content in all three species, and that the conservation of SLiMs classes is higher when considering mouse and chick, rather than vertebrate and *Drosophila* Hox proteins, as expected from the evolutionary distances between these three species.

From a more quantitative point of view, we counted the instances at which SLiMs are found in Hox proteins. To study if there exists a bias towards the usage of particular SLiM motif type, we summed the numbers by motif type, considering *Drosophila*, mouse and chick sequences in isolation or altogether. While the much higher number of RegExp motifs for the LIG motif type (see Fig. 3A and 3B) would predict that the LIG motif provides the major contribution, the data shows that in all cases, the MOD motif type is used with the highest frequency, reaching 52% for the *Drosophila* Hox proteins (Supplementary Fig. 3). LIG type motifs are also frequently used, with a global score over the three species of 27%. The four other motif types have a less pronounced contribution, altogether accounting for 31% (DOC (14%), TRG: (8%), CLV (7%) and DEG (2%)).

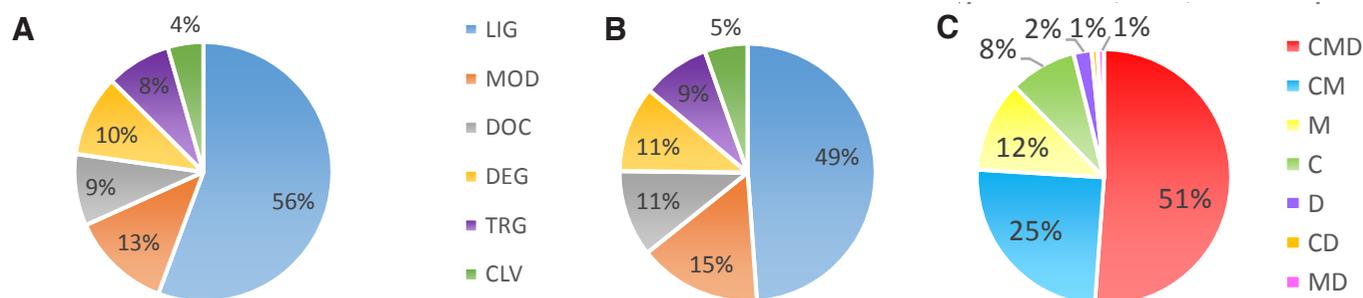


Fig. 3. Distribution of short linear motif (SLiM) classes by motif types. (A) Distribution of SLiM classes by motif types for the whole ELM database. **(B)** Distribution of SLiM classes by motif types for the Hox-associated SLiMs. **(C)** Conservation of SLiM classes by motif types in mouse, chick and fly Hox complements. C, chick; M, mouse; D, *Drosophila*; CM, chick and mouse; CD, chick and *Drosophila*; MD, mouse and *Drosophila*; CMD, chick, mouse and *Drosophila*.

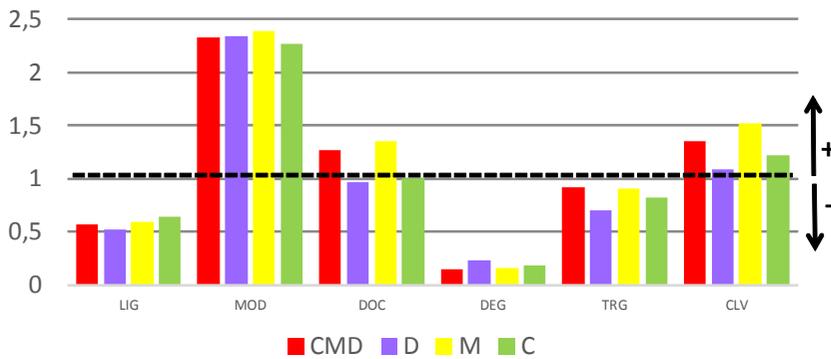


Fig. 4. Short linear motif (SLiM) enrichment in the mouse, chick and *Drosophila* Hox proteins. Enrichment score per motif type in mouse, chick and fly, in isolation or combined. Enrichment score for each type of motif was obtained as a ratio of the number of SLiM instances to the number of SLiM classes by motif types (C, chicken; M, mouse or D, *Drosophila*) or in all three species (CMD).

Fig. 4 displays the SLiM enrichment scores by motif types, which takes into account the number of SLiM classes within each motif type family. This allows a SLiM class not to be scored as overrepresented because it has a large variety of motif types, and hence a greater probability of such motifs to be found in Hox sequences. This highlights the enrichment of the MOD type motif, with the LIG and DEG type motifs being highly underrepresented. This trend is seen when all Hox proteins are taken as a unique data set, as well as when treated in isolation. This collectively suggests that protein complex formation, and its regulation through posttranslational modification is likely central to the activity of Hox

proteins, far beyond its current appreciation through the PBC class protein interaction and the functionally characterized post translational modifications associated to Hox proteins (see Draime *et al.*, this issue).

Position of SLiMs within Hox proteins

We next aimed at exploring if SLiMs tend to accumulate in specific regions of the proteins. We arbitrarily split Hox proteins into 4 sections: the N-terminal region that precedes the HX motif, the linker region that separates the HX motif from the HD, the HD itself, and the region C-terminal to the HD. We then scored for each of the Hox protein SLiM instances within each of these four defined regions (Fig. 5A). Note that the paralogs 11/13 have no HX so the score for the linker region is always 0. Additionally, paralogs 9 and 10 have a very short linker region (5-6 residues) and have no SLiMs in this part of the protein. The mean number of motifs is around 28 in the vertebrates and 22 for flies in the HD; 56/68 in vertebrates and flies respectively for the N-terminal part; 4.5/8 in vertebrates and flies respectively for the linker region; 11.5/30 in vertebrates and flies respectively for the C-terminal part. On closer examination, taking the Hox proteins one by one shows that the number of motifs present depends greatly on the size of each of the four region defined, with small regions harboring a limited amount of SLiMs when compared to larger ones. Fig. 5B displays a normalized representation taking into account the size of the region considered for the *Drosophila* and mouse Hox proteins, indicating a rather wide coverage of Hox protein sequences by SLiMs. Notably however, and maybe unexpectedly

with regard to the fact that the HD is highly structured is that the HD displays a high SLiM content. As SLiMs in structured domains are located at the surface of such globular domains, this indicates that besides providing key contacts toward the DNA, residues at the external surface of the globular HD are likely to contribute as protein interaction interfaces. Similar results were obtained for the chick Hox proteins (not shown).

Refining the study of SLiM positions taking into account the motif type reveals in most instances a preferential localization within Hox sequences (Fig. 6). For example, CLV SLiMs are preferentially located in the HD (Fig. 6A); DEG SLiMs are never found in the linker region and rarely in the C-terminal portion, (Fig. 6B); TRG SLiMs also seem to be located mainly in the HD or in regions close to it (linker and directly C-terminal to it; Fig. 6G). Such preferential localization is however not a rule, as illustrated by MOD SLiMs (Fig. 6E) and more particularly by phosphorylation sites (Fig. 6F).

Illuminating Hox protein sequences through the lens of SLiMs thus suggests that a large proportion of Hox proteins

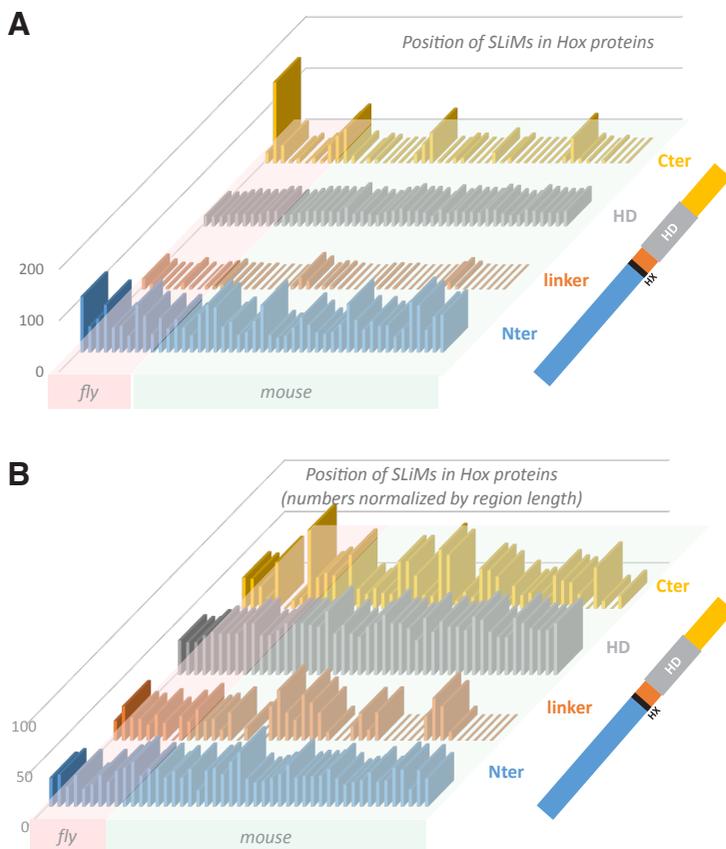


Fig. 5. Short linear motif (SLiM) positions within Hox proteins. (A) Hox proteins have been split in four regions: N-terminal region (blue), linker region (orange), HD (grey) and C-terminal region (yellow) except for the some of the posterior group proteins for which a linker region could not be defined owing to the lack of HX. The number of SLiM by region is showed for the fly (red highlight) and mouse (green highlight) proteins by region. Hox proteins are ordered from anterior to posterior and from cluster A to D for mouse sequences. (B) The number of SLiMs in each region was normalized according to the number of amino acid in each of these regions.

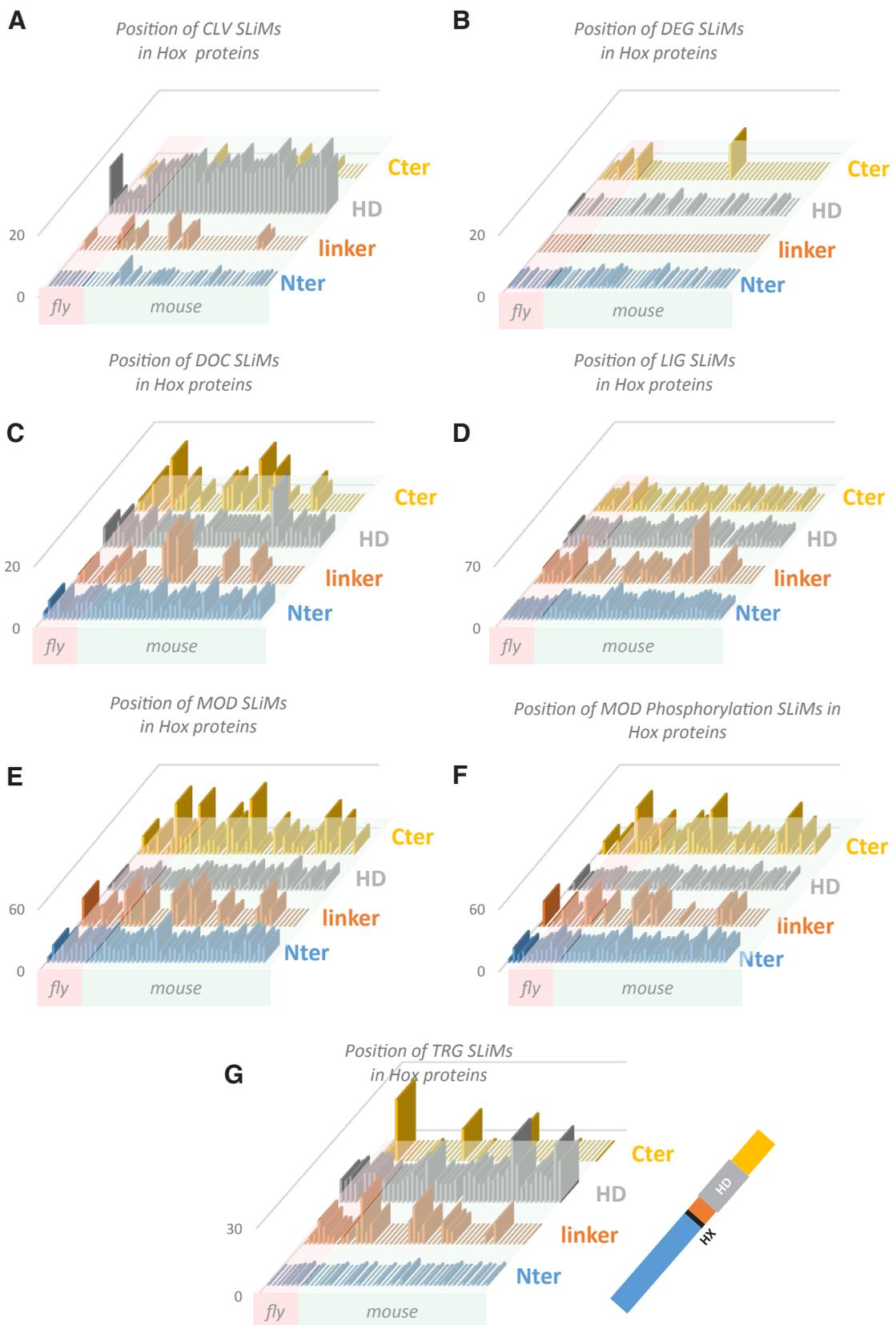


Fig. 6. Short linear motif (SLiM) positions within Hox proteins by motif types. (A-E and G) Number of SLiM by motif type normalized by region length in the fly and mouse Hox proteins, as described in Figure 5. **(F)** Number of MOD phosphorylation SLiMs normalized by region length in the fly and mouse Hox proteins.

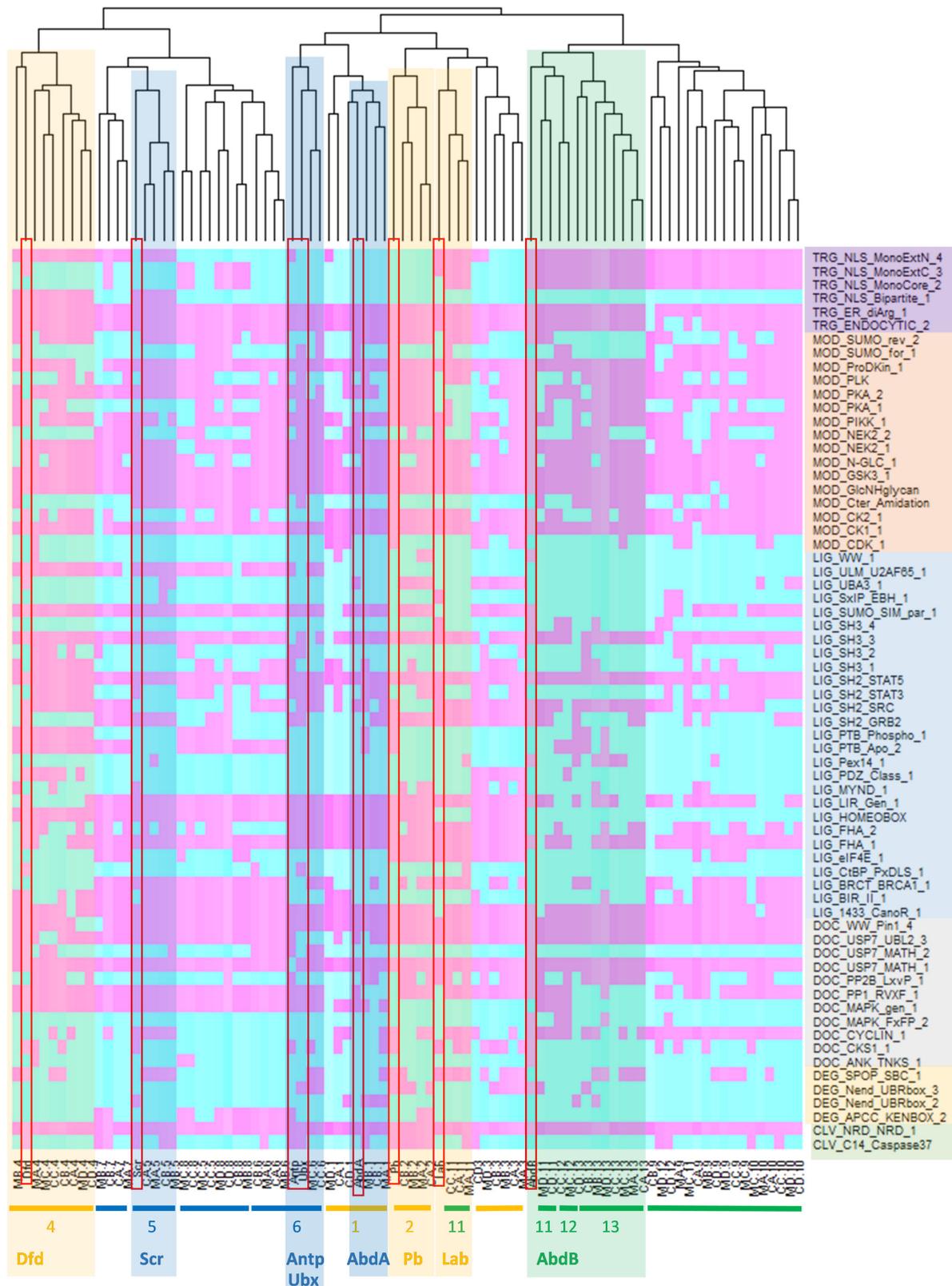


Fig. 7. Heatmap and clustering of Hox proteins by presence or absence of short linear motif (SLiM) classes. Clustering was made using R software with a simple presence/absence matrix. Only SLiM classes present in all three species were taken into account for this analysis (51%, 66 classes, see Figure 3C and Supplementary Figure 1). Paralog groups (PG) are indicated under the clustering with the anterior group highlighted in orange, the central group in blue and posterior group in green. The Drosophila proteins are in red boxes. Proteins names are written as follow for the vertebrates: MA.1, mouse Hoxa1; CA.1, Chicken Hoxa1 and so on. Blue, absence; pink, presence.

sequences still need to be investigated to achieve a better grasp on Hox protein activity and its regulation.

SLiM conservation and paralog specific features

The structural and organizational conservation of Hox genes has allowed grouping them in paralog groups (PG). Members of a same paralog group have a similar relative position in the Hox cluster, a similar expression pattern, and display a high degree of sequence identity (Duboule, 2007). Accordingly, paralogs from different species, even over long evolutionary distances, display stronger sequence identity than Hox proteins within the same species.

Here we explored if SLiMs could contribute in defining paralog specific Hox features. While analyzing SLiMs found in the mouse, chick and *Drosophila* Hox complements, we found that 51% (66 out of 129 total) of SLiM classes present in Hox proteins are found in the three species (Fig. 3C). We decided to explore if the qualitative use of SLiMs (the presence of a given SLiM class, without taking into account how often it is used and where it locates within the protein) is sufficient to reconstruct the Hox paralogy groups, which would then argue that SLiMs may contribute in defining paralog specific features. This was achieved by focusing on the evolutionarily conserved SLiM classes (66), which were taken as the primary dataset for constructing a presence/absence matrix for each *Drosophila*, mouse and chick Hox protein (Fig. 7). This representation, where the clustering of proteins results from similarities in SLiM class usage as defined above, allows to directly visualize if clustering follows Hox paralogy groups. This can easily be evidenced by examining if the *Drosophila* Hox proteins are clustered with their vertebrate paralogs. In general, there seem to be a good fit, with Pbx clustering with PG2, Deformed (Dfd) with PG4, Scr with PG5, Ubx and Antp with PG6 and AbdB with PG11 and 12. However, in some instances the clustering does not fit the expectations: Labial (Lab) clusters with PG11 and AbdA with PG1. Overall, this suggests that SLiMs may in general contribute to define paralog specific features.

Discussion

SLiMs: a perspective for the study of Hox protein activity/specificity

Hox proteins have very specific functions in specifying the anteroposterior axis of bilaterian embryos (Rezsohazy et al., 2015), a property that contrasts with the fact that they contain a similar DNA recognition domain (Bobola and Merabet, 2017; Mann and Chan, 1996). The functional dissection of Hox proteins has been a major avenue to decipher the mechanisms underlying Hox protein specificity. Such approaches have identified protein regions and sometimes specific residues for distinguishing the activity of Hox proteins, often within the HD and immediately adjacent regions (Chan and Mann, 1993; Foos et al., 2015; Lelli et al., 2011; Merabet et al., 2011; Ortiz-Lombardia et al., 2017; Slattey et al., 2011). An important limitation in these approaches has been the capacity to identify protein domains of putative functional importance, which has often relied on evolutionary conservation. This has allowed the identification and subsequent functional characterization of a few protein domains beside the HD and its direct surrounding regions (Merabet et al., 2009; Merabet et al., 2010).

The present survey highlights a large coverage of mouse, chick

and *Drosophila* Hox protein by SLiMs, which in principle should open novel perspectives in studying the mechanisms underlying Hox protein specificity and diversity. From a qualitative point of view (relative contribution of different SLiM types, irrespective of how often they are used) the proportions of different SLiM types found in Hox proteins are very close to the proportions found in the whole ELM database, indicating that Hox proteins apparently do not display a strong bias toward one of the six particular SLiM motif types. From a more quantitative perspective (SLiM instances, ie number of SLiMs from each types are found in Hox proteins), LIG and DEG SLiMs are underrepresented, while the MOD SLiMs displays a noticeable enrichment, suggesting that post translational modification is likely a key aspect of the biology of Hox proteins. The extent to which Hox post translational modification applies to Hox protein function is extensively discussed in an accompanying review (Draime et al., this issue). We also found that when taken irrespective of the type of SLiM, SLiMs do not concentrate in specific regions of Hox proteins. Thus if SLiMs are functional modules, viewing Hox proteins as a HD-anchored functional unit (the HD and its surrounding sequences, (Merabet et al., 2009; Merabet et al., 2011; Merabet et al., 2010)) with the rest of the protein providing minor functional contributions is likely inadequate. However, when the analysis is refined taking into account the motif type, some preferential localization is observed, suggesting a domain specialization of Hox proteins.

Finally, using the SLiM content of Hox proteins to cluster mouse, chick and *Drosophila* Hox proteins allowed a grouping that in most cases (though not all) follows paralogy groups, suggesting that SLiMs may in general contribute to define paralog specificity. Specific functions at the transcriptional level may be achieved by Hox proteins interacting with different partners according to their paralogy groups, thus activating or repressing different sets of genes depending on their interactor and their own DNA binding sequence. Interestingly, some of the SLiMs listed in Fig. 1 and Supplementary Fig. 1 are present only in a few Hox proteins, sometimes even in only one, supporting a SLiM-mediated functional specialization of Hox protein.

SLiMs in Hox proteins: from identification to function

While the description of IDRs and SLiMs are not novel, they have only marginally influenced experimental approaches aimed at understanding Hox protein activity. It was previously noticed that the *Drosophila* Hox protein Ubx is mainly composed of IDRs, influencing the molecular properties of the protein. However the lack of *in vivo* data supporting a biological importance of IDRs likely contributed to minimize the perception of the importance of IDRs in the control of Hox protein activity. The recent highlight that the HX motif is a SLiM provides ample retrospective evidence that SLiMs do indeed contribute to the control of Hox protein activity (Baeza et al., 2015; Dard et al., 2018). Further supporting SLiMs as intrinsic contributors of Hox protein activity, three SLiMs recently identified in the vertebrate Hox proteins HoxB3, HoxA7 and HoxC8 were shown to be required for interaction with the Pbx proteins, including *in vivo* (Dard et al., 2018). Together with the previous finding that a SLiM domain just C-terminal to the HD in Ubx and AbdA also mediates interaction with the PBC class *Drosophila* Exd protein, this indicates that SLiMs likely allow for multiple modes of interaction with PBC class proteins, and as such constitute integral components of Hox protein specificity.

A few extensive mutagenic approaches of Hox proteins allow addressing the issue of SLiM functional importance from a wider perspective. This is the case for the *Drosophila* protein Ubx, for which small deletions were generated upstream of the HD (Tour *et al.*, 2005). This region contains a high number of SLiMs, yet functional analysis using a couple of genes activated (*teashirt* and *decapentaplegic*) or repressed (*Distalless* and *Antp*) by Ubx only allowed the identification of a single functional domain, a highly conserved short sequence that act as a transcriptional activation domain (Tour *et al.*, 2005). Similarly, an extensive mutational approach of the HoxA1 protein, using a pentapeptide insertional strategy, only allowed to map function to 5 of the 18 insertions generated (Lambert *et al.*, 2010). Function was assessed by transcriptional assays driven by enhancers of EphA2 and HoxB1, two known transcriptional target of HoxA1. Three insertions with functional impact mapped in the HD itself, while the two others were at the N- and C-terminal extremities of the protein. The Ubx and HoxA1 studies, with only few functional domains identified, contrasts with the wide coverage of SLiMs found in Hox proteins in general, as well as in Ubx and HoxA1 in particular. A plausible explanation for this likely lies in that the function of SLiMs not being pleiotropic but rather highly context specific, serving only a subset of Hox protein activities. As such, their function would only be identified if the proper activities are examined, a requisite not easy to satisfy since the full spectrum of Hox activity is not known. The weak pleiotropic nature of SLiM mediated interaction/function will certainly be a difficulty to overcome in order to grasp the overall functional contribution of SLiMs to Hox protein function. The function of SLiMs with a high degree of pleiotropy will likely emerge first, while SLiMs with a low degree of pleiotropy will likely stay longer in the shade.

SLiM control of Hox protein activity/specificity: molecular modalities

A striking feature of the few identified functional SLiMs is that their function all seem to relate to interaction with PBC class proteins. It thus already appears that SLiMs play a key role in Hox-PBC interaction, either directly or indirectly. Two SLiMs, the HX found in most Hox proteins, as well as the SLiM found immediately C-terminal to the HD in *Drosophila* Ubx and AbdA, termed UbdA, were shown to directly contact PBC class proteins. Structural characterization showed that the HX-mediated contact allows for a precise positioning of the Hox HD N-terminal arm within the DNA minor groove (Piper *et al.*, 1999), while the interaction mediated by the UbdA motif more likely fine tunes the positioning of the Hox HD recognition helix (helix 3) within the DNA major groove (Foos *et al.*, 2015). This suggests that protein contacts between Hox and PBC class proteins, depending on the SLiM that mediates the contact, may differently impact on Hox DNA binding properties. This is supported by functional studies that highlight the non-pleiotropic contribution of Hox protein domains, including the HX and UbdA SLiMs: distinct contributions of the HX and UbdA SLiMs (Merabet *et al.*, 2011) may allow the protein to harbor distinct specificities, depending on tissue/cell types or according to specific subsets of target genes, which ultimately may drive functional diversity. Functional non-pleiotropy, which also applies to other motifs, is correlated with differential use of these motifs for protein interactions, including but not restricted to PBC class proteins (Baeza *et al.*, 2015).

The potential of SLiMs for the functional diversification of Hox proteins is well illustrated by tissue specific protein interactions: the impact of the HX or UbdA motif mutations on the Hox interaction potential with a large set of transcription factors is distinct in the mesoderm, epidermis and nervous system (Baeza *et al.*, 2015). In addition, some Hox proteins display tissue-specific alternative splicing, which given the wide coverage of Hox sequences by SLiMs likely modifies the SLiM content of the protein (Niklas *et al.*, 2015). For example, the Ubx protein has 6 isoforms arising from alternate splicing and displaying tissue specific expression patterns. (Geyer *et al.*, 2015): Ubx Ia and Ubx Ib are expressed mainly in the epidermis, mesoderm and peripheral nervous system during embryonic development, Ubx IIa and Ubx IIb can be found mainly in the central nervous system but also in epidermis and Ubx IVa is expressed only in the central nervous system during embryonic development (Lopez and Hogness, 1991). Lastly, it seems that the “b” isoforms are expressed at a very low rate in contrast to “a”. These isoforms vary in their linker regions length, between 5 and 48 residues long. This is due to the presence or absence of 3 alternatively spliced micro exons (b, M1 and MII). We noted that the “full” Ubx linker region contains 14 SLiMs, indicating that the alternative splicing generating the Ubx isoforms concomitantly alters the SLiM content of the protein. Supporting a function for linker region SLiM content, it was previously reported that modifying the linker region alters the DNA binding properties of the Ubx-Exd and AbdA-Exd complexes (Saadaoui *et al.*, 2011). The presence, absence or competition of those alternatively spliced SLiMs could thus well account for the differences in downstream effectors as well as different developmental programs they activate or repress.

Available data establish a role for SLiMs beyond Hox-PBC interactions. The HX and UbdA SLiMs were for example shown to be required for AbdA and Ubx activities that do not require the *Drosophila* Exd protein (Merabet *et al.*, 2011). In addition the HX was also shown to impact on the *in vivo* interaction properties of the *Drosophila* Scr, Antp, Ubx, AbdA and AbdB proteins with a large subset of transcription factors, in a tissue specific manner (Baeza *et al.*, 2015). While the molecular modalities of these interactions are still unknown, and their functional relevance still needs to be established, these data argue for SLiMs acting as key Hox interfaces toward a wide spectrum of proteins which will allow to frame the molecular action of Hox proteins in a much wider perspective.

Materials and Methods

Eukaryotic linear motif resource

The main resource used for this analysis is the eukaryotic linear motif (ELM) database (Puntervoll, 2003) (elm.eu.org). ELM is a manually curated SLiM database which have been published and confirmed in at least two different protein families that analyses a query protein sequence for already known motif sequences. ELM was first released in 2003 (Puntervoll, 2003) and has been growing ever since (Dinkel *et al.*, 2014; Dinkel *et al.*, 2016; Gould *et al.*, 2010; Gouw *et al.*, 2018). Our analysis was performed in 2016, while the database contained 246 SLiM classes instead of 268 at present. The sequences of the SLiMs are normalized as Regular Expression (RegExp; (Davey *et al.*, 2012)) for the whole database. The online suite scans the input sequence against the whole ELM database RegExp and outputs the results.

Different filters can be applied when submitting a sequence, such as subcellular localization, species or structural filter. Indeed, as previously discussed, SLiMs are found predominantly in the disordered regions, hence filtering globular domains allows lowering false positive hit. This filter also

takes into account the disordered regions of globular domains (disordered loops, etc) and the position in the proteins (the motif must be physically available to the partner protein) (Davey *et al.*, 2012). The subcellular localization and species filters make sure that the partner protein is present in that localization or species, respectively. Lastly, a conservation score assesses the conservation of the motifs in related proteins. All the filtered results are available but displayed in a different table on the same page. In our analyses, we did not input the subcellular localization to have a clear representation of the SLiM present in each protein.

SLiMPred

SLiMPred analyses a protein sequence for *de novo* SLiM discovery (Davey *et al.*, 2010; Edwards *et al.*, 2007; Mooney *et al.*, 2012). It requires a UniProt ID to retrieve the desired sequence and its orthologs. It scans each residue of the input sequence and compares it with orthologs, taking into account the evolution of the protein and the ordered and disordered regions. It does so by aligning the ortholog proteins and giving back 4 components: the IUPred disorder propensity, the ordered regions, the relative local conservation (RLC) and SLiMPred score. These four components are complementary, as SLiMs occur mainly in disordered regions. IUPred score identify regions that have a high disorder-to-order transition upon binding. In addition, the degree of conservation of these functional motifs is supposedly higher than the rest of the disordered regions, resulting in a higher RLC in SLiMs compared to the rest of the disordered regions. However, SLiMs may also be found in disordered regions of globular protein structures and SLiMPred is able to detect these and output them in the results. The SLiMPred score is comprised between 0 and 1 with 1 representing the most reliable prediction. It is up to the user to analyze the results and identify interesting SLiMs for testing.

Hox sequences

This study uses the full Hox complement from mouse, chick and fruit fly for the ELM database and 14 of them for the analysis with SLiMPred (all 8 fly proteins and one representative of each anterior, central and posterior Hox proteins for the mouse and chick). NCBI accession numbers of Hox sequences used for ELM and Uniprot numbers used for SLiMPred are:

Drosophila melanogaster: Labial NP_476613.1 and LAB_DROME, Proboscipedia CAA45272.1 and HMPB_DROME, Deformed NP_477201.1 and DFD_DROME, Sex comb reduced NP_524248.2 and SCR_DROME, Antennapedia NP_996167.1 and ANTP_DROME, Ultrabithorax NP_536752.1 and UBX_DROME, Abdominal A NP_476693.1 and E1JIM9_DROME and Abdominal B NP_524896.2 and ABDB_DROME.

Mouse: Hoxa1 NP_034579.3, Hoxa2 NP_034581.1, Hoxa3 NP_034582.1, Hoxa4 NP_032291.1, Hoxa5 NP_034583.1, Hoxa6 NP_034584.1, Hoxa7 NP_034585.1, Hoxa9 NP_001264167.1, Hoxa10 NP_032289.2, Hoxa11 NP_034580.1, Hoxa13 NP_032290.1, Hoxb1 NP_032292.3, Hoxb2 NP_598793.2, Hoxb3 NP_034588.2, Hoxb4 NP_034589.3 and HXB4_MOUSE, Hoxb5 NP_032294.2, Hoxb6 NP_032295.1, Hoxb7 NP_034590.2, Hoxb8 NP_034591.1 and HXB8_MOUSE, Hoxb9 NP_032296.2 and HXB9_MOUSE, Hoxb13 NP_0332293.1, Hoxc4 NP_038581.2, Hoxc5 NP_783857.1, Hoxc6 NP_034595.2, Hoxc8 NP_034596.1, Hoxc9 NP_032298.1, Hoxc10 NP_034592.2, Hoxc11 NP_001020013.1, Hoxc12 NP_0345933.1, Hoxc13 NP_034594.1, Hoxd1 NP_034597.2, Hoxd3 NP_034598.2, Hoxd4 NP_034599.2, Hoxd8 NP_032302.2, Hoxd9 NP_038583.1, Hoxd10 NP_038582.2, Hoxd11 NP_032299.1, Hoxd12 NP_032300.2 and Hoxd13 NP_032301.2.

Chick: Hoxa1 XP_003640745.1, Hoxa2 NP_990481.1, Hoxa3 NP_001336689.1, Hoxa4 NP_001025517.1, Hoxa5 NP_001305348.1, Hoxa6 NP_001026158.1, Hoxa7 NP_989926.1, Hoxa9 NP_001292153.1, Hoxa10 NP_001292574.1, Hoxa11 NP_989950.1, Hoxa13 NP_989470.2, Hoxb1 NP_001074328.2, Hoxb2 XP_003642840.1, Hoxb3 NP_990074.2, Hoxb4 NP_990624.1 and HXB4_CHICK, Hoxb5 NP_001020526.1, Hoxb6 XP_015154927.1, Hoxb7 XP_015154928.1, Hoxb8 NP_990242.1 and Q9YH27_CHICK, Hoxb9 XP_015154931.1 and E1BQM6_CHICK, Hoxb13 XP_001234927.1, Hoxc4 XP_015155814.1, Hoxc5 XP_015155813.1, Hoxc6

XP_015155838.1, Hoxc8 NP_990224.1, Hoxc9 NP_001264211.1, Hoxc10 NP_001292189.1, Hoxc11 XP_428250.2, Hoxc12 XP_426957.3, Hoxc13 XP_001235166.3, Hoxd1 XP_00492781.1, Hoxd3 XP_428765.2, Hoxd4 NP_001012293.1, Hoxd8 XP_015144874.1, Hoxd9 XP_001234507.2, Hoxd10 XP_015145079.1, Hoxd11 NP_989951.1, Hoxd12 NP_990580.1 and Hoxd13 NP_990765.1.

Acknowledgments

The AS and YG team is supported by the CNRS and AMU, and grants from FRM, ARC and AMIDEX.

References

- ARAI, M. (2018) Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophys Rev* 10: 163-181.
- BADIS, G., BERGER, M.F., PHILIPPAKIS, A.A., TALUKDER, S., GEHRKE, A.R., JAEGER, S.A., CHAN, E.T., METZLER, G., VEDENKO, A., CHEN, X., KUZNETSOV, H., WANG, C.F., COBURN, D., NEWBURGER, D.E., MORRIS, Q., HUGHES, T.R., BULYK, M.L. (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324: 1720-1723.
- BAEVA, M., VIALA, S., HEIM, M., DARD, A., HUDRY, B., DUFFRAISSE, M., ROGULJA-ORTMANN, A., BRUN, C., MERABET, S. (2015) Inhibitory activities of short linear motifs underlie Hox interactome specificity in vivo. *Elife* 4: e06034
- BOBOLA, N., MERABET, S. (2017) Homeodomain proteins in action: similar DNA binding preferences, highly variable connectivity. *Curr Opin Genet Dev* 43: 1-8.
- BONDOS, S.E., SWINT-KRUSE, L., MATTHEWS, K.S. (2015) Flexibility and Disorder in Gene Regulation: LacI/GalR and Hox Proteins. *J Biol Chem* 290: 24669-24677.
- BURGLIN, T.R., AFFOLTER, M. (2016) Homeodomain proteins: an update. *Chromosoma* 125: 497-521.
- CHAN, S.K., MANN, R.S. (1993) The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. *Genes Dev* 7: 796-811.
- DARD, A., REBOULET, J., JIA, Y., BLEICHER, F., DUFFRAISSE, M., VANAKER, J.M., FORCET, C., MERABET, S. (2018) Human HOX Proteins Use Diverse and Context-Dependent Motifs to Interact with TALE Class Cofactors. *Cell Rep* 22: 3058-3071.
- DAVEY, N.E., HASLAM, N.J., SHIELDS, D.C., EDWARDS, R.J. (2010) SLiMfinder: a web server to find novel, significantly over-represented, short protein motifs. *Nucleic Acids Res* 38: W534-539.
- DAVEY, N.E., VAN ROEY, K., WEATHERITT, R.J., TOEDT, G., UYAR, B., ALTENBERG, B., BUDD, A., DIELLA, F., DINKEL, H., GIBSON, T.J. (2012) Attributes of short linear motifs. *Mol Biosyst* 8: 268-281.
- DINKEL, H., MICHAEL, S., WEATHERITT, R.J., DAVEY, N.E., VAN ROEY, K., ALTENBERG, B., TOEDT, G., UYAR, B., SEILER, M., BUDD, A., JODICKE, L., DAMMERT, M.A., SCHROETER, C., HAMMER, M., SCHMIDT, T., JEHL, P., MCGUIGAN, C., DYMECKA, M., CHICA, C., LUCK, K., VIA, A., CHATRYAMONTRI, A., HASLAM, N., GREBNEV, G., EDWARDS, R.J., STEINMETZ, M.O., MEISELBACH, H., DIELLA, F., GIBSON, T.J. (2012) ELM--the database of eukaryotic linear motifs. *Nucleic Acids Res* 40: D242-251.
- DINKEL, H., VAN ROEY, K., MICHAEL, S., DAVEY, N.E., WEATHERITT, R.J., BORN, D., SPECK, T., KRUGER, D., GREBNEV, G., KUBAN, M., STRUMILLO, M., UYAR, B., BUDD, A., ALTENBERG, B., SEILER, M., CEMES, L.B., GLAVINA, J., SANCHEZ, I.E., DIELLA, F., GIBSON, T.J. (2014) The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res* 42: D259-266.
- DINKEL, H., VAN ROEY, K., MICHAEL, S., KUMAR, M., UYAR, B., ALTENBERG, B., MILCHEVSKAYA, V., SCHNEIDER, M., KUHN, H., BEHRENDT, A., DAHL, S.L., DAMERELL, V., DIEBEL, S., KALMAN, S., KLEIN, S., KNUDSEN, A.C., MADER, C., MERRILL, S., STAUDT, A., THIEL, V., WELTI, L., DAVEY, N.E., DIELLA, F., GIBSON, T.J. (2016) ELM 2016--data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res* 44: D294-300.
- DUBOULE, D. (2007) The rise and fall of Hox gene clusters. *Development* 134: 2549-2560.
- DUNKER, A.K., LAWSON, J.D., BROWN, C.J., WILLIAMS, R.M., ROMERO, P., OH, J.S., OLDFIELD, C.J., CAMPEN, A.M., RATLIFF, C.M., HIPPS, K.W., AUSIO, J., NISSEN, M.S., REEVES, R., KANG, C., KISSINGER, C.R., BAILEY, R.W., GRISWORLD, M.D., CHIU, W., GARNER, E.C., OBRADOVIC, Z. (2001) Intrinsically

- disordered protein. *J. Molec. Graph. Modelling* 119: 26-59.
- EDWARDS, R.J., DAVEY, N.E., SHIELDS, D.C. (2007) SLiMfinder: a probabilistic method for identifying over-represented, convergently evolved, short linear motifs in proteins. *PLoS One* 2: e967.
- FOOS, N., MAUREL-ZAFFRAN, C., MATE, M.J., VINCENTELLI, R., HAINAUT, M., BERENGER, H., PRADEL, J., SAURIN, A.J., ORTIZ-LOMBARDIA, M., GRABA, Y. (2015) A flexible extension of the *Drosophila* ultrabithorax homeodomain defines a novel Hox/PBC interaction mode. *Structure* 23: 270-279.
- GEHRING, W.J., QIAN, Y.Q., BILLETER, M., FURUKUBO-TOKUNAGA, K., SCHIER, A.F., RESENDEZ-PEREZ, D., AFFOLTER, M., OTTING, G., WUTHRICH, K. (1994) Homeodomain-DNA recognition. *Cell* 78: 211-223.
- GEYER, A., KOLTSAKI, I., HESSINGER, C., RENNER, S., ROGULJA-ORTMANN, A. (2015) Impact of Ultrabithorax alternative splicing on *Drosophila* embryonic nervous system development. *Mech Dev* 138 Pt 2: 177-189.
- GOULD, C.M., DIELLA, F., VIA, A., PUNTERVOLL, P., GEMUND, C., CHABANIS-DAVIDSON, S., MICHAEL, S., SAYADI, A., BRYNE, J.C., CHICA, C., SEILER, M., DAVEY, N.E., HASLAM, N., WEATHERITT, R.J., BUDD, A., HUGHES, T., PAS, J., RYCHLEWSKI, L., TRAVE, G., AASLAND, R., HELMER-CITTERICH, M., LINDING, R., GIBSON, T.J. (2010) ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res* 38: D167-180.
- GOUW, M., MICHAEL, S., SAMANO-SANCHEZ, H., KUMAR, M., ZEKE, A., LANG, B., BELY, B., CHEMES, L.B., DAVEY, N.E., DENG, Z., DIELLA, F., GURTH, C.M., HUBER, A.K., KLEINSORG, S., SCHLEGEL, L.S., PALOPOLI, N., ROEY, K.V., ALTENBERG, B., REMENYI, A., DINKEL, H., GIBSON, T.J. (2018) The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Res* 46: D428-D434.
- HOWELL, D.W., DURAN, C.L., TSAI, S.P., BONDOS, S.E., BAYLESS, K.J. (2016) Functionalization of Ultrabithorax Materials with Vascular Endothelial Growth Factor Enhances Angiogenic Activity. *Biomacromolecules* 17: 3558-3569.
- HSIAO, H.C., GONZALEZ, K.L., CATANESE, D.J., JR., JORDY, K.E., MATTHEWS, K.S., BONDOS, S.E. (2014) The intrinsically disordered regions of the *Drosophila melanogaster* Hox protein ultrabithorax select interacting proteins based on partner topology. *PLoS One* 9: e108217.
- HUDRY, B., THOMAS-CHOLLIER, M., VOLOVIK, Y., DUFFRAISSE, M., DARD, A., FRANK, D., TECHNAU, U., MERABET, S. (2014) Molecular insights into the origin of the Hox-TALE patterning system. *Elife* 3: e01939.
- JOSHI, R., PASSNER, J.M., ROHS, R., JAIN, R., SOSINSKY, A., CRICKMORE, M.A., JACOB, V., AGGARWAL, A.K., HONIG, B., MANN, R.S. (2007) Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131: 530-543.
- LAMBERT, B., VANDEPUTTE, J., DESMET, P.M., HALLET, B., REMACLE, S., REZSOHAZY, R. (2010) Pentapeptide insertion mutagenesis of the Hoxa1 protein: mapping of transcription activation and DNA-binding regulatory domains. *J Cell Biochem* 110: 484-496.
- LAMBERT, S.A., JOLMA, A., CAMPITELLI, L.F., DAS, P.K., YIN, Y., ALBU, M., CHEN, X., TAIPALE, J., HUGHES, T.R., WEIRAUCH, M.T. (2018) The Human Transcription Factors. *Cell* 172: 650-665.
- LARONDE-LEBLANC, N.A., WOLBERGER, C. (2003) Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev* 17: 2060-2072.
- LELLI, K.M., NORO, B., MANN, R.S. (2011) Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity. *Proc Natl Acad Sci USA* 108: 21122-21127.
- LIU, Y., MATTHEWS, K.S., BONDOS, S.E. (2008) Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* hox protein ultrabithorax. *J Biol Chem* 283: 20874-20887.
- LIU, Y., MATTHEWS, K.S., BONDOS, S.E. (2009) Internal regulatory interactions determine DNA binding specificity by a Hox transcription factor. *J Mol Biol* 390: 760-774.
- LONGOBARDI, E., PENKOV, D., MATEOS, D., DE FLORIAN, G., TORRES, M., BLASI, F. (2014) Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. *Dev Dyn* 243: 59-75.
- LOPEZ, A.J., HOGNESS, D.S. (1991) Immunochemical dissection of the Ultrabithorax homeoprotein family in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 88: 9924-9928.
- MANN, R.S., CHAN, S.K. (1996) Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet* 12: 258-262.
- MANN, R.S., LELLI, K.M., JOSHI, R. (2009) Hox specificity; unique roles for cofactors and collaborators. *Curr Top Dev Biol* 88: 63-101.
- MCGINNIS, W., KRUMLAUF, R. (1992) Homeobox genes and axial patterning. *Cell* 68: 283-302.
- MERABET, S., GALLIOT, B. (2015) The TALE face of Hox proteins in animal evolution. *Front Genet* 6: 267.
- MERABET, S., HUDRY, B., SAADAOU, M., GRABA, Y. (2009) Classification of sequence signatures: a guide to Hox protein function. *Bioessays* 31: 500-511.
- MERABET, S., LITIM-MECHERI, I., KARLSSON, D., DIXIT, R., SAADAOU, M., MONIER, B., BRUN, C., THOR, S., VIJAYRAGHAVAN, K., PERRIN, L., PRADEL, J., GRABA, Y. (2011) Insights into Hox protein function from a large scale combinatorial analysis of protein domains. *PLoS Genet* 7: e1002302.
- MERABET, S., LOHMANN, I. (2015) Toward a new twist in Hox and TALE DNA-binding specificity. *Dev Cell* 32: 259-261.
- MERABET, S., MANN, R.S. (2016) To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins. *Trends Genet* 32: 334-347.
- MERABET, S., SAMBRANI, N., PRADEL, J., GRABA, Y. (2010) Regulation of Hox activity: insights from protein motifs. *Adv Exp Med Biol* 689: 3-16.
- MOONEY, C., POLLASTRI, G., SHIELDS, D.C., HASLAM, N.J. (2012) Prediction of short linear protein binding regions. *J Mol Biol* 415: 193-204.
- NIKLAS, K.J., BONDOS, S.E., DUNKER, A.K., NEWMAN, S.A. (2015) Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Front Cell Dev Biol* 3: 8.
- ORTIZ-LOMBARDIA, M., FOOS, N., MAUREL-ZAFFRAN, C., SAURIN, A.J., GRABA, Y. (2017) Hox functional diversity: Novel insights from flexible motif folding and plastic protein interaction. *Bioessays* 39, 4, 1600246.
- PASSNER, J.M., RYOO, H.D., SHEN, L., MANN, R.S., AGGARWAL, A.K. (1999) Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* 397: 714-719.
- PATTERSON, J.L., ARENAS-GAMBOA, A.M., WANG, T.Y., HSIAO, H.C., HOWELL, D.W., PELLOIS, J.P., RICE-FICHT, A., BONDOS, S.E. (2015) Materials composed of the *Drosophila* Hox protein Ultrabithorax are biocompatible and nonimmunogenic. *J Biomed Mater Res A* 103: 1546-1553.
- PEARSON, J.C., LEMONS, D., MCGINNIS, W. (2005) Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* 6: 893-904.
- PIPER, D.E., BATCHELOR, A.H., CHANG, C.P., CLEARY, M.L., WOLBERGER, C. (1999) Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* 96: 587-597.
- PUNTERVOLL, P. (2003) ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* 31: 3625-3630.
- QIAN, Y.Q., BILLETER, M., OTTING, G., MULLER, M., GEHRING, W.J., WUTHRICH, K. (1989) The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell* 59: 573-580.
- REZSOHAZY, R., SAURIN, A.J., MAUREL-ZAFFRAN, C., GRABA, Y. (2015) Cellular and molecular insights into Hox protein action. *Development* 142: 1212-1227.
- ROMERO, I.G., RUVINSKY, I., GILAD, Y. (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* 13: 505-516.
- SAADAOU, M., MERABET, S., LITIM-MECHERI, I., ARBEILLE, E., SAMBRANI, N., DAMEN, W., BRENA, C., PRADEL, J., GRABA, Y. (2011) Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity. *Proc Natl Acad Sci USA* 108: 2276-2281.
- SCOTT, M.P., WEINER, A.J. (1984) Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of *Drosophila*. *Proc Natl Acad Sci USA* 81: 4115-4119.
- SLATTERY, M., RILEY, T., LIU, P., ABE, N., GOMEZ-ALCALA, P., DROR, I., ZHOU, T., ROHS, R., HONIG, B., BUSSEMAKER, H.J., MANN, R.S. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147: 1270-1282.
- SLUPSKY, C.M., SYKES, D.B., GAY, G.L., SYKES, B.D. (2001) The HoxB1 hexapeptide is a prefolded domain: implications for the Pbx1/Hox interaction. *Protein Sci* 10: 1244-1253.
- SMITH, N.C., MATTHEWS, J.M. (2016) Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Curr Opin Struct Biol* 38: 68-74.
- TOKURIKI, N., TAWFIK, D.S. (2009) Protein dynamism and evolvability. *Science* 324: 203-207.

- TOMPA, P. (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* 37: 509-516.
- TOMPA, P., DAVEY, N.E., GIBSON, T.J., BABU, M.M. (2014) A million peptide motifs for the molecular biologist. *Mol Cell* 55: 161-169.
- TOMPA, P., SCHAD, E., TANTOS, A., KALMAR, L. (2015) Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struct Biol* 35: 49-59.
- TOUR, E., HITTINGER, C.T., MCGINNIS, W. (2005) Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development* 132: 5271-5281.
- VAN ROEY, K., UYAR, B., WEATHERITT, R.J., DINKEL, H., SEILER, M., BUDD, A., GIBSON, T.J., DAVEY, N.E. (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev* 114: 6733-6778.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

Gdf11/Smad signalling and Cdx proteins cooperate to activate the Hoxc8 early enhancer in HepG2 cells

Stephen J. Gaunt

Int. J. Dev. Biol. (2017) 61: 427-432
<https://doi.org/10.1387/ijdb.170066sg>

ADHFe1: a novel enzyme involved in retinoic acid-dependent Hox activation

Yehuda Shabtai, Natalie Shukrun and Abraham Fainsod

Int. J. Dev. Biol. (2017) 61: 303-310
<https://doi.org/10.1387/ijdb.160252af>

Identification of distal enhancers for Six2 expression in pronephros

Nanoka Suzuki, Kodai Hirano, Hajime Ogino and Haruki Ochi

Int. J. Dev. Biol. (2015) 59: 241-246
<https://doi.org/10.1387/ijdb.140263ho>

Synergistic action in P19 pluripotential cells of retinoic acid and Wnt3a on Cdx1 enhancer elements

Stephen J. Gaunt and Yu-Lee Paul

Int. J. Dev. Biol. (2014) 58: 307-314
<https://doi.org/10.1387/ijdb.140003sg>

The Parahox gene Pdx1 is required to maintain positional identity in the adult foregut

Andrew M. Holland, Sonia Garcia, Gaetano Naselli, Raymond J. MacDonald and Leonard C. Harrison

Int. J. Dev. Biol. (2013) 57: 391-398
<https://doi.org/10.1387/ijdb.120048ah>

Origins of Cdx1 regulatory elements suggest roles in vertebrate evolution

Stephen J. Gaunt and Yu-Lee Paul

Int. J. Dev. Biol. (2011) 55: 93-98
<https://doi.org/10.1387/ijdb.103252sg>

Identification of hoxb1b downstream genes: hoxb1b as a regulatory factor controlling transcriptional networks and cell movement during zebrafish gastrulation

Willem M.R. van den Akker, Antony J. Durston and Herman P. Spalink

Int. J. Dev. Biol. (2010) 54: 55-62
<https://doi.org/10.1387/ijdb.082678wv>

5 yr ISI Impact Factor (2016) = 2.421

