Contrasting DNA-binding behaviour by ISL1 and LHX3 underpins differential gene targeting in neuronal cell specification

Ngaio C. Smith [a], Lorna E. Wilkinson-White [b], Ann H.Y. Kwan [a,c], Jill Trewhella [a], Jacqueline M. Matthews [a,*]

[a] School of Life and Environmental Sciences, University of Sydney, NSW 2006, Australia
[b] Sydney Analytical Core Research Facility, University of Sydney, NSW 2006, Australia
[c] The University of Sydney Nano Institute, University of Sydney, NSW 2006, Australia

ARTICLE INFO

ABSTRACT

The roles of ISL1 and LHX3 in the development of spinal motor neurons have been well established. Whereas LHX3 triggers differentiation into interneurons, the additional expression of ISL1 in developing neuronal cells is sufficient to redirect their developmental trajectory towards spinal motor neurons. However, the underlying mechanism of this action by these transcription factors is less well understood. Here, we used electrophoretic mobility shift assays (EMSAs) and surface plasmon resonance (SPR) to probe the different DNA-binding behaviours of these two proteins, both alone and in complexes mimicking those found in developing neurons, and found that ISL1 shows markedly different binding properties to LHX3. We used small angle X-ray scattering (SAXS) to structurally characterise DNA-bound species containing ISL1 and LHX3. Taken together, these results have allowed us to develop a model of how these two DNA-binding modules coordinate to regulate gene expression and direct development of spinal motor neurons.

1. Introduction

Interneurons provide the connection between sensory neurons, which are activated by sensory input from the environment, and motorneurons, which transmit signals from the spinal cord to skeletal and smooth muscles. During embryonic development, the differentiation of spinal motor neurons and interneurons, which arise from the same precursor cells, is controlled by the actions of transcriptional complexes that contain LIM-homeodomain transcription factors (Hunter and Rhodes, 2005; Srivastava et al., 2010). In those cells destined to become V2 interneurons, the LIM-homeodomain protein LIM homeobox protein 3 (LHX3) is expressed alongside the protein co-factor LIM domain binding protein 1 (LDB1) (Sharma et al., 1998; Zhadanov et al., 1995). The additional expression of the LIM-homeodomain protein Islet 1 (ISL1) in a neighbouring band of cells triggers a transcriptional switch that directs these cells to become spinal motor neurons (Lee et al., 2012; Pfaff et al., 1996). This switch of cell fate is achieved through altered DNA binding preferences of the complexes containing these transcription factors (Lee et al., 2008). The assembly of these different complexes occurs through protein LIM:LIM interaction domain (LID) interactions,

in combination with LDB1 self-association (Fig. 1) (Bhati et al., 2008b; Lee et al., 2008; Matthews and Visvader, 2003). While there is a plethora of structural information available on the individual domains present in this system (Berger et al., 2008; Bhati et al., 2008a, 2008b; Gadd et al., 2011; Ippel et al., 1999; Yin et al., 2017), the interplay of the homeodomains from LHX3 and ISL1 in determining the specificity of the motor neuron complex has not been fully elucidated.

Electrophoretic mobility shift assay (EMSA) and systematic evolution of ligands by exponential enrichment (SELEX) studies have shown that LHX3$_{HD}$ has a strong preference for binding TAATTA sites on DNA (Bridwell et al., 2001; Lee et al., 2008; Roberson et al., 1994). This sequence is very similar to parts of the consensus sequence that has been determined for the motor neuron complex, CATTAXXXAATTA (Lee et al., 2008). Two separate consensus sequences have been proposed for ISL1$_{HD}$, TAATAT from EMSA studies and CATTAG from SELEX (Behravan et al., 1997; Lee et al., 2008). It has also been suggested that ISL1$_{HD}$ may bind DNA sequences in a manner that is dependent on its protein binding partners (Mazzoni et al., 2013).

It was previously shown that fusion proteins containing both the ISL1 and LHX3 homeodomains bind to a sequence found in the promoters of

motorneuron specific genes that is not bound strongly by either individual homeodomain (Lee et al., 2008, 2012; Robertson et al., 2018), suggesting that they collaborate in some manner – perhaps through direct interaction or cooperative binding to DNA. However, it has been demonstrated that the isolated homeodomains do not cooperatively bind DNA containing the cognate binding site (Robertson et al., 2018). Thus, the mechanism by which ternary ISL1-LHX3-DNA complex is formed remains unclear.

Here, we further investigated the DNA-binding behaviour of the homeodomains from ISL1 and LHX3, both individually and in combination, in order to refine our knowledge of their DNA-binding behaviour, including how it is modulated by the presence of additional DNA-binding motifs. We also probed the structures of LHX3 and ISL1 in complex with DNA and with each other. Our results indicate that ISL1 does not bind DNA with high affinity in isolation. Rather, it acts to influence the specificity of LHX3, potentially through transient interactions with DNA made possible by proximity effects from LHX3 binding to DNA, shedding new light on how ISL1 can influence gene expression without itself being able to bind DNA with high affinity.

2. Materials and methods

2.1. Protein expression and purification

Uniprot codes for the mouse protein sequences used are: LHX3 - P50481 (isoform 1), and ISL1- P61372 (isoform 1). Protein encoding constructs are detailed in the Supplemental Information and were expressed from a modified pET-DUET plasmid that encodes an N-terminal GST tag with an HRV 3C protease cleavage site between the tag and the protein. All proteins were expressed using BL21 (DE3) *E. coli* strains, using induction by 0.4 mM IPTG. GST-tagged ISL1$_{HD}$, LHX3$_{HD}$, and 2HD proteins were each expressed for 3 h at 37 °C. For all other constructs, proteins were expressed for 20 h at 25 °C. Proteins were purified using glutathione affinity chromatography followed by HRV-3C cleavage and cation exchange chromatography, as described in (Robertson et al., 2018).

2.2. EMSAs

Oligonucleotides were designed to mimic *in vivo* promoters that bind either ISL1$_{HD}$ or LHX3$_{HD}$ alone, or both homeodomains simultaneously (Table 1). Varying concentrations of protein (≤5 μM) were incubated with a fixed concentration of fluorescein-labelled oligonucleotide (1 nM M100; 5 nM ISL1GA/LHX3GSU/HDC). Samples were incubated in EMSA reaction buffer (10 mM HEPES, pH 7.7, 50 mM KCl, 2.5 mM MgCl$_2$, 10 mM DTT, 67 μg/mL acetylated BSA, 4% (v/v) Ficoll) for 45 min at 4 °C and run on 8% (w/v) acrylamide gels at 110 V for 3 h at room temperature in TBE buffer (2.5 mM Tris (pH 7.8), 4.5 mM boric acid, 0.13 mM EDTA). Gels were imaged using a Typhoon FLA 9500 scanner (GE Healthcare).

2.3. Surface plasmon resonance (SPR)

SPR was performed on a Biacore T200. Biotinylated versions of the oligonucleotides used for EMSAs were synthesised by IDT (Integrated DNA Technologies; Coralville, IA). Oligonucleotides (4 nM) were immobilised at 25 °C on a Series S SA Biacore Sensor Chip (GE Healthcare; Chicago, IL) to a level that would achieve a maximal

Table 1
Sequences used for EMSA binding studies. A fluorescein moiety was present at the 5′ end of each sequence. All oligonucleotides were double-stranded. Putative homeodomain binding sites are shown in bold.

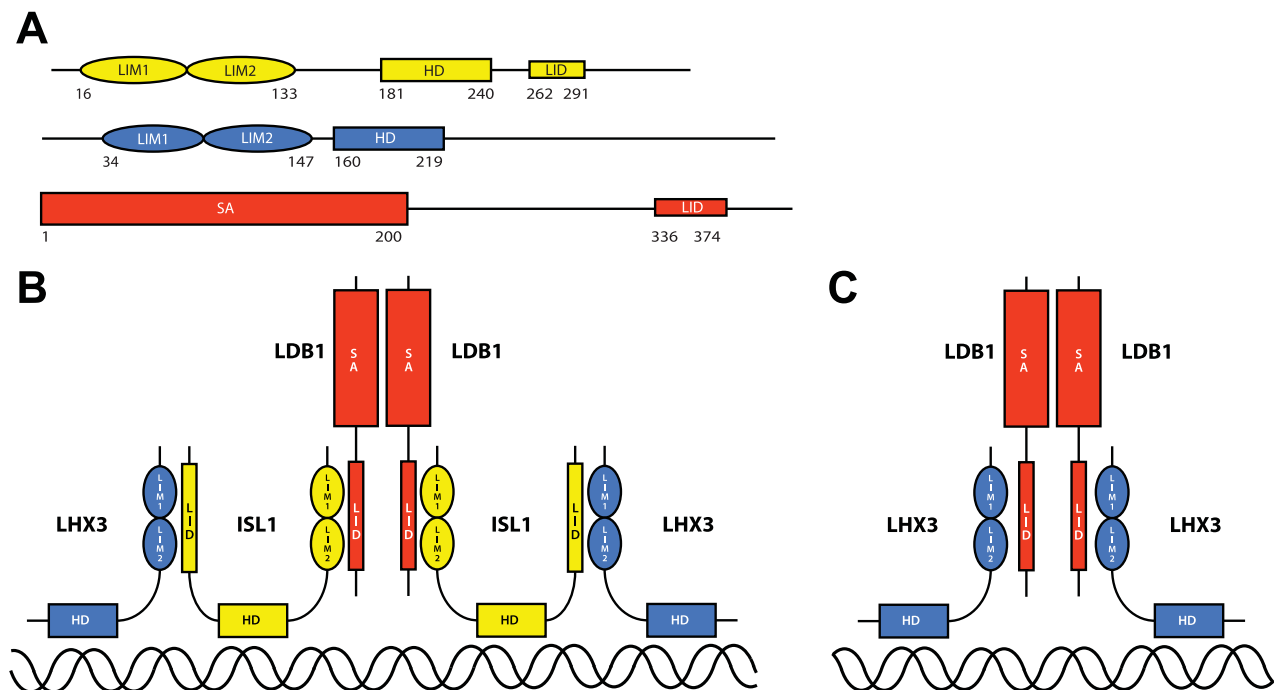| Name	Oligonucleotide sequence	Binding target
ISL1GA	ACCGCG**TAATAT**CTG	ISL1 (Behravan et al., 1997)
LHX3GSU	ACTTAGC**TAATTA**AATGTG	LHX3 (Meier et al., 1999; Roberson et al., 1994)
M100	CGGC**CATTAGCCAAATTA**CGGC	ISL1/LHX3 in complex (Lee et al., 2008)
HDC	CACGTGCCGTCAGCGGTAC	Negative control

Fig. 1. Schematic of the ISL1/LHX3/LDB1 DNA-binding system. **A** Domain structure of ISL1 (yellow), LHX3 (blue), and LDB1 (red), showing LIM (derived from family members Lin11, Isl1, Mec3) domains, homeodomains (HDs), LIM interaction domains (LIDs), and self-association domains (SA). **B** Schematic of the motor neuron developmental complex. **C** Schematic of the interneuron developmental complex. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

response of 80 RU upon titration of the binding partner. Purified homeodomain proteins (0.03 nM – 1 μM in a 1:2 serial dilution) in running buffer (20 mM HEPES, pH 7.4, 150 mM or 250 mM KCl) were injected for 75 sec at a flow rate of 60 μL/min, at 20 °C, over the immobilised oligonucleotides. 1 M NaCl was injected for 60 s between each protein injection to regenerate the surface back to baseline.

Interaction affinities were fitted by a 1:1 specific binding model (Graphpad Prism 8.2.0), using the response intensities at equilibrium. Interaction kinetics were estimated using global fits of association and dissociation (Biacore Insight Evaluation Software).

2.4. Size exclusion chromatography with multi angle laser light scattering (SEC-MALLS)

Protein (2HDLL) and DNA (M100) were dialysed into an appropriate buffer (20 mM sodium phosphate, pH 7.4, 100 mM NaCl, 1 mM DTT). The protein and DNA were then combined in a 1:1 M ratio before being injected onto a Superose 12 10/300 GL column (GE Healthcare) on an Äkta Basic (GE Healthcare) system at 0.5 mL/min in the same buffer. An in-line MiniDAWN™ TREOS multi angle light scattering detector (Wyatt Technology) coupled with an Optilab T-rEX differential refractive index detector (Wyatt Technology) was used to collect data, which was analysed using ASTRA 6.1 software (Wyatt Technology).
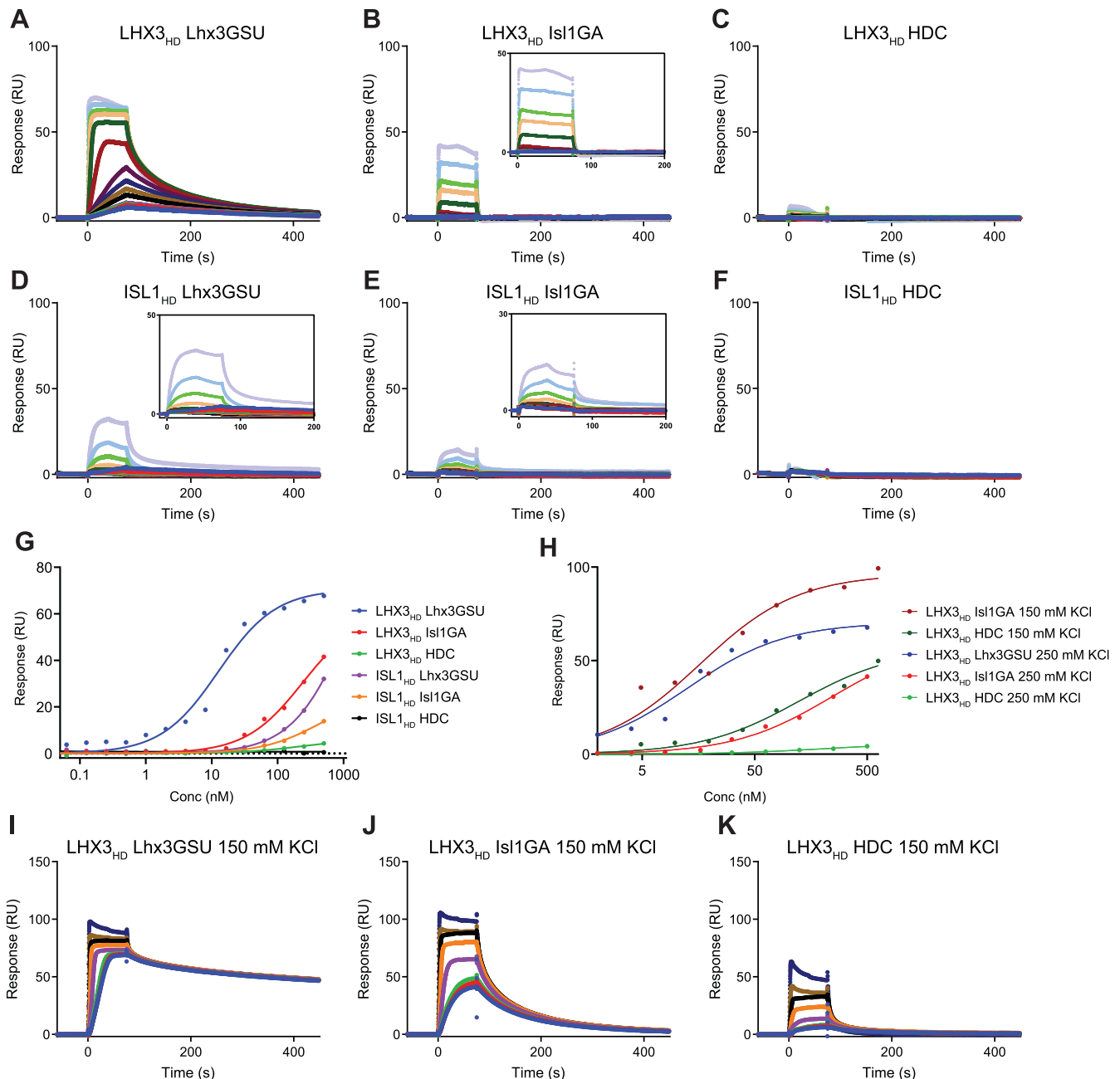

Fig. 2. SPR binding data for LHX3$_{HD}$ and ISL1$_{HD}$. **A-F** Titrations of LHX3$_{HD}$ (A-C, G-I) and ISL1$_{HD}$ (D-F) against the oligonucleotides LHX3GSU (A, D, G), ISL1GA (B, E, H), and HDC (C, F, I), at 250 mM KCl (A-F). **G** Equilibrium binding data for LHX3$_{HD}$ and ISL1$_{HD}$ against the oligonucleotides LHX3GSU, ISL1GA, and HDC, at 250 mM KCl. **H** Comparison of equilibrium binding data between 150 mM KCl and 250 mM KCl for LHX3$_{HD}$ binding. 0.03 nM-500 nM protein was used for experiments in 250 mM KCl; 1.75 nM-625 nM protein was used for experiments in 150 mM KCl. **I-K** Titrations of LHX3$_{HD}$ against the oligonucleotides LHX3GSU (I), ISL1GA (J), and HDC (K), at 150 mM KCl.

2.5. Small angle X-ray scattering (SAXS)

Purified proteins were dialysed against buffer (20 mM sodium phosphate, pH 7.4, 100 mM NaCl, 1 mM DTT), filtered and degassed prior to data collection using the SAXS beamline at the Australian Synchrotron. Four different concentrations in a 1:2 dilution series of each sample were used (see Supplementary information for more detail). Buffer subtractions were automatically performed as part of data processing at the Australian Synchrotron. Data quality was assessed using PrimusQT, and calculating molecular weight from $I(0)$ ((Supplementary information). Modelling was performed using the ATSAS web interface (available at: https://www.embl-hamburg.de/biosaxs/atsas-online/). Full experimental and analytical details, as well as details about modelling parameters and programs, are reported in the Supplementary Information (Tables S1-6 and Figs. S1 and S2) in line with the 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution (Trewhella et al., 2017). SAXS datasets are available at the SASBDB, accession codes: SASDJB9, SASDJC9, SASDJD9, SASDJE9, SASDJF9, SASDJG9, SASDJH9, SASDJJ9, SASDJK9.

3. Results

3.1. Homeodomains from ISL1 and LHX3 have markedly different DNA binding properties

EMSAs previously used to investigate the binding of LHX3$_{HD}$ and ISL1$_{HD}$ to cognate DNA sequences suggested that LHX3$_{HD}$ bound with high affinity to TAATTA sequences but ISL1$_{HD}$ did not bind tightly to any sequence tested (Robertson et al., 2018). Here SPR was used to further probe this binding behaviour with a view to identifying differences in the kinetics of binding. The binding of ISL1$_{HD}$ and LHX3$_{HD}$ against immobilised ISL1GA, LHX3GSU, and HDC was investigated.

LHX3$_{HD}$ showed a strong binding response to LHX3GSU (Fig. 2A), with a reduced binding response to ISL1GA (Fig. 2B) and very low responses to HDC (Fig. 2C). Fitting of the responses at equilibrium gave an estimated affinity of 20 nM for the LHX3$_{HD}$:LHX3GSU interaction, in good agreement with previously published EMSA data (Fig. 2G) (Robertson et al., 2018). The affinities of LHX3$_{HD}$ for ISL1GA and for HDC could not be measured, as a complete binding curve could not be obtained over the assayable concentration range (Fig. 2G). However, these interactions appeared to be at least ten-fold weaker than the LHX3$_{HD}$: LHX3GSU interaction.

The sensorgrams indicated that the LHX3$_{HD}$:LHX3GSU interaction has a fast on-rate and a slow off-rate (Fig. 2A), although these could not be fitted for kinetic quantification as the association rates were outside the limits that could be fitted by the instrument software. In agreement with the observed weaker responses, the LHX3$_{HD}$:ISL1GA interaction demonstrated fast on- and off-rates (Fig. 2B). Again, fits for kinetic measurements could not be obtained.

Overall, the observed responses were stronger when the concentration of KCl was reduced from 250 mM KCl to 150 mM (Fig. 2H, I-K). The interaction between LHX3$_{HD}$ and LHX3GSU at 150 mM KCl demonstrated significantly slower off-rates (Fig. 2A and I). Similarly, the LHX3$_{HD}$: ISL1GA interaction had an observable off-rate at 150 mM KCl, which was not present at 250 mM KCl (Fig. 2B and J). A response was also observed for LHX3$_{HD}$:HDC, whereas barely any interaction was observed at 250 mM KCl (Fig. 2C and K). Based on the combination of their observed off-rates and their salt-dependent binding behaviour, the interactions between LHX3$_{HD}$ and the oligonucleotides ISL1GA and HDC appear to be non-specific and electrostatically driven (Lohman and Von Hippel, 1986). In contrast, while the LHX3$_{HD}$:GSU interaction showed differing kinetic properties between 150 mM KCl and 250 mM KCl, the affinity remained similar (13 nM in 250 mM KCl vs 5 nM in 150 mM KCl).

Over the protein concentrations tested, ISL1$_{HD}$ did not show high levels of binding to any oligonucleotide (Fig. 2D-F). It was not possible to increase protein concentrations beyond 500 nM, or use KCl concentrations lower than 250 mM, as the protein showed a high level of background binding to the reference cell.

3.2. Exploring the influence of the ISL1:LHX3 interaction on DNA-binding

The protein:protein interaction between ISL1 and LHX3 involves regions adjacent to the two homeodomains that could influence the DNA-binding behaviour of the proteins. The possible contribution to DNA-binding of residues outside of the homeodomains was investigated through EMSAs using fusion constructs incorporating different protein domains (Fig. 3A). 2HDLL contains native ISL1 sequence (from the N-terminus of the homeodomain to the C-terminus of the LID) fused, via a short glycine/serine linker, to native LHX3 sequence (from the N-terminus of the LIM domains to the C-terminus of the homeodomain). 2HD was designed to test the effect of homeodomain-adjacent residues that were not part of the LHX3$_{LIM}$:ISL1$_{LID}$ subcomplex, and incorporated the two homeodomains and immediately adjacent residues but excluded the ISL1$_{LID}$ and LHX3$_{LIM}$ domains. 2HD23 is identical, except that a 23-residue portion of native sequence was replaced by a glycine/serine linker. LLHD3 is a shorter form of 2HDLL, beginning with ISL1$_{LID}$, and was designed to probe the effect of the protein interaction domains on LHX3$_{HD}$ DNA binding.

The binding of these constructs was tested against the LHX3GSU sequence and the sequence targeted by the ISL1:LHX3 complex, M100 (Fig. 3). LLHD3 bound strongly to LHX3GSU, but not M100 (Fig. 3B). This behaviour is very similar to that of LHX3$_{HD}$ alone (Fig. 3C), as previously reported (Robertson et al., 2018). These similarities indicate that the presence of the LHX3$_{LIM}$:ISL1$_{LID}$ subcomplex does not directly influence the binding specificity of LHX3$_{HD}$. 2HD23 and 2HD bound equally strongly to M100 (Fig. 3D, Table 2). This pattern of binding suggests that that the sequence immediately adjacent to the homeodomains does not directly influence the DNA binding of the tethered complex.

Combined, these data are consistent with previously observed binding behaviour of LHX3$_{HD}$ and ISL1$_{HD}$ (Robertson et al., 2018). That is, that LHX3$_{HD}$ binds with high affinity (~20 nM) and specificity to a TAATTA sequence but binds more weakly to other sequences, with affinities in the micromolar range. In comparison, ISL1$_{HD}$ exhibits micromolar affinity binding to a range of DNA sequences, with no apparent high affinity target sequence among those tested. Additionally, the data presented here indicate that, provided the homeodomains are tethered through protein:protein interactions or by a physical linker, it is the interplay of the two homeodomains that influences the binding preferences of the overall complex.

3.3. Structural investigations of the 2HDLL + M100 complex

Multiple attempts were made to determine the structures of DNA-bound ISL1:LHX3 complexes. A 2HDLL + M100 complex was stable in solution and shown by SEC-MALLS to form a 1:1 complex (Fig. 4A). The average molecular weight of the complex was measured as 50.5 kDa, in good agreement with the theoretical molecular weight of 49.4 kDa. Attempts to crystallise the protein:DNA complex have not yielded diffracting crystals to date. Using small-angle X-ray scattering (SAXS), acquired in both batch mode (to optimise signal-to-noise given the higher protein concentrations achievable) and with in-line size exclusion chromatography (SEC-SAXS; to optimise sample homogeneity), we have probed the structure of the complex in solution.

SEC-SAXS data revealed that the 2HDLL + M100 complex is homogeneous with a radius of gyration (R_g) of 38–42 Å (Fig. 4B). For reference, the M100 duplex should be approximately 65 Å long, with the LIM: LID interaction region also being roughly 65 Å long in its longest dimension. A sphere with a radius of 38–42 Å should be able to fit all the components expected in the 2HDLL + M100 complex. Static SAXS data
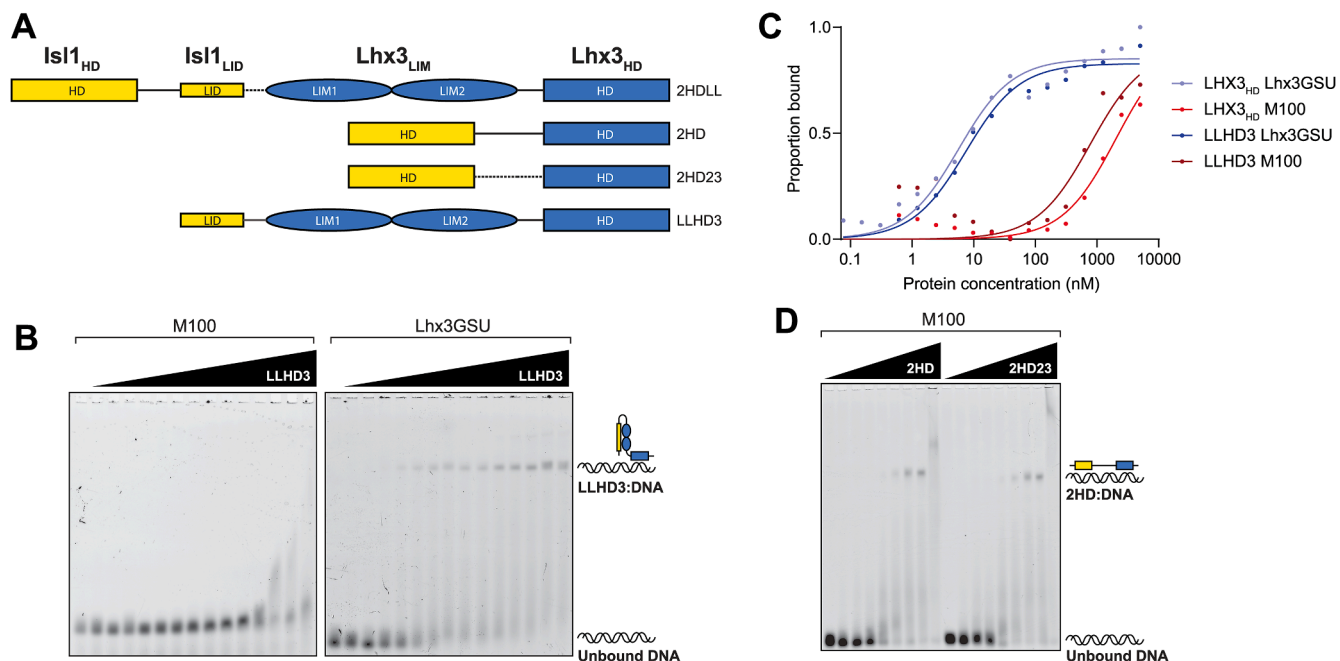
Fig. 3. Behaviour of LHX3 and ISL1 protein constructs in EMSAs. Binding measurements for ISL1$_{HD}$, LHX3$_{HD}$, 2HDLL, and 2HD taken from (Robertson et al., 2018). **A** Schematic of protein constructs used in EMSAs. **B** Examples of EMSA titrations of LLHD3 (0, 0.08–5000 nM LLHD3) against the M100 and LHX3GSU oligonucleotides. **C** Binding affinity curves for LHX3$_{HD}$ and LLHD3 against M100 and LHX3GSU. **D** Examples of EMSA titration of 2HD and 2HD23 (0, 39–5000 nM protein) against M100.

Table 2

Dissociation constants (M) of homeodomain-DNA interactions, ± 1 SD. K$_d$s were determined through densitometry analysis of EMSAs, where binding was observed. N/A denotes no binding observed. * denotes n = 2; # denotes n = 1. All other measurements are n = 3. Figures reported for ISL1$_{HD}$ (HD1), LHX3$_{HD}$ (HD3), 2HDLL, and 2HD taken from previously published data (Robertson et al., 2018).

	Oligonucleotide		
Protein	M100 CATTAGCCAATTA	ISL1GA TAATAT	LHX3GSU TAATTA
HD1	$3 \pm 2 \times 10^{-6}$*	$5.9 \pm 0.3 \times 10^{-6}$	$2 \pm 0.9 \times 10^{-6}$
HD3	$4 \pm 3 \times 10^{-6}$*	$2 \pm 1 \times 10^{-6}$*	$9 \pm 3 \times 10^{-9}$
2HDLL	$2.4 \pm 0.3 \times 10^{-8}$	$7 \pm 5 \times 10^{-7}$*	$4 \pm 2 \times 10^{-6}$
2HD	$1.6 \pm 0.4 \times 10^{-7}$	$5 \pm 2 \times 10^{-7}$*	$4.4 \pm 0.9 \times 10^{-7}$
2HD23	$1.8 \pm 2.5 \times 10^{-7}$	N/A	$4.4 \pm 5 \times 10^{-7}$
LLHD1	N/A	N/A	N/A
LLHD3	1×10^{-6}#	N/A	1.5×10^{-8}#

for detailed structural analysis were collected for 2HD, LLHD3, HD3 and 2HDLL, as well as HD3 + M100, LLHD3 + M100, 2HD + M100, and 2HDLL + M100 complexes, at a range of concentrations. (See Supplementary Information for full details).

P(*r*) analysis using GNOM as implemented in PrimusQT yielded the radius of gyration (R_g) and maximum dimension (d_{max}) values of the observed species (Table 3) indicating that the single domain HD3 and 20-mer DNA (M100) are similarly compact, while the 2-, 4- and 5-domain constructs 2HD, LLHD3, and 2HDLL, respectively, are more extended with each approximately double the maximum dimension of HD3. M100 binding results in somewhat long values for d_{max}, except for 2HD where we interpret that as a contraction.

The flexibility of each species was assessed using dimensionless Kratky plots (Fig. 4C and D, also Fig. S3) (Bernadó and Svergun, 2012; Semisotnov et al., 1996; Trewhella et al., 2017). A well-ordered, globular structure will give a dimensionless Kratky plot that is bell-shaped with a peak at $qR_g = \sqrt{3}$, while for increasingly elongated structures the peak will become increasingly flattened, extending to higher qR_g values. In the case of multidomain structures the plot may show some

additional features such as a secondary peak or shoulders. Plots that do not show a peak, but instead display a continuous increasing intensity with increasing qR_g, indicate disorder. A peak followed by a steady increasing trend at higher qR_g values indicates a combination of ordered and disordered structure - as might be expected for multi-domain proteins connected with flexible linkers.

The dimensionless Kratky plot for M100 shows the expected shape for a somewhat elongated, mostly ordered species consistent with a 20-mer DNA in solution. The plot for the 1-domain HD3 construct rises approximately to where you expect a peak for a compact folded domain but, as is the case for all the protein constructs, the low-signal-to-noise at high qR_g prohibits definitive assessment of relative flexibility from the protein data alone. However, a comparison of the plots for the multi-domain protein constructs alone and in complex with M100 is revealing. The shape of the Kratky plot for the 4-domain LLHD3 shows no discernible change upon M100 binding. The 5-domain 2HDLL, which also shows some evidence for a secondary peak in the protein only plot, shows a substantial sharpening of the peak on M100 binding accompanied by a possible reduction in intensity at high qR_g that suggests reduction in flexibility upon complex formation. The most dramatic changes are observed for the 2-domain 2HD. Binding to M100 for this domain results in a peak shift and change in shape with a significant reduction in intensity at high qR_g indicating substantially reduced flexibility and decreased elongation, consistent with the observed reduction in d_{max}. Overall, the Kratky analyses suggest that the ISL1$_{LID}$:LHX3$_{LIM}$ region, which is present in both LLHD3 and 2HDLL but absent in 2HDN, is not directly involved in DNA binding.

One dataset of each species was chosen for *ab initio* shape modelling (See Supplementary Information for modelling parameters and programs used, Table S5) (Franke and Svergun, 2009; Petoukhov and Svergun, 2015; Svergun, 1999a, 1999b). AMBIMETER modelling indicates a moderate-to-high degree of ambiguity in all of the datasets, consistent with a flexible DNA-bound ISL1:LHX3 complex. Given this ambiguity, the number of variables resulting from the flexible linkers between domains and the information content of the scattering data, MONSA modelling (Table S6), which allows different parts of a complex
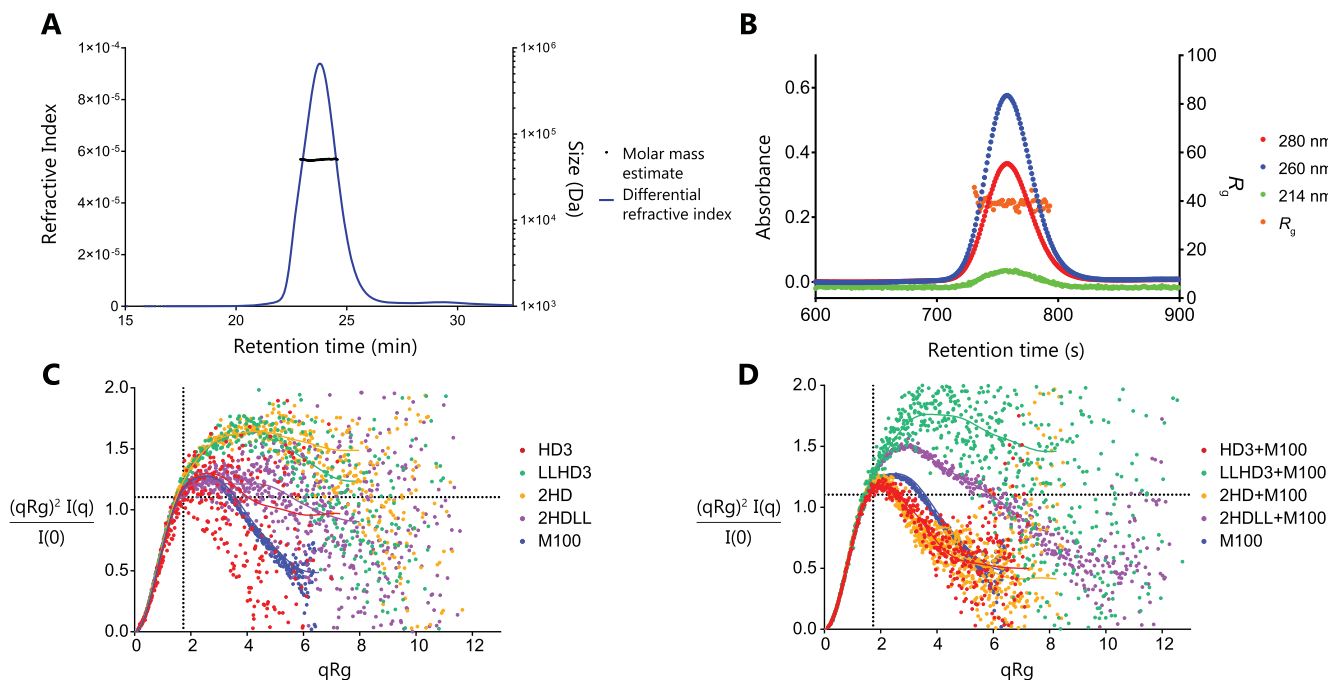
Fig. 4. Structural characterisation of LIM-HD protein constructs and their DNA binding behaviour. **A** MALLS trace of 2HDLL + M100. **B** SEC-SAXS trace of 2HDLL + M100. **C** Dimensionless Kratky plots of HD3 (0.17 mg/mL), LLHD3 (0.33 mg/mL), 2HD (0.63 mg/mL), 2HDLL (0.52 mg/mL) and M100 (3 mg/mL) datasets. **D** Dimensionless Kratky plots of HD3 + M100 (0.2 mg/mL), LLHD3 + M100 (0.48 mg/mL), 2HD + M100 (1 mg/mL), 2HDLL + M100 (0.69 mg/mL), and M100 (3 mg/mL). For **C** and **D** the Solid curves are the GNOM fits to the data and dotted lines indicate $qR_g = \sqrt{3}$.

to be modelled (i.e., protein and DNA) was used to gain insight into the arrangement of domains in the overall complex. Each data set was separately modelled several (6) times. The d_{max} determined from GNOM in PrimusQT was used to constrain modelling but the R_g value was not fixed in the input. The output R_g and the difference between the GNOM and MONSA derived R_g values were used as indicators of variability between generated models (Table 3).

Irrespective of the protein present, modelling of M100 consistently produced an elongated shape consistent with that expected for a 20-mer DNA, with R_g values between 19.72 Å and 20.23 Å in close agreement with the calculated value of 19.84 Å from $P(r)$ analysis and low R_g difference values (<0.25 Å). (Fig. 5A, See also Supplementary information Figs. S4 and S5).

Unsurprisingly, more variation was seen in the modelled R_g values of most of the protein constructs (Table 3; difference R_gs 0.2–1.01 Å). HD3 and M100 reported similar R_g values. HD3 reported a similar R_g to M100 based on $P(r)$ analysis (GNOM), as expected for a single, folded, homeodomain, but models generated for HD3 were seen to have a smaller R_g, with the average being 1 Å less than the R_g from GNOM. Unlike the models of M100, the models for HD3 do not share any defining features (Fig. S4). This might be explained by the flexible nature of the homeodomain fold when not bound to DNA (Dragan et al., 2006).

The $P(r)$-derived R_g and d_{max} values for the multi-domain 2HD, LLHD3, and 2HDLL were all similar, despite these proteins having a range of molecular weights (Table 3), consistent with individual folded domains with flexible linkers Modelling of these constructs gave R_g values that are in good agreement with the $P(r)$ analysis (Table 3). 2HD showed the largest difference, with the average modelled R_g being 0.5 Å smaller than that obtained from GNOM. The models themselves show an array of potential conformations for each protein, with no distinct features beyond being a more elongated species than HD3 or M100 (Fig. S4). These models probably reflect the disordered sequence between the domains, as it is well established that the LHX3$_{LIM}$:ISL1$_{LID}$ subcomplex remains folded and associated in solution; circular dichroism also shows that the individual homeodomains are primarily folded in solution (Fig. S5) (Gadd et al., 2011).

As these models did not show any clear structural trends, the protein alone datasets were not used as inputs for modelling larger complexes. For modelling protein + DNA complexes, both DNA alone and the protein + DNA datasets were used as inputs, allowing specification of both the protein and DNA component of each complex. Note that the MONSA R_g values for these data (Table 3) are those of the specified components (i.e., one protein and one dsDNA species), compared with the overall complex R_g reported by GNOM. The difference R_gs tend to be larger for the protein components (0.51–12.27 Å) than the DNA components (0.06–0.27 Å).

With the exception of 2HDLL, the R_g values of all protein constructs reported through MONSA tended to be smaller in the protein + DNA models than in the protein alone models (Table 3). This difference was particularly large for 2HD, supporting the notion that conformational restriction of the homeodomains occurs upon DNA binding, with the protein:protein interaction regions remaining unconstrained and flexible in solution.
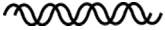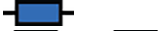
The models for HD3 + M100 show the protein localised to one end of the oligonucleotide (Fig. 5B), most likely bound to the AAATTA sequence. Models of LLHD3 + M100 show more variability (Fig. 5C), but a substantial amount of protein density, not associated with the DNA, is common to all of the models generated (Fig. S4). This is consistent with LHX3$_{HD}$ binding the AAATTA site, with the remainder of the protein (the LHX3$_{LIM}$:ISL1$_{LID}$ subcomplex) not associating with DNA.

Models of 2HD + M100 also show the majority of the protein localised to one end of the DNA (Fig. 5D). However, the positions of the respective homeodomains vary from model to model. It is notable that 2HD showed the largest change in R_g when comparing the protein alone and protein + DNA datasets, with a difference of 12.27 Å. This, in combination with the Kratky analysis, is consistent with restricted conformation and loss of flexibility in the presence of M100, suggesting that both homeodomains interact with the DNA.

There is much less consistency in the models generated for 2HDLL + M100 (Fig. 5E, Fig. S4). Some show binding along the whole length of M100, whereas some show one end of the DNA fragment exposed (Fig. S4). All models show some portion of protein not in association

Table 3

Radii of gyration (R_g) for species analysed by SAXS. This includes both values calculated from data analysis in PrimusQT (GNOM) and those generated through iterations of *ab initio* MONSA modelling (n = 6). The reported difference is the difference between the GNOM R_g and the average modelled MONSA R_g. *The difference in values between (A) & (B) for complex samples is per component, comparing to the GNOM R_g for the individual component samples (reported in the first 5 rows of the table).

Species	M_w (kDa) based on sequence	P(r) derived parameters (using GNOM)		Ab initio shape modelling results (using MONSA)			
		d_{max}	R_g (Å) (A)	Average modelled R_g (Å) (B)	Error (SD)	Error (SE)	Difference between values in (A) & (B) (Å)
M100	12.2 kDa	68	19.84	M100: 19.74	0.01	0.01	0.1
HD3	9.6 kDa	70	18.76	HD3: 17.75	0.08	0.19	1.01
2HD	21 kDa	127	34.01	2HD: 33.51	0.05	0.11	0.5
LLHD3	27.4 kDa	120	33.15	LLHD3: 32.95	0.02	0.06	0.2
2HDLL	37.2 kDa	130	35.3	2HDLL: 35.22	0.07	0.18	0.08
HD3 + M100	21.8 kDa	70	20.94	HD3: 15.22	0.58	1.42	3.54*
				M100: 19.79	0.00	0.01	0.06*
2HD + M100	33.2 kDa	75	25.12	2HD: 21.75	1.14	2.80	12.27*
				M100: 20.07	0.01	0.23	0.23*
LLHD3 + M100	39.6 kDa	130	38.09	LLHD3: 37.58	1.68	4.12	0.51*
				M100: 20.09	0.03	0.07	0.25*
2HDLL + M100	49.4 kDa	140	35.97	2HDLL: 37.66	1.61	3.94	2.36*
				M100: 20.11	0.04	0.09	0.27*

with the DNA, but the extent of this varies, as reflected in the R_g values for 2HDLL from the 2HDLL + M100 models which show the largest error of any modelled component. However, for all 2HDLL:M100 models the protein density close to the DNA indicates more than one homeodomain interacting with the DNA. Given the evidence that the LHX3$_{LIM}$:ISL1$_{LID}$ subcomplex does not interact directly with DNA, this additional interacting density is likely due to ISL1$_{HD}$ interacting with the DNA, albeit transiently and in a manner that cannot be fully resolved using this technique.

4. Discussion

The data presented here provides new insight into the role of ISL1 in the regulation of gene expression. It is evident that ISL1$_{HD}$ in isolation does not bind DNA strongly, whereas LHX3 binds strongly and specifically to TAATTA sequences in DNA. Intriguingly, this specificity is influenced by the presence of ISL1. Previous data indicates that there is no cooperative binding between the two homeodomains, raising the question of how this change in specificity is achieved (Robertson et al., 2018).

The sequence of the LHX3GSU contains several features that may enhance the binding of LHX3$_{HD}$. Firstly, the six base consensus sequence of LHX3 (TAATTA) is palindromic. This allows LHX3$_{HD}$ to bind to the

oligonucleotide regardless of orientation, and could affect the overall kinetics of binding. Additionally, the sequence contains an additional AAT immediately downstream of the TAATTA binding site. This AT-rich region may again enhance the on-rate for binding. While EMSA data does not indicate that multiple copies of LHX3$_{HD}$ can bind simultaneously to the LHX3GSU oligonucleotide, it is still possible that transient interactions with adjacent sequence helps to maintain the occupancy of the LHX3 consensus sequence, as has been seen in other systems (Iwahara et al., 2004).

Efforts to determine the structure of the LHX3/ISL1/M100 complex through X-ray crystallography have been unsuccessful, consistent with the flexibility of the complex as indicated by SAXS data. Although the 2HDLL:M100 SAXS data presented here do not provide a high-resolution structure that explains the molecular basis of specificity, the combined data presented here contribute to a model of the binding mechanism (Fig. 6).

First, EMSA data indicates that neither the protein sequences adjacent to the homeodomains of ISL1 and LHX3 nor the LHX3$_{LIM}$:ISL1$_{LID}$ subcomplex play a direct role in contributing to DNA binding of the ISL1:LHX3 transcriptional complex. The SAXS data supports this, with 2HDLL:M100 models showing a large portion of protein not in association with the DNA, in comparison with the 2HD:M100 models that consistently show tight association between protein and DNA. These
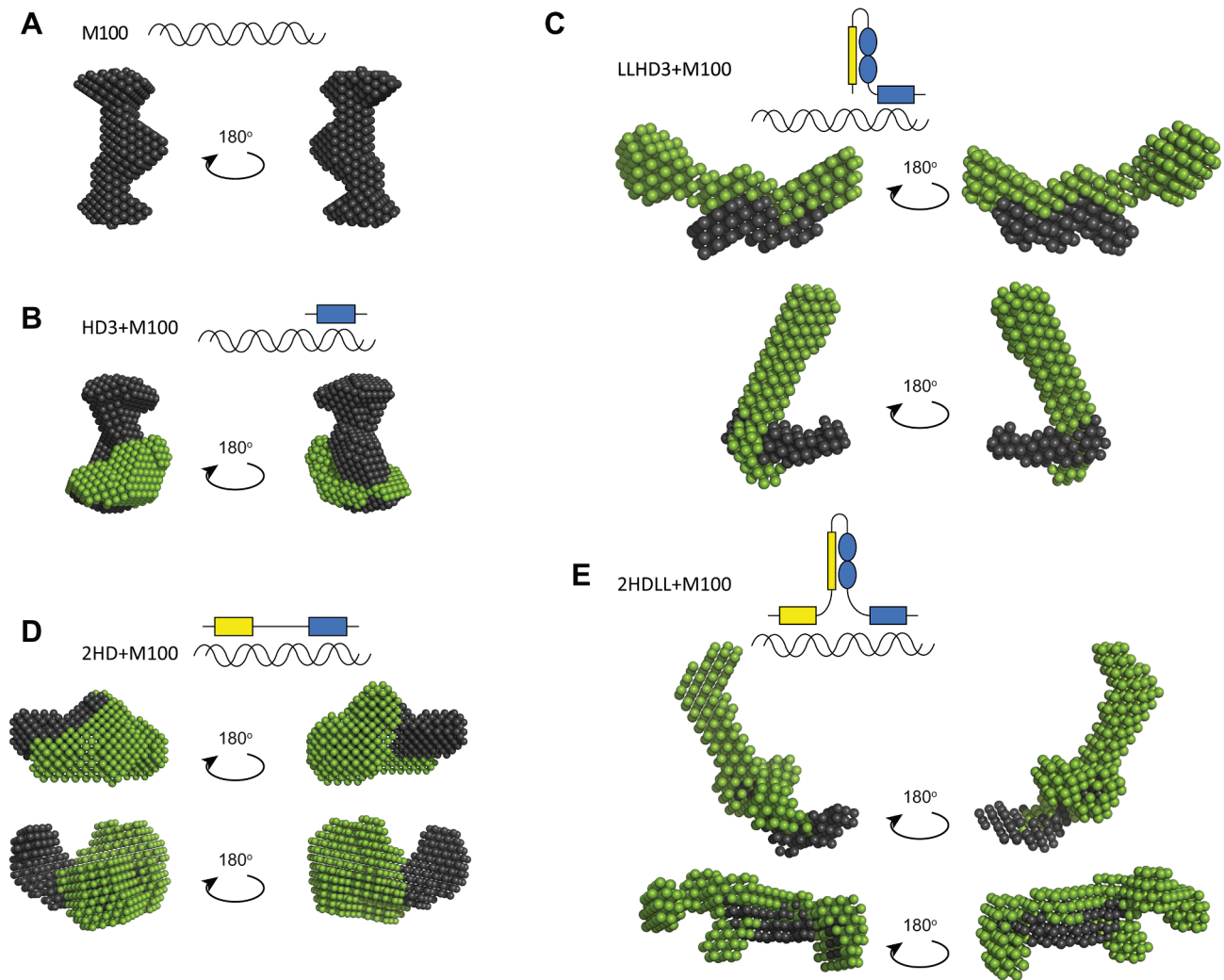
Fig. 5. Representative structures of LIM-HD:DNA complexes generated through MONSA *ab initio* modelling. Yellow: ISL1; Blue: LHX3; Green: protein; Grey: DNA. **A** M100 oligonucleotide. **B** HD3 + M100 complex. **C** Two iterations of LLHD3 + M100 modelling, showing the variation in binding along the length of the oligonucleotide. **D** Two iterations of 2HD + M100 modelling, showing the variation in binding along the length of the oligonucleotide. **E** Two iterations of 2HDLL + M100 modelling, illustrating the variation in protein position. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data suggest that it is the interactions that occur between the homeodomains and DNA that determines the overall binding specificity of the complex.

Second, SAXS data indicates that when ISL1 and LHX3 homeodomains are brought into close proximity, both homeodomains can interact with DNA. This can be seen in the 2HD:M100 models (Fig. 5D), where the protein density is consistently in contact with the DNA. This is in contrast to previously published EMSA data, which did not report binding between ISL1$_{HD}$ and sequences it targets *in vivo* (Robertson et al., 2018). These data suggest that ISL1$_{HD}$ influences the specificity of LHX3$_{HD}$ without having a high affinity binding target itself. This effect could be achieved through the formation of transient interactions between ISL1$_{HD}$ and DNA, probably driven by electrostatic forces (Lohman and Von Hippel, 1986). In such a model, the presence of ISL1 bound to LHX3 is sufficient to displace LHX3 from its TAATTA consensus sequence on DNA (Fig. 6A-C). The ISL1:LHX3 protein complex then moves around, forming transient interactions with other parts of DNA, until it reaches the consensus sequence for the complex (Fig. 6D-E). At such locations, LHX3$_{HD}$ can bind stably to an AAATTA site, while ISL1 is localised by binding to LHX3, with ISL1$_{HD}$ interacting with DNA in a more dynamic manner.

This proposed specificity for ISL1/LHX3 DNA-binding complexes is

consistent with observed behaviour of many other transcriptional complexes containing homeodomains (Laughon, 1991; Mann et al., 2009). Homeodomains commonly target AT-rich DNA sequences that are six bases long, in a manner that is primarily driven by electrostatic forces (Lohman and Von Hippel, 1986); this low level of specificity is compensated for by the formation of higher order transcriptional complexes, often containing multiple homeodomain-containing proteins, to regulate gene expression more precisely during development (Dragan et al., 2006; Gehring et al., 1994; Laughon, 1991; Noyes et al., 2008; Wolberger, 1996). Most of these changes in DNA binding specificity arise from either the interplay of two high-affinity binders, or from allosteric changes to the conformation of the homeodomain, arising from a protein:protein interaction (Slattery et al., 2011; Zandvakili and Gebelein, 2016). However, neither of these mechanisms fully explains the actions of ISL1 and LHX3 binding.

It was previously suggested that ISL1 binds DNA in a manner dependent on its protein binding partners (Mazzoni et al., 2013). Our data additionally suggests that the presence of such protein binding partners is necessary for ISL1 to associate with DNA with appreciable affinity. The notion that ISL1 does not independently bind DNA with higher affinity is consistent with studies that have found non-specific DNA-binding to precede specific DNA-binding in other systems
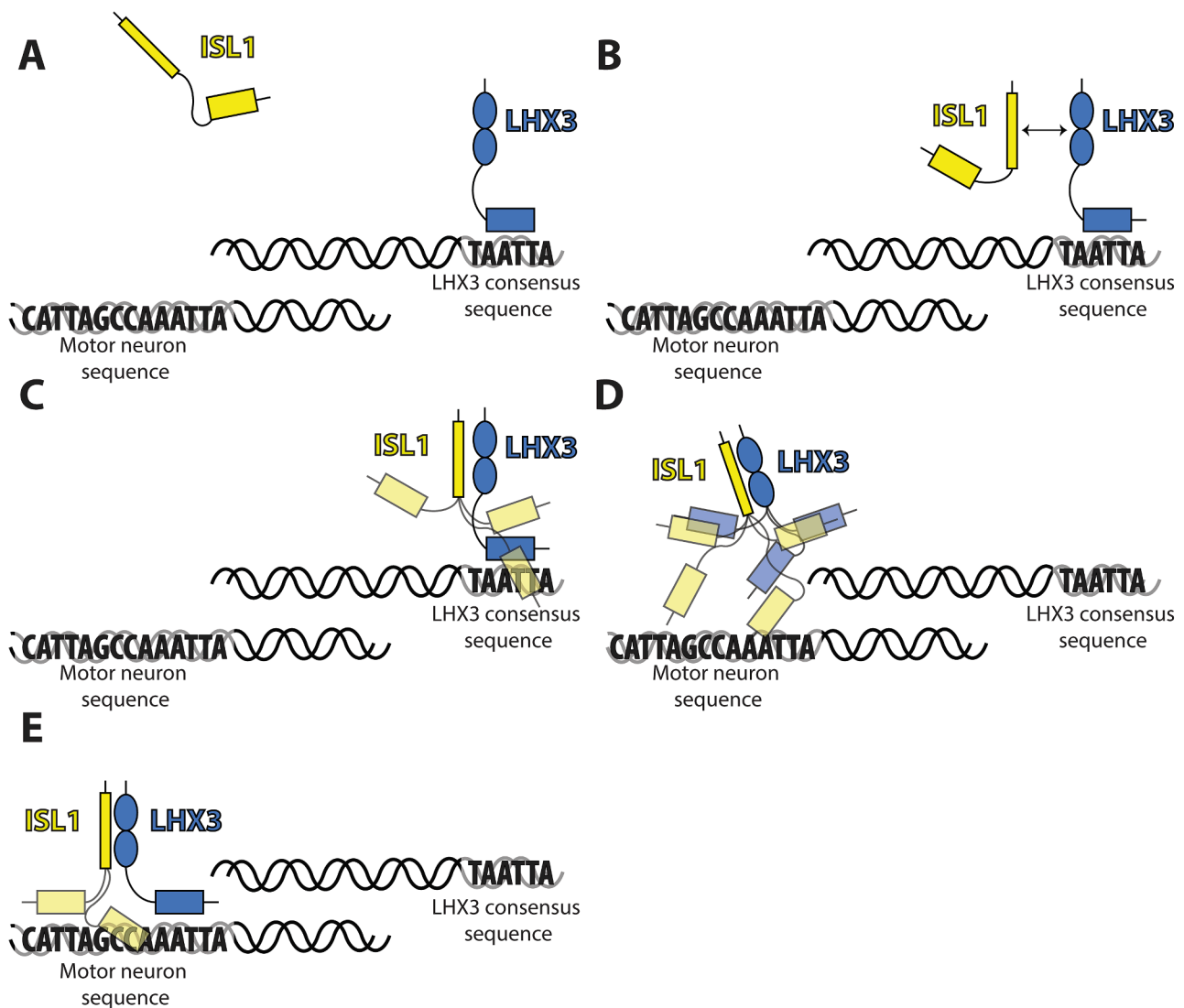
Fig. 6. Model of the proposed mechanism by which ISL1:LHX3 DNA binding specificity is determined. **A** When not associated to ISL1 (yellow), LHX3 (blue) can stably bind its consensus sequence, and ISL1 does not bind DNA strongly. **B** When ISL1 is introduced an association is formed between ISL1 and LHX3. **C** This association brings $ISL1_{HD}$ into proximity of the DNA, allowing it to displace $LHX3_{HD}$. **D** The ISL1:LHX3 complex translocates around DNA. **E** Upon finding the appropriate sequence, $LHX3_{HD}$ is able to bind stably, with $ISL1_{HD}$ forming more dynamic non-specific interactions with the DNA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

([Hendrix et al., 2010; Vukojević et al., 2010](#)). This mode of action may provide an evolutionary advantage in the context of gene regulation, by ensuring that gene expression is only affected when both ISL1 and its protein binding partner/s are present. Having two DNA binding modules working in concert can also result in a more efficient searching algorithm for binding sites in the genome, especially in the case of one weak DNA binder in combination with one strong DNA binder ([Vuzman et al., 2010a, 2010b](#)).

In a broader context, this may also provide an insight as to why ISL1 has been shown to play a role in transcriptional regulation in a multitude of tissues, with numerous different binding partners ([Cai et al., 2003; Galloway et al., 2015; Li et al., 2014; Takeuchi et al., 2005; Yang et al., 2015](#)). By modulating the specificity of protein binding partners, as opposed to having DNA binding specificity of its own, ISL1 could to target a more diverse range of sequences, as dictated by its protein binding partner in the context of each cell type.

CRediT authorship contribution statement

Ngaio C. Smith: Methodology, Investigation, Writing - original

draft. **Lorna E. Wilkinson-White:** Methodology, Visualization. **Ann H. Y. Kwan:** Methodology, Supervision. **Jill Trewhella:** Methodology, Formal analysis, Visualization. **Jacqueline M. Matthews:** Conceptualization, Supervision.

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.

Acknowledgements

Appendix A.　Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yjsbx.2020.100043.

References

Behravan, G., Lycksell, P.-O., Larsson, G., 1997. Expression, purification and characterization of the homeodomain of rat ISL-1 protein. Protein Eng. 10, 1327–1331.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S.A., Morris, Q.D., Bulyk, M.L., Hughes, T.R., 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell 133, 1266–1276.

Bernadó, P., Svergun, D.I., 2012. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol. BioSys. 8, 151–167.

Bhati, M., Lee, M., Nancarrow, A.L., Bach, I., Guss, J.M., Matthews, J.M., 2008a. Crystallization of an Lhx3-Isl1 complex. Acta Crysallogr. Sect. F 64, 297–299.

Bhati, M., Lee, C., Nancarrow, A.L., Lee, M., Craig, V.J., Bach, I., Guss, J.M., Mackay, J.P., Matthews, J.M., 2008b. Implementing the LIM code: the structural basis for cell type-specific assembly of LIM-homeodomain complexes. EMBO J. 27, 2018–2029.

Bridwell, J.L., Price, J.R., Parker, G.E., McCutchan Schiller, A., Sloop, K.W., Rhodes, S.J., 2001. Role of the LIM domains in DNA recognition by the Lhx3 neuroendocrine transcription factor. Gene 277, 239–250.

Cai, C.-L., Liang, X., Shi, Y., Chu, P.-H., Pfaff, S.L., Chen, J., Evans, S., 2003. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Dev. Cell 5, 877–889.

Dragan, A.I., Li, Z., Makeyeva, E.N., Milgotina, E.I., Liu, Y., Crane-Robinson, C., Privalov, P.L., 2006. Forces driving the binding of homeodomains to DNA. Biochem 45, 141–151.

Franke, D., Svergun, D.I., 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. 42, 342–346.

Gadd, M.S., Bhati, M., Jeffries, C.M., Langley, D.B., Trewhella, J., Guss, J.M., Matthews, J.M., 2011. Structural basis for partial redundancy in a class of transcription factors, the LIM homeodomain proteins, in neural cell type specification. J. Biol. Chem. 286, 42971–42980.

Galloway, J.R., Bethea, M., Liu, Y., Underwood, R., Mobley, J.A., Hunter, C.S., 2015. SSBP3 interacts with islet-1 and Ldb1 to impact pancreatic β-cell target genes. Mol. Endocrinol. 29, 1774–1786.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., Wüthrich, K., 1994. Homeodomain-DNA recognition. Cell 78, 211–223.

Hendrix, J., Gijsbers, R., De Rijck, J., Voet, A., Hotta, J.-I., McNeely, M., Hofkens, J., Debyser, Z., Engelborghs, Y., 2010. The transcriptional co-activator LEDGF/p75 displays a dynamic scan-and-lock mechanism for chromatin tethering. Nucleic Acids Res. 39, 1310–1325.

Hunter, C., Rhodes, S., 2005. LIM-homeodomain genes in mammalian development and human disease. Mol. Biol. Rep. 32, 67–77.

Ippel, H., Larsson, G., Behravan, G., Zdunek, J., Lundqvist, M., Schleucher, J., Lycksell, P.-O., Wijmenga, S., 1999. The solution structure of the homeodomain of the rat insulin-gene enhancer protein Isl-1. Comparison with other homeodomains. J. Mol. Biol. 288, 689–703.

Iwahara, J., Schwieters, C.D., Clore, G.M., 2004. Characterization of nonspecific protein−DNA interactions by 1H paramagnetic relaxation enhancement. J. Am. Chem. Soc. 126, 12800–12808.

Laughon, A., 1991. DNA binding specificity of homeodomains. Biochem 30, 11357–11367.

Lee, S., Lee, B., Joshi, K., Pfaff, S.L., Lee, J.W., Lee, S.-K., 2008. A regulatory network to segregate the identity of neuronal subtypes. Dev. Cell 14, 877–889.

Lee, S., Cuvillier, J.M., Lee, B., Shen, R., Lee, J.W., Lee, S.-K., 2012. Fusion protein Isl1–Lhx3 specifies motor neuron fate by inducing motor neuron genes and concomitantly suppressing the interneuron programs. Proc. Natl. Acad. Sci. 109, 3383–3388.

Li, R., Wu, F., Ruonala, R., Sapkota, D., Hu, Z., Mu, X., 2014. Isl1 and Pou4f2 form a complex to regulate target genes in developing retinal ganglion cells. PLoS One 9, e92105.

Lohman, T.M., Von Hippel, P.H., 1986. Kinetics of protein-nucleic acid interactions: use of salt effects to probe mechanisms of interactio. Crit. Rev. Biochem. 19, 191–245.

Mann, R.S., Lelli, K.M., Joshi, R., 2009. Hox specificity: unique roles for cofactors and collaborators. Curr. Topics Dev. Biol. 88, 63–101.

Matthews, J.M., Visvader, J.E., 2003. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. EMBO Rep. 4, 1132–1137.

Mazzoni, E.O., Mahony, S., Closser, M., Morrison, C.A., Nedelec, S., Williams, D.J., An, D., Gifford, D.K., Wichterle, H., 2013. Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. Nat. Neurosci. 16, 1219–1227.

Meier, B.C., Price, J.R., Parker, G.E., Bridwell, J.L., Rhodes, S.J., 1999. Characterization of the porcine Lhx3/LIM-3/P-Lim LIM homeodomain transcription factor. Mol. Cell. Endocrinol. 147, 65–74.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., Wolfe, S. A., 2008. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell 133, 1277–1289.

Petoukhov, M.V., Svergun, D.I., 2015. Ambiguity assessment of small-angle scattering curves from monodisperse systems. Acta Crystallogr. Sect. D 71, 1051–1058.

Pfaff, S.L., Mendelsohn, M., Stewart, C.L., Edlund, T., Jessell, T.M., 1996. Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. Cell 84, 309–320.

Roberson, M.S., Schoderbek, W.E., Tremml, G., Maurer, R.A., 1994. Activation of the glycoprotein hormone alpha-subunit promoter by a LIM-homeodomain transcription factor. Mol. Cell. Biol. 14, 2985–2993.

Robertson, N.O., Smith, N.C., Manakas, A., Mahjoub, M., McDonald, G., Kwan, A.H., Matthews, J.M., 2018. Disparate binding kinetics by an intrinsically disordered domain enables temporal regulation of transcriptional complex formation. Proc. Natl. Acad. Sci. 115, 4643–4648.

Semisotnov, G.V., Kihara, H., Kotova, N.V., Kimura, K., Amemiya, Y., Wakabayashi, K., Serdyuk, I.N., Timchenko, A.A., Chiba, K., Nikaido, K., 1996. Protein globularization during folding. A study by synchrotron small-angle X-ray scattering. J. Mol. Biol. 262, 559–574.

Sharma, K., Sheng, H.Z., Lettieri, K., Li, H., Karavanov, A., Potter, S., Westphal, H., Pfaff, S.L., 1998. LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. Cell 95, 817–828.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. Cell 147, 1270–1282.

Srivastava, M., Larroux, C., Lu, D.R., Mohanty, K., Chapman, J., Degnan, B.M., Rokhsar, D.S., 2010. Early evolution of the LIM homeobox gene family. BMC Biol. 8, 4.

Svergun, D., 1999a. Ab initio shape determination by simulated annealing using a single phase dummy atom model. Biophys. J. 76, 2879–2886.

Svergun, D.I., 1999b. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. 76, 2879–2886.

Takeuchi, J.K., Mileikovskaia, M., Koshiba-Takeuchi, K., Heidt, A.B., Mori, A.D., Arruda, E.P., Gertsenstein, M., Georges, R., Davidson, L., Mo, R., Hui, C.-C., Henkelman, R.M., Nemer, M., Black, B.L., Nagy, A., Bruneau, B.G., 2005. Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development. Development 132, 2463–2474.

Trewhella, J., Duff, A.P., Durand, D., Gabel, F., Guss, J.M., Hendrickson, W.A., Hura, G. L., Jacques, D.A., Kirby, N.M., Kwan, A.H., Perez, J., Pollack, L., Ryan, T.M., Sali, A., Schneidman-Duhovny, D., Schwede, T., Svergun, D.I., Sugiyama, M., Tainer, J.A., Vachette, P., Westbrook, J., Whitten, A.E., 2017. 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. Acta Crystallogr. Sect. D 73, 710–728.

Vukojević, V., Papadopoulos, D.K., Terenius, L., Gehring, W.J., Rigler, R., 2010. Quantitative study of synthetic Hox transcription factor–DNA interactions in live cells. Proc. Natl. Acad. Sci. 107, 4093–4098.

Vuzman, D., Azia, A., Levy, Y., 2010a. Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. J. Mol. Biol. 396, 674–684.

Vuzman, D., Polonsky, M., Levy, Y., 2010b. Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. Biophys. J. 99, 1202–1211.

Wolberger, C., 1996. Homeodomain interactions. Curr. Opin. Struct. Biol. 6, 62–68.

Yang, Z., Zhang, Q., Lu, Q., Jia, Z., Chen, P., Ma, K., Wang, W., Zhou, C., 2015. ISL-1 promotes pancreatic islet cell proliferation by forming an ISL-1/Set7/9/PDX-1 complex. Cell Cyc. 14, 3820–3829.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., Nitta, K.R., Taipale, M., Popov, A., Ginno, P.A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., Taipale, J., 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356.

Zandvakili, A., Gebelein, B., 2016. Mechanisms of specificity for Hox factor activity. J. Dev. Biol. 4, 16.

Zhadanov, A.B., Bertuzzi, S., Taira, M., Dawid, I.B., Westphal, H., 1995. Expression pattern of the murine LIM class homeobox gene Lhx3 in subsets of neural and neuroendocrine tissues. Dev. Dyn. 202, 354–364.