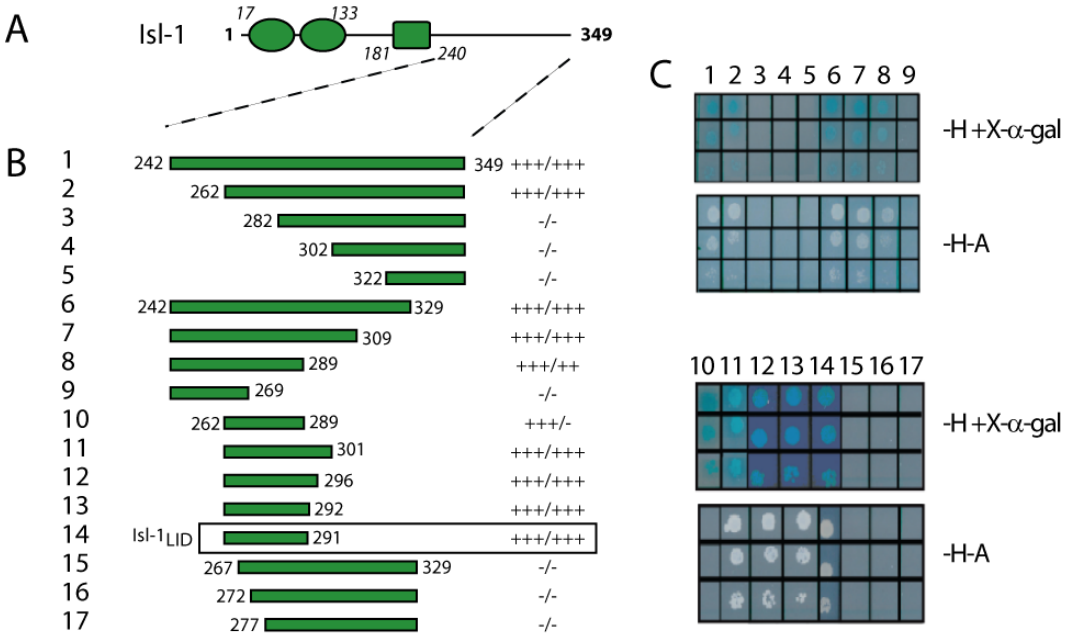
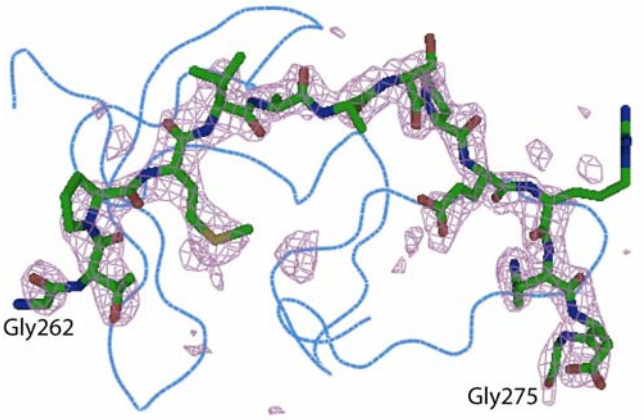
Supplementary data

Supplementary data 1. Definition of $Isl1_{LBD}$

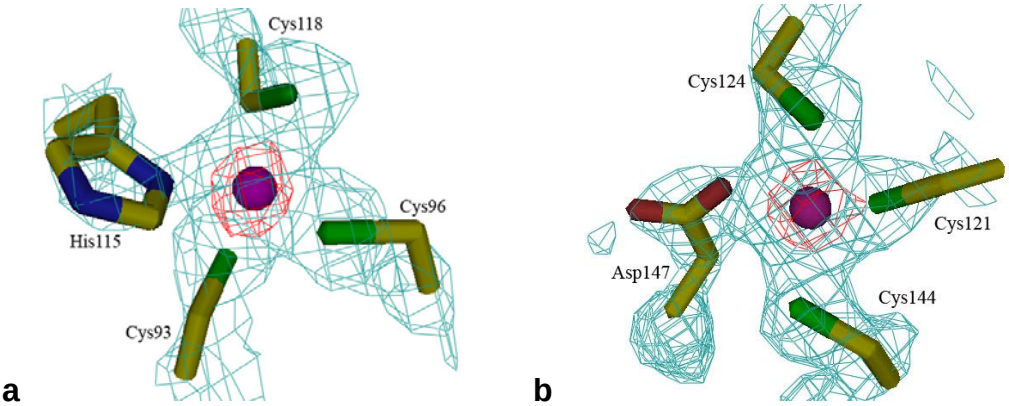
$Isl1_{LBD}$ was defined using yeast-two-hybrid analysis of deletion mutants as Isl1 residues 262–291. The stringency of selection conditions is shown as low (-L-W-H, 1 mM 3-AT + X-α-gal)/ high (-L-W-H-A). Interactions between $Isl1_{LBD}$ and Lhx3 constructs were evidenced by growth of yeast colonies. +++ indicates growth on the 10^0, 10^{-1} and 10^{-2} dilutions, ++ indicates growth only in 10^0 and 10^{-1}, + only in 10^0, and – indicates no growth at any dilution used. **A** is a schematic of Isl1 indicating the region that was tested. **B** shows the deletion mutants and a summary of the results. **C** shows an example of the yeast-two-hybrid data for all mutants on different selection media.



Supplementary data 2. X-ray analysis of $Lhx3_{LIM1+2}$-$Isl1_{LBD}$



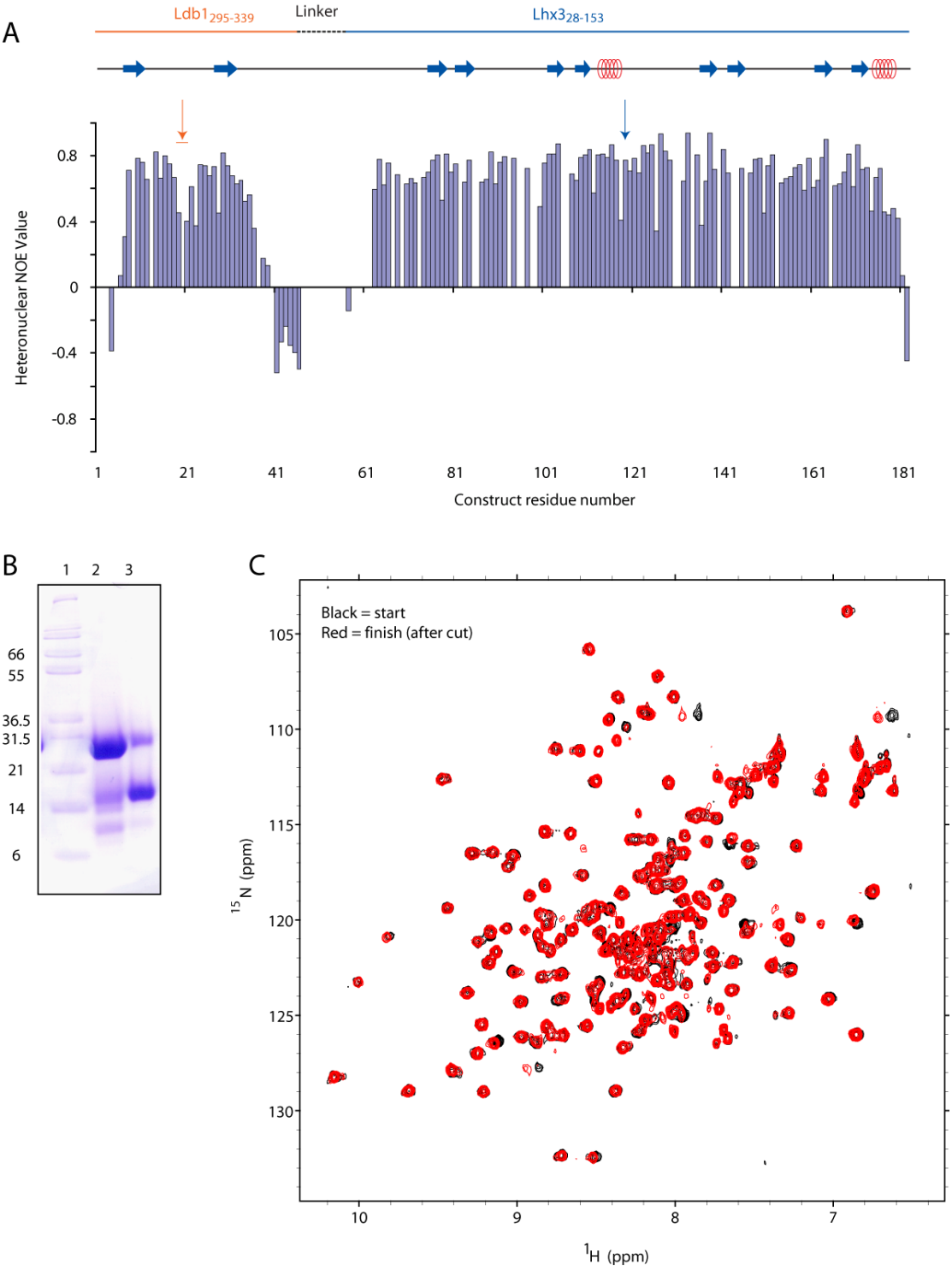
A Electron density over $Isl1_{LBD}$ at LIM2. An F_o-F_c electron density difference (3σ) map of $Isl1_{LBD}$ over $Lhx3_{LIM2}$ (blue) is shown. The map was calculated after energy minimization without NCS restraints and excluding $Isl1_{LBD}$.



a **b**

B Zn co-ordination at the Zn3 and Zn4 ions of $Lhx3_{LIM2}$ in the $Lhx3_{LIM1+2}$-$Isl1_{LBD}$ complex. **a** Zn3, and **b** Zn4. Electron density over the zinc ions and Zn ligands is shown in blue on a F_o-F_c map (contour level 5σ). The map was calculated after energy minimization without NCS restraints and excluding the zinc ligands. An anomalous difference map is shown in red (contour level 7σ).

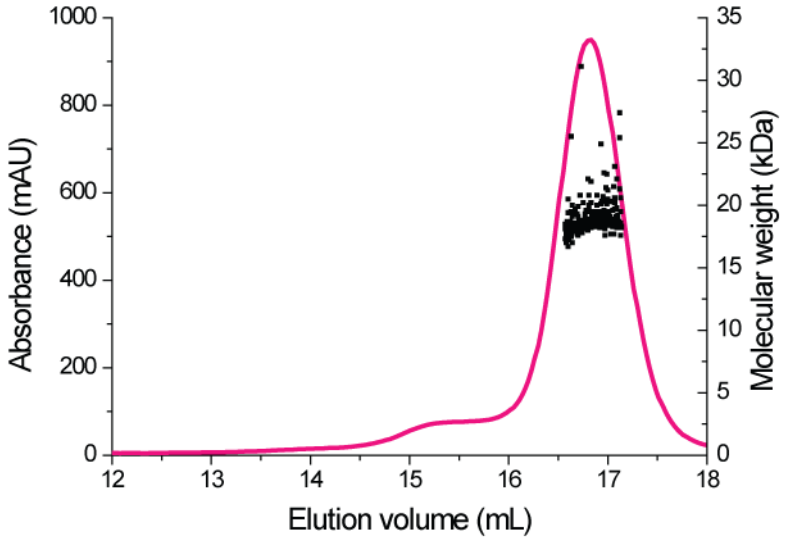
Supplementary Data 3. NMR data for the Lhx3-Ldb1$_{LID}$ complex.



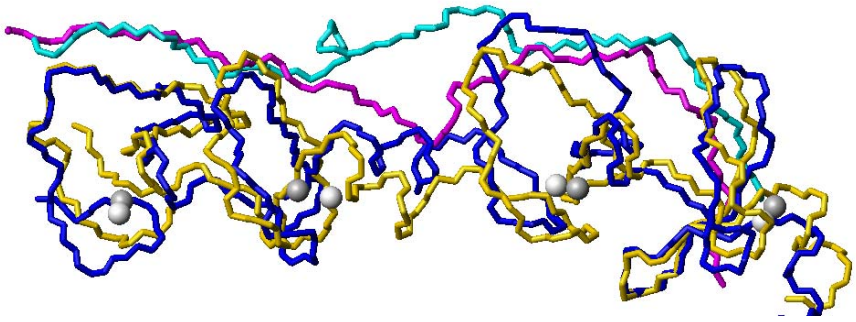
A ^{15}N-^{1}H-Heteronuclear NOE data. Arrows show the positions of Ldb1$_{313–315}$ (orange arrow) and Lhx3$_{119}$(blue arrow). Contributions from Ldb1 the linker and Lhx3 are indicated, as are secondary structural elements. **B,C** Demonstration of the integrity of the Lhx3-Ldb1$_{LID}$ complex. **B** SDS page showing a ^{15}N-labelled sample of Ldb1$_{LID}$-Xa-Lhx3$_{LIM1+2}$ (which contains a Factor Xa site in the linker) before (lane 2) and after (lane 3) extended treatment with Factor Xa**. C** Overlaid ^{1}H-^{15}N-HSQC spectra of the same protein samples before (black) and after (red) Factor-Xa treatment.

Supplementary Data 4. MALLS analysis of $Lhx3_{LIM1+2}$-$Isl1_{LBD}$ complex

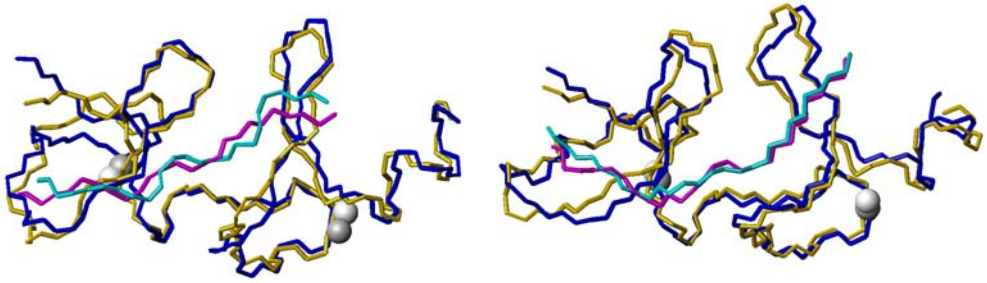
A sample of the complex was run on a Suprose12™ size exclusion column in 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM dithiothreitol. The eluent was detected by absorbance at 280 nm (pink trace) and an in-line MALLS detection system (Wyatt Refractometer and MiniDawn MALLS detector) and analysed using the software provided by the manufacturer. Bovine serum albumin was used as a standard.



Supplementary Data 5. Overlays of Lhx3-$Ldb1_{LID}$ and Lhx3-$Isl1_{LBD}$ structures.



A



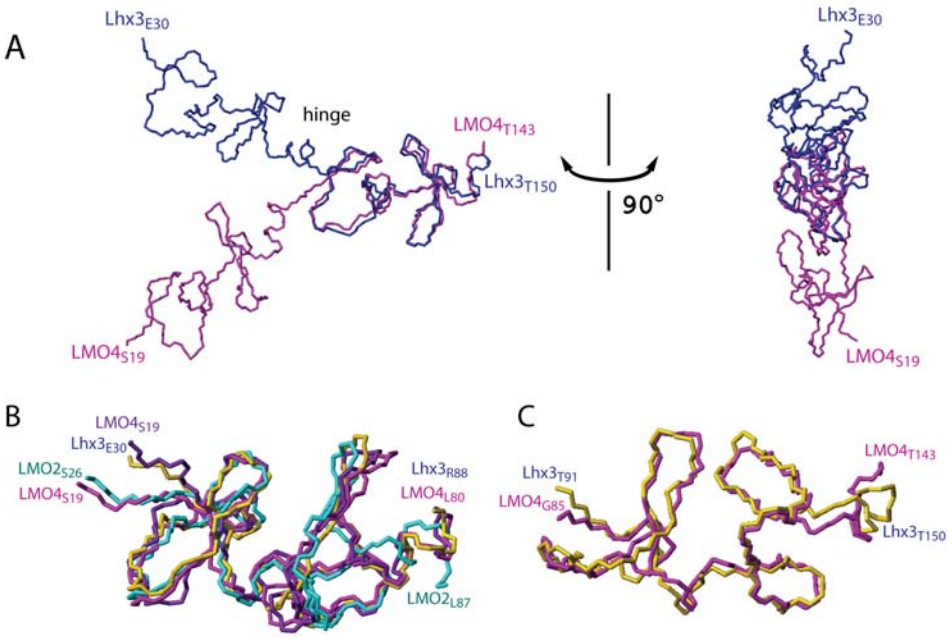
B **C**

Comparison of the $Lhx3_{LIM1+2}$-$Ldb1_{LID}$ (blue/cyan) and the monomeric $Lhx3_{LIM1+2}$-$Isl1_{LBD}$ (yellow/magenta) complexes. The backbone residues of the **A** LIM1+2 domains, **B** LIM1 and **C** LIM2 domains are overlaid. The zinc atoms are grey for $Lhx3_{LIM1+2}$-$Isl1_{LBD}$ and white for $Lhx3_{LIM1+2}$-$Ldb1_{LID}$.

Each LIM domain from the NMR and crystal structures conforms to the typical LIM domain topology of two Zn-ligating modules (Zn1 and Zn2 for LIM1 and Zn3 and Zn4 for LIM2), in which each module comprises two β-hairpins followed by an α-helix of varying length (Fig 2A,D). The helices within Zn1 and Zn3 appear as poorly defined single turns, whereas the helices in Zn2 and Zn4 contain several turns. The zinc ions in each module of Lhx3 are ligated by the sidechains of two Cys residues in the first β-hairpin and one Cys residue plus one His (Zn1, Zn3), or Asp (Zn2, Zn4) in the helical segment (Supp. Data 2B). All of the secondary structure elements tend to be short, and as such are not all recognized by the Kabsch/Sander algorithm (Kabsch & Sander, 1983). In terms of comparison between the structures, RMSDs over the backbone residues of Lhx3 from the lowest energy member of the NMR ensemble of Lhx3-$Ldb1_{LID}$ and the Lhx3-$Isl1_{LBD}$ monomer are 3.5 Å for $Lhx3_{LIM1+2}$, 1.8 Å for $Lhx3_{LIM1}$, and 1.3 Å for $Lhx3_{LIM2}$.

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577-2637.

Supplementary Data 6. Comparison of LIM-$Ldb1_{LID}$ complexes. A Comparison of the relative orientation of the LIM domains from LMO4 (1RUT, magenta) and the lowest energy structure from the Lhx3 ensemble (blue) overlaid over the backbone atoms of the LIM2 domains. Comparison of **B** LIM1 and **C** LIM2 domains from complexes containing LMO2 (1J2O, cyan), LMO4 (1M3V, violet; IRUT, magenta) and Lhx3 (yellow) overlaid over the backbone atoms.



Supplementary Data 7. Modelling of relative species distributions of Lhx3-, Isl1-, and Ldb1-containing complexes

In order to understand the relative populations of complexes that form within cells, a series of simple modelling experiments was carried out using the program DynaFit3 (Biokin Inc., USA) (Kuzmic, 1996). Because a number of unknown variables exist these experiments were designed to look for trends in complex formation rather than deriving absolute values for the concentrations of the various transcription factor complexes.

To minimise the complexity of the modelling, the self-association of Ldb1 was not specifically considered. Instead, self-association of Ldb1 was treated as a means of enhancing DNA binding to multiple sites in the single molecule DNA-binding (S) model and the single molecule DNA-binding advantage (SA) model. The association of Isl1 and Lhx3 was additionally treated as a means of enhancing DNA binding to multiple sites.

The real effective concentrations of the components are unknown. Thus, a range of concentrations was used. Estimates of numbers of transcription factors per cell and the volume of eukaryotic nuclei suggest that nuclear protein concentrations of 1–100 nM are reasonable estimates (Ryan et al., 2007); however, higher concentrations (up to 1 mM) were also used to mimic high local concentrations of proteins. The total concentrations of each protein and DNA species were all assigned the same value ranging from 1 nM to 1 mM.

Chemical denaturation data from Fig 4B indicated that the Lhx3:Isl1 interaction is significantly weaker than that for the Lhx3:Ldb1 interaction. Data from our laboratory indicates that a lower limit of K_d ~ 100 nM applies to interactions detected by yeast-two hybrid analysis under stringent conditions (-L-W-H-A) (Ryan et al., 2006), and the Lhx3:Isl1 interaction appears to lie close to this limit; the interaction is detected under these conditions by yeast-two hybrid analysis when $Lhx3_{LIM1+2}$ is fused to the GAL4-activation domain, but not when it is fused to the GAL4-DNA-binding domain (Table III). Thus the K_d for the interaction between Lhx3 and Isl1 was set at 200 nM. This represents an ~5-fold weaker interaction than that between Lhx3 and Ldb1. The K_d for the interaction between Ldb1/Isl1 and Lhx3 was given the same estimated value.

The dissociation constants for homeodomain-DNA interactions for binding to single sites were set as 100 nM in all cases. In the (S) model the dissociation constants for double site binding (*e.g.*, Ldb1:Isl1, Ldb1:Lhx3, or Isl1:Lhx3 complexes binding to two closely spaced sites on a single DNA molecule) was set as 1 nM, and the dissociation constant for quadruple site binding (Ldb1:Isl1:Lhx3 binding to four closely spaced sites on DNA) was set as 0.01 nM (See Table S7 for details).

The dissociation constant for a protein:protein interaction involving the same interfaces was assigned the same value (*e.g.*, 3 and 4 in Table S6 each involve the formation of an Isl1:Lhx3 interface), except in the SA model in reactions where the addition of a protein could make multiple interfaces with either protein or DNA (**12–17** in Table S7). In those cases the K_d was assigned a 100-fold lower value to reflect effectively tighter binding.

When considering only the relative competition of Lhx3 and Isl1 for Ldb1 in the absence of any other interactions, the simultaneous equilibria 1 and 2 (Table S7) were used (Binary B series; Fig S7.2A; Fig4C).

When considering the relative competition of Lhx3 and Isl1 for Ldb1 and Isl1 and Ldb1 for Lhx3 in the absence of binding to DNA, the simultaneous equilibria 1–5 were used (Ternary T series; Fig S7.2A; Fig 4C).

When considering the relative competition of Lhx3 and Isl1 for Ldb1 and Isl1 and Ldb1 for Lhx3 in the presence of binding to DNA, two different models were used. In the Independent (I) model each DNA site was considered to bind in an independent fashion and DNA sites were considered to be specific for Lhx3 (D3) or Isl1 (D1). The simultaneous equilibria 1–27 were used. In the single molecule DNA-binding (S) model and the single molecule DNA-binding advantage (SA) the simultaneous equilibria 1–5 and **6–17** were used (Table S7; Figure S7.1).

The concentration populations of key species under various conditions were estimated and are shown as relative populations (to Ldb1:Lhx3; or Ldb1:Lhx3:DNA) in Figure S7.2.

Table S7: Simultaneous equilibria and values of dissociation constants used in modelling.

Reaction	**Reaction**	***Kd* (nM)**	
Protein only, B and T series			
Ldb1 + Lhx3 ↔ Ldb1:Lhx3	1	35	
Ldb1 + Isl1 ↔ Ldb1:Isl1	2	90	
Isl1 + Lhx3 ↔ Isl1:Lhx3	3	200	
Ldb1:Isl1 + Lhx3 ↔ Ldb1:Isl1:Lhx3	4	200	
Isl1:Lhx3 + Ldb1 ↔ Ldb1:Isl1:Lhx3	5	90	
Protein plus DNA - Independent Model (I)			
Lhx3 + D3 ↔ Lhx3:D3	6	100	
Isl1 + D1 ↔ Isl1:D1	7	100	
Ldb1:Lhx3 + D3 ↔ Ldb1:Lhx3:D3	8	100	
Ldb1:Isl1 + D1 ↔ Ldb1:Isl1:D1	9	100	
Isl1:Lhx3 + D1 ↔ Isl1:Lhx3:D1	10	100	
Isl1:Lhx3 + D3 ↔ Isl1:Lhx3:D3	11	100	
Lhx3:D3 + Ldb1 ↔ Ldb1:Lhx3:D3	12	35	
Lhx3:D3 + Isl1 ↔ Isl1:Lhx3:D3	13	200	
Isl1:D1 + Ldb1 ↔ Ldb1:Isl1:D1	14	200	
Isl1:D1 + Lhx3 ↔ Isl1:Lhx3:D1	15	200	
Lhx3:D3 + Isl1:D1 ↔ Isl1:Lhx3:D1:D3	16	200	
Ldb1:Isl1 + Lhx3:D3 ↔ Ldb1:Isl1:Lhx3:D3	17	200	
Isl1:Lhx3:D3 + D1 ↔ Isl1:Lhx3:D1:D3	18	100	
Isl1:Lhx3:D1 + D3 ↔ Isl1:Lhx3:D1:D3	19	100	
Ldb1:Isl1:D1 + Lhx3 ↔ Ldb1:Isl1:Lhx3:D1	20	200	
Ldb1:Isl1:Lhx3 + D1 ↔ Ldb1:Isl1:Lhx3:D1	21	100	
Ldb1:Isl1:Lhx3 + D3 ↔ Ldb1:Isl1:Lhx3:D3	22	100	
Isl1:Lhx3:D1 + Ldb1 ↔ Ldb1:Isl1:Lhx3:D1	23	90	
Isl1:Lhx3:D3 + Ldb1 ↔ Ldb1:Isl1:Lhx3:D3	24	90	
Ldb1:Isl1:Lhx3:D1 + D3 ↔ Ldb1:Isl1:Lhx3:D1:D3	25	100	
Ldb1:Isl1:D1 + Lhx3:D3 ↔ Ldb1:Isl1:Lhx3:D1:D3	26	35	
Ldb1:Isl1:Lhx3:D3 + D1 ↔ Ldb1:Isl1:Lhx3:D1:D3	27	100	
Isl1:Lhx3:D1:D3 + Ldb1 ↔ Ldb1:Isl1:Lhx3:D1:D3	28	90	
Protein plus DNA – Single Molecule DNA Models		**S**	**SA**
Lhx3 + DNA ↔ Lhx3:DNA	**6**	100	100
Isl1 + DNA ↔ Isl1:DNA	**7**	100	100
Ldb1:Lhx3 + DNA ↔ Ldb1:Lhx3:DNA	**8**	1	1
Ldb1:Isl1 + DNA ↔ Ldb1:Isl1:DNA	**9**	1	1
Isl1:Lhx3 + DNA ↔ Isl1:Lhx3:DNA	**10**	1	1
Ldb1:Isl1:Lhx3 + DNA ↔ Ldb1:Isl1:Lhx3:DNA	**11**	0.01	0.01
Lhx3:DNA + Ldb1 ↔ Ldb1:Lhx3:DNA	**12**	200	2
Lhx3:DNA + Isl1 ↔ Isl1:Lhx3:DNA	**13**	200	2
Isl1:DNA + Ldb1 ↔ Ldb1:Isl:DNA	**14**	35	0.35
Isl1:DNA + Lhx3 ↔ Isl1:Lhx3:DNA	**15**	200	2
Isl1:Lhx3:DNA + Ldb1 ↔ Ldb1:Isl1:Lhx3:DNA	**16**	90	0.9

Ldb1:Isl1:DNA + Lhx3 ↔ Ldb1:Isl1:Lhx3:DNA	**17**	200	2

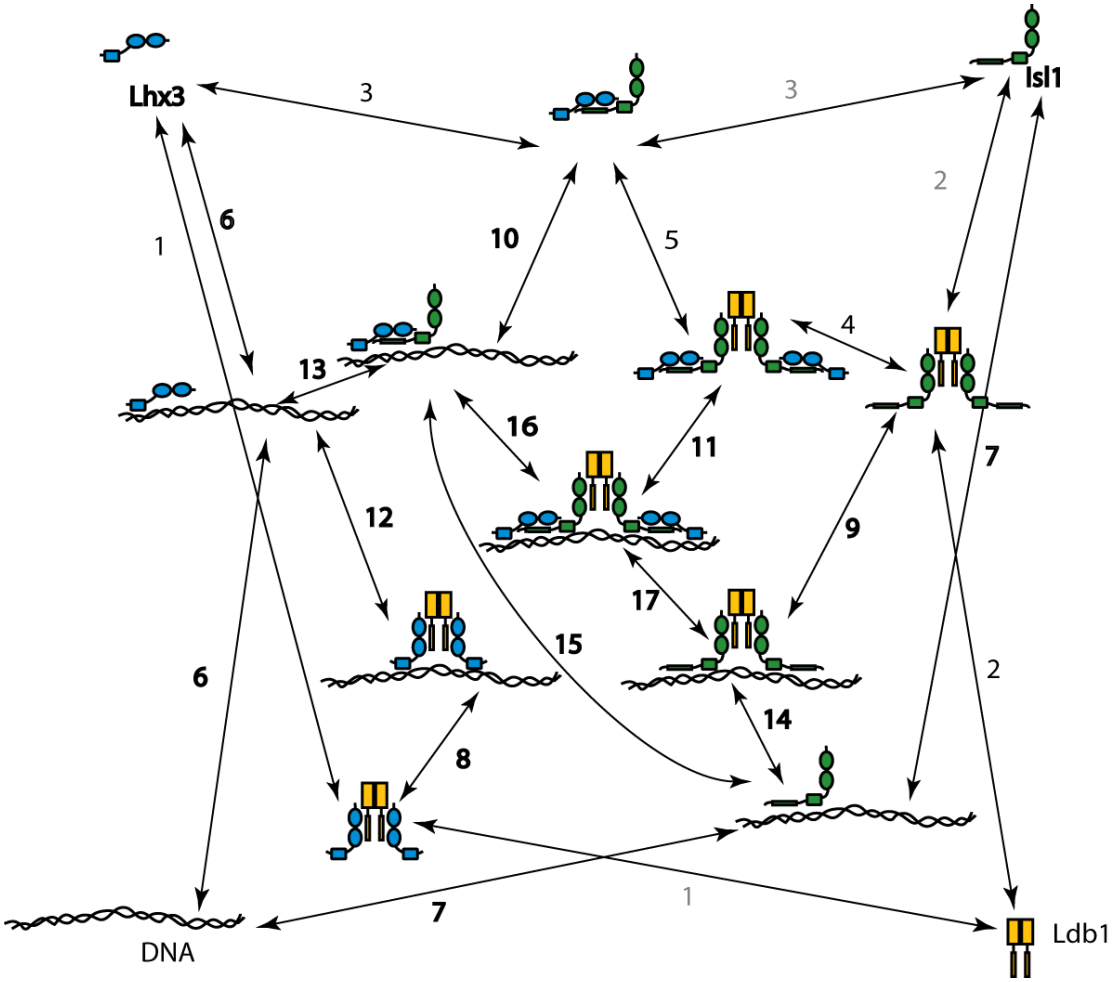


Figure S7.1: Schematic showing network of species and simultaneous equilibria used in the S and SA models. The numbers of the reactions correspond to those in Table S7. In some instances the same reactions appear more than once in the schematic. These duplicated reactions are indicated with grey numbers.

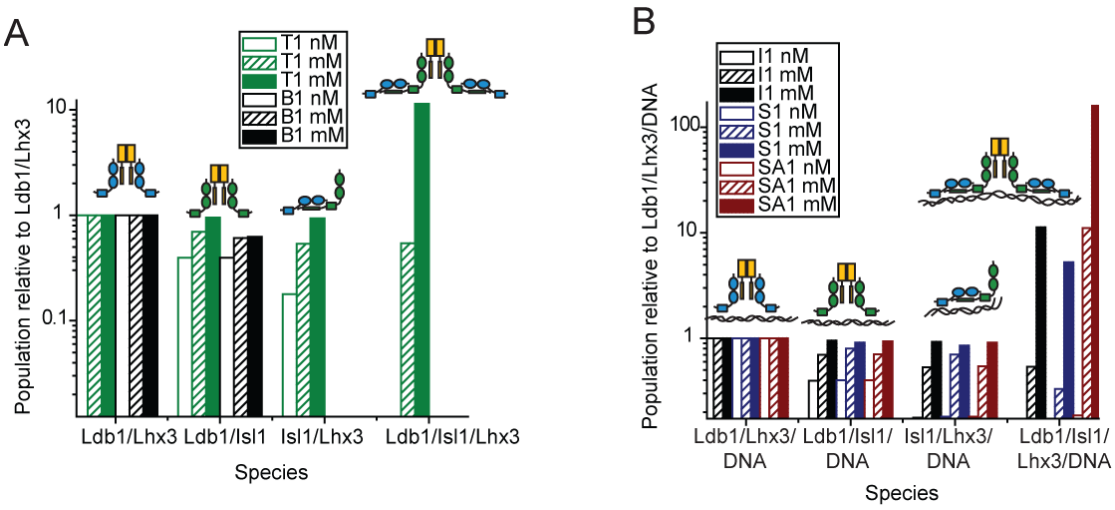


Figure S7.2. Trends for the formation of protein complexes containing Ldb1, Lhx3 and Isl1 at low to high protein concentrations. Concentrations of all components were set at 1 nM, 1μM, and 1 nM **A** Distribution of protein-only complexes relative to Ldb1:Lhx3. The B (binary) series considers only the competition of Lhx3 and Isl1 for Ldb1. The T (Ternary) series also considers the competition between Ldb1 and Isl1 for Lhx3. **B** Distribution of key protein:DNA complexes relative to Ldb1:Lhx3:DNA. The I series show the results of using the independent model, the S series shows the results of the single molecule DNA-binding model and the SA series shows the results of the single molecule DNA-binding advantage model. Note that the Y-axes have log scales whereas the graphs in Fig 4C,D have linear scales.

Supplementary Data 8. Sequence comparison of Isl_{LBD} from multiple species.

```
Vertebrate Isl1LBD
Blast submission (no gaps)      GT-PMVAASPERHDGGLQAN-PVEVQSYQPP--WK Mouse/human/chick/hamster
tr|Q08B34|Q08B34_XENLA          ..-....S...........-.....T....--.. Xenopus
sp|P53405|ISL1_DANRE            ..-....T...........-Q.........--.. Zebrafish

Vertebrate Isl2/3LBD
sp|P53408|ISL2A_ONCTS           ..-.L..G..I...NTV.G.-.....T....--.. Chinook Salmon
sp|P50212|ISL2B_ONCTS           ..-....G..I...NTVLG.-.....T....--.. Chinook Salmon
sp|P53406|ISL2A_DANRE           ..-.L..G..I...TTV.G.-.....T....--.. Zebrafish
tr|Q4SM40|Q4SM40_TETNG          ..-A...G..I..ENTV.G.-.....T....--.. Green Pufferfish
sp|P53409|ISL3_ONCTS            ..-.L..R..I...NTV.G.-S....T....--.. Chinook Salmon
sp|Q96A47|ISL2_HUMAN            ..-.L..G..I..ENAV.GS-A....T....--.. Human
sp|Q9CXV0|ISL2_MOUSE            ..-.L..G..IG.ENAV.GS-A....T....--.. Mouse
sp|P53410|ISL2_CHICK            ..-.L..G..V..ESAV.GS-A....T....--.. Chicken
tr|Q6JTI9|Q6JTI9_AMBME          ..-.L..G..I..ESAVHGA-A....T....--.. Axolotl
sp|P53407|ISL2B_DANRE           ..-.L..G..I..NPSVPGH-.....A....--.. Zebrafish
sp|P50480|ISL2_RAT              ..-LL..G..SA.ENAV.GS-A....T....--.. Rat

Arthropod dmISLET
tr|Q7PMI4|Q7PMI4_ANOGA          .I-....S..V...SP.NLH-GL..TA....--.. African Malaria Mosquito
tr|Q17L14|Q17L14_AEDAE          .I-....S..V...SP.NMH-GL..TA....--.. Yellowfever Mosquito
tr|P92031|P92031_DROME          .I-..I.S..V...SP.NLQ-GL...T....--.. Drosophila

Urochordata/Cephalochordata Isl
tr|Q9NHC8|Q9NHC8_BRAFL          .V-....QE.V..ESQM...-........Q.PP.. Lancelet (Amphioxus)
tr|Q9BLJ0|Q9BLJ0_HALRO          .V-....SE.V.N.NAVNIK-....RN....-A.. Sea squirt
tr|Q4H3A5|Q4H3A5_CIOIN          .V-....SE.V.N.NSVSVA-....RN....-A.. Transparent Sea Squirt

Nematode LIM-7
*tr|Q94160_CAEEL                 .IG.LMVQPATP.IDNTLGG-.IDIQHFAQ---.N Nematode (C. elegans)
*BP:CBP06623                     .IG.LMVQPPTP.IDTTLGGG.IDIHSFAQ---.S Nematode (C. Briggsae)
```

Sequence alignment of Isl_{LBD} from diverse species as identified from a standard pBLAST search of the Swiss-Prot(sp) and TrEMBL(tr) databases plus equivalent regions from nematode homologs. Sequence identity with the $Isl1_{LBD}$ (.) and gaps in the alignment (-) are indicated. * Sequences were not identified through the BLAST search but were manually identified and aligned. The linear binding motifs identified in this study are in bold on the BLAST submission sequence.