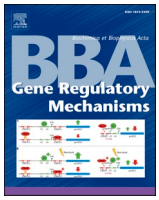
BBA - Gene Regulatory Mechanisms 1867 (2024) 194993

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm



A poly-histidine motif of HOXA1 is involved in regulatory interactions with cysteine-rich proteins

Damien Marchese, Florent Guislain, Tamara Pringels, Laure Bridoux, René Rezsohazy *

Louvain Institute of Biomolecular Science and Technology, UCLouvain, Place Croix du Sud 5 (L7.07.10), B-1348 Louvain-la-Neuve, Belgium

ARTICLE INFO

Keywords:
HOX transcription factors
Activity regulation
Protein-protein interaction
Histidine repeat
Histidine-cysteine interaction

ABSTRACT

Homopolymeric amino acid repeats are found in about 24 % of human proteins and are over-represented in transcriptions factors and kinases. Although relatively rare, homopolymeric histidine repeats (polyH) are more significantly found in proteins involved in the regulation of embryonic development. To gain a better understanding of the role of polyH in these proteins, we used a bioinformatic approach to search for shared features in the interactomes of polyH-containing proteins in human. Our analysis revealed that polyH protein interactomes are enriched in cysteine-rich proteins and in proteins containing (a) cysteine repeat(s). Focusing on HOXA1, a HOX transcription factor displaying one long polyH motif, we identified that the polyH motif is required for the HOXA1 interaction with such cysteine-rich proteins. We observed a correlation between the length of the polyH repeat and the strength of the HOXA1 interaction with one Cys-rich protein, MDFI. We also found that metal ion chelators disrupt the HOXA1-MDFI interaction supporting that such metal ions are required for the interaction. Furthermore, we identified three polyH interactors which down-regulate the transcriptional activity of HOXA1. Taken together, our data point towards the involvement of polyH and cysteines in regulatory interactions between proteins, notably transcription factors like HOXA1.

1. Introduction

Homopolymeric amino acid repeats are very common in eukaryotic proteins [1]. Such repeats are present in 24 % of human proteins compared to 1.9 % in *Escherichia coli* [2]. Their expansions are known to play a role in protein evolution but also to cause diseases [3]. The most abundant repeats in the human proteome are, in decreasing order, alanine (A), glycine (G), proline (P), serine (S) arginine (R), glutamic acid (E), glutamine (Q), and histidine (H) repeats [4]. These regions are known to be intrinsically disordered, they are therefore underrepresented in protein structure databases. As a corollary, repeats provide flexibility between individual folded domains, supporting proteins to associate in large, multimolecular complexes [5,6]. These disordered regions also define short linear motifs which by themselves are accessible for multiple weak-affinity protein binding [7].

The implication of amino acid repeats in protein-protein interactions has been reported for poly-glutamine (polyQ[1]) and poly-alanine (polyA) repeats mediating homophilic interactions, i.e. interactions in-between polyQ and polyA containing proteins, respectively. These repeats can form coiled coil super-secondary structures which can regulate the oligomerization and the functions of proteins [8,9]. In contrast, polyproline (polyP) motifs as well as proline-rich sequences have been identified to mediate protein-protein interactions by binding to non-repetitive domains [10].

Homopolymeric repeat-containing proteins are overrepresented among transcription factors and kinases [5,11]. It has been suggested that such repeats are abundant in transcription factors and signal transduction proteins because their functions rely on their ability to establish multiple, diverse and versatile context-dependent interactions, for which weak affinity protein binding capacities is a necessary property [7,12]. Homopolymeric repeat-containing proteins are also over-represented among proteins involved in the regulation of embryonic

* Corresponding author.
E-mail address: rene.rezsohazy@uclouvain.be (R. Rezsohazy).

[1] Abbreviations: GST: glutathione s-transferase; hHOXA1: human HOXA1; His1: first polyH of HOXA1; His2: second polyH of HOXA1; hMDFI: human MDFI; $HOXA1^{\Delta HD}$: HOXA1 without the homeodomain; $HOXA1^{\Delta His1}$: HOXA1 without the first polyH; $HOXA1^{\Delta His2}$: HOXA1 without the second polyH; $HOXA1^{\Delta His1+2}$: HOXA1 without; the polyHs and the region in-between; $HOXA1^{\Delta Nter}$: HOXA1 without the 40 first amino acids; $HOXA1^{\Delta 71-199}$: HOXA1 without the 71 to 199 amino acids; mHOXA1: mouse HOXA1; mMDFI mouse MDFI; PolyH: homopolymeric histidine repeats; PolyQ: homopolymeric glutamine repeats; PolyA: homopolymeric alanine repeats; PolyP: homopolymeric proline repeats; PolyG: homopolymeric glycine repeats.

https://doi.org/10.1016/j.bbagrm.2023.194993
Received 5 June 2023; Received in revised form 5 October 2023; Accepted 2 November 2023
Available online 11 November 2023

development, with a 2- to 10-times higher occurrence than in the overall proteome. More than two thirds of developmental proteins from families such as HOX, FOX, and SOX proteins, present amino acid repeats. Polyhistidine (polyH) is the most overrepresented repeat in developmental proteins, followed by, in decreasing order, polyG, -A, -R, -Q, -P, and -S [4]. Outside this category of development regulators, polyH are relatively rare [5] and on average they are longer than the other amino acid repeats, most polyH being more than five histidine-long [11].

On an evolutionary time scale, modifications of repeated sequences might be relatively rapid because of the high mutation rates associated with repetitive DNA sequences [13]. The expansion and shortening of repeats within transcription factors have been linked to major morphological changes in vertebrates [14]. Therefore, considering the relatively high occurrence of polyH in developmental proteins, mutations in histidine repeats may have played significant roles in the evolution of developmental processes. Many polyH proteins have paralogues devoid of histidine repeats. It has been inferred that polyH-containing paralogues derived from gene duplications that occurred in the early vertebrate evolution and that the polyH coding capacity has been gained after the duplication of the parental gene. [15].

Acquisition of polyH might have contributed to the reorganization of protein-protein interaction networks at work during development [16]. In Cyclin T1, a polyH has indeed been described as a protein interacting interface. Cyclin T1 is involved in the positive regulation of RNA polymerase II by stimulating the phosphorylation of its carboxy terminal domain. This activity of cyclin T1 is inhibited by its polyH-mediated physical interaction with granulin, a protein displaying cysteine-rich repeats [17]. PolyH also addresses proteins to nuclear speckles. Addition of a polyH to the green fluorescent protein GFP relocates it at nuclear speckles and the length of the polyH has been correlated to the proportion of relocated GFP [16]. The functional implications of the polyH-mediated addressing of proteins towards nuclear speckles have been described in nuclear speckle biogenesis, for DYRK1A (dual specificity tyrosine-phosphorylation-regulated kinase) [18], or in the regulation of polII phosphorylation, for cyclin T1 and DYRK1A [19,20].

HOX proteins are key transcription factors involved in animal development. They critically contribute in patterning the anteroposterior axis and the limbs. They also participate in most organogenesis processes. In mammals, HOXA1 is involved in the development of the heart, the hindbrain, cranial nerves as well as of the middle and inner ear [21–25]. *HOXA1* expression has also been associated to different types of cancer, including liver, stomach, lung, prostate or endometrium cancer, and it has been identified as an oncogene in breast cancers [26].

Several high throughput approaches have contributed to lift the veil on how HOXA1 might fulfill its functions: RNA-seq on mouse embryos [21], ChIP-seq analyses on ES cells [27] and interactomic screens have been carried out [28]. However, from a structural viewpoint, the molecular determinants which underlie HOXA1 protein activities and interactions remain largely unexplored. Like all HOX proteins, HOXA1 contains a homeodomain (HD) known to be responsible for DNA binding. This is the most conserved domain shared by the HOX family. A hexapeptide (HX) motif N-terminal to the homeodomain has been identified to be critical for the HOXA1 interaction with PBX proteins [29–31]. PBX proteins are Three Amino acid Loop Extension homeodomain proteins. These are the best characterized HOX interactors known to enhance the selectivity, affinity and specificity of HOX DNA binding [32–34]. In addition to these widely shared homeodomains and hexapeptide motif, HOXA1 is the only HOX protein showing polyH motifs. In human, HOXA1 has two polyH, one of 10 histidines (His1) and one of 5 histidines (His2). Although conserved among vertebrate HOXA1 orthologues, these polyH have undergone changes in length through evolution (Fig. 1). Within the placental mammal lineage, the His1 motif varies from 6- (in the northern greater galago, *Otolemur garnettii*) to 12 residues in length (in the rabbit, *Oryctolagus cuniculus*). Surprisingly in the northern greater galago the shorter His1 repeat is correlated with a longer His2 motif. In non-placental mammals, His1 is shorter (about 6 histidines) with a conserved alanine insertion. In other

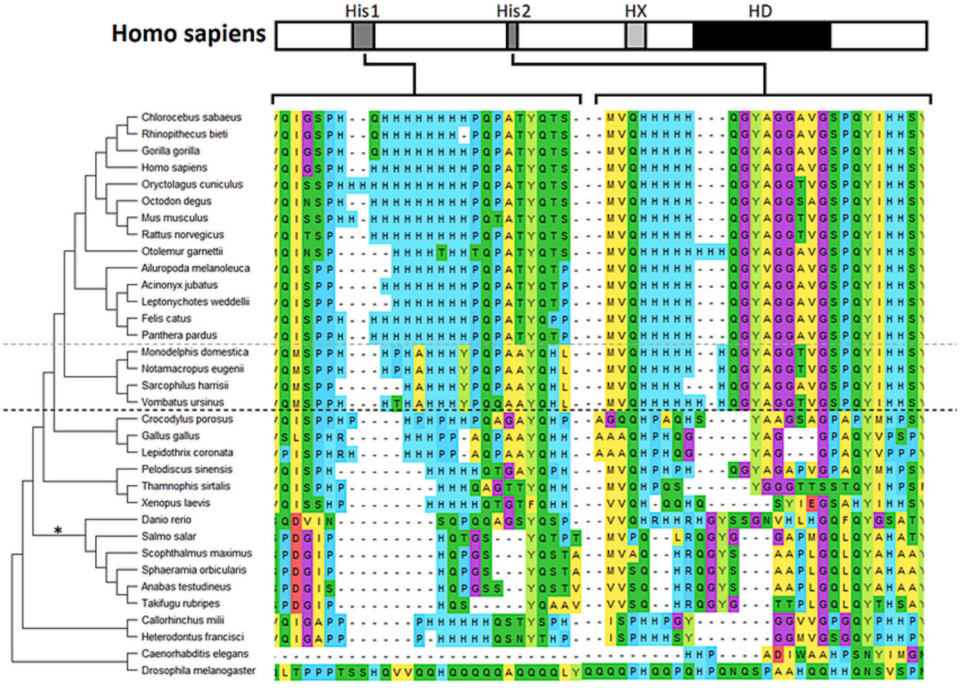


Fig. 1. Alignment of the His1- and His2-containing regions of the human HOXA1 protein sequence with orthologs. The alignment and the phylogenetic tree were generated using MEGA5. The black dotted line separates mammalian species (above) from other vertebrates and protostomes (*Caenorhabditis elegans* and *Drosophila melanogaster*) species. The grey dotted line separates non-placental (below) and placental (above) mammals, the asterisk indicates the actinopterygians branch. The aligned sequences correspond to the His1- and His2- containing protein regions, as depicted on the schematic representation of the HOXA1 protein. HX: hexapeptide; HD, homeodomain.

vertebrates, His2 is generally absent and His1 consists in about five histidines. Most significantly, the ray-finned fishes (actinopterygians) do not show polyH motifs in their HOXA1 orthologues while the cartilaginous fishes do. The actinopterygian phylum has undergone a whole genome duplication with respect to the other vertebrates, suggesting that the deletion of the polyH is related to the neofunctionalization of HOXA1 paralogues associated with this whole genome duplication.

The integrity of the HOXA1 polyH motifs has been shown to be important for the HOXA1 functions in distinct biological contexts. In human, length modification in His1 is associated with autism [35]. More recently, length variations of the His1 motif has been identified to be associated with cardiac malformations in humans and to cause similar cardiac malformations in the mouse and zebrafish [36]. HOXA1 has previously been shown to be involved in heart development [21,24,37–39], the His1 motif would therefore define an important molecular determinant of the functional interactions which would take place in this particular context. His1 has also been identified to be involved in HOXA1 protein interactions [40,41]. Through the interaction with RBCK1 and TRAF2, His1 has been shown to be required for the HOXA1-mediated regulation of the NF-$_{K}$B pathway which would be relevant to the oncogenic properties of HOXA1 in breast cancer [42].

To get a better understanding of the role played by polyH in transcription factors, we first used a bioinformatic approach to look for common features shared by the interactomes of polyH-containing proteins in human. This led us to identify that polyH proteins share interactomes enriched in cysteine-rich proteins and in proteins displaying (a) cysteine repeat(s). Then, focusing on HOXA1, we revealed a correlation between the length of the polyH repeat and the strength of the interactions, the requirement of metal ions for the interaction with cysteine-rich proteins, and we highlighted a down-regulatory impact of some interactions on the HOXA1 transcriptional activity.

2. Results

2.1. *In silico analysis of interactomes of proteins with histidine repeats*

Previous human HOXA1 (hereafter referred to as hHOXA1) interaction data revealed the importance of polyH for protein-protein interactions [42]. To address the possible involvement of polyH motifs in protein-protein interactions in the human proteome, we selected all the human proteins containing a repeat of >7 histidines which define a subset of 37 proteins (from UniProt) (Fig. 2). A gene ontology (GO) analysis confirmed the previously reported prevalence of transcription factors (p-value = 1.1 *10^{-6}) and developmental proteins (p-value = 7.6

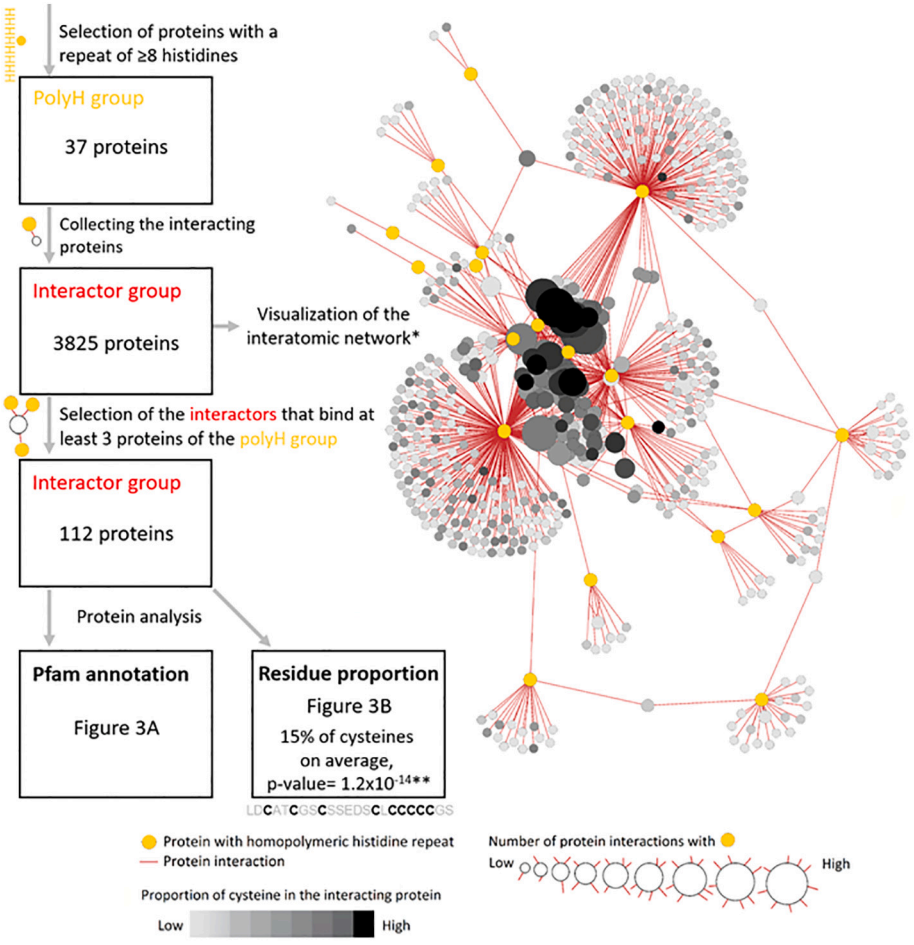


Fig. 2. Workflow of the bioinformatic analysis. Sequences of proteins with repeat of 8 or more histidines were collected from UniProt, and their interatomic information extracted from the BioGRID database. The grey arrows represent the analysis process, and results are highlighted in the black boxes. For the statistical analyses, filters were applied to identify interactors that bind to at least 3 proteins of the polyH group, aiming to identify common features in proteins that interact with polyH proteins (Pfam annotation (see Fig. 3A) and residue proportion (see Fig. 3B)). Before filtering, the interaction network was visualized. Red bars represent physical interactions between proteins. It is important to note that the length of the red bar does not convey any specific information but is adjusted to facilitate visualization. Interestingly, the darker dots, representing proteins with higher proportion of cysteines, are mainly located in the center of the network, suggesting their predominant ability to interact with polyH proteins. *To ensure clarity of visualization, 19 polyH proteins were randomly selected. ** Comparing this subset to the entire proteome (which has on average of 2.6 % cysteines).

*10^{-3}) among polyH proteins [4]. The proteins in this subset are reported to interact on average with 103 proteins (median at 43; BioGRID database), for a total of 3825 interactions. We then crossed the interactomes of the 37 polyH proteins and extracted proteins that have been identified in three or more interactomes, i.e. proteins reported to interact with at least three polyH proteins. We next tried to find common features shared by the polyH protein interactors. A Pfam analysis (using DAVID platform) highlighted an enrichment in keratins and keratin associated proteins (Fig. 3A). A common characteristic of the keratins and keratin associated proteins is their high cysteine content (used for protein interaction by forming disulfide bridges [43]). To determine if this is a shared feature of polyH interacting proteins, the proportion in particular amino acids in the interactor sequences has been addressed. Most significantly, the interactors of polyH proteins contain proportionally more cysteines than other proteins, as revealed when comparing these interactors to the whole proteome (15 % versus 2 % cysteines, respectively, p-value = $1.2*10^{-14}$), which appear drastically more enriched than serine, the second most enriched amino-acid in the interactors of polyH proteins (12 versus 8 %, p-value = $1*10^{-6}$; Fig. 3B).

To challenge the significance of this observation, we repeated this analysis with proteins harboring the two most common homopolymeric amino acid repeats, polyA and polyG, as well as with polyQ containing proteins as the proportion of polyQ proteins in the human proteome is close to that of polyH [4]. Interactors of polyH, polyA, polyQ and polyG containing proteins have been sorted into four groups according to the number of proteins they interact with: interactors showing two, three, four or more than four interactions with polyH, polyA, polyQ or polyG proteins, respectively. Finally, the different groups of proteins were compared for their abundance in cysteines (Fig. 3C).

For the proteins that interact with more than four polyH, polyA, polyQ or polyG proteins, those that interact with polyH proteins are significantly more enriched in cysteines as compared to the polyA, polyQ or polyG protein interactors (Fig. 3C). This higher proportion of cysteine is also observed, although to a lesser extent, if compared to proteins interacting with two, three or four polyH proteins. It is not observed for the control groups, i.e. polyA, polyQ or polyG protein interactors.

Keratins and keratin-associated proteins are only present in the polyH protein interactomes (in comparison to polyA, polyG, and polyQ interactomes). To ensure that the mere occurrence of keratins in interactomes does not bias the analysis, it has been repeated excluding keratins and keratin-associated proteins. In the subset of proteins that interact with at least three polyH proteins, enrichment in cysteines was still observed as significant when compared to the whole proteome (p-value = 0.0053; Fig. 3D). When comparing polyH protein interactors with those of polyA, polyQ and polyG proteins, the difference in cysteine abundance appeared significant only for proteins interacting with more than four polyH proteins versus proteins interacting with more than four

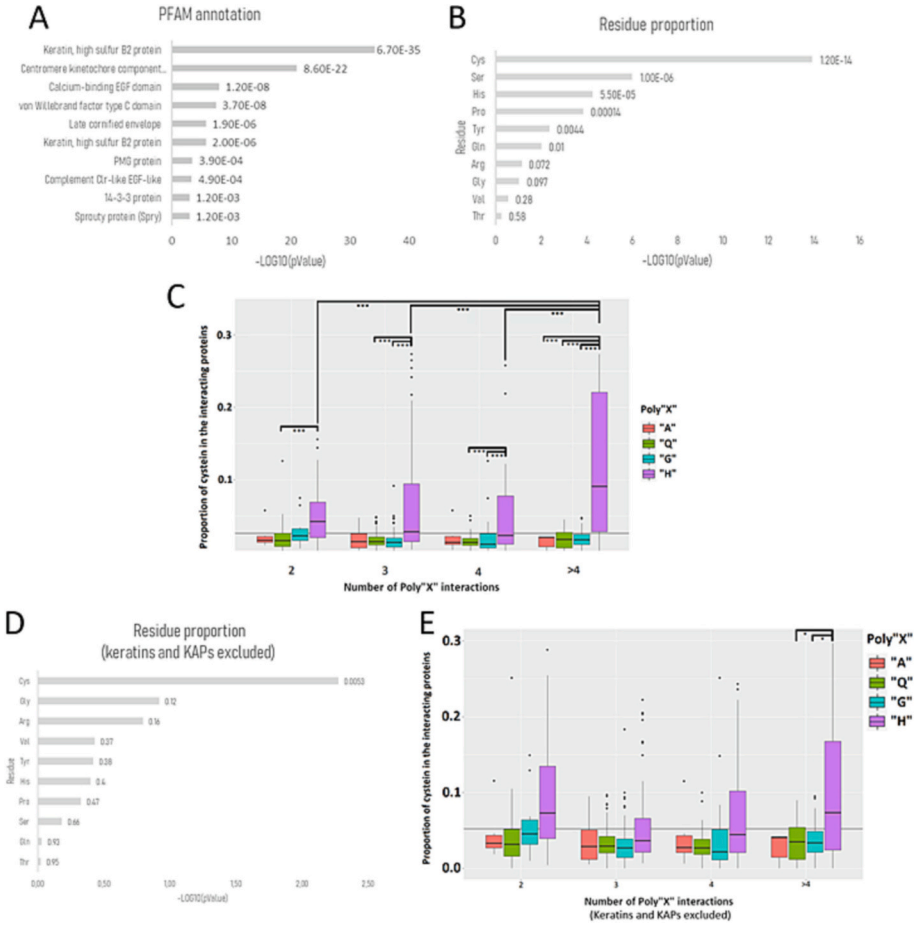


Fig. 3. A high content in cysteines is a common feature of polyH interacting proteins. (A) Pfam analysis on proteins that interact with at least three polyH proteins. The corresponding p-values for Pfam signatures are indicated. (B and D) Amino acid proportion in proteins that interact with at least three polyH proteins. The corresponding p-values for comparative residue proportion between interactors and the whole proteome are indicated as label on the histogram. (C and E) Comparative analysis of polyA, polyQ, polyG, polyH interacting proteins. Proteins that interact with polyA, polyQ, polyG, and polyH containing proteins have been divided in four groups, according to the number of polyA, polyQ, polyG, and polyH proteins they respectively interact with: two, three, four or more than four. The different groups have been compared based on the proportion of cysteines contained in the proteins they include. The statistical analysis was performed using a Wilcoxon test (* = p value <0,05; *** = p value <0.001). (D and E) The keratins and keratin-associated proteins (KAPs) were excluded from the analysis.

polyG and polyQ proteins (Fig. 3E). For polyA protein interactors, comparisons were not statistically robust because only seven proteins defined the group of proteins interacting with more than four for polyA proteins. In contrast, 20 proteins have been reported to interact with more than four polyQ proteins, 56 with more than four polyG proteins, and 20 with more than four polyH proteins.

Taken together, the significant correlations drawn by the interactome analyses prompted us to address the involvement of the polyH motif in the interaction with cysteine-rich proteins.

2.2. In silico analysis of interactomes of proteins with cysteine repeats

To confirm this correlation, we reversed our analysis and combined the interactomes of all proteins that contain a repeat of cysteines (polyC). These repetitions are shorter compared to the previously studied repetitions (i.e. polyH, -A, -G, -Q). Indeed, proteins containing polyC motifs of 6 residues or more represent 11 proteins, of 5 residues or more, 25 proteins, and of 4 residues or more, 60 proteins. Consequently, the group of 60 proteins was selected to have a number of proteins equivalent to the polyH, -A, -G, -Q groups (40 proteins on average) and to have enough proteins to generate a robust analysis. This analysis revealed the polyH proteins OTX1 (p-value = $1.2 *10^{-6}$) and HOXA1 (p-value = $6.2 *10^{-6}$) as the proteins showing the greatest number of polyC interactors (6 interactions).

2.3. Involvement of polyH in the HOXA1 interactome

To investigate the correlation observed in silico between polyH proteins and cysteine-rich interactors, we decided to focus on HOXA1 which has a well-characterized interactome and for which some regulatory interactions have already been studied [26,28,40–42].

We used bimolecular fluorescence complementation (BiFC) to observe the protein-protein interactions mediated by the wildtype and a mutant version of hHOXA1 lacking the longest histidine tract (hHOXA1$^{\Delta His1}$). The C-terminal (VC) or N-terminal (VN) complementary moieties of the Venus fluorescent protein were fused to hHOXA1 or its interactors, respectively (Fig. 4A). We selected interactors described in Lambert et al. [28]. Plasmids coding for these fusion proteins were co-transfected in COS7 cells. Controls consisted in co-expressing each fusion protein with its complementary unfused Venus fragment (VC-hHOXA1 with VN, VN-interactor with VC) or to combine both unfused VN and VC Venus polypeptides.

BiFC assays revealed that the hHOXA1$^{\Delta His1}$ mutant was able to establish interactions with low proportion of cysteine proteins (Fig. 4B, green dots) but unable to interact with high cysteine proportion proteins (Fig. 4B, red dots), except for FHL5 (Fig. 4A and B, all additional BiFC images available in Supplementary Fig. 1). For this interactor, removal of both His1 and the second, shorter, histidine repeat, His2, (HOXA1$^{\Delta His1+2}$) resulted in the loss of interaction (Supplementary Fig. 1). Taken together, these data tend to confirm the *in-silico* correlations. The interaction with cysteine-rich proteins is impacted by the removal of the long hHOXA1 histidine repeat, while interaction with proteins containing less cysteines is less affected.

2.4. Comparison of protein-protein interaction strength by co-precipitation

To better characterize the contribution of histidines and cysteines in polyH protein-mediated interactions, we carried out co-precipitation assays and focused on the interaction between HOXA1 and MDFI (Fig. 5). MDFI was selected among the HOXA1 interactors because it is one of the richest in cysteines and it loses the interaction with hHOXA1$^{\Delta His1}$. The wildtype and previously generated mutant mouse HOXA1 proteins (hereafter referred to as mHOXA1) (Fig. 5A) [40] were fused to Glutathione-S-Transferase (GST) and co-expressed with a Flag-tagged mouse MDFI (hereafter referred to mMDFI).

The GST-fused proteins were then purified on glutathione-sepharose beads. Cells transfected with unfused GST and Flag-mMDFI were used to

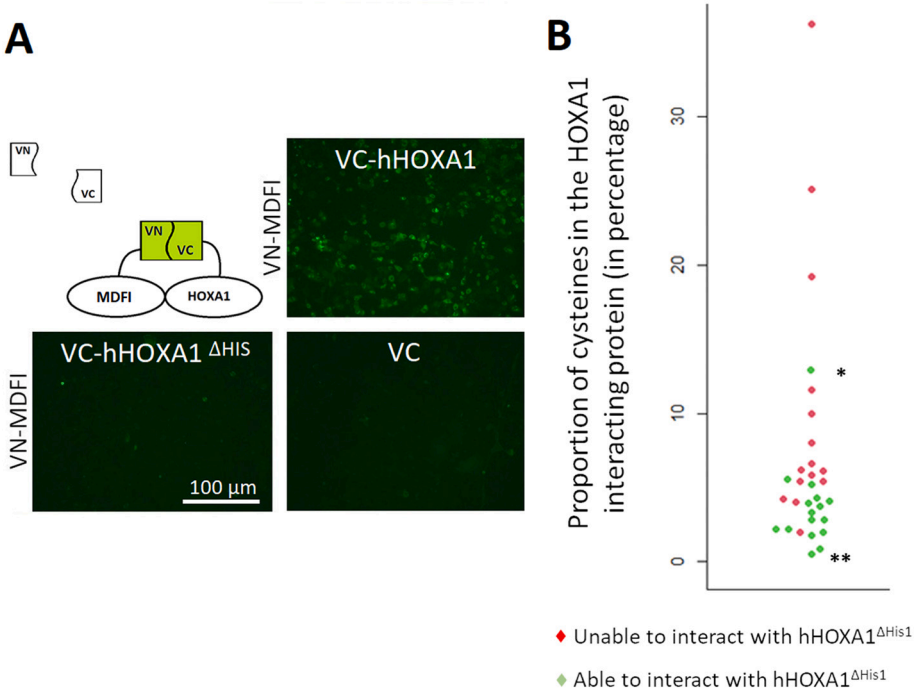


Fig. 4. Deletion of the His1 repeat of hHOXA1 impairs interaction with cysteine-rich proteins. (A) Representation of the BiFC experiments using MDFI as an illustrative example. COS7 cells were transfected with expression vectors coding for VC or VC-hHOXA1 (WT or hHOXA1$^{\Delta His1}$) together with VN or VN fused to an interacting protein. Green fluorescence corresponds to the reassembly between the two moieties of the Venus protein subsequent to the interaction between the two proteins. (B) Loss of interactions in BiFC assays (images presented in Supplementary Fig. 1). * and ** indicate FHL5 and BAT2L1, respectively, for these proteins the interaction between with HOXA1$^{\Delta His2}$ and HOXA1$^{\Delta His1-2}$ also has been tested.

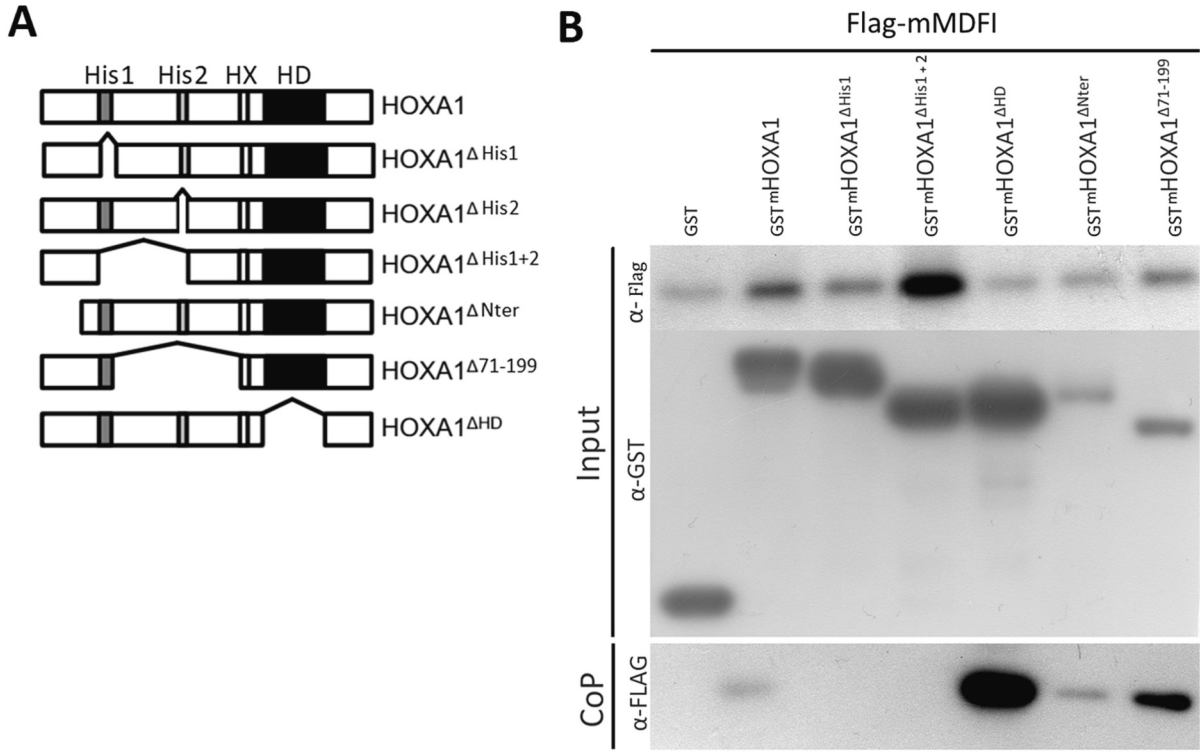


Fig. 5. Deletion of the His1 repeat of mHOXA1 impairs interaction with mMDFI. (A) Schematic representation of HOXA1 deletion mutants adapted form Draime, et al. 2018 [41]. (B) Determination of mouse HOXA1 (mHOXA1) regions involved in the interaction with mouse MDFI (mMDFI) by co-precipitation assays. GST or GST-mHOXA1 fusion proteins (wildtype and deletion derivatives, see details in the text) were transfected with Flag-mMDFI in HEK293T cells. Forty-eight hours after transfection, cells were lysed and expression of Flag- and GST-fused proteins were analyzed by western blotting ('input'). The interactions were evaluated by the presence of Flag-mMDFI after co-precipitation on glutathione beads ('CoP'). mHOXA1 interacts with mMDFI and the polyH of mHOXA1 is involved in this interaction. Provided blots are representative of at least three independent experiments. Western blots were revealed with anti-Flag ("α-Flag") or anti-GST ("α-GST") antibodies.

control non-specific interaction. mHOXA1 and its different mutant versions influenced the abundance of Flag-mMDFI in cells. These variations were present in every biological replicate but did not influence the general conclusion of the experiment. Flag-mMDFI was successfully detected after precipitation of GST-mHOXA1 but not with GST alone. As observed with BiFC, deletion of His1 impaired the interaction with mMDFI (Fig. 5B). mHOXA1$^{\Delta His1+2}$, deleted from the long and short histidine motifs, was also not able to interact with mMDFI. mHOXA1$^{\Delta HD}$ (deleted of the complete homeodomain), mHOXA1$^{\Delta 71–199}$ (lacks the central region extending from the 11-His repeat to the hexapeptide) and mHOXA1$^{\Delta Nter}$ (deleted of its 40 N-terminal residues) were still able to interact with mMDFI. In conclusion, all the mHOXA1 mutant versions that did not lack His1 were still able to co-precipitate mMDFI, confirming the involvement of this polyH in mediating the interaction.

In order to get insight on the possible relationship between the length of histidine repeats and the affinity for cysteine-rich interactors, we took advantage of the elongated version His1 present in the mouse HOXA1 which displays 11 histidines instead of 10 in the human orthologue. The mouse and human HOXA1 orthologues otherwise share 93.4 % sequence identity. We observed a stronger interaction between the human MDFI

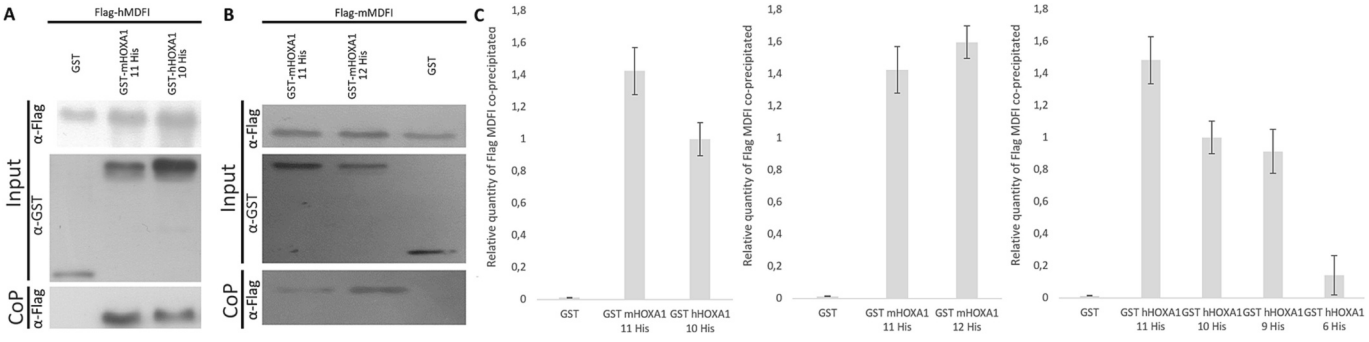


Fig. 6. HOXA1 interacts with MDFI and the length of the His1 polyH motif is correlated with the strength of the interaction. Forty-eight hours after transfection, cells were lysed and GST as well as Flag- and GST-fused proteins were analyzed by western blotting ('input'). The interactions were evaluated by the presence of Flag-MDFI after co-precipitation on glutathione beads ('CoP'). (A) Co-precipitation assay for the interaction between hMDFI and the mHOXA1 or hHOXA1 proteins expressed in HEK293T cells. (B) Co-precipitation assay for the interaction between mMDFI and mHOXA1 with 11- or 12-histidine long His1 motif expressed in HEK293T cells. Provided blots are representative of at least three independent experiments. Western blots were revealed with anti-Flag ("α-Flag") or anti-GST ("α-GST") antibodies. (C) Quantification of the relative amount of hMDFI or mMDFI co-precipitated, normalized with the Flag and GST input signals. Co-precipitation data form A and B as well as data from experiments involving hMDFI and hHOXA1 derivatives with His1 variants of 6, 9, 10 or 11 residues.

(hereafter referred to hMDFI) and mHOXA1, compared to hHOXA1 (Fig. 6A). The interaction has been conserved during evolution as mMDFI is able to interact with mHOXA1 and hHOXA1 (Supplementary Fig. 2). The sequence of mHOXA1 was next modified to add one histidine in the His1 motif. By comparing the WT mHOXA1 containing 11 histidines and the mutant version with 12 histidines, we observed a correlation between the number of residues in the His1 of mHOXA1 and the amount of mMDFI co-precipitated. The mHOXA1 protein with 12 histidines resulted in a better interaction with mMDFI as observed upon co-precipitation (Fig. 6B). We then modified the length of His1 in hHOXA1 to get 6, 9, and 11 histidines. As shown in Fig. 6C, the same correlation appeared between the length of His1 and the strength of the interaction with hMDFI. Taken together, these data clearly highlight a link between the strength of the interaction with MDFI and the length of the His1 repetition in the HOXA1 protein.

2.5. Implication of metal ions in the HOXA1-MDFI interaction

Based on structural data from cysteine- and histidine-constrained zinc fingers, as well as on the capacity histidine repeats have to coordinate metal ions (like Nickel for example), we hypothesized that metal ions could contribute to the protein-protein interaction between the cysteine-rich MDFI and the polyH protein HOXA1. We repeated the MDFI-HOXA1 co-precipitation assay with buffers containing cation scavengers imidazole or phenanthroline. We separately produced hHOXA1 and hMDFI in transfected cells. Proteins were collected in buffers containing, or not, imidazole or phenanthroline. They were next pooled and incubated with glutathione beads. In the control condition, without scavenger, the hHOXA1-hMDFI interaction was recovered. However, GST-hHOXA1 was not able to co-precipitate Flag-hMDFI in the presence of imidazole or phenanthroline. The presence of those competitors impaired the hHOXA1-hMDFI co-precipitation, supporting the possible involvement of a metal ion in establishing the interaction (Fig. 7). Noteworthy, when the proteins were co-expressed in cells, the interaction was maintained in presence of imidazole, i.e. once the interaction is established it cannot be disrupted by a cation scavenger (Supplementary Fig. 3).

2.6. Role of the histidine-mediated interactions on HOXA1 transcriptional activity

To address the functional implication of the interactions involving cysteine-rich interactors, the transcriptional activity of hHOXA1 was characterized by luciferase reporter assays, in the presence or absence of interactors. Luciferase expression was driven by a HOXA1 target enhancer, the *TSEII* element [30]. Ten interactors were tested for their possible effect on the hHOXA1 activity (Fig. 8). These proteins were shown above to lose their interaction with hHOXA1 upon deletion of the hHOXA1 His1 motif. As previously reported, HOXA1 was active on *TSEII-luc*, but required the co-transfection of the TALE proteins PBX1 and PREP1 expression vectors [30]. We next tested the functional interaction between the interactors and hHOXA1. Three of them, TRAF1, RBCK1 and SPRY1, showed an inhibitory effect on the hHOXA1 activity (Fig. 8A). No impact of the other seven was detected. Next, the transcriptional activity of hHOXA1$^{\Delta His1}$ appeared similar to that of the wildtype protein (Fig. 8B). Most significantly, the inhibitory effect of TRAF1, RBCK1 and SPRY1 on hHOXA1 was lost when removing the His1 repeat from hHOXA1 (Fig. 8B). This supports a significant role of histidine-mediated interactions in the activity modulation of HOXA1 by these cysteine-rich interactors.

3. Discussion

In this paper, focusing on the HOXA1 transcription factor as paradigm, we provided evidence that polyH and cysteine-rich domains can serve as interacting interface in protein-protein interactions. In addition,

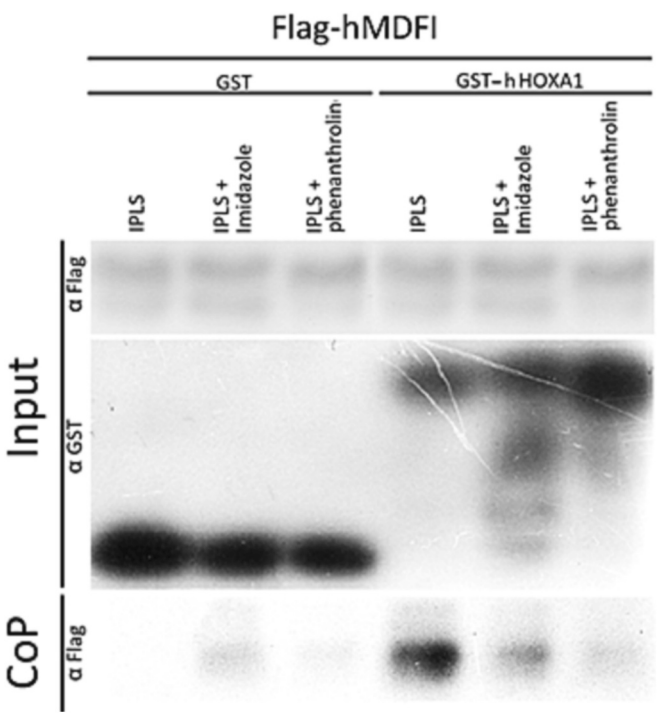


Fig. 7. Metal ion is required for hHOXA1 - hMDFI interaction. Flag-hMDFI and GST or GST-hHOXA1 were separately transfected in HEK293T cells. Forty-eight hours after transfection, cells were lysed with IPLS containing or not imidazole or phenanthroline. Expression of Flag- and GST-fused proteins was analyzed by western blotting (Input). Cell extracts were pooled before co-precipitation of protein complexes on glutathione beads. The interaction between hHOXA1 and hMDFI was evaluated by the presence of Flag-hMDFI after co-precipitation on glutathione beads (CoP). Provided blots are representative of at least three independent experiments. Western blots were revealed with anti-Flag ("α-Flag") or anti-GST ("α-GST") antibodies.

co-precipitation assays indicated a relationship between the length of the polyH motif and the strength of interaction with one selected HOXA1 interactor, MDFI. Our results also support that this interaction is coordinated around a metal ion. Finally, functional luciferase assays supported that polyH serves as interacting surface for the regulation of the HOXA1 transcriptional activity. Taken together, our data identify polyH-cysteine interactions as important determinants for the intermolecular modulation of protein activities.

The analysis of BioGRID data on polyH proteins highlighted an enrichment in physical interaction with cysteine-rich proteins. Conversely, the most enriched proteins among the polyC interactors are polyH-containing proteins. The data in BioGRID however contain some biases that need to be taken into consideration. First, not all the proteins received the same effort of investigation. For some proteins, more than a hundred interactions have been reported, where others have less than five. To reach conclusions from comparative analyses of such interaction datasets, however, we chose to overcome this bias by the use of control groups of proteins, namely proteins showing distinct homopolymeric repeats (polyA, polyG and polyQ proteins). Second, the interactions recorded in databases rarely come from experiments performed in physiological contexts. Most high-throughput interatomic data have been generated in heterologous systems like in yeast two hybrid assays. Thus, interactions that can take place in such systems could actually never take place in a physiological situation. More specifically for polyH proteins, interatomic data in BioGRID display a strong enrichment in keratins and keratin- associated proteins, which is not the case for the polyA, polyG and polyQ interatomic data. As most polyH proteins are transcription factors, like HOXA1, the physiological implication of their interaction with keratins and keratin-associated proteins is

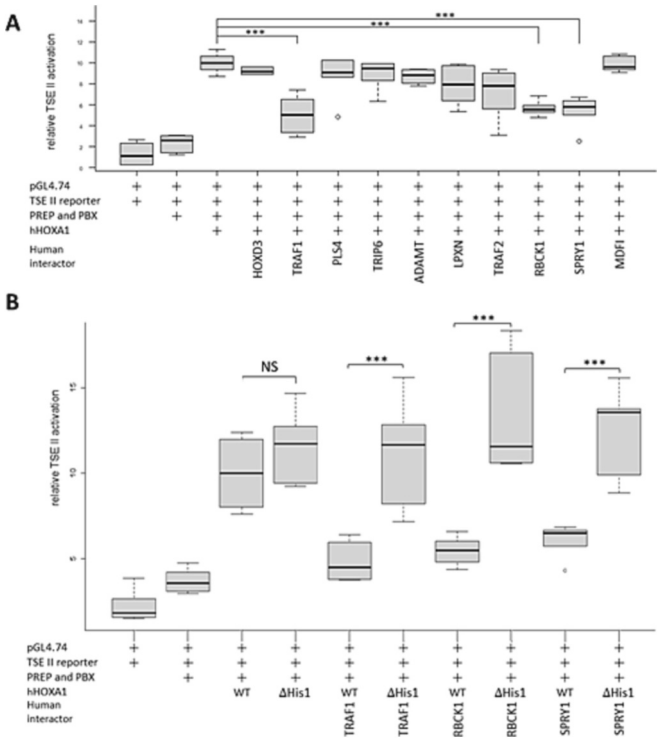


Fig. 8. Consequences of the interactions mediated by polyH on the hHOXA1 transcriptional activity. (A) Effect of HOXA1 polyH interacting proteins on the activation of the TSEII-luc reporter by hHOXA1. Luciferase activity assays were performed in HEK293T cells transfected with luc reporters and expression vectors for hHOXA1, HOXA1 polyH interacting protein, and the cofactors PREP and PBX, as indicated. (B) Effect of HOXA1 polyH interacting proteins on the activation of TSEII-luc by hHOXA1 or hHOXA1$^{\Delta His1}$. A constitutively active luciferase expression vector (pGL4.74) was included in each transfection for assay normalization. Data are presented as TSEII-luc/pGL4.74 activity ratios. Provided data correspond to six biological replicates. The statistical analysis was performed with a Wilcoxon test (*** = p value ⟨0,001), NS (non-significant).

questionable. Nonetheless, although it can be hypothesized that polyH and cysteine-rich proteins generate artefactual outcomes in high-throughput interatomic screenings, this still supports the involvement of the cysteine-histidine interfaces in intermolecular interactions. Notwithstanding, while removing keratins and keratin-associated proteins from datasets, enrichment in cysteine-rich proteins among polyH interactors remains clearly significant.

Interaction between histidines and cysteines has already been reported but, to our knowledge, only in intramolecular configurations, like in zinc finger proteins. In their context, the interaction is stabilized by the coordination of a metal ion between histidine(s) and cysteine(s) [44]. This cation-mediated histidine-cysteine interaction is required for the proper folding of the structural Zn-finger domain. A similar cation coordination is used by proteins involved in the storage or in the resistance to metal ions. Ccsp (Cytosolic copper storage protein), for example, uses cysteines and histidines for the cytosolic stabilization of metal ions [45]. In our data, co-precipitation of MDFI and HOXA1 supports that metal ion is also required for protein-protein interactions mediated by histidines and cysteines. The nature of this ion as well as the conformation of the coordination appeals for additional structural investigation.

The hypothesis of the involvement of metal ions in the HOXA1-MDFI interaction and the capacity of cysteines and histidines to coordinate metal ions is reinforced by two additional observations. Firstly, the conserved interaction between HOXA1 and MDFI seems to rely on the cysteine-rich domain of MDFI. Indeed, both the human and murine MDFI have been shown to interact with HOXA1 in co-precipitation assays (Fig. 5). Interestingly, while MDFI sequences of the mouse and human proteins strikingly differ in their N-terminal regions, their C-terminal half, which contains most cysteine residues (26/29), is conserved with 98 % of identity (Supplementary Fig. 2). This points towards the MDFI C-terminal region and its cysteines as supporting the interaction with HOXA1. Secondly, regarding the interaction between FHL5 and HOXA1, the deletion of both His1 and His2 motifs in hHOXA1 was required to affect the interaction. This could be due to the very high density of cysteines in FHL5. Indeed, except from keratins, FHL5 is the most cysteine-dense HOXA1 interactor used in this analysis. The abundance in cysteines would support the interaction with the few histidines remaining in the hHOXA1$^{\Delta His1}$ mutant.

It has been hypothesized that polyH have emerged along with the whole-genome duplications that occurred during vertebrate evolution [16]. The vertebrate *HOX* complexes, four in number, exemplify the two whole-genome duplications which are inferred to have taken place in this phylum. HOXA1 is the only HOX protein, and the only representative among its paralogues, to display a polyH motif of >7 residues. Consistently with the hypothesis that polyH appeared in proteins after gene diversification through gene duplication in the vertebrate history, polyH are absent in the HOXA1 orthologues from non-vertebrate species. Next, as any repetitive sequence, polyH is prone to expansion or shortening. Indeed, the HOXA1 polyH region is relatively short in non-mammalian species and dramatically expanded in placental mammals. The role of repeated sequence length modification during evolution has been highlighted is several reports [4,14,16]. Data we obtained by co-precipitation suggest that modification in length of the polyH region modifies the strength of interaction with cysteine-rich proteins. Variations in polyH length may have been selected during the evolution while reinforcing or reducing some interactions thereby allowing the evolution of interatomic networks at work, in particular in developmental processes. Evolution of these networks is a driving process towards the evolution of living forms.

Among the cysteine-rich interactors of HOXA1 which have been identified here to establish polyH-dependent interactions, three of them have been shown to inhibit the transcriptional activity of HOXA1: RBCK1, TRAF1 and SPRY1. The other interactors tested did not influence the activity of HOXA1 and in particular no interactor appeared to stimulate HOXA1 activity. It might be hypothesized that some proteins interact with HOXA1 in other molecular contexts than in transcription regulation or that their possible effect on HOXA1 activity is masked in our experimental system based on transfection and protein overexpression. The molecular and functional interactions with RBCK1, TRAF1 and SPRY1 might be biologically meaningful. The interaction between RBCK1 and HOXA1 has been identified to be relevant to the modulation of the TNFα/NF-κB pathway in the context of breast oncogenesis [46]. Other interactors like TRAF1 or SPRY1 which are known modulators of signaling pathways might influence HOXA1 as a downstream response to such signaling by promoting post-translational modification or destabilization of HOXA1 (as already shown for other interactors [47,48]), by modulating HOXA1 access to the nucleus or HOXA1 access to the chromatin and enhancers within the nucleus. These possibilities however deserve further investigation. RBCK1, TRAF1 and HOXA1 can be co-expressed in breast cancers and have been associated with reduced survival prognostic [26,42,49–52]. SPRY1 and HOXA1 are both involved in regulating neural crest cell migration towards the outflow tract of the heart during development, and they both participate to the formation of the aortic valve. It would therefore worth investigating how HOXA1 and these interactors influence each other in the physiological context of development or in the etiology of cancers.

The polyH and other homopolymeric repeats are known to be disordered regions. It has been shown that these regions become more structurally ordered upon protein-protein interaction [53]. This structural organization taking place upon interaction might be interactor-dependent as well as context-dependent, which can be related to the

hypothetically versatile repertoire of interactions proteins like transcription factors are supposed to establish. Obtaining the structure of distinct protein-protein complexes involving histidine-to-cysteine interactions would therefore be extremely informative to make the connection between disordered and linear motifs residing in transcription factors and the versatility of their intermolecular interactions.

4. Materials and methods

4.1. Protein sequence alignment

Sequences from HOXA1 were retrieved from the UniProt database (10/08/2022; https://www.uniprot.org/). Only sequences manually annotated and reviewed by UniProtKB curators were used. The alignment and the phylogenetical tree were generated using MEGA5.

4.2. Interactome analysis

In UniProt database, the reviewed proteins containing a polyH motif of at least 7 histidines were selected for analysis (37 proteins) (11/08/2022). The interactors of the selected proteins were collected from the BioGRID database (https://thebiogrid.org/). The sequences of the interacting proteins were also retrieved from UniProt. Sequence analyses and statistical analyses were performed on R. For the control groups, proteins containing polyQ of 7 glutamines or more (30 proteins), polyA of 12 alanines or more (56 proteins), and polyG of 11 glycines or more (39 proteins) were selected. Gene ontology and PFAM analyses were performed on the DAVID platform (https://david.ncifcrf.gov/).

4.3. Genetic constructs

The plasmid coding for mouse HOXA1 and PBX1A [30]; hHOXA1 [26]; mHOXA1$^{\Delta Nter}$, mHOXA1$^{\Delta 71–199}$, mHOXA1$^{\Delta His1}$, mHOXA1$^{\Delta HD}$ and mHOXA1$^{\Delta His1+2}$ [40]; hHOXA1^{11His}, hHOXA1^{9His}, and hHOXA1^{6His} [36]; PREP1 [54]; and hMDFI, were describe previously.

The mHOXA1^{12His} mutant mHOXA1 was generated using the following primers: 5′- CTCGCCCCACCACCACCACCACCACCACCACCATCACCACCCCCAGACGGCTACTTACCA-3′ and 5′- TCTGGGGGTGGTGATGGTGGTGGTGGTGGTGGTGGTGGGGCGA GCTGATCTGCACCC -3′. The mMDFI vectors were produced starting from mouse teratocarcinoma P19 cells RNA. RNA was extracted using an RNA extraction kit (Roche #11828665001) following the manufacturer's instructions. Reverse transcription reaction was carried out with 2 µl of 10 mM dNTPs (Thermo Fisher Scientific #R0192), 1 µl of Random Hexamer Primer (Thermo Fisher Scientific #SO142), 0.25 µl of RiboLock RNAse Inhibitor (Thermo Fisher Scientific #EO0381), 0.5 µl of RevertAid reverse transcriptase, and with 4 µl of 5× Reaction Buffer (Thermo Fisher Scientific #EP0441), 12.25 µl of H_20, and 1 µg of RNA. The mMDFI coding sequence was introduced in pDONR223™ Gateway™ by BP clonase®-mediated recombination (Life Technologies #11789020) using the following primers: attB forward, 5′- GGGGACAACTTTGTACAAAAAAGTTGGCATGTCCCAGGTGAGCGGTCAG −3′; attB reverse, 5′-GGGGACAACTTTGTACAAGAAAGTTGGGTTCAGGAGGAGAAACAGAGTCC -3′. LR clonase® (Life Technologies #11791020) reactions towards vectors pDEST-FLAG(Nter), pDEST-GST(Nter) and pDEST-VN^{173}(Nter) vectors [28] were used to fuse mMDFI with Flag-, VN- and GST- coding sequences, respectively. hHOXA1$^{\Delta His1}$, hHOXA1$^{\Delta His2}$ and hHOXA1$^{\Delta His1+2}$ mutant constructs were obtained by triple PCR using the following primers: 5′- CTGGGGGGGCGAACCGATCTGC-3′ and 5′-GTTCGCCCCCCCAGCCGGCTACCTAC-3′, for hHOXA1$^{\Delta His1}$; 5′- CATAACCCTGCTGGACCATGGGAGATG-3′ and 5′- GGTCCAGCAGGGTTATGCTGGGG-3′, for hHOXA1$^{\Delta His2}$; 5′- CTGGGGGGGCGAACCGATCTGC-3′ and 5′- GGTCCAGCAGGGTTATGCTGGGG-3′, for hHOXA1$^{\Delta His1+2}$. These versions of hHOXA1 were introduced in pDONR223™ Gateway™ by BP clonase®-mediated recombination (Life Technologies #11789020) using the following primers: attB forward, 5′-GGGGACAACTTTGTACAAAAAAGTTGGCATGGACAATGCAAGAATGAACTCC -3′; attB reverse, 5′- GGGGACAACTTTGTACAAGAAAGTTGGGTAGTGGGAGGTAGTCAGAGTGTC -3′. LR clonase® (Life technologies #11791020) reaction towards pDEST-VC^{155}(Nter) [28] was used to fuse the variant hHOXA1 and VC Venus moiety coding sequences.

4.4. Cell culture, transfection and treatment

HEK293T, P19 and COS7 cell lines were grown at 37 °C under 5 % CO_2 in DMEM medium (Thermofisher #61965 for HEK293T and P19 cells, and Thermofisher #31885–023 for COS7 cells) supplemented with 10 % fetal bovine serum (Fisher scientific #10270–106), 100 U/ml of penicillin–streptomycin (Thermofisher #15140122). Sodium pyruvate (1 mM, Thermofisher # 11360–070,) was added to the HEK293T cell medium. JetPrime transfection reagent (Polyplus-transfection #114–07) was used 24 h after plating to transfect plasmid constructs, according to the manufacturer's instructions.

4.5. Bimolecular fluorescence complementation

Seventy-five thousand COS7 cells were seeded on glass coverslips in 24-well plates and were transfected with 250 ng of pEXP-VC^{155}(Nter) and 250 ng of pEXP-VN^{173}(Nter) plasmids according to the experiment. Twenty-four hours post-transfection, the cells were rinsed once in PBS and fixed for 20 min with 4 % PFA-PBS (Sigma #441244) at room temperature. Cells were then rinsed twice (5 min) in TBS (Tris 0.05 M; NaCl 0.15 M; pH 7.5) -0.1 % TritonX100, and once (10 min) with TB (Tris 0.05 M; pH 7.5). Coverslips were mounted and stained with a solution containing DAPI and Vectashield (Labconsult #H-1200) to be observed under an epifluorescence microscope (Zeiss Axioskop 2). Fluorescence quantification was performed using the ImageJ software and tested interactions were considered as positive when the fluorescence emitted in the test condition was at least three times more intense than in the three control conditions involving empty pDEST-VC^{155}(Nter) and/or pDEST-VN^{173}(Nter) vectors.

4.6. Glutathione co-precipitation

Three hundred fifty thousand HEK293T cells were seeded per well of a 6-well plate and transfected with 1 µg of DNA (500 ng for Flag- and 500 ng for GST-tagged coding plasmid). The empty pDEST-GST(Nter) vector was used as a negative control. Forty-eight hours after transfection, cells were rinsed once with PBS and then lysed in cold IPLS buffer (20 mM TrisHCl, pH 7.5; 120 mM NaCl; 0.5 mM EDTA; 0.5 % NP40; 10 % glycerol) supplemented with 1× Complete™ protease inhibitor (Sigma #11873580001), during 20 min on ice under gentle agitation. Cell lysates were centrifuged 5 min at 16.000 *g* at 4 °C. Expression of Flag- or GST-fused proteins was analyzed by western blotting with a mouse anti-Flag antibody (Sigma #F1804), and a mouse anti-GST antibody (Sigma #G1160), respectively. Thirty µl of glutathione-sepharose beads (Sigma #GE17–0756-01) were washed three times with cold IPLS and added to protein lysates and incubated overnight at 4 °C on a rotating wheel. Beads were centrifuged and the supernatant was kept to control the abundance of unbound GST-fusion protein. Beads were then washed three times with cold IPLS buffer, resuspended in Laemmli loading buffer and boiled 5 min at 95 °C. Western blots were performed to detect Flag-tagged proteins (detection of the interaction) and GST-fused proteins (determination of the abundance of GST-protein attached to the beads).

To test the importance of metal ion in protein-protein interactions, imidazole 200 mM (Merck #515876) or phenanthroline 200 mM (Merck #77510) were added to the IPLS buffer. Cells were separately transfected for the Flag-tagged and GST-tagged interactors, and cells lysates were mixed afterwards on the beads.

4.7. Luciferase assays

HEK293T cells seeded on 24-well plates (175.000 cells per well) were transfected with 250 ng of firefly luciferase reporter plasmid (*TSEII-luc*) [30], 25 ng of *Renilla* standard reporter plasmid, 50 ng of PREP1, 50 ng of PBX1A, and 100 ng of HOXA1 or HOXA1$^{\Delta His1}$ expression vectors [40], and 100 ng of the pEXP-VN^{173}(Nter) interactor (MDFI; SPRY1; RBCK1; TRAF2; LPXN; ADAMT; TRIP6; PLS4; TRAF1; HOXD3) coding vectors [28]. The total amount of DNA was kept equal for all conditions by the addition of carrier pCAT vector when required. Forty-eight hours after transfection, cells were harvested and luciferase assays were performed following the manufacturer's instructions (Dual-Glo® Luciferase Assay System, Promega #E2920).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2023.194993.

CRediT authorship contribution statement

DM: made experiments, wrote the manuscript.
FG: made experiments.
TP: made experiments.
LB: made experiments, revised manuscript.
RR: supervised the experiments, revised the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by Research Credits from the “Fund for Scientific Research-FNRS”, Belgium (CDR # J.0157.21, # J.0152.23) and the “Fonds Spéciaux de Recherche” from UCLouvain. LB is Post-doctoral Researcher from the “Fund for Scientific Research -FNRS”.

References

[1] M.A. Huntley, G.B. Golding, Simple sequences are rare in the protein data Bank, Proteins 48 (2002) 134–140.
[2] M.M. Albà, P. Tompa, R.A. Veitia, Amino acid repeats and the structure and evolution of proteins, Genome Dyn. 3 (2007) 119–130.
[3] R. Wetzel, Physical chemistry of polyglutamine: intriguing tales of a monotonous sequence, J. Mol. Biol. 421 (2012) 466–490.
[4] I. Pelassa, M. Cibelli, V. Villeri, E. Lilliu, S. Vaglietti, F. Olocco, et al., Compound dynamics and combinatorial patterns of amino acid repeats encode a system of evolutionary and developmental markers, Genome Biol. Evol. 11 (2019) 3159–3178.
[5] N.G. Faux, S.P. Bottomley, A.M. Lesk, J.A. Irving, J.R. Morrison, M.G. de la Banda, et al., Functional insights from the distribution and role of homopeptide repeat-containing proteins, Genome Res. 15 (2005) 537–551.
[6] D. Karlin, F. Ferron, B. Canard, S. Longhi, Structural disorder and modular organization in Paramyxovirinae N and P, J. Gen. Virol. 84 (2003) 3239–3252.
[7] A.K. Dunker, M.S. Cortese, P. Romero, L.M. Iakoucheva, V.N. Uversky, Flexible nets. The roles of intrinsic disorder in protein interaction networks, FEBS J. 272 (2005) 5129–5148.
[8] F. Fiumara, L. Fioriti, E.R. Kandel, W.A. Hendrickson, Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins, Cell 143 (2010) 1121–1135.
[9] I. Pelassa, D. Corà, F. Cesano, F.J. Monje, P.G. Montarolo, F. Fiumara, Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction, Hum. Mol. Genet. 23 (2014) 3402–3420.
[10] H. Yu, J.K. Chen, S. Feng, D.C. Dalgarno, A.W. Brauer, S.L. Schreiber, Structural basis for the binding of proline-rich peptides to SH3 domains, Cell 76 (1994) 933–945.
[11] M.M. Albà, R. Guigó, Comparative analysis of amino acid repeats in rodents and humans, Genome Res. 14 (2004) 549–554.
[12] P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions, Trends Biochem. Sci. 33 (2008) 2–8.
[13] J.L. Weber, C. Wong, Mutation of human short tandem repeats, Hum. Mol. Genet. 2 (1993) 1123–1128.
[14] K. Anan, N. Yoshida, Y. Kataoka, M. Sato, H. Ichise, M. Nasu, et al., Morphological change caused by loss of the taxon-specific polyalanine tract in Hoxd-13, Mol. Biol. Evol. 24 (2007) 281–287.
[15] L. Abi-Rached, A. Gilles, T. Shiina, P. Pontarotti, H. Inoko, Evidence of en bloc duplication in vertebrate genomes, Nat. Genet. 31 (2002) 100–105.
[16] E. Salichs, A. Ledda, L. Mularoni, M.M. Alba, S. de la Luna, Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment, PLoS Genet. 5 (2009), e1000397.
[17] M. Hoque, T.M. Young, C.G. Lee, G. Serrero, M.B. Mathews, T. Pe'ery, The growth factor granulin interacts with cyclin T1 and modulates P-TEFb-dependent transcription, Mol. Cell. Biol. 23 (2003) 1688–1702.
[18] M. Alvarez, X. Estivill, S. de la Luna, DYRK1A accumulates in splicing speckles through a novel targeting signal and induces speckle disassembly, J. Cell Sci. 116 (2003) 3099–3107.
[19] C.H. Herrmann, M.A. Mancini, The Cdk9 and cyclin T subunits of TAK/P-TEFb localize to splicing factor-rich nuclear speckle regions, J. Cell Sci. 114 (2001) 1491–1503.
[20] H. Lu, D. Yu, A.S. Hansen, S. Ganguly, R. Liu, A. Heckert, et al., Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II, Nature 558 (2018) 318–323.
[21] N. Makki, M.R. Capecchi, Cardiovascular defects in a mouse model of HOXA1 syndrome, Hum. Mol. Genet. 21 (2012) 26–31.
[22] T. Lufkin, A. Dierich, M. LeMeur, M. Mark, P. Chambon, Disruption of the Hox-1.6 homeobox gene results in defects in a region corresponding to its rostral domain of expression, Cell 66 (1991) 1105–1119.
[23] M. Mark, T. Lufkin, J.L. Vonesch, E. Ruberte, J.C. Olivo, P. Dolle, et al., Two rhombomeres are altered in Hoxa-1 mutant mice, Development 119 (1993) 319–338.
[24] M. Roux, B. Laforest, M. Capecchi, N. Bertrand, S. Zaffran, Hoxb1 regulates proliferation and differentiation of second heart field progenitors in pharyngeal mesoderm and genetically interacts with Hoxa1 during cardiac outflow tract development, Dev. Biol. 406 (2015) 247–258.
[25] E.M. Carpenter, J.M. Goddard, O. Chisaka, N.R. Manley, M.R. Capecchi, Loss of Hox-A1 (Hox-1.6) function results in the reorganization of the murine hindbrain, Development 118 (1993) 1063–1075.
[26] M. Belpaire, B. Ewbank, A. Taminiau, L. Bridoux, N. Deneyer, D. Marchese, et al., HOXA1 is an antagonist of ERα in breast cancer, Frontiers in Oncology (2021) 11.
[27] B. De Kumar, H.J. Parker, A. Paulson, M.E. Parrish, J. Zeitlinger, R. Krumlauf, Hoxa1 targets signaling pathways during neural differentiation of ES cells and mouse embryogenesis, Dev. Biol. 432 (2017) 151–164.
[28] B. Lambert, J. Vandeputte, S. Remacle, I. Bergiers, N. Simonis, J.C. Twizere, et al., Protein interactions of the transcription factor Hoxa1, BMC Dev. Biol. 12 (2012) 29.
[29] B. Hudry, S. Remacle, M.C. Delfini, R. Rezsohazy, Y. Graba, S. Merabet, Hox proteins display a common and ancestral ability to diversify their interaction mode with the PBC class cofactors, PLoS Biol. 10 (2012), e1001351.
[30] S. Remacle, C. Shaw-Jackson, C. Matis, X. Lampe, J. Picard, R. Rezsöhazy, Changing homeodomain residues 2 and 3 of Hoxa1 alters its activity in a cell-type and enhancer dependent manner, Nucleic Acids Res. 30 (2002) 2663–2668.
[31] S. Remacle, L. Abbas, O. De Backer, N. Pacico, A. Gavalas, F. Gofflot, et al., Loss of function but no gain of function caused by amino acid substitutions in the hexapeptide of Hoxa1 in vivo, Mol. Cell. Biol. 24 (2004) 8567–8575.
[32] R.S. Mann, S.K. Chan, Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins, Trends Genet. 12 (1996) 258–262.
[33] R.S. Mann, M. Affolter, Hox proteins meet more partners, Curr. Opin. Genet. Dev. 8 (1998) 423–429.
[34] R.S. Mann, K.M. Lelli, R. Joshi, Hox specificity unique roles for cofactors and collaborators, Curr. Top. Dev. Biol. 88 (2009) 63–101.
[35] R.C. Paraguison, K. Higaki, Y. Sakamoto, O. Hashimoto, N. Miyake, H. Matsumoto, et al., Polyhistidine tract expansions in HOXA1 result in intranuclear aggregation and increased cell death, Biochem. Biophys. Res. Commun. 336 (2005) 1033–1039.
[36] G. Odelin, A. Faucherre, D. Marchese, A. Pinard, H. Jaouadi, S. Le Scouarnec, et al., Variations in the poly-histidine repeat motif of HOXA1 contribute to bicuspid aortic valve in mouse and zebrafish, Nat. Commun. 14 (2023) 1543.
[37] N. Makki, M.R. Capecchi, Hoxa1 lineage tracing indicates a direct role for Hoxa1 in the development of the inner ear, the heart, and the third rhombomere, Dev. Biol. 341 (2010) 499–509.
[38] N. Bertrand, M. Roux, L. Ryckebusch, K. Niederreither, P. Dolle, A. Moon, et al., Hox genes define distinct progenitor sub-domains within the second heart field, Dev. Biol. 353 (2011) 266–274.
[39] M. Roux, B. Laforest, N. Eudes, N. Bertrand, S. Stefanovic, S. Zaffran, Hoxa1 and Hoxb1 are required for pharyngeal arch artery development, Mech. Dev. 143 (2017) 1–8.
[40] A. Draime, L. Bridoux, M. Belpaire, T. Pringels, J. Tys, R. Rezsohazy, PRDM14, a putative histone methyl-transferase, interacts with and decreases the stability and activity of the HOXA1 transcription factor, Biochim Biophys Acta Gene Regul Mech. 1861 (2018) 534–542.
[41] A. Draime, L. Bridoux, M. Belpaire, T. Pringels, H. Degand, P. Morsomme, et al., The O-GlcNAc transferase OGT interacts with and post-translationally modifies the transcription factor HOXA1, FEBS Lett. 592 (2018) 1185–1201.

[42] A. Taminiau, A. Draime, J. Tys, B. Lambert, J. Vandeputte, N. Nguyen, et al., HOXA1 binds RBCK1/HOIL-1 and TRAF2 and modulates the TNF/NF-κB pathway in a transcription-independent manner, Nucleic Acids Res. 44 (2016) 7331–7349.
[43] X. Feng, P.A. Coulombe, A role for disulfide bonding in keratin intermediate filament organization and dynamics in skin keratinocytes, J. Cell Biol. 209 (2015) 59–72.
[44] W. Li, J. Zhang, J. Wang, W. Wang, Metal-coupled folding of Cys2His2 zinc-finger, J. Am. Chem. Soc. 130 (2008) 892–900.
[45] M.L. Straw, A.K. Chaplin, M.A. Hough, J. Paps, V.N. Bavro, M.T. Wilson, et al., A cytosolic copper storage protein provides a second level of copper tolerance in Streptomyces lividans, Metallomics 10 (2018) 180–193.
[46] A. Taminiau, A. Draime, J. Tys, B. Lambert, J. Vandeputte, N. Nguyen, et al., HOXA1 binds RBCK1/HOIL-1 and TRAF2 and modulates the TNF/NF-kappaB pathway in a transcription-independent manner, Nucleic Acids Res. 44 (2016) 7331–7349.
[47] A. Draime, L. Bridoux, M. Belpaire, T. Pringels, H. Degand, P. Morsomme, et al., The O-GlcNAc transferase OGT interacts with and post-translationally modifies the transcription factor HOXA1, FEBS Lett. 592 (2018) 1185–1201.
[48] A. Draime, L. Bridoux, M. Belpaire, T. Pringels, J. Tys, R. Rezsohazy, PRDM14, a putative histone methyl-transferase, interacts with and decreases the stability and activity of the HOXA1 transcription factor, Biochim. Biophys. Acta 1861 (2018) 534–542.
[49] N.M. Patel, S. Nozaki, N.H. Shortle, P. Bhat-Nakshatri, T.R. Newton, S. Rice, et al., Paclitaxel sensitivity of breast cancer cells with constitutively active NF-kappaB is enhanced by IkappaBalpha super-repressor and parthenolide, Oncogene 19 (2000) 4159–4169.
[50] Z. Niu, J. Fan, F. Chen, H. Yang, X. Li, T. Zhuang, et al., RBCK1 regulates the progression of ER-positive breast cancer through the HIF1α signaling, Cell Death Dis. 13 (2022) 1023.
[51] C.C. Hong, L.E. Sucheston-Campbell, S. Liu, Q. Hu, S. Yao, K.L. Lunetta, et al., Genetic variants in immune-related pathways and breast Cancer risk in African American women in the AMBER consortium, Cancer Epidemiol. Biomarkers Prev. 27 (2018) 321–330.
[52] M. Belpaire, A. Taminiau, D. Geerts, R. Rezsohazy, HOXA1, a breast cancer oncogene, Biochim. Biophys. Acta Rev. Cancer 1877 (2022), 188747.
[53] P. Tompa, E. Schad, A. Tantos, L. Kalmar, Intrinsically disordered proteins: emerging interaction specialists, Curr. Opin. Struct. Biol. 35 (2015) 49–59.
[54] G. Goudet, S. Delhalle, F. Biemar, J.A. Martial, B. Peers, Functional and cooperative interactions between the homeodomain PDX1, Pbx, and Prep1 factors on the somatostatin promoter, J. Biol. Chem. 274 (1999) 4067–4073.