Investigating the role of protein-protein and protein-DNA interactions in the function of Isl1

Ngaio Charlotte Smith

A thesis submitted to fulfil requirements for the degree of Doctor of Philosophy

School of Life and Environmental Sciences
Faculty of Science
University of Sydney

2019

Declaration

The work described in this thesis was conducted between March 2015 and November 2018 in the School of Life and Environmental Sciences (formerly the School of Molecular Bioscience) at the University of Sydney. I performed all experiments unless otherwise specified. This work has not been submitted, in part or in full, for the purpose of obtaining any other degree, or for any other purpose.

Ngaio C. Smith
February 2019

Abstract

LIM-homeodomain (LIM-HD) transcription factors act as key developmental regulators, both through their ability to bind DNA through homeodomain-DNA interactions, and through their ability to form larger complexes through protein-protein interactions. Many interactions that have been characterised are formed using their N-terminal LIM domains, but likely also involve other regions, which have not yet been described for many LIM-HD proteins.

The LIM-HD protein Isl1 has been implicated in the development of many tissues, such as the nervous system, heart, and pancreas. However, relatively little detail is known about how Isl1 functions in these systems and the pathways in which it acts. The first part of this thesis aimed to identify and characterise novel binding partners for Isl1. Close to 180 potential binding partners were isolated through use of yeast two-hybrid mating screens in an earlier project; over the course of this thesis, further methodology was developed to identify additional proteins in a medium throughput manner. Downstream validation protocols were then applied to determine which interactors were likely to represent biologically relevant interaction partners for Isl1.

The second part of this thesis focussed on the mechanisms by which Isl1 and another LIM-HD protein, Lhx3, play a role in cell fate determination in the developing central nervous system. These proteins, along with Ldb1, interact via LIM:LID interactions to form cell-specific transcriptional complexes that target genes different to those targeted by either LIM-HD protein alone. It was not known if the homeodomains target these different sites solely because of the LIM:LID interactions or if the homeodomains themselves bind cooperatively to DNA. The DNA-binding behaviour of various iterations of the Lhx3/Isl1/Ldb1 complex are described, and structural characterisation of the Isl1/Lhx3 DNA-binding unit has been pursued. These data provide new insights into the mechanisms by which Isl1 and Lhx3 work together in regulating gene expression.

Acknowledgements

Firstly I would like to acknowledge and thank Jacqui for supervising me over the duration of my PhD and giving me the opportunity to work on this project. You allowed me to pursue something that evolved much further than it was planned to, and I've learned and grown immensely from the experience.

Huge thanks also go to the people who helped to edit this thesis: Mario, Lorna, Ana, Jill, and Jacqui. I know you all had a lot of important things you had to do besides reading this, so I appreciate the energy and effort you gave me.

Thank you to Ann, Jill, Tim, and the other SAXS beamline scientists at the Australian Synchrotron, for helping me with all the SAXS data collection and analysis.

I'd also like to thank all the people of the Structural Biology discipline, for their advice and support. Every person that's been through the lab while I've been here has had something to teach me, and I really appreciate that. Special thanks to Mario, Lorna, Ana, Jason, and Flyp for many morning and afternoon teas talking about experiments and troubleshooting. Thanks also to my older Matthews brother Neil, for always being willing to chat and throw ideas around.

I'd also like to thank Angela, Ben, Trudie, and the autoclaving/glassware team, without whom most of this thesis would not have been possible. You make working in a lab so much easier and more pleasant, I think we'd all fall apart without you.

I acknowledge the support of an Australian Postgraduate Award/Research Training Program scholarship.

On a personal note, a lot of people have helped me along this journey, and I owe them a lot. As these are personal thanks that are owed, however, I will thank them in person, and not here.

Contents

List of Figures and Tables

Publications

Below are listed publications that arose from the work described in this thesis. While no content is taken directly from these publications, they form the basis for some of the work presented here.

Smith, N.C., Matthews & Matthews, J.M. *Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors*. Current Opinion in Structural Biology, 2016. **38**: p. 68-74.

This is a review researched by me and co-written by me and Professor Matthews. The content of this review contributes to various chapter introductions and discussions, as well as portions of Chapter 1.

Robertson, N.O., Smith, N.C., Manakas, A., Mahjoub, M., McDonald, G., Kwan, A. H. & Matthews, J.M. *Disparate binding kinetics by an intrinsically disordered domain enables temporal regulation of transcriptional complex formation*. Proceedings of the National Academy of Science, 2018. **115(**18): p. 4643-4648.

This paper describes efforts to characterise the kinetic interplay of LIM-HD transcription factors during spinal motor neuron development. DNA-binding experiments presented in this paper were performed by me, and can be found in Chapter 5, specifically Sections 5.5-5.8.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Ngaio Smith 14 March 2019

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Jacqueline Matthews 14 March 2019

Abbreviations

Abbreviation	Meaning
3-AT	3-amino-1,2,4-triazole
AD	Activation domain
APS	Ammonium persulfate
Art3	Ecto-ADP-ribosyltransferase 3
Asrgl1	Isoaspartyl peptidase/L-asparaginase
Bicine	2-(Bis(2-hydroxyethyl)amino)acetic acid
Bis-Tris	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
BLAST	Basic Local Alignment Search Tool
Bmp	Bone morphogenic protein
Bmp4	Bone morphogenetic protein 4
BSA	Bovine serum albumin
CD	Circular dichroism
cDNA	complementary DNA
ChIP	Chromatin immunoprecipitation
Chx10	Visual system homeobox 2
Cited2	Cbp/p300-interacting transactivator 2
Co-IP/MS	Co-immunoprecipitation coupled with mass spectrometry
Copb1	Coatomer subunit beta
Cops5	COP9 signalosome complex subunit 5
CRAPome	Contaminant Repository for Affinity Purification
CSM	Complete supplement media
CTLH	C-terminal to LisH
Cuz1	Cdc48-associated UBL/zinc finger protein-1
Cyc1	Cytochrome C1
DBD	DNA-binding domain
Ddx20/Dp103/Gemin3	DEAD-box helicase DDX20
Dfna5/Gsdme/Dfna5h	Gasdermin-E
Dkk1	Dickkopf-related protein 1
Dkk2	Dickkopf-related protein 2

DMSO	Dimethyl sulfoxide
dNTPs	Deoxyribonucleotide triphosphates
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylene ditetraacetic acid
Egr	Early growth response
EMSA	Electrophoretic mobility shift assay
Fgf10	fibroblast growth factor 10
Fhl1	Four and a half LIM domains protein 1
Foxa1	Hepatocyte nuclear factor 3-alpha
Foxa2	Hepatocyte nuclear factor 3-beta
GA	Proglucagon
Gata4	GATA binding protein 4
GST	Glutathione S-transferase
GSU	Glycoprotein hormone subunit α
h	hour
HA	Hemagglutinin
HEK	Human embryonic kidney
HEPES	4-(2-hydroxylethyl)-1-piperazineethanesulfonic acid
Hopx	homeodomain-only protein
HRP	Horseradish peroxidase
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl β-D-thiogalactopyranoside
Isl1/2	Islet-1/2
ITC	Isothermal titration calorimetry
Jmjd3	Lysine-specific demethylase 6B
kb	kilobase
Kctd9	BTB/POZ domain-containing protein KCTD9
K_d	Dissociation constant
kDa	kilodalton
Lace1/Afg1l	AFG1-like ATPase
LB	Luria-Bertani broth

Ldb1/2	LIM domain-binding protein 1/2
Lhx	LIM/homeobox protein
LID	LIM interaction domain
LIM	Domain found in Lin11, Isl1, Mec3
LIM-HD	LIM-homeodomain
Limk2	LIM kinase 2
LisH	Lissencephaly-1 homology
Lmo4	LIM domain only protein 4
Lmx1a/b	LIM homeobox transcription factor 1 alpha/beta
Lrrc51	Leucine-rich repeat-containing protein 51
Ly6c1	Lymphocyte antigen 6C1
mAb	monoclonal antibody
MALLS	Multi-angle laser light scattering
MBP	Maltose binding protein
MCS	Multiple cloning site
Mec3	Mechanosensory protein 3
Mef2c	Myocyte-specific enhancer factor 2C
Mep1b	Meprin A subunit beta
MES	2-ethanesulfonic acid
min	minute
Mkln1	Muskelin
Mnx1	Motor neuron and pancreas homeobox protein 1
MPD	(+/−)-2-Methyl-2,4-pentanediol
MQW	MilliQ® water
Ms4a5	Membrane-spanning 4-domains, subfamily A, member 5
MST	Microscale thermophoresis
MWCO	Molecular weight cut-off
NeuroD1	Neurogenic differentiation factor 1
Nkiras1	NF-kappa-B inhibitor-interacting Ras-like protein 1
Nkx2.5	Homeobox protein Nkx-2.5
NLS	Nuclear localisation sequence
NMR	Nuclear magnetic resonance
Nup50	Nuceloporin 50

Oct1	Octamer-binding protein 1
Oct4	Octamer-binding protein 4
OD_{600}	Optical density at 600 nm
Oscp1	Oxidored-nitro domain-containing protein 1
Pax6	Paired box protein 6
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline supplemented with 0.1% Tween
PCR	Polymerase chain reaction
Pbx	Pre-B-cell leukaemia transcription factor
Pdx1	Pancreas/duodenum homeobox protein 1
PEG	Polyethylene glycol
PEI	Polyethylimine
Phox2a	Paired mesoderm homeobox 2A
Pias4	Protein inhibitor of activated STAT protein gamma
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
PMSF	Phenylmethanesulfonyl fluoride
poly dG	Poly deoxyguanine
poly dT	Poly deoxythymine
Pou4f2	POU domain, class 4, transcription factor 2
RanBPM	Ran-binding protein M
R_g	Radius of gyration
Rnf167	E3 ubiquitin-protein ligase RNF167
Rps18	40S ribosomal protein S18
Rps23	20S ribosomal protein S23
Rps26	20S ribosomal protein S26
Rps29	20S ribosomal protein S29
s	second
S. cerevisiae	Saccharomyces cerevisiae
SA	Self-association domain
SAXS	Small angle X-ray scattering
SD	Synthetic dextrose broth
SDS	Sodium dodecyl sulfate
Scpep1	Retinoid-inducible serine carboxypeptidase

SEC-MALLS	Size exclusion chromatography coupled to multi angle laser light scattering
SEC-SAXS	Size exclusion chromatography coupled to small angle X-ray scattering
SELEX	Selective evolution of ligands by exponential enrichment
Set7/9	Pancreas/duodenum homeobox protein 1
Shh	Sonic Hedgehog
snRNP	small nuclear ribonucleoprotein
Sparc	Secreted protein, acidic and rich in cysteine
Sparcl1/Ecm2/Sc1	SPARC-like protein 1
Spata7	Spermatogenesis-associated protein 7 homolog
SPR	Surface plasmon resonance
Ssbp3	Single-stranded DNA-binding protein 3
Stam1	Signal transducing adapter molecule 1
Stat3	Signal transducer and activator of transcription 2
SV40	Simian virus 40
Tbx	T-box
Tbx20	T-box transcription factor 20
TCEP	tris(2-carboxyethyl)phosphine
TEMED	N,N,N′,N′-Tetramethylethylenediamine
Tigd2	Tigger transposable element-derived protein 2
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
Trpc1	Transient receptor potential channel 1
Ubd/Fat10	Ubiquitin D
Usp8	Ubiquitin carboxyl-terminal hydrolase 8
UTR	Untranslated region
Wif1	Wnt inhibitory factor 1
X-α-gal	5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside
YPD	Yeast peptone dextrose broth
Zdhhc20	Palmitoyltransferase ZDHHC20
Zfand1	AN1-type zinc finger protein 1

1 Introduction

1.1 Developmental gene regulation

In order to correctly control the development of an organism, the genomes of higher order eukaryotic organisms have evolved to rely on many layers of gene regulation [1, 2]. One of the most prevalent of these is the interplay of transcription factors that promote or repress expression of specific genes at distinct times during development [3]. Many transcription factors have individual DNA-binding preferences, but most commonly the action of multiple transcription factors acting in a complex is required to precisely modulate gene expression, especially in the context of development [4, 5]. Dysregulation of these systems leads to developmental defects, and anomalies of the temporal control of these systems can also lead to disease states such as cancer in fully developed organisms [6, 7]. Understanding systems that regulate gene expression can thus give us deep insights into how organisms develop, and how to prevent or correct instances of dysregulation.

1.2 The LIM-homeodomain family of transcription factors

The LIM-homeodomain (LIM-HD; LIM derived from Lin11, Isl1, Mec3) proteins are a family of transcription factors found in animals. This family is highly conserved, with all proteins implicated in mammalian development (Table 1.1) [8]. The knock out or knock down of many of the LIM-HD proteins have severe impact, and can be lethal, or lead to developmental disorders [9]. Many LIM-HD proteins play a role in development in multiple tissues (Table 1.1). This is especially the case with the LIM-HD protein Isl1, which is the focus of this thesis.

Table 1.1: The role of LIM-HD proteins in mammalian development.

LIM-HD protein	Developmental tissue	Knockout phenotype
Isl1	Discussed in detail in Section 1.4	Embryonic lethal [10]
Isl2	Motor neurons [11] Retina [12]	Death <24 hours after birth [13]
Lmx1a	Forebrain [14] Spinal cord [15]	Behavioural and cognitive defects [16]
Lmx1b	Eye [17] Spinal cord [15] Limb patterning, kidneys [18]	Nail-Patella syndrome [19]
Lhx1	Cerebellum [20] Early definition of body axes [21] Forebrain [22]	Embryonic lethal [23]
Lhx2	Eye, cerebral cortex, erythrocytes [24] Olfactory neurons [25]	Embryonic lethal [24]
Lhx3	V2a Interneurons [26] Inner ear [27] Pituitary gland [28, 29] Motor neurons [30]	Death <24 hr after birth [31]
Lhx4	Pituitary gland [28, 29] Motor neurons [30]	Death <15 min after birth [32]
Lhx5	Cerebellum [20] Forebrain [22]	Death a few days after birth, or severe cognitive defects [33, 34]
Lhx6	Forebrain, teeth [35]	Unknown
Lhx7/8	Teeth [35, 36] Forebrain [35]	Death within 24 hours after birth, or cleft palate [37]
Lhx9	Gonads [38] Cerebral cortex [39] Forebrain [40]	Urogenital defects, entirely female phenotype [38]

1.2.1 Similarities within the LIM-HD family

Phylogenetic analysis reveals distinct paralogue pairs (i.e., closely related genes that probably arose through gene duplication) within the family of LIM-HD proteins (Figure 1.1) [41]. These pairs are found to have similar, but not entirely redundant, functional roles [11, 15, 30, 35, 39, 42].

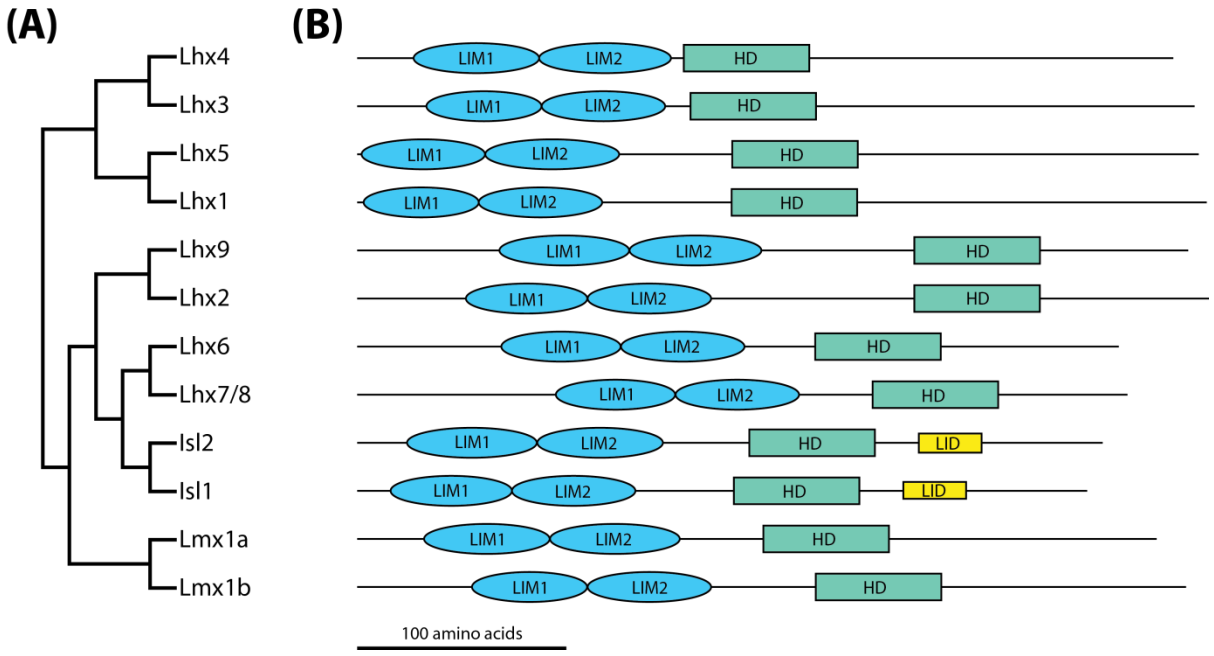


Figure 1.1: The family of LIM-HD proteins. (A) Phylogenetic tree of LIM-HD proteins. (B) Domain organisation of the LIM-HD proteins from (A).

The paralogous pairs of LIM-HD proteins have a high level of evolutionary conservation, with all mammals possessing six paralogous pairs of LIM-HD proteins (Figure 1.1A) [43]. This suggests that both paralogues are necessary for normal growth or function, and therefore it is possible that the redundant pairs have evolved to act in spatially or temporally separate environments, but with very similar functions.

1.2.2 Structural features of LIM-HD proteins

The LIM-HD family of proteins derives its name from the two domains contained in all family members: an N-terminal pair of closely spaced LIM domains, and a downstream homeodomain (Figure 1.1B) [44, 45]. These motifs are crucial for forming interactions that facilitate the function of the proteins. The LIM domains are protein-protein interaction motifs, allowing association with other regulatory proteins [45]. The homeodomain is a well-characterised DNA binding domain, often found in developmental transcription factors [46].

1.2.2.1 LIM domains and LIDs

A LIM domain is a zinc finger that contains two tandemly arrayed zinc-coordinating motifs [47]. Each zinc-coordinating motif chelates two zinc ions, through cysteine and histidine residues (Figure 1.2). Unlike classical DNA-binding zinc fingers, LIM domains appear to act solely as protein-protein interaction domains [45, 48]. Working with isolated LIM domains from LIM-HD and related LIM-only (Lmo) proteins has proved difficult to date, as recombinantly expressed forms of these proteins tend to aggregate when not in the presence of a binding partner, suggesting they may be obligate binding domains [49, 50].

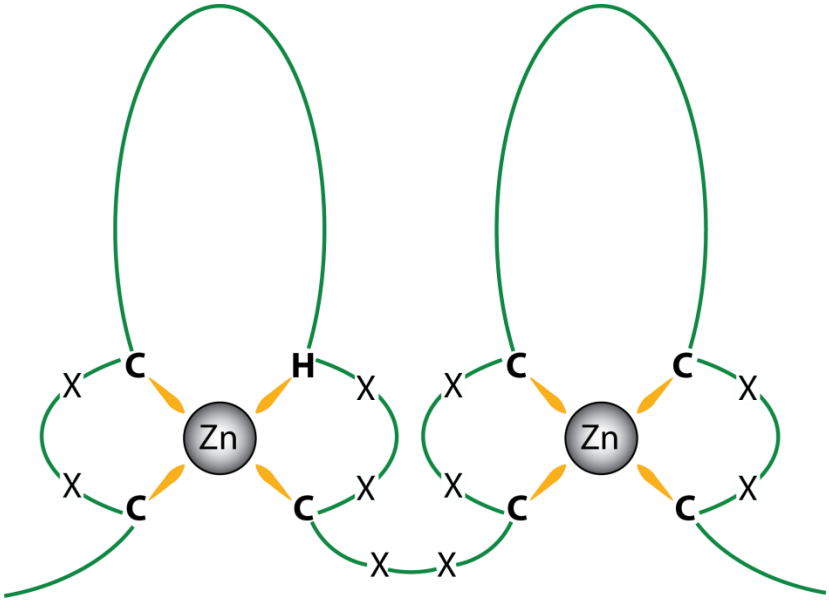


Figure 1.2: Structural schematic of a LIM domain, showing residues important for zinc ion coordination.

The motif that has been found to most commonly interact with LIM domains from the LIM-HD family of proteins is the LIM interaction domain (LID). The most ubiquitous of these is the LID found in LIM domain-binding protein 1 (Ldb1) [51]. Many structures of LIM-HD/Lmo LIM domains in complex with a LID from Ldb1 or another protein have been published, revealing a very consistent binding mechanism, despite low levels of sequence conservation (Figure 1.3) [52-57]. Isolated LIDs tend to be intrinsically disordered, and fold upon binding to LIM domains [58]. When bound, the LID forms an extended structure, lying across the surface of the LIM domains, contributing an extra strand to the beta-sheets of the zinc binding modules that make up the LIM domains [53, 58]. The LID binding surface on the LIM domains is referred to as the peptide-binding interface [50].

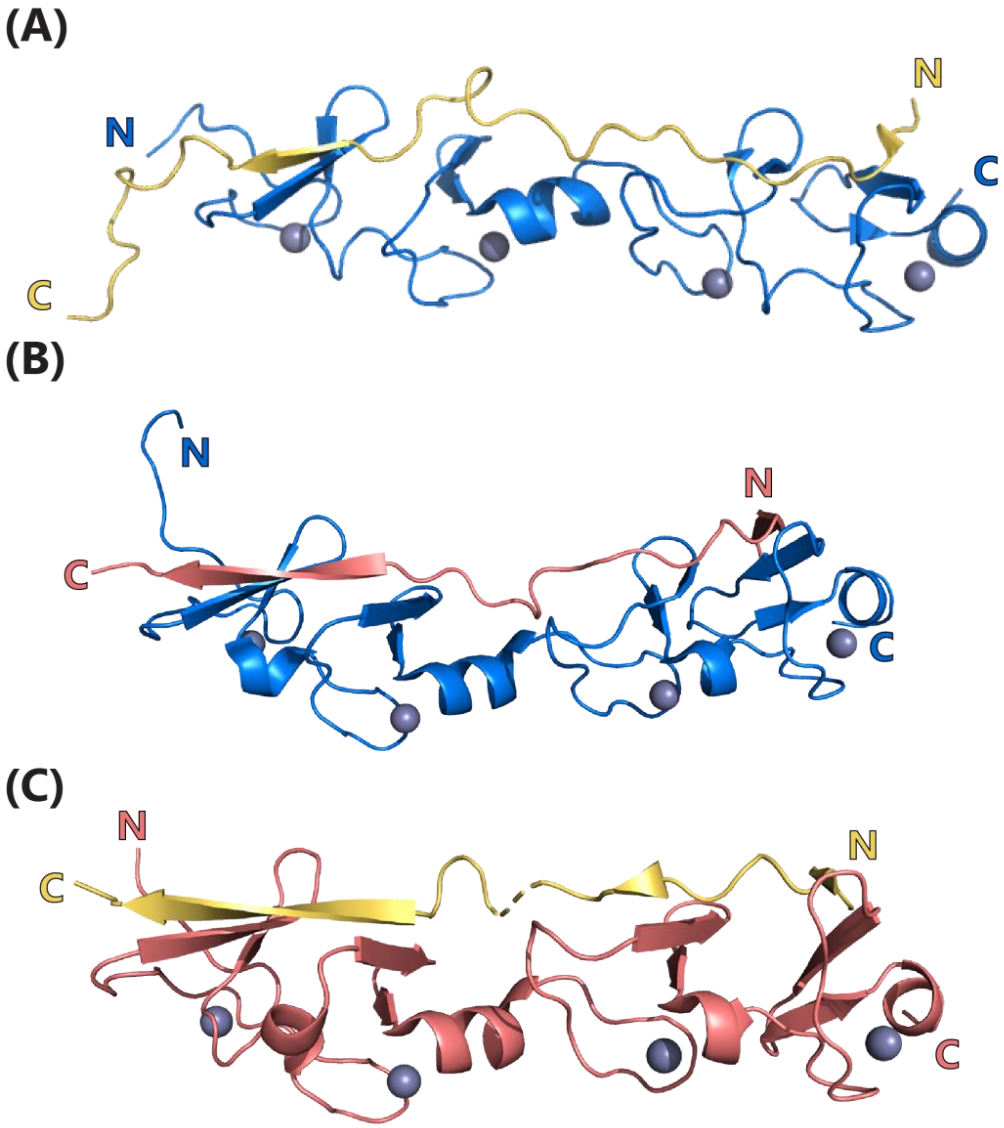


Figure 1.3: Structures of LIDs bound to LIM domains. N- and C-termini are labelled N and C respectively. (A) $Ldb1_{LID}$ (yellow) bound to $Lhx3_{LIM}$ (blue) (PDB: 2JTN); (B) $Isl1_{LID}$ (salmon) bound to $Lhx3_{LIM}$ (blue) (PDB: 2RGT); (C) $Ldb1_{LID}$ (yellow) bound to $Isl1_{LIM}$ (salmon) (PDB: 4JCJ). Zinc ions are shown in grey.

1.2.2.2 Conservation of the LIM domain

LIM-HD proteins are the focus of this thesis, but there are many other proteins that contain LIM domains [48, 59]. There are an additional 13 families of LIM domain containing proteins, with diverse functions and subcellular localisations [48, 59, 60]. Many of these proteins are involved in regulating actin dynamics in the cytoskeleton, but several families can also translocate between the cytoskeleton and the nucleus, depending on their protein binding partners [59, 60]. LIM-HD proteins are conserved across metazoan (animals), but

some families of LIM domain containing proteins are conserved across filozoa (animals and their nearest unicellular relatives) [60].

1.2.2.3 Homeodomains

The homeodomain is a highly conserved ~60 amino acid DNA-binding domain, found throughout the kingdom of animals, as well as in fungi and plants [46, 61-64]. It contains three helices that pack in a helix-turn-helix motif, with the third helix being the DNA-binding recognition helix (Figure 1.4) [65]. The N-terminal residues of the homeodomain also contribute to DNA binding, through interacting with the minor groove of the DNA (Figure 1.4 B) [63]. These residues are disordered when not in complex with DNA [66].

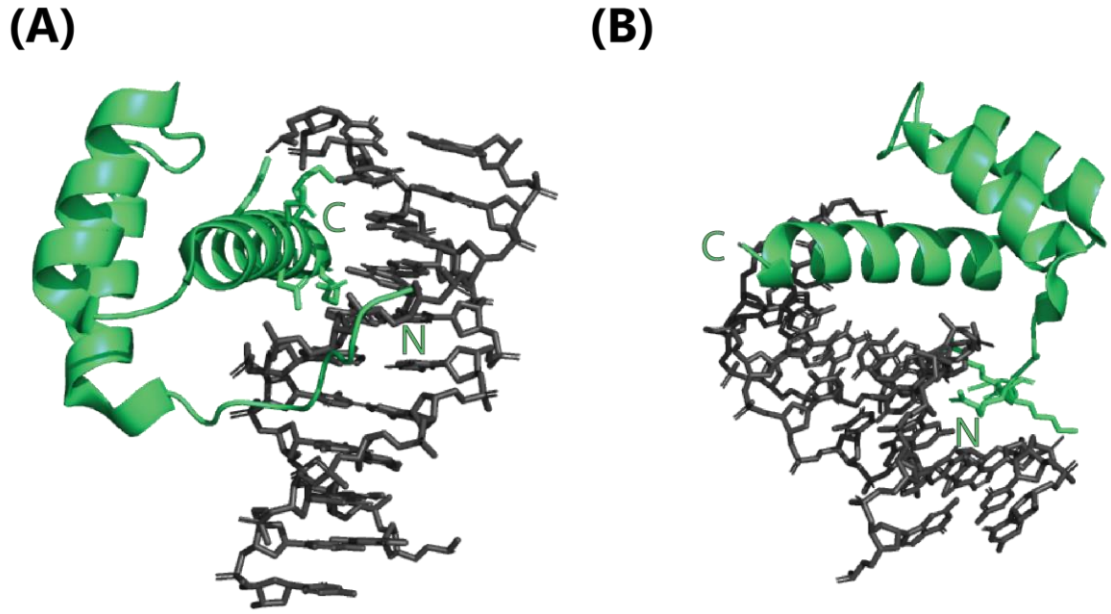


Figure 1.4: Structure of a homeodomain bound to DNA. The homeodomain from LIM-HD protein Lhx4 is shown, bound to its consensus sequence. N- and C-termini are labelled N and C respectively (PDB: 5HOD). (A) Interaction between Helix 3 of the homeodomain and DNA. (B) Interaction between the N-terminal tail of the homeodomain and the minor groove of DNA.

The majority of homeodomain structures available show the protein bound to DNA and were determined through the use of X-ray crystallography [67]. The presence of DNA can aid in crystallographic structure determination of this type of protein by providing a favoured conformation for the protein and facilitating packing in crystals, due to the opposing charges of the two macromolecules.

Homeodomains from different proteins tend to have very similar sequence preferences for DNA binding, due to their highly conserved structure and binding mechanism [68]. Generally, homeodomains target a TAATXX sequence, with the final two bases varying between individual homeodomains [69]. Electrostatic forces are the primary driving force for DNA binding by the recognition helix [70]. There is relatively little distortion of the DNA double helix upon homeodomain binding [63].

As homeodomains target sequences that are only six base pairs long, it is very common to observe homeodomain-containing proteins in higher order complexes with other DNA-binding proteins, presumably to achieve specific binding of chromatin [66]. The LIM-HD family of proteins are a prime example of this, as they can make both protein-DNA and protein-protein interactions to form higher order transcriptional complexes.

1.3 Isl1

Islet-1 (Isl1) was first discovered in the pancreatic islets [44]. It was one of the first LIM domain containing proteins discovered, contributing to the naming of the domain along with two other LIM-HD proteins: **l**in-11 (LIM homeobox gene lin-11), **I**sl1, and **m**ec-3 (mechanosensory protein 3) [71]. In addition to the LIM domains and homeodomain, Isl1 contains a C-terminal LID (Figure 1.5A) [44, 53, 58]. There is also a short glutamine-rich region immediately downstream of the homeodomain with no known function (not shown in Figure 1.5).

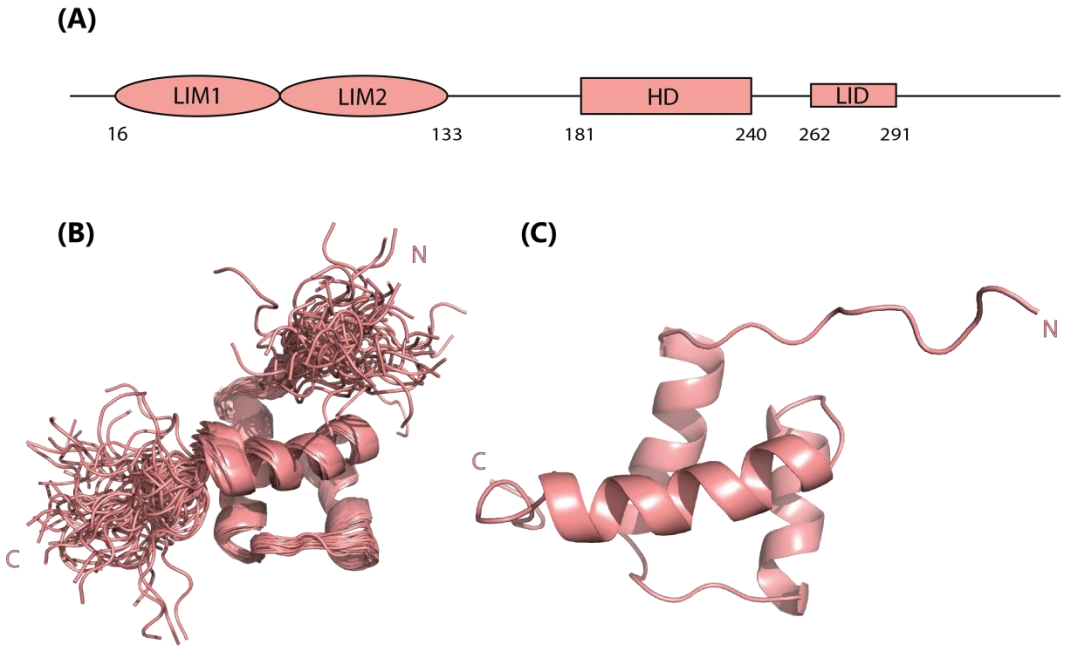


Figure 1.5: Structural features of Isl1. (A) Schematic of domain organisation of Isl1. (B) Full ensemble of 50 models of the NMR solution structure of $Isl1_{HD}$ (PDB:1BW5). (C) Representative structure of $Isl1_{HD}$ (PDB: 1BW5).

The solution structure of the homeodomain from Isl1 was determined by NMR (nuclear magnetic resonance) techniques (Figure 1.5B and 1.5C) [72]. This structure depicts $Isl1_{HD}$ in the absence of DNA. Regions normally involved in DNA binding (The N-terminus, as well as the C-terminal end of helix 3) appear more flexible in this structure than in homeodomain-DNA structures (Figure 1.5B). Apart from these differences, $Isl1_{HD}$ shows a canonical homeodomain structure.

Isl1 and its paralog Isl2 are the only LIM-HD proteins known to possess both a LID and a pair of LIM domains [53]. This combination potentially allows intramolecular LIM:LID binding to occur. Such an interaction could prevent non-specific, or weak, binding of the LIM domains to inappropriate partner proteins [54]. In this way, an extra layer of regulation may exist within Isl1/2, ensuring that only appropriate interactions occur in the cell. Structures that contain each of the LIM domains and the LID of Isl1 in complex with partner proteins have been determined (Figure 1.3B and 1.3C) [54, 58]. The presence of a LID in the sequence of Isl1 provides a second motif for interaction with other proteins, allowing formation of higher order transcriptional complexes. An example of this can be seen in the development of motor neurons, as discussed in detail in Section 1.4.1 [13].

1.3.1 Conservation of Isl1

Isl1 is highly conserved in animals, with well characterised homologs present in mice (100% identical to the human Isl1), *Drosophila melanogaster* (59% identical to the human Isl1) and *Caenorhabditis elegans* (38% identical to the human Isl1), as well as lesser characterised homologs in basal metazoan organisms [8, 73, 74]. An alignment of the human, *D. melanogaster*, and *C. elegans* sequences is shown in Figure 1.6.

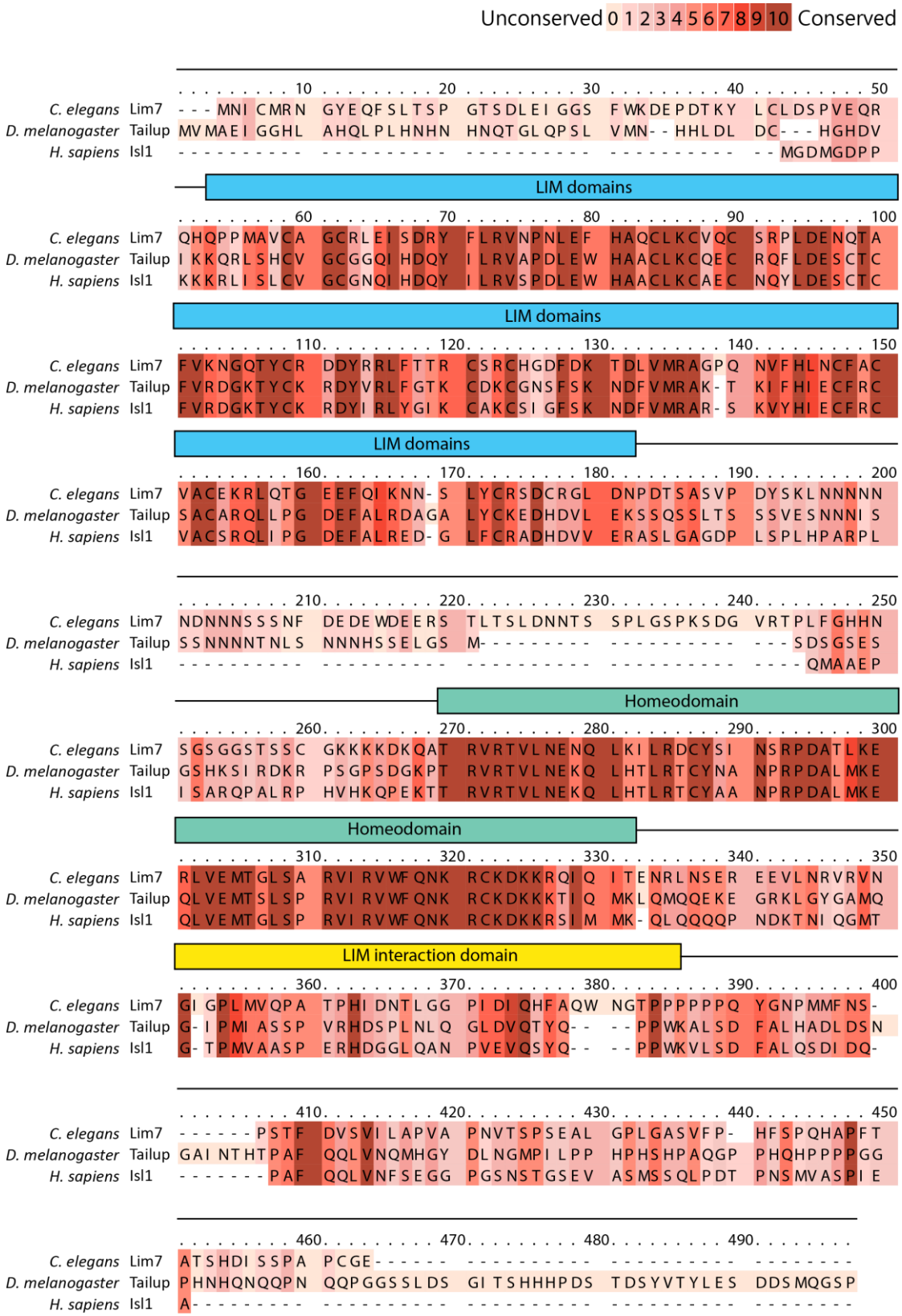


Figure 1.6: Sequence alignment of Isl1 homologs. Alignment generated using Praline [75].

Very few genetic variants of Isl1 have been identified in humans. The majority of those observed are silent mutations, with only two producing a change in the amino acid sequence of the protein [76, 77]. The first, Q310X, is a mutation that was discovered in diabetes patients that results in the production of truncated Isl1 [77]. The C-terminal 40 amino acids are lost in this mutant, but there is no effect on the LIM domains, homeodomain, or LID. The effect of this mutation appears to be minimal, although it may contribute to a diabetic phenotype. The second mutation, R252S, was observed in cardiomyopathy patients. This substitution lies immediately C-terminal to the homeodomain, and some evidence suggests the mutant may activate transcription more potently [76].

1.4 The role of Isl1 in development

Isl1 expression was detected in adult pancreatic, retinal, and neuronal tissues [44, 78, 79]. However, it is expressed in a wide variety of tissues during development, where its role has been more comprehensively characterised [80]. Isl1 is crucial for embryonic development; Isl1 knockout mice embryos exhibit arrested development at E9.5, followed by necrosis around E11.5 [10]. Further evidence for the early requirement of Isl1 is the presence of Isl1 expression in the zebrafish as early as gastrulation, the stage at which cells differentiate into the endoderm, mesoderm, and ectoderm [81].

In most instances, Isl1 functions to promote the differentiation of tissues during development. In human embryonic stem cells, the promoter of *Isl1* is occupied by transcription factors crucial for maintaining pluripotency: Oct4 (Octamer-binding protein 4), Sox2, and homeobox protein Nanog [5]. In such cells, expression of Isl1 is repressed by this trio of transcription factors. Aberrant expression of Isl1 is associated with several types of cancer, with both upregulation and downregulation of expression being observed [82-85]. Isl1 has roles in the development of many organs. Each of these roles has been characterised to varying extents, as discussed below.

1.4.1 Isl1 in motor neuron development

The action of Isl1 in the development of motor neurons is the most well understood of its roles. Indeed, Isl1 is the first molecular indicator of spinal motor neuron differentiation [86, 87]. Motor neuron development begins with the formation of the floor plate in the embryonic neural tube (Figure 1.7). Shh (Sonic Hedgehog) is secreted in the notochord at the ventral end

of the neural tube, creating a signalling cascade that triggers the differentiation of these cells into the floor plate [88, 89]. These cells then secrete more Shh, in a second signalling cascade [90]. Meanwhile, BMP (bone morphogenic protein) cytokines are secreted at the dorsal end of the neural tube, [88]. The diffusion of these signalling proteins along the dorsoventral axis establishes a gradient that dictates the specification of cells along the length of the neural tube (Figure 1.7B) [88].

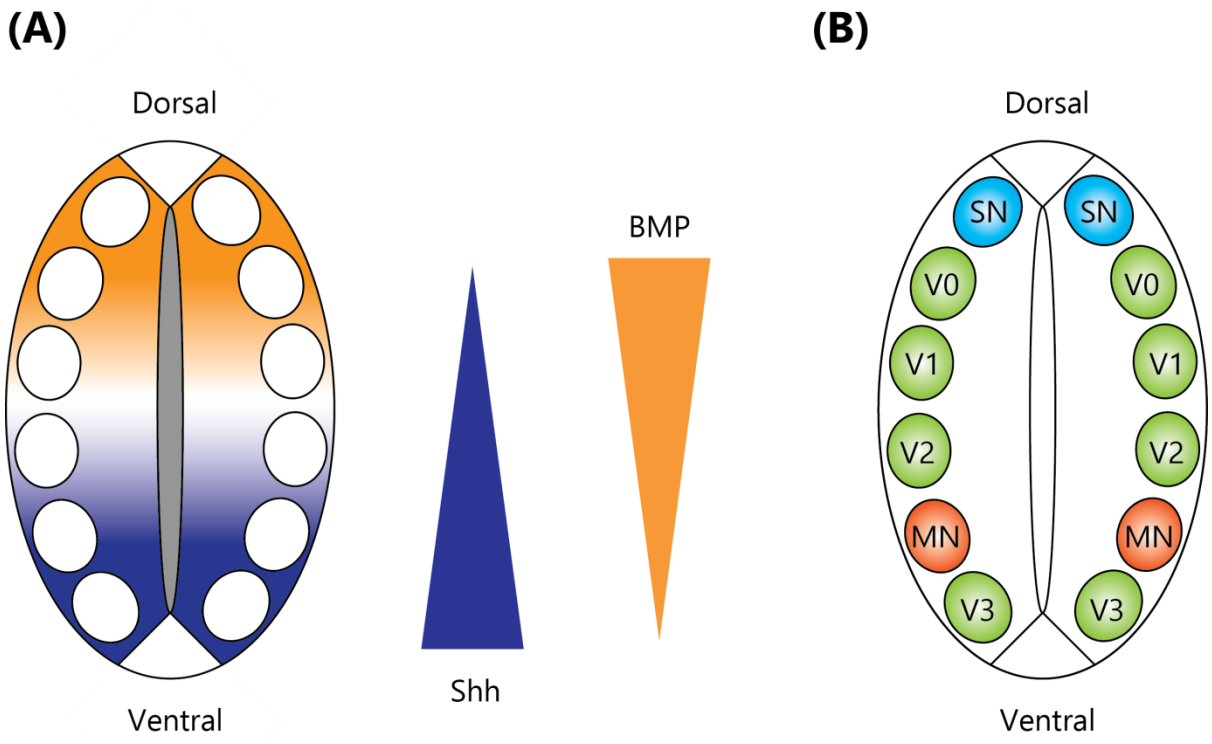


Figure 1.7: Schematic of the developing notochord. (A) Concentration gradients of Sonic hedgehog (Shh) and bone morphogenic proteins (BMP) establish cell populations along the neural tube. (B) Differentiated identities of neural tube neurons. SN: sensory neurons; V0-3: ventral interneurons; MN: motor neurons. Figure based on Figure 8 from [88].

Cells destined to become V2 interneurons express Ldb1 and Lhx3 (LIM homeobox protein 3) [30, 91]. In the band of cells immediately ventral to this population, Isl1 expression is also induced, which is sufficient to redirect cell fate towards a ventral motor neuron identity [10, 86, 92, 93]. The exact mechanism through which Isl1 expression is induced remains unknown, although Onecut proteins have been shown to be necessary for maintaining its expression, and Phox2a (Paired mesoderm homeobox protein 2A) binds to *Isl1* enhancer elements in developing neurons [94, 95]. Direct Shh signalling represses expression of Isl1, indicating that intermediate transcription factors must be produced prior to induction of Isl1

expression [96]. As only a twofold difference in Shh concentration is required to result in Isl1 expression, this additional level of regulation may be necessary to correctly direct Isl1 expression [97].

In cells destined to become motor neurons, Isl1 forms a ternary transcriptional complex with Lhx3 and Ldb1 (Figure 1.8A) [98-100]. This complex is characterised by sets of LIM:LID interactions, the self-association of Ldb1, and homeodomain-DNA interactions. When Isl1 is absent, Lhx3 and Ldb1 form a binary complex, promoting the expression of genes that induce differentiation into V2 spinal interneurons (Figure 1.8B) [13].

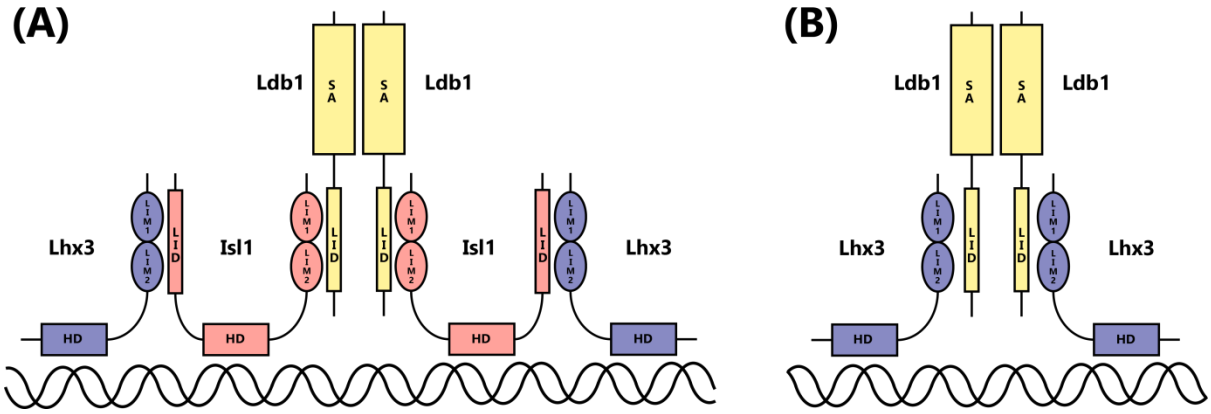


Figure 1.8: Schematics of LIM-HD neuronal transcriptional complexes. (A) Ternary Lhx3/Isl1/Ldb1 complex formed in cells destined to become motor neurons. $Isl1_{LID}$ interacts with $Lhx3_{LIM}$, $Isl1_{LIM}$ interacts with $Ldb1_{LID}$, and the homeodomains from Isl1 and Lhx3 bind DNA. The Ldb1 self-association domain (SA) facilitates higher order complex formation. (B) Binary Lhx3/Ldb1 complex formed in cells destined to become V2 interneurons, in which $Lhx3_{LIM}$ interacts with $Ldb1_{LID}$.

The evidence for the existence of these complexes comes primarily from *in vivo* experiments using various mutants and chimeras [93, 98, 101]. Although structural information is available, or can be inferred by homology, for the individual domains and the LIM:LID subcomplexes, questions remain about the assembly of the ternary complex and how the different complexes target different DNA sequences to direct development.

1.4.1.1 DNA-binding of LIM-HDs in spinal neuron development

It is known that the ternary complex targets sites with a CATTAXXXAAATTA consensus sequence [99]. This sequence was initially identified in the promoters of key downstream targets of the ternary complex, and was later confirmed as a consensus sequence through

SELEX (selective evolution of ligands by exponential enrichment) experiments [98, 99]. The individual binding preferences of Lhx3 and Isl1 have also been explored. Lhx3 has a preference for binding to TAATTA sites, and this sequence is also bound by Lhx4, as shown in the crystal structure of Lhx4 bound to DNA (Figure 1.4) [102, 103]. Two different consensus sequences have been determined for Isl1: CATTAG from SELEX [99], and TAATAT from *in vitro* promoter binding studies [104]. It was suggested that the DNA binding preferences of Isl1 depend on the protein binding partners it is associated with [105].

1.4.1.2 Downstream targets of the motor neuron ternary complex

Several known downstream targets of the Isl1/Lhx3/Ldb1 motor neuron complex have been shown to be essential for the next step towards motor neuron development. *Isl1* and *Lhx3* are both targeted by the complex, generating a positive feedback loop [106]. The ternary complex also represses expression of Chx10 (Visual system homeobox 2), a protein that is important in further V2 interneuron development [107]. Promoted targets include *mir-218*, Lmo4 (LIM domain only protein 4), Stam1 (Signal transducing adapter molecule 1), and Mnx1 (Motor neuron and pancreas homeobox protein 1) [98, 106, 108-110]. Lmo4 is involved in modulating the formation of the ternary complex, by competing for binding to Ldb1 [106, 111]. Stam1 is involved in axonal projections, and Mnx1 has been shown to repress genes related to interneuron development, helping direct cells towards a motor neuron identity [112]. Stat3 (Signal transducer and activator of transcription 3) was identified as a coactivator and a potential binding partner of the Lhx3/Isl1/Ldb1 transcriptional complex [113].

1.4.2 Isl1 is necessary for correct heart development

The role of Isl1 in heart development was discovered when mouse *Isl1* deletion mutants were shown to have impaired vascular development, especially in the formation of the dorsal aorta [10]. These developmental defects were thought to be the cause of the embryonic lethality of the mutant mice. Upon further study, Isl1 expression was found to be necessary for the correct formation and survival of the outflow tract, right ventricle, and both atria of the heart [114, 115].

Heart development is an extremely intricate process (comprehensively reviewed in [116, 117]). During early development of the heart, two distinct cell populations are established in

the cardiac mesoderm, called the first and second heart fields. The first heart field consists of cells that begin to differentiate early, to initially form the cardiac crescent, which later becomes the linear heart tube (Figure 1.9). Second heart field cells differentiate later, migrating into the developing heart tube after its initial formation (Figure 1.9B). In the mature heart, the left ventricle is made primarily from first heart field cells, but most other structures originate from both first and second heart field cells [118, 119]. Isl1 expression is used as a marker of cells from the second heart field, as it is expressed in all second heart field cells [114, 120-122]. It has also been observed in some cells from the first heart field, but does not appear to be required for their development [123].

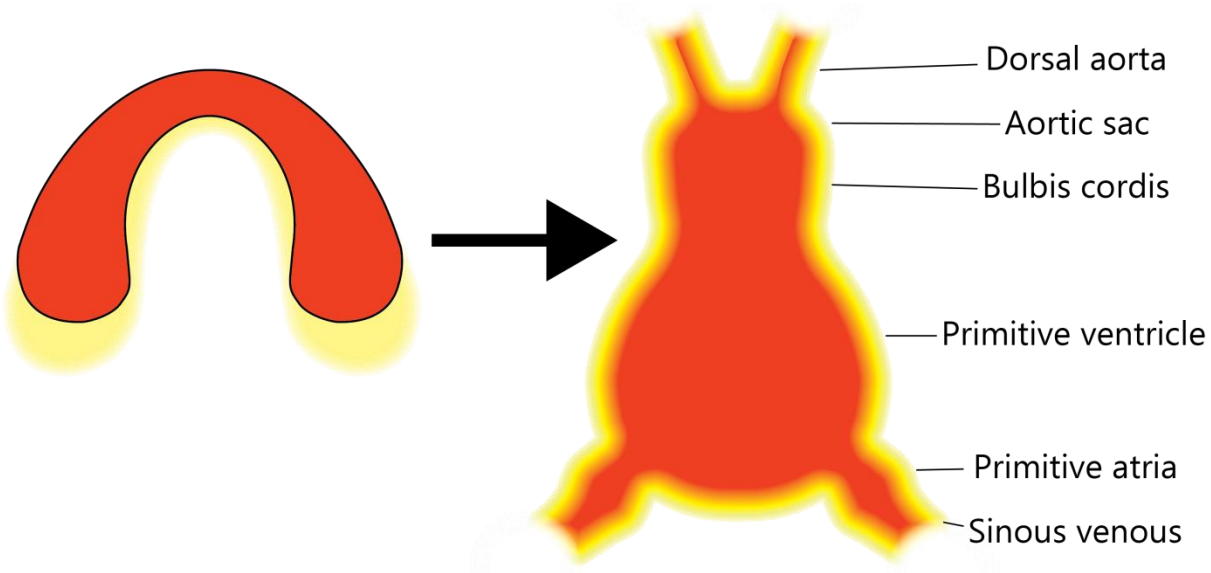


Figure 1.9: Early stages of heart development. The first (red) and second (yellow) heart fields are shown. (A) The cardiac crescent is formed from first heart field cells with second heart field cells located below it. (B) The cardiac crescent forms the heart tube. Second heart field cells (where Isl1 is expressed) migrate into this structure and begin to differentiate alongside first heart field cells, with both contributing to most mature heart tissues. Figure based on Figure 1 from [116] and Figure 1 from [124].

Differentiated heart tissue shows no Isl1 expression, indicating that Isl1 plays a role only during the development of the heart, and not in its ongoing function [114]. The exact role of Isl1 in heart development is still unclear, but various downstream and upstream effectors have been identified that can help to establish the role Isl1 is playing in these tissues. In contrast to motor neuron development, there is evidence that Isl1 acts upstream of Shh signalling during heart development [125].

Mef2c (Myocyte-specific enhancer factor 2C) was found to be a direct target of Isl1 in the anterior heart field, the tissue that later contributes to the outflow tract and right ventricle of the developing heart [126]. *Mef2c* null mutant mice show an almost identical phenotype to *Isl1* null mutants, indicating that the deletion of Isl1 could be exerting an affect primarily through the lack of expression of Mef2c [127]. An enhancer element has been identified that activates expression of Mef2c when bound by Isl1, Tbx20 (T-box transcription factor 20), and GATA zinc finger transcription factors [114, 126, 128]. Tbx20 has also been implicated in motor neuron development, potentially cooperating with Isl1 to induce expression of Mnx1 and Isl2 [128]. *Fgf10* (fibroblast growth factor 10) and *Nkx2.5* (Homeobox protein Nkx-2.5) are direct targets of Isl1, in concert with GATA binding protein 4 (Gata4) and Tbx proteins [128-130]. A potential interaction between Isl1 and Jmjd3 (Lysine-specific demethylase 6B), a histone demethylase protein, may also promote expression of downstream targets [131].

The expression of Isl1 itself appears to be regulated in heart tissues by forkhead transcription factors [132]. An evolutionarily conserved enhancer element from the *Isl1* gene was shown to contain multiple binding sites for several different forkhead transcription factors [132]. Other proteins that directly promote Isl1 expression in developing heart tissues are Octamer-binding protein 1 (Oct1), β-catenin, and Cbp/p300-interacting transactivator 2 (Cited2) [133-136]. Isl1 can interact with Cited2, further driving cardiac differentiation [136].

Recently, the Isl1/Ldb1 complex was demonstrated to be relevant to heart development [137]. Given the ability of Isl1 to simultaneously interact with two binding partners through both its LIM and LID domains, together with the absence of Lhx3 expression in the developing heart, an alternate binding partner could form a ternary transcription factor complex with Isl1 and Ldb1 that targets genes associated with heart development [138]. One potential binding partner could be from the Ajuba family of LIM domain proteins, which regulate Isl1 activity in developing heart cells [139].

1.4.3 Development of the pancreas requires Isl1

As noted above (Section 1.3), the discovery of Isl1 in pancreatic islets gave the protein its name [44]. During embryonic development, the pancreas forms as two buds along the gut tube, one of the first endodermic structures formed in embryogenesis (Figure 1.10) [140]. A series of signalling cascades then assign three different cell types to the growing pancreas:

endocrine cells, exocrine cells, and duct cells [141, 142]. The endocrine cells differentiate further into different types of islet cells, which are responsible for secreting pancreatic hormones like insulin, glucagon, and somatostatin.

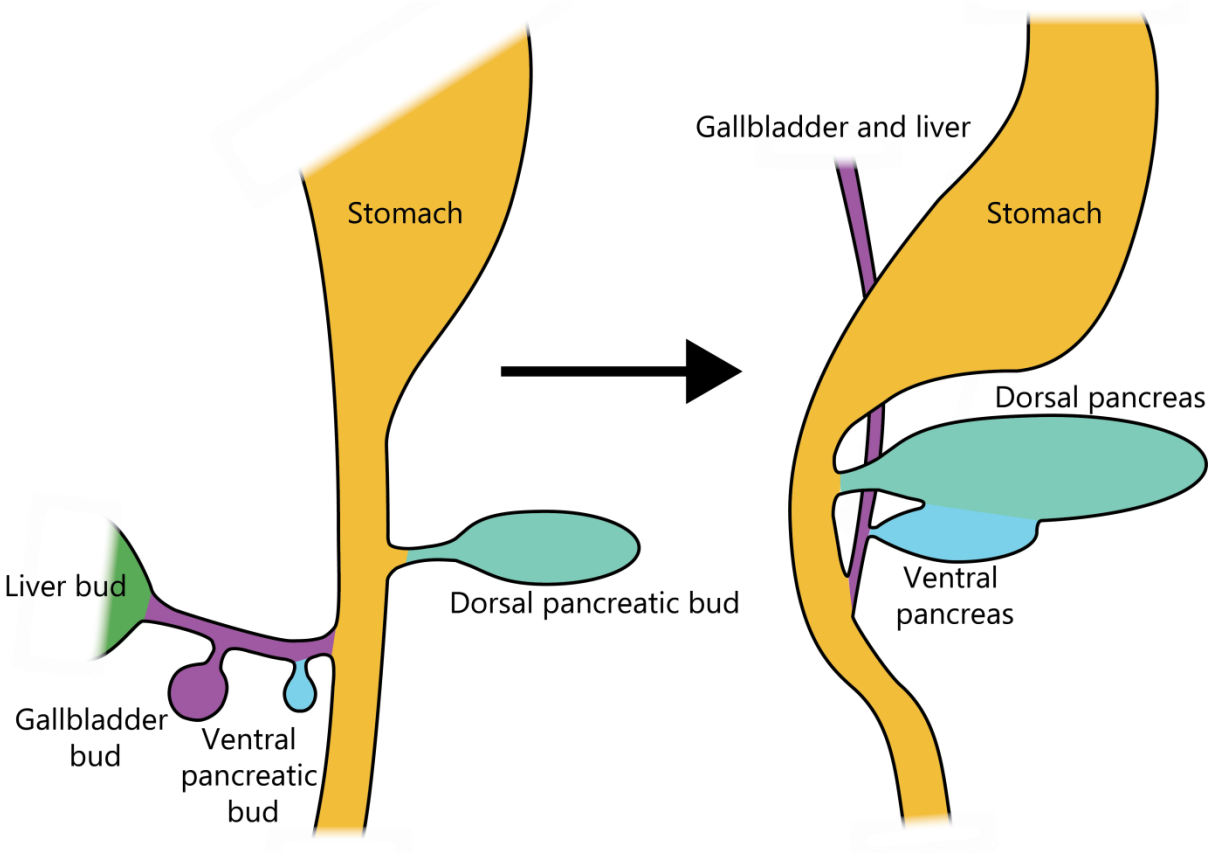


Figure 1.10: Formation of the pancreas. Along the gut tube (gold) the pancreas forms as two buds, one dorsal (aqua) and one ventral (blue). As embryogenesis progresses, rotation of the gut tube occurs, allowing the ventral and dorsal pancreatic buds to merge into one structure.

Pancreatic Isl1 deletion embryos exhibited two main pancreatic abnormalities: malformed dorsal pancreatic mesenchyme, and an absence of glucagon-expressing islet cells in the dorsal pancreatic epithelium [143, 144]. Once born, these mutants also show a reduction in islet cells [144]. Further studies revealed that Isl1 plays an important role in differentiating islet cells, immediately after they exit the cell cycle [44, 78, 143]. There are conflicting reports as to whether Isl1 plays a role in the expansion of this cell population, or only its maintenance [145-148]. Functionally, Isl1 was shown to regulate glucagon expression in mature islet cells [104].

More details about the mechanisms of action of Isl1 in pancreatic development remain elusive. Shh was shown to play a role in pancreatic development, and it appears to be responsible for initiating Isl1 expression in the developing pancreatic islets [149]. This may indicate a similar mechanism for inducing Isl1 expression as the one found in the developing central nervous system (Section 1.4.1).

Several potential downstream targets of Isl1 in pancreatic cells have been identified. Paired box protein 6 (Pax6), another homeodomain transcription factor required for both the development of islet cells and the production of glucagon, is not detected when Isl1 expression is knocked down, and deletion of either *Isl1* or *Pax6* results in a similar phenotype [150, 151]. Isl1 can also bind to an enhancer of *MafA*, a key transcription factor involved in the downstream development of pancreatic endocrine tissue [144, 152].

Several candidates for pancreatic binding partners of Isl1 have also been identified. These include Neurogenic differentiation factor 1 (NeuroD1) [153], Single-stranded DNA-binding protein 3 (Ssbp3) [154], Protein inhibitor of activated STAT protein gamma (Pias4) [155], Pancreas/duodenum homeobox protein 1 (Pdx1) and Histone-lysine N-methyltransferase SETD7 (Set7/9) [145], as well as Ldb1 [156, 157]. Whether or not these proteins all interact directly with Isl1 is not clear. For example, NeuroD1 was implicated as an Isl1 interaction partner involved in regulating insulin expression [153], but studies by other groups did not see the same result [154, 156]. There is more confidence in the role of Ldb1 acting with Isl1 in the pancreas, with evidence that the two act to regulate MafA expression [156]. Ssbp3 is also likely to be a genuine interaction partner of the Isl1/Ldb1 complex, as an interaction between Ssbp proteins and Ldb1 within such complexes has already been well characterised; however it is unknown whether there is a direct interaction between Isl1 and any Ssbp protein [158-160].

1.4.4 Brain development requires Isl1

To date, characterising the role of Isl1 in brain development has proved difficult, due to the complexity of the organ. Isl1 has been determined to play a role in the development of several structures in the brain. These include the pituitary gland, striatum, hypothalamus, and thalamus (Figure 1.11) [78, 161, 162].

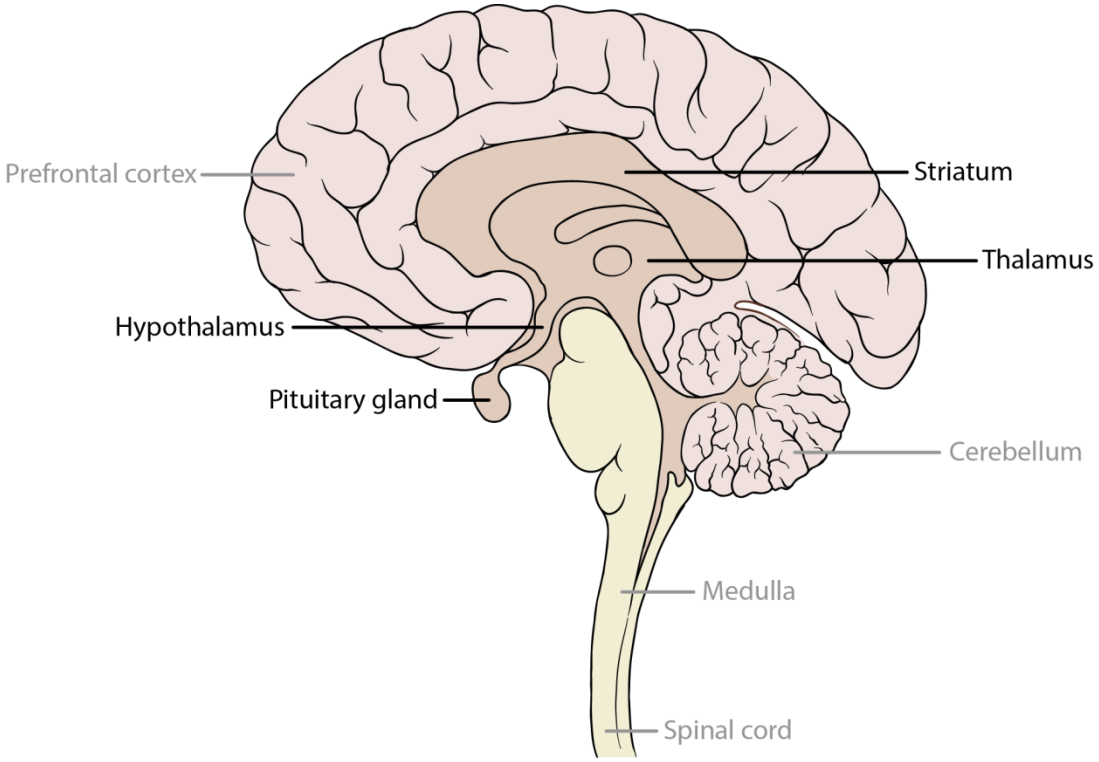


Figure 1.11: Schematic of the adult brain. Regions of Isl1 expression are labelled in black, with other brain structures labelled in grey for reference. Image adapted from "Brain human sagittal section" by Patrick J. Lynch, medical illustrator; C Carl Jaffe (MD), cardiologist.

Low levels of expression of Isl1 were initially identified in the adult brain, specifically in the pituitary gland and other endocrine tissues [78]. Further studies have revealed that Isl1 plays a role in both hormone expression in the mature pituitary gland, and in pituitary development [163]. Shh signalling has been found in the developing pituitary gland, but it was later observed that removing the Shh signalling had no effect on Isl1 expression, indicating a different signalling cascade to that seen in developing motor neurons [164, 165]. Interestingly, Shh signalling can induce Lhx3 expression in the developing pituitary, and there is evidence that Isl1 and Lhx3 are co-expressed in this tissue for a brief period of time [29, 166, 167].

Isl1 is crucial for the development of cholinergic neurons in the striatum of the brain [168-170]. Expression of Lhx6 and Lhx7/8 are required for correct development of these neurons, and evidence from co-immunoprecipitation and chromatin immunoprecipitation (ChIP) studies suggests that Lhx/Isl1/Ldb complexes contribute to this process [171, 172]. Beyond these studies, there is limited knowledge of downstream targets of such a ternary complex, where Lhx3 is substituted with Lhx6 or Lhx7/8.

The role of Isl1 in the developing thalamus is similarly under-characterised. In the developing thalamus, expression of several LIM-HD proteins, including Isl1, is required for correct development [173]. Isl1 is expressed in the hypothalamus as well, where it may play a role in modulating hormones [174, 175]. However, there is limited knowledge of binding partners of Isl1 in these tissues, how Isl1 expression is induced, and how Isl1 influences the transcriptional environment.

1.4.5 Isl1 is involved in development of the kidney and urinary tract

Expression of Isl1 in the kidney was first detected shortly after the initial discovery of Isl1. This expression was observed specifically in tubules proximal to the glomerulus in adult rats (Figure 1.12) [176]. It has only been recently that the role of Isl1 in renal tissues has been investigated. Conditional Isl1 knockout mutant mice displayed developmental defects in the kidney and urinary tract, indicating Isl1 plays a developmental role in these tissues [177, 178].

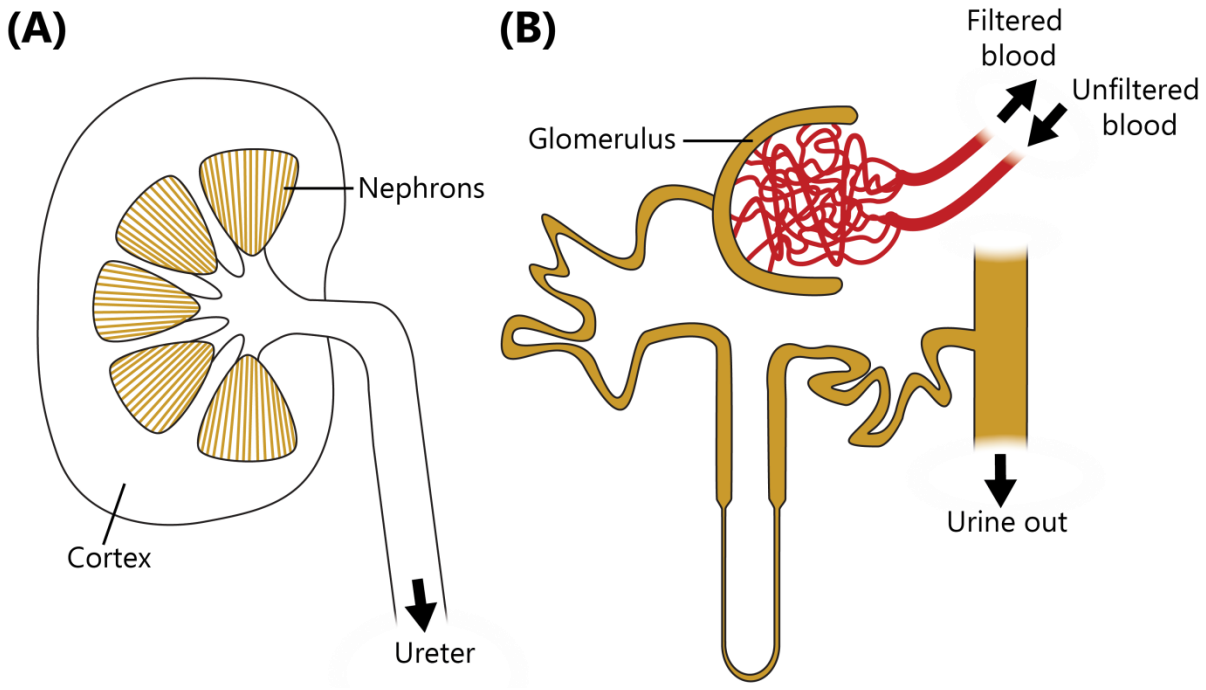


Figure 1.12: Schematics of the kidney. (A) Structure of the kidney and its connection to the ureter. (B) Schematic of a nephron.

Isl1 expression was observed in developing tissues between E10.5 and E14.5, and when this expression was not present, Bone morphogenic protein 4 (Bmp4) levels were reduced [177]. Bmp4 plays a major role in development of the kidneys and urinary tract, and deletion of

Bmp4 results in a phenotype similar to that of Isl1 deletion [177, 179, 180]. It is likely that Isl1 stimulates Bmp4 expression in the developing kidney and urinary tract. Recently this stimulation was observed in developing genital tissue, where Isl1 was demonstrated to promote expression of Bmp4, Fgf10, and Wnt5a [181]. All three of these proteins form components of major developmental signalling cascades that could drive the differentiation of this tissue.

1.4.6 Isl1 in the developing retina

Isl1 expression was initially detected in the adult retina and was later detected in developing retinal cell populations [79, 175, 182, 183]. In these cells, Isl1 has been shown to act in combination with the homeodomain protein POU domain, class 4, transcription factor 2 (Pou4f2) [184, 185]. Isl1 is necessary for maintaining the expression of Pou4f2, and the two proteins form complexes to regulate expression of genes relevant for retinal ganglion differentiation [186, 187].

1.4.7 Other roles of Isl1 in development

Many additional instances of Isl1 influencing development (Table 1.2) have not been explored in as much depth as those discussed above. Note that in many cases it has not been established whether the regulatory relationships listed in Table 1.2 are direct or indirect.

Table 1.2: Expression of Isl1 in development of a multitude of other tissues.

Tissue	Role	Knockout phenotype	Targets of Isl1 regulation	Isl1 expression regulated by	References
Ear	Inner ear development	-	-	Lmo4	[188-190]
Gut	Expressed in stomach and intestinal cells	Impaired lipid absorption, and glucose homeostasis	-	Foxa1, Foxa2	[191-193]
Jaw	Growth of the lower jaw	Extremely misshapen lower jaw	*Wif1, Dkk1 Dkk2,* β-catenin	Bmp4	[194-196]
Limbs	Hindlimb development	Severe malformation of the hindlimb	β-catenin	-	[195, 197]
Muscle	Head muscle development	-	-	Bmp4	[198]
Teeth	Tooth type specification	-	*Bmp4*	Bmp4	[196, 199]

1.5 Known binding partners of Isl1

Isl1 can be found associated with many different protein binding partners across different tissues, as discussed above. In many cases, it has not been established whether these interactions are direct protein-protein interactions, indirect interactions bridged by other proteins, or indirect interactions facilitated through DNA-binding. Summarised in Table 1.3 are the most well-described binding partners of Isl1, and the tissues in which they interact.

Table 1.3: Known protein interaction partners of Isl1.

Binding partner	Tissue of interaction	References
Lhx3	Developing spinal motor neurons	[93, 99]
Ldb1	Developing spinal motor neurons	[99]
	Developing heart tissue	[137]
	Pancreas	[156, 157]
Phox2a	Developing cholinergic neurons (brain)	[105]
Pou4f2	Developing retina	[187]
Stat3	Developing spinal motor neurons	[113, 200]
	Non-Hodgkin lymphoma	[201]
c-Jun	Non-Hodgkin lymphoma	[201]
Tbx20	Developing heart	[128]
	Developing motor neurons	
Ssbp3	Pancreas	[154]
Cited2	Developing heart	[136]
Pdx1	Developing pancreas	[145]
Set7/9	Developing pancreas	[145]

1.6 Alternative forms of Isl1

1.6.1 Isl1-β

An isoform of Isl1, named Isl1-β, lacks residues 256-278, meaning it is missing the first 17 residues of the LID [202]. The role of this isoform has yet to be elucidated. It is likely to be unable to form higher order complexes through the binding of the LID, which would affect many functions of Isl1, but this has not yet been investigated in depth.

1.6.2 Phosphorylation of Isl1

The same study that identified Isl1-β also identified a phosphorylation site in the sequence of Isl1 [202]. However, it remains to be seen if there is a specific phosphorylation site on Isl1 that is relevant to its function. Additionally, the role of this modification has yet to be elucidated, though there is evidence it plays a role in insulin production in the liver [203].

1.6.3 Isl2

As mentioned in Section 1.2.1, all mammals possess a paralogue of Isl1 called Isl2. Isl2 is very similar to Isl1 in terms of sequence and function, with the two murine sequences being 72% identical [57]. The high degree of conservation between the two paralogues makes it difficult to study them independently, as most antibodies targeted to one paralogue will recognise both [30, 204]. The most distinct difference observed between the two paralogues is their expression patterns. For example, Isl2 is expressed later than Isl1 in motor neuron development [205], exhibits more prolonged expression during inner ear development [206], and is expressed in different areas of the developing mouse retina [207].

1.7 Aims of this thesis

As discussed in Section 1.4, Isl1 is necessary for the correct development of many different tissues, but in most cases there is little molecular insight into how the protein functions. This is especially the case with the many potential binding partners of Isl1 reported in the literature, where there is a lack of evidence to indicate whether interactions are direct, or mediated by DNA and/or other proteins. However, it is clear that the interplay of protein-protein and protein-DNA interactions act to help Isl1 fulfil a function in a variety of different cellular contexts. By characterising the protein-protein and protein-DNA interactions Isl1 can make, this thesis aims to further our understanding of how Isl1 can fulfil such a broad range of functions, in temporally and spatially distinct settings.

In order to do this, two broad aims were defined. Firstly, to search for novel direct protein binding partners of Isl1. Chapter 3 describes efforts to achieve this through the use of yeast two-hybrid library screening. Chapter 4 then analyses the potential binding partners identified, assessing the likelihood of the identified interactions being biologically relevant to the known functions of Isl1.

The second broad aim was to investigate the molecular mechanisms associated with the DNA-targeting of the Lhx3/Isl1/Ldb1 ternary complex. Chapter 5 explores the DNA-binding behaviour of the Isl1 and Lhx3 homeodomains both in isolation and in combination. Chapter 6 then pursues the structural characterisation of the Isl1/Lhx3 DNA-binding complex.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals and reagents

A complete index of all chemicals and reagents referred to in this thesis can be found in Table 2.1. All solutions were prepared in ultrapure MilliQ® water (MQW), unless otherwise specified.

Table 2.1: Chemicals and reagents used throughout this thesis.

Material	Manufacturer
(+/−)-2-Methyl-2,4-pentanediol (MPD)	Hampton Research (Aliso Viejo, CA)
1,6-hexanediol	Sigma-Aldrich (St Louis, MO)
25:24:1 phenol:chloroform:isoamyl alcohol	Sigma-Aldrich (St Louis, MO)
2-log ladder DNA ladder	New England BioLabs (Ipswich, MA)
3-amino-1,2,4-triazole (3-AT)	Sigma-Aldrich (St Louis, MO)
4-(2-hydroxylethyl)-1-piperazineethanesulfonic acid (HEPES)	Biochemicals.com.au (Gymea, NSW)
4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS)	Sigma-Aldrich (St Louis, MO)
5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-gal)	Gold Bio Technology (St Louis, MO)
Acetic acid	Chem-Supply (Port Adelaide, SA)
Acrylamide/Bis acrylamide 29:1, 40% w/v	Biochemicals.com.au (Gymea, NSW)
Adenine hemisulfate	Sigma-Aldrich (St Louis, MO)
Agarose	Amyl Media (South Dandenong, VIC)
Ammonium chloride (NH_4Cl)	Ajax Finechem (Taren Point, NSW)
Ammonium persulfate	Sigma-Aldrich (St Louis, MO)
Ampicilin sodium salt	Gold Bio Technology (St Louis, MO)
Arginine	Sigma-Aldrich (St Louis, MO)
Barium chloride	Merck (Darmstadt, Germany)
2-(Bis(2-hydroxyethyl)amino)acetic acid (Bicine)	Sigma-Aldrich (St Louis, MO)
2-[Bis(2-hydroxyethyl)amino]-2-	Amresco (Solon, OH)

(hydroxymethyl)propane-1,3-diol (Bis-Tris)	
β-mercaptoethanol	Sigma-Aldrich (St Louis, MO)
Boric Acid	Sigma-Aldrich (St Louis, MO)
Bolt^{TM} 4-12% polyacrylamide gels	Thermo Fisher Scientific (Waltham, MA)
Bovine serum albumin (BSA) (acetylated)	Sigma-Aldrich (St Louis, MO)
Bovine serum albumin (BSA) (monomeric)	Sigma-Aldrich (St Louis, MO)
Bromophenol blue	Sigma-Aldrich (St Louis, MO)
Calcium chloride ($CaCl_2$)	Univar (Downers Grove, IL)
Carbenicilin sodium salt	Gold Bio Technology (St Louis, MO)
Choramphenicol	Gold Bio Technology (St Louis, MO)
Complete supplement media (CSM)	MP biomedical (Santa Ana, CA)
cOmplete^{TM} EDTA-free protease inhibitor	Sigma-Aldrich (St Louis, MO)
Coomassie Brilliant Blue R-250	Life Technologies (Carlsbad, CA)
Deoxyribonucleotide triphosphates (dNTPs)	Boehringer Mannheim (Mannheim, Germany)
Deuterium oxide (D_2O)	Sigma-Aldrich (St Louis, MO)
D-Glucose	Chem-Supply (Port Adelaide, SA)
Difco yeast extract	Bacto Laboratories (Mt Pritchard, NSW)
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich (St Louis, MO)
Disodium hydrogen phosphate (Na_2HPO_4)	Ajax Finechem (Taren Point, NSW)
Dithiothreitol (DTT)	Gold Bio Technology (St Louis, MO)
Ethanol	Chem-Supply (Port Adelaide, SA)
Ethylene ditetraacetic acid (EDTA)	Ajax Finechem (Taren Point, NSW)
Ficoll	Sigma-Aldrich (St Louis, MO)
Glass beads (5 mm diameter)	Sigma-Aldrich (St Louis, MO)
Glutathione Sepharose 4B resin	GE Healthcare (Chicago, IL)
Glycerol	Chem-Supply (Port Adelaide, SA)
Guanidine hydrochloride	Biochemicals.com.au (Gymea, NSW)
Hexammine cobalt chloride	Sigma-Aldrich (St Louis, MO)
Histidine	Sigma-Aldrich (St Louis, MO)
HydraGreen^{TM}	ACTGene (Piscataway, NJ)
Hydrochloric acid	Univar (Downers Grove, IL)
Igepal CA-630	Sigma-Aldrich (St Louis, MO)

Immobilon Western Chemiluminescent HRP substrate	Merck (Darmstadt, Germany)
Isoleucine	Sigma-Aldrich (St Louis, MO)
Isopropanol	Univar (Downers Grove, IL)
Isopropyl β-D-thiogalactopyranoside (IPTG)	Progen (Heidelberg, Germany)
Leucine	Sigma-Aldrich (St Louis, MO)
Lithium acetate (LiAc)	Ajax Finechem (Taren Point, NSW)
Lysine	Sigma-Aldrich (St Louis, MO)
Magnesium chloride ($MgCl_2$)	Ajax Finechem (Taren Point, NSW)
Magnesium sulfate ($MgSO_4$)	Ajax Finechem (Taren Point, NSW)
Manganese chloride ($MnCl_2$)	Sigma-Aldrich (St Louis, MO)
Mark12™ protein standard	Thermo Fisher Scientific (Waltham, MA)
MES (2-ethanesulfonic acid) powder	Sigma-Aldrich (St Louis, MO)
MES (2-ethanesulfonic acid) buffer 20 ×	Thermo Fisher Scientific (Waltham, MA)
Methanol	Chem-Supply (Port Adelaide, SA)
Methionine	Sigma-Aldrich (St Louis, MO)
N,N,N′,N′-Tetramethylethylenediamine (TEMED)	Sigma-Aldrich (St Louis, MO)
NuPAGE® LDS Sample buffer (4X)	Sigma-Aldrich (St Louis, MO)
Peptone	Affymetrix (Santa Clara, CA)
Phenylalanine	Sigma-Aldrich (St Louis, MO)
Phenylmethanesulfonyl fluoride (PMSF)	Sigma-Aldrich (St Louis, MO)
Piperazine-N,N′-bis(2-ethanesulfonic acid) (PIPES)	Sigma-Aldrich (St Louis, MO)
Polyethylene glycol, MW 3350 (PEG 3350)	Sigma-Aldrich (St Louis, MO)
Polyethylene glycol, MW 4000 (PEG 4000)	Hampton Research (Aliso Viejo, CA)
Polyethylene glycol, MW 8000 (PEG 8000)	Thermo Fisher Scientific (Waltham, MA)
Polyethylimine (PEI) 50% solution	Sigma-Aldrich (St Louis, MO)
Polyethylimine (PEI) Mw 25,000, Transfection Grade	Polysciences Inc (Warrington, PA)
Potassium acetate	Merck (Darmstadt, Germany)
Potassium chloride (KCl)	Ajax Finechem (Taren Point, NSW)
Potassium dihydrogen phosphate (KH_2PO_4)	Chem-Supply (Port Adelaide, SA)

Proline	Sigma-Aldrich (St Louis, MO)
Salmon sperm DNA	Sigma-Aldrich (St Louis, MO)
Skim milk powder	Woolworths (Bella Vista, NSW)
Sodium cacodylate	Sigma-Aldrich (St Louis, MO)
Sodium chloride (NaCl)	Chem-Supply (Port Adelaide, SA)
Sodium dihydrogen phosphate (NaH_2PO_4)	Ajax Finechem (Taren Point, NSW)
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich (St Louis, MO)
Sodium hydroxide (NaOH)	Univar (Downers Grove, IL)
Spermine tetrahydrochloride	Sigma-Aldrich (St Louis, MO)
Streptactin HRP (horseradish peroxidase) conjugate	BioRad (Hercules, CA)
Strontium chloride	Sigma-Aldrich (St Louis, MO)
SYPRO™ Ruby Protein Gel Stain	Thermo Fisher Scientific (Waltham, MA)
Thiamine	Sigma-Aldrich (St Louis, MO)
Threonine	Sigma-Aldrich (St Louis, MO)
2-Amino-2-(hydroxymethyl)propane-1,3-diol (Tris)	Chem-Supply (Port Adelaide, SA)
tris(2-carboxyethyl)phosphine (TCEP)	Soltec Biosciences (Gloucester, MA)
Triton X-100	Sigma-Aldrich (St Louis, MO)
Tryptone T	Oxoid (Hampshire, England)
Tryptopan	Sigma-Aldrich (St Louis, MO)
Tween-20	Amresco (Solon, OH)
Tyrosine	Sigma-Aldrich (St Louis, MO)
Uracil	Sigma-Aldrich (St Louis, MO)
Valine	Sigma-Aldrich (St Louis, MO)
WesternC™ protein standard	BioRad (Hercules, CA)
Yeast Extract	Affymetrix (Santa Clara, CA)
Yeast nitrogen base	Bacto Laboratories (Mt Pritchard, NSW)

2.1.2 Enzymes

All enzymes were used with the buffers provided by the manufacturer, and used according to manufacturers' instructions.

Table 2.2: Enzymes used throughout this thesis.

Enzyme		Manufacturer
BamHI	EC 3.1.21.4	New England BioLabs (Ipswich, MA)
Dnase I	EC 3.1.21.1	Sigma-Aldrich (St Louis, MO)
DpnI	EC 3.1.21.4	New England BioLabs (Ipswich, MA)
EcoRI	EC 3.1.21.4	New England BioLabs (Ipswich, MA)
HRV-3C protease	EC 3.4.22.28	Mackay & Matthews laboratory, University of Sydney
Lysozyme	EC 3.2.1.17	BioMatik (Cambridge, Ontario)
Lyticase	EC 3.2.1.4	Sigma-Aldrich (St Louis, MO)
Phusion DNA polymerase	EC 2.7.7.7	Dr Jason Low, University of Sydney
QuickStick DNA ligase	EC 6.5.1.1	Bioline (London, UK)
RNase A	EC 3.1.27.5	Sigma-Aldrich (St Louis, MO)
RNase (DNase free)	EC 3.1.27.5	Thermo Fisher Scientific (Waltham, MA)

2.1.3 Media

A full list of media used in this project can be found in Table 2.3. Luria-Bertani broth (LB) and yeast peptone dextrose broth (YPD) were prepared in distilled water; all other media were prepared in MilliQ® water. All media were sterilised by autoclaving before use. Sterile glucose was prepared and autoclaved separately, and added to media after autoclaving where appropriate. For solid media, agarose was added to a final concentration of 1.5% w/v.

EXPI293™ expression medium was purchased as a complete media (Life Technologies; Carlsbad, CA).

Table 2.3: Media used throughout this thesis.

Media	Components
LB	1% w/v peptone, 0.5% w/v yeast extract, 0.5% w/v NaCl
SOC media	2% w/v tryptone, 0.5% w/v yeast extract, 2 mM NaCl, 2

	mM KCl, 2 mM $MgCl_2$, 2 mM $MgSO_4$, 0.4% w/v glucose
M9-L minimal medium	0.4% w/v glucose, 30 mg/L isoleucine, 150 mg/L valine, 20 mg/L arginine, 30 mg/L lysine, 20 mg/L methionine, 50 mg/L phenylalanine, 200 mg/L threonine, 30 mg/L tyrosine, 20 mg/L uracil, 20 mg/L histidine, 20 mg/L tryptophan, 20 mg/L adenine, 40 mg/L proline, 10 mg/L thiamine, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 50 mM Na_2HPO_4, 25 mM KH_2PO_4, 1 mM NaCl, 20 mM NH_4Cl
YPD	2% w/v tryptone, 1% w/v yeast extract, 2% w/v glucose, 0.2% w/v adenine
Synthetic dextrose broth (SD)	0.65% w/v yeast nitrogen base, 0.4% w/v glucose, 1 × CSM
EXPI293™ expression medium	Gibco™ EXPI293™ expression medium (Thermo Fisher Scientific; Waltham, MA)

20 µg/mL X-α-gal and additional nutrients were added to SD media as required. Additional components are listed in Table 2.4.

Table 2.4: Additional nutrients present in selective yeast media.

Media	Additional components
SD-L	2 g/L adenine, 2 g/L histidine, 2 g/L tryptophan
SD-W	2 g/L adenine, 2 g/L histidine, 10 g/L leucine
SD-L-W	2 g/L adenine, 2 g/L histidine
SD-L-W-H	2 g/L adenine
SD-L-W-H+0.5 mM 3-AT	2 g/L adenine, 0.5 mM 3-AT
SD-L-W-H+2 mM 3-AT	2 g/L adenine, 2 mM 3-AT
SD-L-W-H-A	None

2.1.4 Oligonucleotides and plasmids

Single-stranded oligonucleotides used for cloning and DNA binding studies were synthesised by IDT (Integrated DNA Technologies; Coralville, IA). A list of all oligonucleotides used for binding studies can be found in Appendix A. All oligonucleotides used for binding studies

were annealed to a complementary strand through heating to 90 °C for 10 min, followed by gradual cooling.

The yeast plasmids used throughout this project were pGAD10 (*LEU2*, *ampR*) (BD Biosciences Clontech; Mountain View CA), pGADT7-RecAB (*LEU2*, *ampR*) (BD Biosciences Clontech; Mountain View, CA) and NpGBT9 (*TRP1*, *ampR* or *kanR*) (Professor Merlin Crossley, University of Sydney). NpGBT9 is a derivative of pGBT9, in which the EcoRI and BamHI cut sites in the multiple cloning site have been swapped. Throughout this work, NpGBT9 will be referred to as pGBT9.

Bacterial expression was conducted using a modified pMAL vector (Dr Ivan Nisevic, University of Sydney) or a modified pET-DUET vector (Herman Fung, University of Sydney). The pMAL vector encodes an MBP (maltose binding protein) tag with a downstream factor Xa cleavage site immediately upstream of the multiple cloning site. The modified pET-DUET vector encodes a GST (glutathione S-transferase) tag with a downstream HRV-3C cleavage site immediately upstream of the first multiple cloning site. Protein-encoding DNA fragments were subcloned into only the first multiple cloning site.

pcDNA3.1, encoding either an N-terminal FLAG or HA tag, was used for all work in mammalian cells (Dr Jason Low, University of Sydney). The amino acid sequences of protein fusion tags can be found in Appendix B.

2.1.5 Antibodies and peptides

Antibodies used throughout this thesis are listed in Table 2.5.

Table 2.5: Antibodies used throughout this thesis.

Antibody	Manufacturer	Concentration used
Rabbit α-FLAG mAb-HRP	Cell Signaling Technology (Danvers, MA)	1:4000
Mouse α-HA mAb-HRP	Cell Signaling Technology (Danvers, MA)	1:40000

Additionally, 3×FLAG peptide (APExBio; Houston, TX) was used during FLAG affinity purification. The sequence of this peptide is: MDYKDHDGDYKDHDIDYKDDDDK.

2.1.6 Organisms

Table 2.6 lists the genotypes of bacteria and yeast used throughout this work. DH5α cells were used for cloning and plasmid propagation; HB101 cells were used for isolation of yeast plasmids and plasmid propagation; BL21 Gold (DE3) cells were used for protein expression. *Escherichia coli* (*E. coli*) strains were maintained in sterile LB media. *Saccharomyces cerevisiae* (*S. cerevisiae*) strains were maintained in YPD media.

Table 2.6: Organism strain genotypes.

Species	Organism strain	Genotype
S. cerevisiae	AH109 (BD Biosciences Clontech; Mountain View, CA)	*MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-HIS3, GAL2*$_{UAS}$*-GAL2*$_{TATA}$*-ADE2, URA3::MEL1*$_{UAS}$*-MEL1*$_{TATA}$*-lacZ, MEL1*
S. cerevisiae	Y187 (BD Biosciences Clontech; Mountain View, CA)	*MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met–, gal80Δ, URA3::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-lacZ*
E. coli	DH5α (Bethesda Research Laboratories, Gaithersburg, MD)	F$^-$*, Φ80lacZΔM15, Δ(lacZYA-argF), U169, recA1, endA1, hsdR17,* (*rK–, mK+*)*, phoA, supE44, λ–, thi-1, gyrA96, relA1*
E. coli	BL21 Gold (DE3) (Integrated Sciences, Chatswood, NSW)	F$^-$, *ompT, hsdS*$_B$(r_B–, m_B–), *gal, dcm,* (DE3)
E. coli	Rosetta™ 2 (DE3) (Merck; Darmstadt, Germany)	F$^-$, *ompT, hsdS*$_B$(r_B^- m_B^-), *gal, dcm,* pRARE2 (Cam^R), (DE3)
E. coli	HB101 (CGC, University of Minnesota)	F$^-$, *mcrB, mrr, hsdS20*(r_B^- m_B^-), *recA13, leuB6, ara-14, proA2, lacY1, galK2, xyl-5, mtl-1, rpsL20*(Sm^R), *glnV44,* λ^-

A pre-transformed Mate & Plate Universal Mouse yeast library (Clontech; Mountain View, CA) was used for yeast two-hybrid mating screens.

Suspension-adapted HEK (human embryonic kidney) Expi293F™ cells (Thermo Fisher Scientific; Waltham, MA) were used for mammalian cell work.

2.2 Cloning

All constructs produced during this thesis were sequenced at the Australian Genome Research Facility (Westmead, NSW) to confirm the correct sequence was present. A list of sequencing primers used can be found in Appendix C. Throughout this thesis, a slash will be used to signify protein fusions, with a colon being used to represent specific protein-protein or protein-DNA interactions.

2.2.1 Gibson cloning

Gibson cloning was used to produce new protein constructs or edit vectors (Table 2.7).

Table 2.7: Constructs produced through Gibson cloning.

Vector	Insert	Purpose
pGBTK9	*kanR*	Changing ampicillin resistance to kanamycin resistance (Section 2.3.3)
pGBT9	$LIMK1_{LIM}$	Yeast two-hybrid analysis
pGBT9	$LIMK2_{LIM}$	Yeast two-hybrid analysis
pcDNA3.1-FLAG	$Isl1_{LIM}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Isl1_{\Delta LIM}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Isl1_{FL}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Ldb1_{LID}/Isl1_{LIM}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Isl1_{LIM}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Isl1_{\Delta LIM}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Isl1_{FL}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Ldb1_{LID}/Isl1_{LIM}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Mkln1_{FL}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Mkln1_{ND}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Mkln1_{NL}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Mkln1_{NK}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Mkln1_{LC}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Mkln1_{CC}$	Mammalian co-immunoprecipitation

pcDNA3.1-FLAG	$Mkln1_{KC}$	Mammalian co-immunoprecipitation
pcDNA3.1-FLAG	$Mkln1_{FC}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Mkln1_{FL}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Mkln1_{ND}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Mkln1_{NL}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Mkln1_{NK}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Mkln1_{LC}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Mkln1_{CC}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Mkln1_{KC}$	Mammalian co-immunoprecipitation
pcDNA3.1-HA	$Mkln1_{FC}$	Mammalian co-immunoprecipitation
pGADT7-RecAB	$Mkln1_{FL}$	Yeast two-hybrid analysis
pGADT7-RecAB	$Mkln1_{ND}$	Yeast two-hybrid analysis
pGADT7-RecAB	$Mkln1_{NL}$	Yeast two-hybrid analysis
pGADT7-RecAB	$Mkln1_{NK}$	Yeast two-hybrid analysis
pGADT7-RecAB	$Mkln1_{LC}$	Yeast two-hybrid analysis
pGADT7-RecAB	$Mkln1_{CC}$	Yeast two-hybrid analysis
pGADT7-RecAB	$Mkln1_{KC}$	Yeast two-hybrid analysis
pGADT7-RecAB	$Mkln1_{FC}$	Yeast two-hybrid analysis

2.2.1.1 Polymerase chain reaction (PCR)

Phusion polymerase was used for all PCRs. PCRs were conducted in Phusion buffer (10 mM Tris-HCl (pH 8.8), 50 mM KCl, 2.5 mM $MgSO_4$, 0.1% Triton X-100, 0.2 mg/mL BSA), with 0.5 µM forward and reverse primers, 0.5 mM dNTPs, and 37.5-75 ng template DNA. PCRs were conducted using a Biometra T3000 Thermocycler (Analytik Jena; Jena, Germany). PCRs were run for 35 cycles, using the following program: denaturation at 98 °C (15 s), annealing at 65 °C (20 s), extension at 72 °C (30 s/kb amplicon).

2.2.1.2 PCR product purification

After PCR, reaction mixtures were purified using an ISOLATE II PCR and Gel Kit (Bioline; London, UK). Samples were incubated with 10 mU DpnI to ensure removal of parental DNA, in CutSmart® buffer (New England Biolabs; Ipswich, MA). Reactions were incubated for 1-4 h at 37 °C.

2.2.1.3 Ligation

Gibson assembly [208] was used to ligate amplified products into vectors. These reactions contained 50 ng vector, with threefold molar equivalents of insert. To this mixture was added Gibson Assembly Master Mix (New England Biolabs; Ipswich, MA). Reactions were incubated for at 50 °C, 4 h, in a Biometra T3000 Thermocycler, before being transformed into competent DH5α *E. coli* (Section 2.5.1).

2.2.2 Restriction cloning

Restriction cloning was used to subclone already cloned constructs into new vectors (Table 2.8).

Table 2.8: Constructs produced through restriction cloning.

Source Vector	Destination Vector	Insert
pGAD10	pET-DUET	LLHD1
pGEX	pET-DUET	2HDLL
pGEX	pET-DUET	2HDN
pGEX	pET-DUET	NHD1
pGEX	pET-DUET	NHD3
pGEX	pET-DUET	LLHD3
pGBT9	pGBTK9	$Isl1_{LIM}$
pGBT9	pGBTK9	$Isl1_{\Delta LIM}$

Source and destination vectors were digested by incubating the appropriate plasmids in CutSmart® buffer with 20 U each of high fidelity BamHI and EcoRI (New England Biolabs; Ipswich, MA). These reactions were incubated for 4 h at 37 °C.

Digested DNA was run on a 1.5% w/v agarose gel, supplemented with HydraGreen™ (1:60,000), made in TAE buffer (40 mM Tris pH 8.0, 20 mM acetic acid, 1 mM EDTA) for 35 min, before desired species were visualised, excised, and purified using the ISOLATE II PCR and Gel Kit (Bioline; London, UK). Inserts were ligated into the desired destination vectors using Quick-Stick ligase (Bioline; London, UK). Vector and insert were combined in the provided buffer at both a 1:6 and a 1:3 ratio before enzyme was added. Reactions were

incubated for 30 min to allow ligation to occur. This reaction mix was used for transformations into DH5α *E. coli* cells (Section 2.5.1).

2.3 Yeast handling procedures

2.3.1 Preparation and transformation of competent yeast

Yeast competency and transformation protocols were already established in the Matthews laboratory [209, 210]. AH109 yeast were grown in rich YPD media overnight with shaking (150 rpm) at 30 °C. Cells were pelleted by centrifugation (5 min, 1000 × g, room temperature). Cells were washed in 25 mL MQW, before being subjected to centrifugation (5 min, 1000 × g, room temperature). Supernatant was removed, and cells resuspended in competency buffer (10 mM Tris (pH 8.0), 1 mM EDTA, 100 mM LiAc (pH 7.5)), in preparation for transformations.

An aliquot of competent yeast (100 µL) was added to plasmid DNA (1 µL of each plasmid to be transformed or co-transformed, at a typical concentration of 500 ng/µL for each plasmid) and carrier salmon sperm DNA (10 µL). Sterile PEG/LiAc solution (600 µL; 10 mM Tris (pH 8.0), 40% PEG 4000, 100 mM LiAc, 1 mM EDTA) was added and the solution mixed gently. This mixture was incubated for 30 min with shaking (150 rpm) at 30 °C, before DMSO (70 µL) was added. Cells were heat shocked for 15 min at 42 °C, being mixed gently every 5 min. Following this, cells were chilled on ice water for 2 min. The mixture was subjected to centrifugation (5 min, 2000 × g, 4 °C), and the supernatant removed. The cell pellet was resuspended in resuspension buffer (300 µL; 10 mM Tris pH 8.0, 1 mM EDTA). This mixture was inoculated onto selective media appropriate for the plasmids being transformed (see Section 2.1.3 for complete media formulation):

- SD-L, SD media with no leucine supplementation, for transformations with pGAD10/pGADT7-RecAB.
- SD-W, SD media with no tryptophan supplementation, for transformations with pGBT9.
- SD-L-W, SD media with no leucine or tryptophan supplementation, for co-transformations with one pGBT9 plasmid and one of pGAD10 or pGADT7-RecAB plasmid.

Plates were incubated for 65-72 h at 30 °C before being scored for colonies.

For high throughput applications, yeast transformations were carried out in sterilised 96-well plates, with a well volume of 2 mL. All the transformation solutions used were halved in volume, with two exceptions: the plasmid DNA, which was kept the same as in the original transformation protocol, and the final resuspension volume, which was increased to 750 µL per sample. An aliquot (50 µL) of this suspension was inoculated onto appropriate agar prepared in 24-well tissue culture plates. Plates were allowed to dry for 1-2 h under a bunsen burner flame after inoculation, to reduce the amount of residual liquid present. They were then incubated for 65-72 h at 30 °C before being scored for colonies.

2.3.2 Yeast two-hybrid spot test assays

Yeast two-hybrid spot test assay protocols were already established in the Matthews laboratory [209, 210]. Transformed yeast were grown overnight with shaking (150 rpm) at 30 °C in selective SD-L-W media (1 mL). Cell densities were normalised using the optical density of the solution at 600 nm with a 1-cm pathlength (OD_{600}). Two 1-in-10 serial dilutions were prepared from the normalised cell suspension, resulting in solutions with OD_{600} values of 0.2, 0.02 and 0.002. Aliquots (2 µL) of each sample were spotted onto each selection condition, as well as growth control SD-L-W plates. Selection conditions used from least to most stringent included:

- SD-L-W-H
- SD-L-W-H+0.5 mM 3-AT
- SD-L-W-H+1 mM 3-AT
- SD-L-W-H+2 mM 3-AT
- SD-L-W-H+5 mM 3-AT
- SD-L-W-H-A

Section 2.1.3 contains full compositional details of these media (Table 2.4).

2.3.3 Mating library screens

Library mating screens were performed as laid out in the Clontech Matchmaker® Gold Yeast Two-Hybrid System User Manual [211]. Briefly, AH109 yeast transformed with the selected bait plasmid were grown overnight, before being added to an aliquot (1 mL) of Y187 library yeast resuspended in fresh 2 × YPD media. This mix was incubated for 24 h at 30 °C with shaking (40 rpm). Cells were pelleted by centrifugation (5 min, 1000 × g, room temperature)

and resuspended in a smaller volume, before 150-mm selective media agar plates were inoculated with the mixture. Selection conditions were chosen to minimise auto-activation growth by the bait and are listed in Table 2.9. Plates were incubated for 3-5 days (30 °C), before colonies were picked for analysis.

Table 2.9: Composition of screening plates used for different bait in mating screens.

Bait construct	Media composition
$Isl1_{LIM}$	SD-L-W-H+0.5 mM 3-AT
$Ldb1_{LID}$/$Isl1_{LIM}$	SD-L-W-H
$Isl1_{\Delta LIM}$	SD-L-W-H+2 mM 3-AT SD-L-W-H+2.5 mM 3-AT

Screens conducted before the commencement of this thesis used baits subcloned into pGBT9 with ampicillin resistance (*ampR*). Screens conducted during the timeframe of this thesis used baits subcloned into pGBT9 with kanamycin resistance (*kanR*). This simplified downstream recovery of the prey plasmids (Section 2.3.5).

2.3.3.1 Library screening quality control

To assess the viability of each mating partner and ensure the efficiency of library screening, samples of the individual mating partners and the mated solution were taken, diluted (1:10, 1:100, 1:1000, and 1:10000), and inoculated onto appropriate selective media (SD-L and SD-W for single plasmid selection, SD-L-W for dual plasmid selection). After incubation at 30 °C for 72 h the numbers of colonies for each condition were counted. Full details can be found in Appendix D.

2.3.4 Extraction of plasmid DNA from mated yeast

Over the course of this thesis, two protocols were used to extract plasmid DNA from yeast. The first used phenol:chloroform extraction to lyse cells and separate DNA from proteins and RNA [209], followed by isopropanol precipitation to purify DNA away from any remaining phenol. The second and more successful protocol used enzymatic digestion and alkaline lysis to lyse cells, in the presence of RNase to remove RNA. Both are described below.

2.3.4.1 Lysis by phenol:chloroform extraction with isopropanol precipitation

Aliquots (10 mL) of SD-L media were inoculated with mated yeast and allowed to grow overnight at 30 °C, with shaking (200 rpm). Cells were pelleted by centrifugation (5 min, 5000 × g, room temperature) before being resuspended in lysis buffer (200 µL; 10 mM Tris, pH 8.0, 2% v/v Triton X-100, 1% SDS, 100 mM NaCl, 1 mM EDTA). Glass beads (0.5 mm, amount corresponding to 200 µL volume) were added. An equal volume of 25:24:1 phenol:chloroforom:isoamyl alcohol, 1 mM EDTA, equilibrated to pH 8.0, was added, and the samples vortexed at 3500 rpm for 3 min to encourage cell lysis and extraction of DNA. The samples were centrifuged (5 min, 11000 × g, room temperature) to separate the phases. Samples (200 µL) of each aqueous phase were taken for isopropanol precipitation. NaCl was added to a concentration of 350 mM, and isopropanol added to a concentration of 40% (v/v). After the addition of isopropanol, samples were immediately subjected to centrifugation (30 min, 11000 × g, 4 °C). The supernatant was removed, and the pellet washed with 70% ethanol. Samples were subjected to centrifugation (30 min, 11000 × g, 4 °C), the supernatant removed, and the pellets dried. Once dry, the pellets were resuspended in MilliQ® water. HB101 bacteria were transformed using these DNA solutions.

2.3.4.2 Lysis by enzymatic digestion and alkaline conditions

Aliquots (10 mL) of SD-L media were inoculated with mated yeast and allowed to grow for 24 h with shaking (180 rpm) at 30 °C. Cells were pelleted by centrifugation (15 min, 1000 × g, room temperature) and resuspended in resuspension buffer (100 µL; 25 mM Tris, pH 8.0, 10 mM EDTA, 50 mM glucose, 0.1 mg/mL RNase A). To this was added an equal volume of 1000 U/mL lyticase and the sample incubated for 2 h at 37 °C. Lysis solution (200 µL; 0.2 M NaOH, 1% SDS) was added, and the samples incubated for a further 10 min at room temperature. Neutralisation buffer (280 µL; 3 M potassium acetate, 4 M guanidine HCl) was added to stop lysis, and the sample incubated on ice for 30 min, before being subjected to centrifugation (10 min, 10000 × g, 4 °C) to clarify the lysate. HB101 bacteria were transformed with these clarified lysates.

2.3.5 Isolation and identification of prey plasmid

HB101 bacteria were transformed with extracted yeast DNA to isolate the pGADT7-RecAB prey plasmid only from mated yeast samples. Initially this was done through leucine complementation on M9-L media [209], as the HB101 strain of *E. coli* has a deficiency in

leucine biosynthesis, which can be complemented by the presence of the *LEU2* gene on the pGADT7-RecAB plasmid. In processing the screens conducted during the course of this thesis, this system was not necessary, as the ampicillin resistance gene was replaced with one for kanamycin (Section 2.2.1), so bait and prey plasmids carried differing antibiotic resistance genes (kanamycin and ampicillin, respectively). LB media supplemented with the appropriate antibiotic was sufficient to select for the prey plasmid.

E. coli were transformed as described in Section 2.5.1. However, due to the low abundance of the prey plasmid in the transformation input solution, freshly prepared competent cells were required to achieve sufficient transformation efficiency. These were prepared according to the Inoue method [212]. Briefly, SOC media (50 mL) was inoculated with HB101 cells to an OD_{600} of 0.05 and incubated at 37 °C until the culture reached an OD_{600} between 0.40 and 0.44, with shaking (150 rpm). The culture was cooled in ice water for 10 min, pelleted by centrifugation (15 min, 1000 × g, 4 °C), and resuspended in TB (8 mL; 10 mM PIPES, pH 6.7, 250 mM KCl, 15 mM $CaCl_2$, 55 mM $MnCl_2$). This solution was incubated on ice for 15 min, before being subjected to centrifugation (15 min, 1000 × g, 4 °C) and resuspended in TB (2 mL). DMSO (140 µL) was added in dropwise while gently agitating the solution. Cells were subsequently transformed with the appropriate plasmids. This protocol was scaled up to 400 mL of SOC culture when >40 transformations were required, with all reagents being used at scale.

Successfully transformed colonies were observed after overnight incubation. Individual colonies containing the prey plasmid were grown overnight in LB liquid media (10 mL) supplemented with ampicillin to 100 µg/mL. Plasmid DNA was extracted using the BioLine Isolate II Plasmid Mini Kit and sequenced using primers that flanked the multiple cloning site to identify the prey gene of interest. Sequencing was conducted by the Australian Genome Research Facility (Westmead, NSW). Sequencing primer sequences can be found in Appendix C.

2.3.6 Extraction of protein from transformed yeast

Yeast were lysed according to the method described by von der Haar (2007) [213]. Briefly, cultures of yeast were grown in either SD-L or SD-L-W-H-A (1 mL) at 30 °C with shaking (150 rpm) until they reached maximum density (~24 h; roughly 1×10^8 cells). Cells were

harvested and resuspended in lysis solution (200 µL; 0.1 M NaOH, 0.05 M EDTA, 2% SDS, 2% β-mercaptoethanol), before being incubated for 10 min at 90 °C. 4 M acetic acid (5 µL) was added, and samples were vortexed for 30 s, before being incubated for a further 10 min at 90 °C. Lysis was halted through the addition of a neutralising loading buffer (50 µL; 0.25 M Tris pH 6.8, 50% glycerol, 0.05% bromophenol blue). Samples were stored at -20 °C until analysed through SDS-PAGE and Western blotting (Section 2.4.5).

2.4 Validation of interactions from screening through FLAG pulldowns

2.4.1 Cell growth conditions

Suspension-adapted HEK Expi293F™ cells were used to produce proteins in a mammalian system. These cells were grown in Expi293™ Expression Medium at 37 °C with shaking (130 rpm) and 5% CO_2, until reaching a density of 2×10^6 cells/mL, at which point they were transfected with two pcDNA3.1 plasmids encoding proteins of interest, fused to either a FLAG or a HA tag.

2.4.2 Transfections

Aliquots (1.9 mL) of cells were transfected with two plasmids each. An equimolar mix of plasmids totalling 3.8 µg was prepared in PBS (205 µL; Phosphate buffered saline; 137 mM NaCl, 27 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4), before filter-sterilised PEI (7.6 µL; 1 mg/mL) was added. The mixture was immediately vortexed for 10 s, before being incubated for 20 min at room temperature. This DNA mix was added to aliquoted cells, which were incubated in the same conditions as described previously (Section 2.4.1) for a further 65 h. Cells were harvested in three aliquots (two of 1 mL, one of 100 µL) by washing twice with PBS (1.9 mL), being subjected to centrifugation (5 min, 300 × g, 4 °C) between washes to pellet cells. Once washed, cells were frozen with liquid nitrogen, and pellets stored at -80 °C until needed.

2.4.3 Expression checks

An aliquot (corresponding to 100 µL of culture) of cell pellet was resuspended in NuPAGE® LDS Sample buffer and PBS, boiled for 5 min, and subjected to centrifugation through a 0.8 mL Pierce™ spin filter (Thermo Fisher Scientific; Waltham, MA) for 1 min at 6000 × g

at room temperature, to remove viscous material. Samples were analysed for protein expression through Western blotting with appropriate antibodies (Section 2.4.5).

2.4.4 Cell lysis and FLAG-affinity immunoprecipitation

Cell pellet aliquots (corresponding to 1 mL of culture) were resuspended in lysis buffer (500 µL; 50 mM Tris, pH 7.9, 150 mM NaCl, 1% (v/v) Triton X-100, 1 × cOmplete® EDTA-free protease inhibitor, 0.2 mM DTT), before being lysed with a probe sonicator (3 cycles, each 20 s, 10-20% amplitude, 0.5 s duty cycle). Lysates were incubated on ice for 30 min to allow chromatin to precipitate before being clarified through centrifugation (20 min, 16000 × g, 4 °C). Clarified lysates were incubated with pre-equilibrated anti-FLAG Sepharose 4B beads (20 µL; BioTool, Houston, TX) overnight at 4 °C with rotation (10 rpm). Beads were washed five times with chilled wash buffer (1 mL; 50 mM Tris pH 7.5, 150 mM NaCl, 0.5% (v/v) Igepal CA-630, 0.2 mM DTT), with samples being subjected to centrifugation (5 min, 1000 × g, 4 °C) to sediment the beads each time. Bound proteins were eluted with FLAG elution buffer (20 µL; 10 mM HEPES pH 7.5, 150 mM NaCl, 300 µg/mL 3×FLAG peptide), with samples being incubated for 30 min before being subjected to centrifugation (5 min, 300 × g, 4 °C) and the supernatant removed. Three elution steps were carried out for each sample, with the eluates being pooled for Western blot analysis.

2.4.5 Western blot analysis

Samples, along with WesternC standards (3 µL; BioRad; Hercules, CA) were subjected to SDS-PAGE on Bolt™ 4-12% Bis-Tris Plus gels run in 1 × MES buffer (Table 2.1) at 180 V for 27 min. Samples were transferred from the gel to PVDF membranes through electroblotting at 20 V, 1 h, in transfer buffer (25 mM Bicine, 25 mM Bis-Tris pH 7.2, 1 mM EDTA, 10% methanol). The gel was stained through conventional Coomassie staining to confirm that protein transfer to the membrane was successful. The membrane was washed five times with PBS-T (~30 mL; 137 mM NaCl, 27 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 0.1% Tween), rocking for 5 min at room temperature each time. After washing, the membrane was blocked with 5% skim milk in PBS-T (~30 mL), incubating for 1 h with rocking at room temperature, before being incubated overnight with rocking at 4 °C with the desired HRP-conjugated antibody/antibodies in PBS-T/milk (antibody dilutions as specified in Table 2.5), and with StrepTactin-HRP Conjugate (1:20000) for visualisation of the WesternC ladder. The membrane was washed five times with PBS-T in the same conditions

as above. Blots were developed using Immobilon™ Western Chemiluminescent HRP Substrate, according to the manufacturer's instructions, and were scanned using a LI-COR C-DiGit Blot Scanner (LI-COR Biotechnology; Lincoln, NE).

2.5 Production of homeodomain containing protein constructs

2.5.1 Bacterial transformations

Transformation protocols were adapted from previously used protocols in the Matthews laboratory [209, 210, 214]. Plasmid DNA (100 ng) was added to sterile KCM buffer (50 µL; 0.1 M KCl, 30 mM $CaCl_2$, 50 mM $MgCl_2$) or sterile TFB buffer (17 µL; 100 mM KCl, 10 mM MES, pH 6.2, 45 mM $MnCl_2$, 10 mM $CaCl_2$, 3 mM hexammine cobalt chloride). Resuspended competent *E. coli* cells (50 µL) were added, and the mixture incubated for 30 min at 4 °C. The mixture was incubated at 42 °C for 30 s (for HB101 cells) or 60 s (for all other strains), before being transferred to ice water for 2 min. Sterile SOC media was added (200 µL) and the solution was incubated for 1 h at 37 °C with shaking (150 rpm). The transformation mixture was inoculated onto LB agar plates containing appropriate selection conditions to ensure retention of the plasmid.

2.5.2 Protein overexpression

Overexpression protocols were adapted from previously used protocols in the Matthews laboratory [209, 214]. Several colonies of BL21 Gold PLysS (DE3), transformed with the appropriate plasmid, were inoculated into LB broth (10 mL per litre of expression culture) supplemented with both 100 µg/mL carbenicilin and 34 µg/mL chloramphenicol, and incubated overnight at 37 °C with shaking (150 rpm). This culture was used to inoculate fresh media (1 L) to an OD_{600} of 0.05, and the resulting culture was incubated at 37 °C with shaking (120 rpm), until an OD_{600} of 0.7 was reached. Protein expression was induced through the addition of IPTG (Isopropyl β-D-thiogalactopyranoside) to a concentration of 0.4 mM. Cultures were incubated further to allow protein expression to proceed. For the constructs NHD1, NHD3, 2HDN, and 2HD23, expression cultures were incubated for an induction period of 3 h at 37 °C, before being harvested. For 2HDLL, LLHD3, and LLHD1, the induction period was 16-20 h at 25 °C. After the induction period, cultures were harvested by centrifugation (30 min, 5000 × g, 4 °C) to sediment cells. Supernatant was removed, and pellets were snap frozen with liquid nitrogen and stored at -80 °C, for future lysis and purification.

2.5.3 Protein Purification

Purification protocols for all constructs followed the same procedure, based on previously established protocols from the Matthews laboratory [209, 214]. Due to differences in isoelectric point, buffers used throughout the purification of LLHD1 contained 20 mM Bis-Tris, pH 6.6, rather than 20 mM Tris, pH 8.0.

2.5.3.1 Cell lysis

Cell pellets were resuspended in lysis buffer, at a density of ~1 g cells/10 mL lysis buffer (20 mM Tris, pH 8.0, 300 mM NaCl, 1 mM PMSF, 1 mM β-mercaptoethanol, 10 µg/mL lysozyme, 10 µg/mL DNase). Once resuspended, cells were lysed using a French Press, pre-chilled at 4 °C, under a pressure of 10000-12000 psi. Soluble material was separated from insoluble material through centrifugation (40 min, 30000 × g, 4 °C).

2.5.3.2 PEI precipitation of nucleic acids

A solution of 10% PEI in lysis buffer was added dropwise to the clarified lysate at 4 °C with stirring, until a concentration of 0.8% was reached. This solution was incubated for a further 15 min at 4 °C with stirring, before being subjected to centrifugation (20 min, 10000 × g, 4 °C) to sediment precipitated nucleic acids and PEI.

2.5.3.3 Glutathione affinity purification of target protein

The supernatant recovered from PEI precipitation was applied to pre-equilibrated Glutathione Sepharose 4B resin (1 mL per L of expression culture) and incubated for 1 h at 4 °C with rotation (~40 rpm), to maximise binding to the resin. The resin was washed with 10 column volumes of wash buffer (20 mM Tris, pH 8.0, 50 mM NaCl, 10% glycerol, 1 mM PMSF, 1 mM β-mercaptoethanol) to remove contaminants. The resin was washed with cleavage buffer (10 column volumes; 20 mM Tris, pH 8.0, 50 mM NaCl, 1 mM PMSF, 1 mM β-mercaptoethanol). The resin was resuspended in 3 column volumes of cleavage buffer supplemented with HRV-3C protease, and proteolytic cleavage allowed to progress overnight at 4 °C with rotation (3 rpm).

2.5.3.4 Cation exchange chromatography

Cation exchange chromatography was used to further purify target proteins from contaminant proteins, as well as reducing any remaining nucleic acid contamination. A UnoS1 cation exchange column (BioRad; Hercules, CA) was used on a BioLogic DuoFlow FPLC system (BioRad; Hercules, CA) for this. Cation exchange chromatography was conducted in 20 mM Tris pH 8.0 (2HDN, 2HD23, 2HDLL, NHD1, NHD3, LLHD3) or 20 mM Bis-Tris pH 6.6 (LLHD1), supplemented with 1 mM DTT. A 0.05-1 M NaCl linear gradient was applied to elute proteins.

2.5.4 Protein concentration, dialysis and storage

Proteins were concentrated using Vivaspin® Centrifugal Concentrators (GE Healthcare; Chicago, IL) with a 3 kDa or 10 kDa molecular weight cut-off (MWCO), blocked with 1% Tween. Concentrators were subjected to centrifugation at 2000-3000 × g in a spin bucket centrifuge. Protein concentrations were measured using the absorbance at 280 nm (A280), as measured by an ND-1000 NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Extinction coefficients for all proteins are listed in Table 2.10 as estimated through ProtParam [215].

Table 2.10: Extinction coefficients of purified proteins.

Protein	Extinction coefficient (ε) (M^{-1} cm^{-1})
2HDN	20970
2HD23	19480
2HDLL	37930
NHD1	6990
NHD3	13980
LLHD1	21430
LLHD3	30940

Proteins were either dialysed for experimental use immediately after purification, or frozen in liquid nitrogen, and stored at -80 °C for future use. Dialysis was conducted overnight at 4 °C with stirring, using SnakeSkin™ 3.5 kDa molecular weight cut-off dialysis tubing (Thermo Fisher Scientific; Waltham, MA), except in the case of 2HDLL, where 10 kDa molecular weight cut-off dialysis tubing was used.

2.6 Homeodomain characterisation and DNA binding studies

2.6.1 Far-UV Circular dichroism (CD)

Freshly dialysed protein samples at 5-6 µM in buffer (20 mM sodium phosphate monobasic/dibasic, pH 7.4, 50 mM NaCl, 1 mM DTT) were used. Measurements were recorded using a Jasco J-815 spectropolarimeter equipped with a peltier-controlled sample holder (JASCO; Easton, MD), using 1-mm pathlength stoppered quartz cells. All reported temperatures refer to the sample temperature. For individual spectra, three scans were collected at 20 °C with a data pitch of 0.5 nm, a scanning speed of 20 nm/min, a digital integration time of 1 s, and a bandwidth of 1 nm and averaged. For spectra recorded over the course of a thermal denaturation/melt experiment, two scans per temperature point were collected, with a 1-nm data pitch, a digital integration time of 0.5 s, and a bandwidth of 1 nm. Spectra were recorded in increments of 2 °C, with the temperature increasing at a rate of 2 °C/min, with a 1 min delay time before collecting data. More simple melts were also recorded, measuring data only at 208 nm, 222 nm, and 247 nm. For these measurements, samples were heated at a rate of 1 °C/min, collecting data points every 2 °C. A bandwidth of 2 nm was used, with a digital integration time of 4 s. Data were corrected for buffer contribution and smoothed by the Savitzky Golay method [216] using Spectra Manager version 2.08 (JASCO; Easton, MD).

2.6.2 Nuclear magnetic resonance (NMR)

Freshly dialysed protein samples at 100 µM (NHD1) or 70 µM (NHD3) in buffer (20 mM sodium phosphate monobasic/dibasic, pH 7.0, 100 mM NaCl, 1 mM DTT) were supplemented with 5% D_2O and 2 µM DSS. An 800 MHz Bruker Avance 3 spectrometer, fitted with a cryogenic TCI probehead (Bruker BioSpin; Billerica, MA) was used to record one dimensional ^1H spectra of samples at 25 and 37 °C, with 128 scans taken for each acquisition. These spectra were processed using TOPSPIN 3.5PL7 (Bruker BioSpin; Billerica, MA). ^1H shifts were directly referenced to DSS at 0 ppm.

2.6.3 Fluorescent electrophoretic mobility shift assay (EMSA)

Protein was dialysed into EMSA buffer (10 mM HEPES pH 7.7, 50 mM KCl, 2.5 mM $MgCl_2$, 1 mM DTT) overnight, and used to construct a twofold dilution series with a top concentration of 6.8 µM. These protein samples were incubated for 45 min on ice with a specified concentration of fluorescein-labelled oligonucleotide, either 1 nM (M100) or 5 nM

(all other oligonucleotides), in reaction buffer (10 mM HEPES pH 7.7, 50 mM KCl, 2.5 mM $MgCl_2$, 10 mM DTT, 67 µg/mL acBSA, 4% Ficoll), in a final reaction volume of 30 µL. Acrylamide gels (8%) made in TBE buffer (5 mM Tris pH 7.8, 9 mM boric acid, 0.25 mM EDTA) were pre-run in 0.5 × TBE buffer for 1 h at 110V, before samples were loaded and subjected to electrophoresis at 110V for 3 h at room temperature. Gels were imaged using a Typhoon fluorescence imager (GE Healthcare; Chicago, IL).

2.6.4 Microscale thermophoresis (MST)

Protein was dialysed into reaction buffer (20 mM HEPES, pH 7.4, 100 mM NaCl, 1 mM DTT). A 1:2 dilution series of protein was prepared, with a top final concentration of 75 µM. Fluorescein-labelled oligonucleotide was added to these solutions, to a concentration of 50 nM, in a total volume of 30 µL. Samples were loaded into capillary tubes before undergoing microscale thermophoresis in a Monolith NT.115 instrument (NanoTemper Technologies; Munich, Germany). Thermophoresis was conducted at 95% LED power, and MST power at both 20% and 40%.

2.6.5 Isothermal titration calorimetry (ITC)

Protein and oligonucleotides being studied were dialysed individually into the same reservoir of buffer (20 mM Tris, pH 7.4, 150 mM NaCl, 1 mM TCEP) and their concentrations measured. A MicroCal i200 ITC calorimeter (Malvern Panalytical; Malvern, UK) was used to record data, injecting an oligonucleotide of choice into the cell, which contained a solution of the protein of choice (either 9.3 µM NHD3 or 9 µM NHD1). Injections of oligonucleotide into buffer (not containing protein) were used to correct for any changes in temperature not arising from protein-DNA interactions. Twenty injections of 2 µL each were recorded, with 120 s between injections and an injection speed of 0.5 µL/s. Data were analysed using Origin 7.0, using the ITC data analysis software package (Malvern Panalytical; Malvern, UK).

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

SEC-MALLS protocols were adapted from previously used protocols from the Matthews laboratory [209, 210]. Freshly dialysed protein samples were adjusted to a concentration of 2-3 mg/mL in buffer (20 mM sodium phosphate monobasic/dibasic, pH 7.0, 100 mM NaCl, 1 mM DTT). Protein solution (150 µL) was applied at 0.5 mL/min to a Superose 12 10/300 GL

column (GE Healthcare; Chicago, IL) on an Äkta Basic, fitted with in-line MiniDAWN™ TREOS multi angle light scattering and Optilab T-rEX differential refractive index detectors (both from Wyatt Technology, Santa Barbara, CA). ASTRA 6.1 from Wyatt software was used to normalise data and calculate the average molecular weight of the species of interest in the sample.

2.6.7 X-ray crystallography

Proteins were dialysed into EMSA buffer overnight (10 mM HEPES pH 7.7, 50 mM KCl, 2.5 mM $MgCl_2$, 1 mM DTT). The protein concentration was measured, and an equimolar amount of oligonucleotide (Table 6.1) added. The protein-DNA mix was concentrated (Section 2.5.4) until the volume had reduced to a volume corresponding to 15-20 mg/mL. The protein-DNA mix was subjected to centrifugation (5 min, 10000 × g, room temperature) to sediment any aggregates. A Freedom EVO 100 liquid dispensing robot (Tecan; Männedorf, Switzerland) was used to dispense 100-µL aliquots of crystallisation solutions into 96-well MRC plates (Molecular Dimensions; Suffolk, UK), before a Mosquito positive displacement liquid handling robot (TTP Labtech Ltd.; Cambridge, UK) was used to generate drops containing 200 nL mother liquor mixed with either 200 nL protein-DNA or 200 nL DNA-only samples. Crystal trays were incubated at 18 °C, and observed for signs of crystallisation beginning 24 h after initial inoculation.

Initial screens for protein-DNA crystals were conducted using commercial screens: PEG Rx, PEG Ion, Salt Rx, Index HT, Natrix HT, and Crystal Screen suites from Hampton Research (Aliso Viejo, CA), and JCSG+ HT and PACT HT suites from Molecular Dimensions (Suffolk, UK). These crystallisation screens were used to find conditions favourable for crystallisation of the desired species.

Four-gradient screening trays used for optimisation of crystallisation conditions were dispensed using the Freedom EVO 100 liquid dispensing robot. For each gradient, four solutions were prepared representing the most extreme conditions being screened. These solutions were then combined at different ratios using the Freedom EVO 100 liquid dispensing robot in order to produce screening gradients.

2.6.7.1 Preparation of crystals

Crystals were recovered from crystallisation drops and transferred into cryoprotection solution of the same composition as the crystallography solution, but with an additional 25% glycerol present. They were either dissolved or cryo-cooled in liquid nitrogen and stored in liquid nitrogen until being exposed to X-rays.

2.6.7.2 Characterisation of crystals

Crystals were analysed by SDS-PAGE to determine whether or not protein was present in the crystal. To achieve this, crystals were recovered from crystallisation wells and dissolved in NuPAGE® LDS Sample buffer, before being loaded onto Bolt™ 4-12% polyacrylamide gels and subjected to SDS-PAGE. Mark12™ protein standard (2 µL) was included to estimate protein size. Proteins were visualised using SYPRO Ruby, according to the manufacturer's instructions. Briefly, the gel was fixed (7% acetic acid, 50% methanol), before being stained with SYPRO™ Ruby gel stain. The gel was destained (7% acetic acid, 10% methanol) before being visualised using a Typhoon fluorescence imager (GE Healthcare; Chicago, IL). Cryoprotected crystals were sent to the Australian Synchrotron for X-ray diffraction experiments, on the MX2 beamline, using an ADSC Quantum 315r detector.

2.6.8 Small angle X-ray scattering (SAXS)

SAXS experiments were conducted at the Australian Synchrotron SAXS beamline (Clayton, VIC), using a 1M Pilatus detector and a 900-mm camera length. Experiments were recorded in both static mode using a 96-well plate autoloader, and in line with size exclusion chromatography, using a Superose 12 10/300 GL column (GE Healthcare; Chicago, IL) with a 1 mL/min flow rate.

Protein (2HDLL) and DNA (M100c20b) samples were dialysed overnight into buffer (20 mM sodium phosphate monobasic/dibasic, pH 7.4, 100 mM NaCl, 1 mM DTT), before concentrations were adjusted to a desired range. For complex samples, DNA and protein were mixed together in a 1:1 ratio. For SEC-SAXS, one sample each of protein (2.55 mg/mL), DNA (3 mg/mL), and protein-DNA (3.9 mg/mL) was injected. For static SAXS, 1:2 serial dilutions were constructed, diluting the sample in a matched buffer, for a total of 6 samples per species. Concentrations for protein-DNA samples ranged between 125 µg/mL and 4 mg/mL; concentrations for DNA-only samples ranged between 30 µg/mL and 1

mg/mL. A single 2HDLL-only sample was analysed, at a concentration of 2.55 mg/mL. Samples were subjected to centrifugation (10000 × g, 5 min, room temperature) to sediment any aggregates, and degassed, before being subjected to SAXS.

Data were processed using scatterBrain version 2.82 (Australian Synchrotron; Clayton, VIC) and PrimusQT (ATSAS, EMBL Hamburg; Hamburg, Germany). Modelling was conducted using MONSA (ATSAS, EMBL Hamburg; Hamburg, Germany). Further information, including references for analysis programs can be found in Appendix E.

Part I

Protein-protein interactions

3 Screening for novel Isl1-interacting proteins through yeast two-hybrid library screening

3.1 Introduction

As discussed in Chapter 1, Isl1 plays a role in the development of many tissues. It is likely that Isl11 is acting in combination with other developmentally relevant transcription factors, where those factors vary according to tissue type, giving rise to diverse functions of Isl1. This chapter describes the use of yeast two-hybrid library screening to discover novel binding partners for Isl1 that could help to broaden our knowledge about the modes of action of Isl1 in different tissues.

3.2 The Yeast two-hybrid system for probing interactions

The yeast two-hybrid system has been established as a simple and efficient way of screening for direct protein-protein interactions [217]. It utilises the yeast transcription factor GAL4, which contains independently folding DNA binding (DBD) and activation (AD) domains as part of the same polypeptide chain. In a yeast two-hybrid experiment, two potential binding partners are subcloned into plasmids encoding either the $GAL4_{DBD}$ or the $GAL4_{AD}$ (Figure 3.1). An interaction between the two resulting fusion proteins will bring the domains of GAL4 into proximity, promoting the expression of reporter genes downstream of the *GAL* promoter. If there is no interaction between the binding partners, the $GAL4_{AD}$ will not be brought close to the $GAL4_{DBD}$, and will not induce gene expression. Stronger interactions will promote higher levels of expression of the downstream genes than weaker interactions and will persist under high stringency conditions.

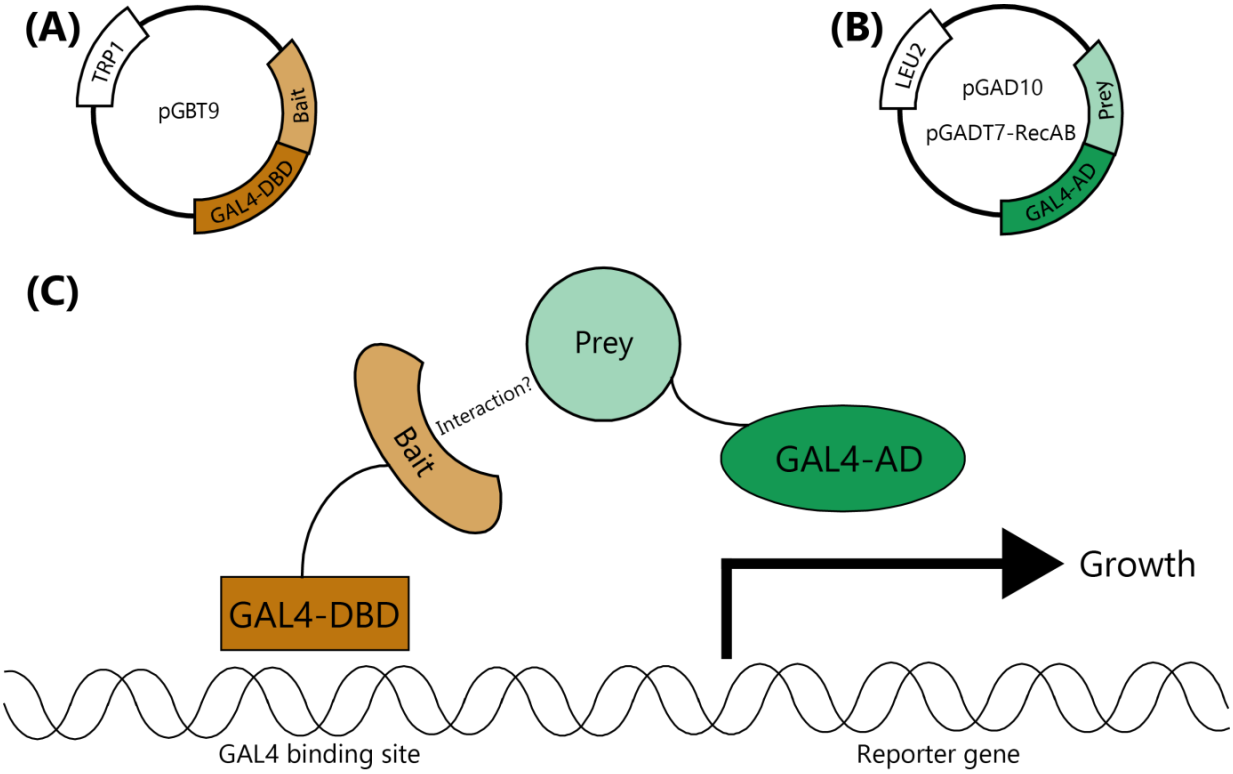


Figure 3.1: Schematic representation of the yeast two-hybrid system. Schematics of the plasmids used: (A) pGBT9 was used for all bait constructs; (B) pGAD10 was used for prey constructs made in the Matthews laboratory, and the prey library used pGADT7-RecAB. (C) Molecular mechanism behind yeast two-hybrid.

These interaction screens are conducted in auxotrophic strains of *Saccharomyces cerevisiae*. These yeast strains have been modified to grow only on media that contains specific nutrients, or when they are transformed with plasmids that encode genes to complete specific nutrient biosynthesis pathways. For example, the plasmids encoding the bait protein (pGBT9) and prey protein (pGAD10/pGADT7-RecAB) also encode one nutrient biosynthesis gene each for tryptophan (*TRP1*) and leucine (*LEU2*) respectively. Various strains of *S. cerevisiae* have been engineered to knock out either one or both of these genes, so that maintaining these yeast on media lacking leucine (-L) and/or tryptophan (-W) will select for only those yeast possessing the plasmids.

Further, these yeast strains have been modified such that the histidine biosynthesis gene *HIS3* is under the control of a *GAL* promoter. Hence, when in the absence of histidine, only the yeast expressing a pair of proteins that interact will be able to grow, as the presence of the interaction will bring the $GAL4_{AD}$ into proximity of the $GAL4_{DBD}$, bound to a *GAL* promoter, enabling expression of the *HIS3* gene.

3.2.1 Classifying yeast two-hybrid interactions according to strength

Many *S. cerevisiae* strains optimised for yeast two-hybrid analysis contain multiple reporter genes under the control of *GAL* promoters. In the work described here, three reporter genes were used: *HIS3*, *ADE2*, and *LacZ*. Each gene is under the control of a different GAL4 responsive promoter: *GAL1*, *GAL2*, and *MEL1* respectively. By varying the composition of the selective media used, the strength of an interaction can be estimated (Table 3.1).

Table 3.1: Selective media used for yeast two-hybrid screening. Exact media compositions can be found in Section 2.1.3.

Media	Nutrients of note	Use
SD-L-W	Excludes tryptophan and leucine	Growth control
SD-L-W-H	Excludes tryptophan, leucine, and histidine	Detects interactions of any strength
SD-L-W-H +0.5 mM 3-AT	Excludes tryptophan, leucine, and histidine. Includes 0.5 mM 3-AT	Eliminates weakest interactions
SD-L-W-H +2 mM 3-AT	Excludes tryptophan, leucine, and histidine. Includes 2 mM 3-AT	Detects medium and strong interactions
SD-L-W-H-A	Excludes tryptophan, leucine, histidine, and adenine	Detects only strong interactions

3.2.1.1 Assessing interaction strength using ADE2

The product of the *ADE2* gene is an enzyme in the adenine biosynthesis pathway in yeast. In general, weak interactions between bait and prey are insufficient to simultaneously activate expression of both *HIS3* and *ADE2*, making it useful as a selectable marker for identifying strong interactions (Table 3.1).

3.2.1.2 Colourimetric assessment of interactions using LacZ

LacZ encodes a galactosidase enzyme that can be used as a colourimetric marker when combined with media containing X-α-gal (5-Bromo-4-chloro-3-indoxyl-α-D-galactopyranoside): an interaction between the bait and prey will result in production of an α-galactosidase, which cleaves the X-α-gal, forming galactose and an indole that spontaneously dimerises to form a blue product [218, 219]. A more intense blue appearance correlates with

increased expression of the galactosidase, and so indicates a stronger interaction. Blue colouration may not always be seen, as the *MEL1* promoter is more weakly bound by GAL4 than the *GAL1* and *GAL2* promoters [220].

3.2.1.3 Discerning weak and moderate interactions using HIS3 expression

The *HIS3* gene can also be used to indicate the strength of an interaction, through addition of the competitive inhibitor 3-AT (3-amino 1,2,4-triazole). 3-AT inhibits the protein product of the *HIS3* gene, and so inhibits growth. Adding increasing amounts of 3-AT will allow moderately strong interactions to be distinguished from weak interactions, as only stronger interactions will induce higher levels of *HIS3* expression, overcoming the effect of 3-AT. The levels of 3-AT used throughout this work are described in Section 2.1.3 and Table 3.1.

3.3 Yeast two-hybrid library screens were conducted using Isl1 as bait

The yeast two-hybrid system was originally developed to screen for novel binding partners in a library format. This is achieved through mating compatible strains of yeast (one with the allele *MATa* and the other with the allele *MATα*), where one strain is transformed with the bait plasmid encoding the protein of interest, and where the other strain is transformed with a library of potential binding partners subcloned into a prey vector [221]. When the two strains of yeast mate, the resulting hybrid yeast will contain both the bait protein plasmid and one prey plasmid from the library. The progeny can then be screened on nutritionally selective media to select for bait-prey interactions.

In this thesis, yeast two-hybrid mating screens were used to identify putative Isl1-interacting proteins. These screens used constructs of Isl1 fused to the $GAL4_{DBD}$ as bait, and a commercial library of pre-transformed yeast as prey.

3.3.1 Three constructs of Isl1 were chosen as baits for screening

Full length Isl1 protein could not be used for screening, as it has been found by the Matthews laboratory to show high levels of auto-activation in the yeast two-hybrid system [222]. That is, yeast transformed with pGBT9 encoding full length Isl1 and pGAD10 encoding only the $GAL4_{AD}$, are able to grow on media lacking histidine and adenine. To avoid this problem, three constructs containing portions of the Isl1 protein were chosen for screening: $Isl1_{LIM}$,

$Isl1_{\Delta LIM}$, and $Ldb1_{LID}/Isl1_{LIM}$ (Figure 3.2). These constructs had previously been subcloned into the pGBT9 vector (Amy Nancarrow, University of Sydney), and as part of my Honours work in 2014 were shown to have little to no auto-activation (Figure 3.2B) [209]. Library screens were conducted on selective media designed to eliminate auto-activation, and so minimise the occurrence of false positives where yeast growth occurs in the absence of an interaction between Isl1 and a bait protein.

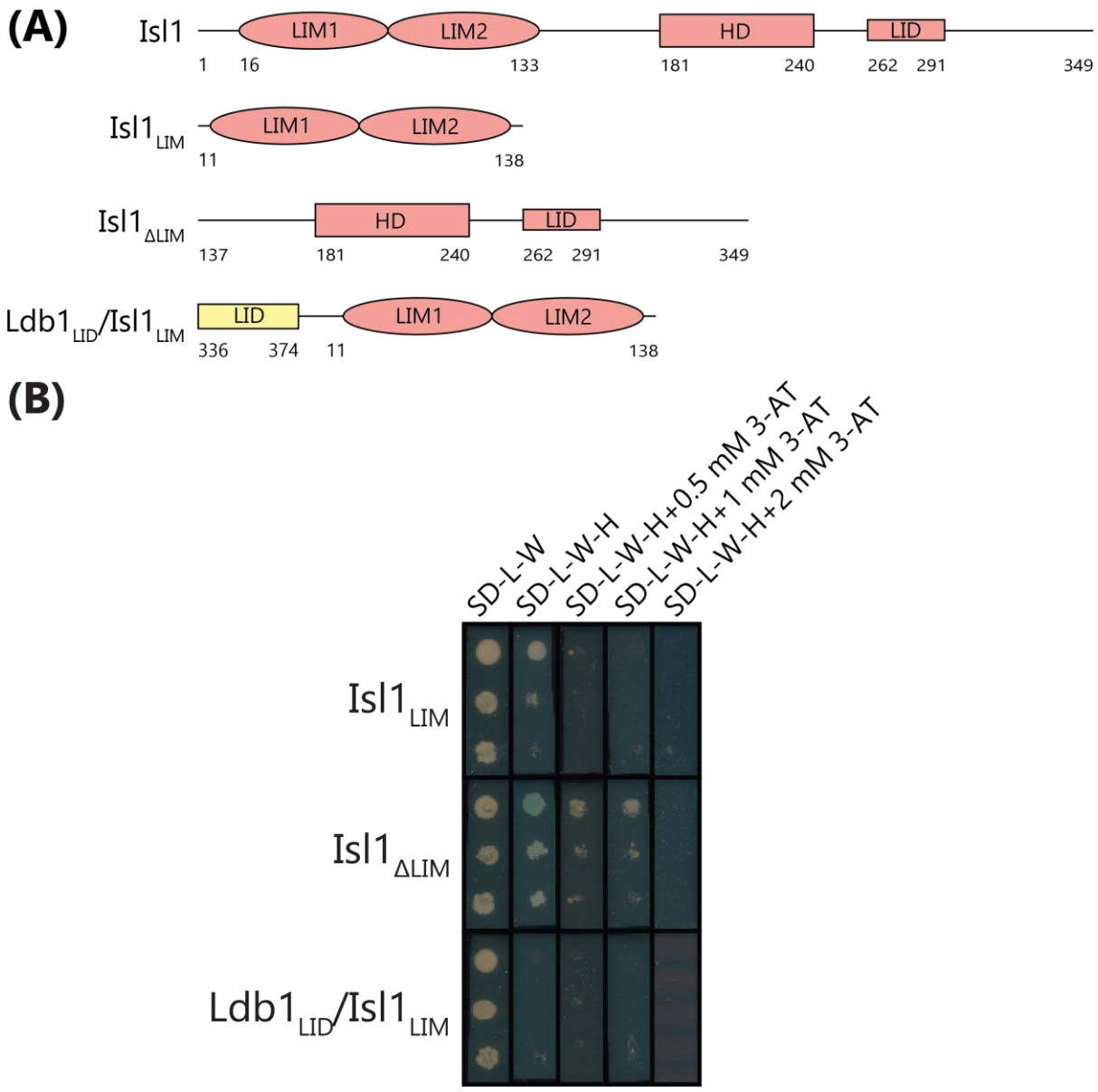


Figure 3.2: Schematic representations of Isl1 constructs used for library screening. (A) Domains of Isl1 bait constructs. (B) Previously determined auto-activation levels of Isl1 bait constructs, using yeast co-transformed with an empty pGAD10 plasmid and an Isl1-encoding pGBT9 plasmid. Each set of three spots contains a 1:10 serial dilution of yeast, from OD_{600} 0.2 to 0.002.

3.3.2 Construction of the Clontech Mate & Plate Library

A Clontech Normalised Universal Mouse Mate & Plate library was purchased in 2014 for conducting these screens, providing enough material for performing five screens. Two of these screens were conducted in 2014, with the other three being undertaken during the course of this thesis [209].

The commercial library was generated through isolating RNA from mouse tissues and producing complementary DNA (cDNA) from this, using primers that anneal to the poly-adenylation signal on mature mRNAs (Figure 3.3) [223]. In this way, the library was designed to be biased towards proteins being expressed in the source tissue. Following cDNA production, a normalisation step was undertaken to reduce the prevalence of abundant transcripts. Normalisation used a duplex specific nuclease that selectively cleaves double stranded DNA [223]. The pool of cDNAs was heated to separate the two strands of each transcript, and then slowly cooled. More abundant transcripts should anneal at higher temperatures due to their higher concentration in solution. Those transcripts were digested by the duplex specific nuclease, while low abundance transcripts should have remained single stranded and have been protected from digestion. All remaining cDNAs were then cloned into a prey vector (pGADT7-RecAB) using restriction digestion cloning that cuts at sites introduced into the cDNA during the conversion of mRNA into cDNA. The resulting library of plasmids was transformed into a *MATα* strain of *S. cerevisiae*, Y187.

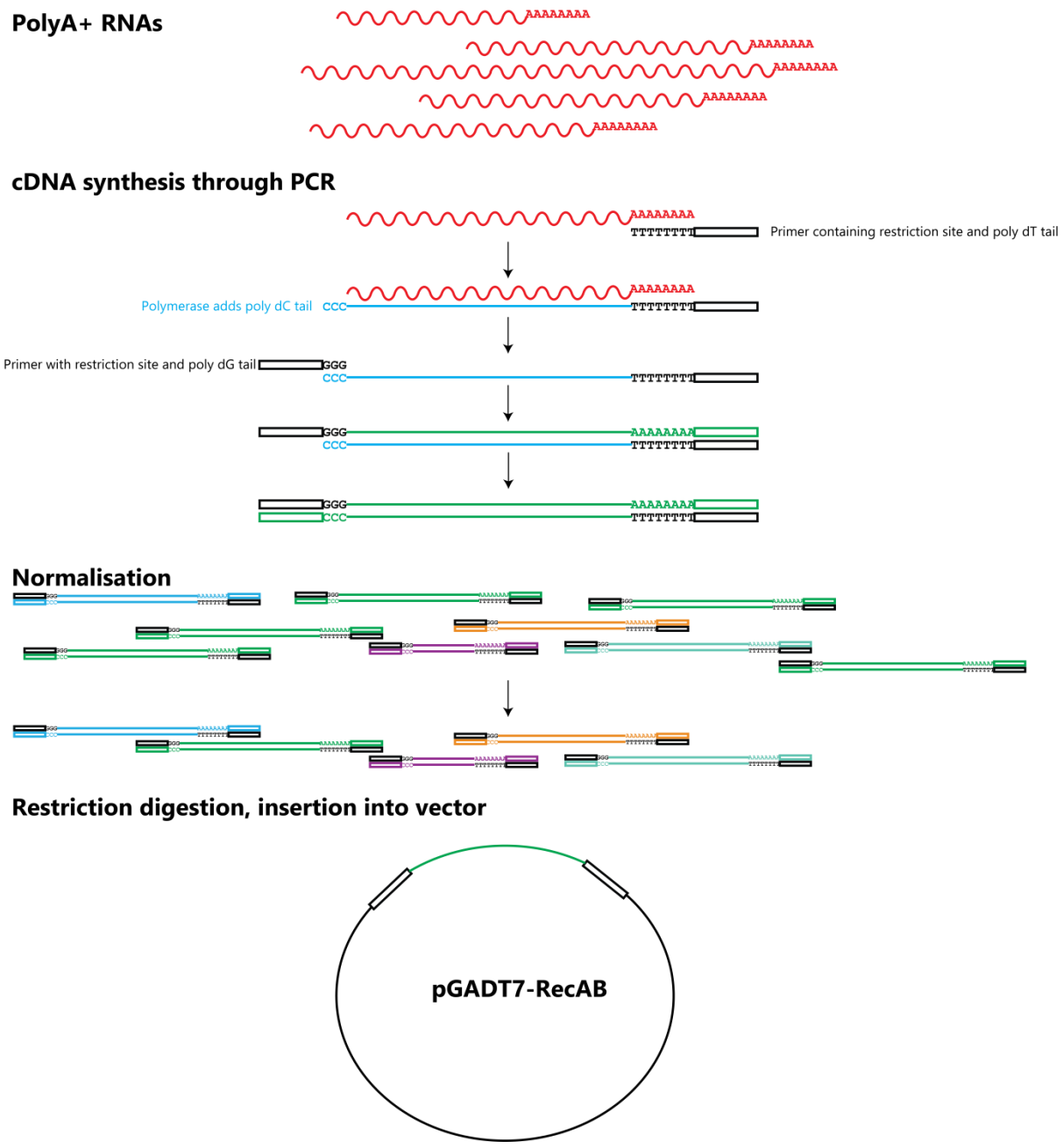


Figure 3.3: Schematic of the production of the cDNA library used for interaction partner screening.

3.3.3 Mating screens against Isl1 were conducted successfully

The bait construct of interest was first transformed into AH109, a MATa strain of yeast. This bait strain was mated to Y187 yeast containing the library prey. The mated yeast were then inoculated onto agar plates using nutritional selection to search for bait-prey interactions. In total, five screens were conducted, with two each against $Isl1_{LIM}$ and $Isl1_{\Delta LIM}$, and one against $Ldb1_{LID}/Isl1_{LIM}$. Two of these screens (one against $Isl1_{LIM}$ and one against $Ldb1_{LID}/Isl1_{LIM}$) were conducted in 2014, as part of my Honours project; however, the resulting pools of prey

were only partially processed in that time [209]. As further processing of these screens contributed to this thesis, all five screens will be discussed.

All screening experiments were conducted correctly. This was judged both by the appearance of colonies on the screening plates after incubation, and by titrations of all strains of yeast to quantify viability of both binding partners, mating efficiency, and screening efficiency (reported in Appendix D). While some inconsistencies in mating efficiencies were encountered, these tests indicate that more than 20 million clones were screened in each mating screen performed.

Initially, one screen against each Isl1 construct was planned. However, due to significant sample loss occurring during the processing of the first $Isl1_{LIM}$ pool of interactors, and concerns that the nutritional selection was too harsh in the first $Isl1_{\Delta LIM}$ screen, additional screens were carried out for these two constructs (Table 3.2).

Table 3.2: Comparison of yeast isolated from mating screens.

Isl1 construct used as bait	Selective media used	Colonies observed
$Ldb1_{LID}/Isl1_{LIM}$	SD-L-W-H	74
$Isl1_{LIM}$ screen A	SD-L-W-H+0.5 mM 3-AT	>1000
$Isl1_{LIM}$ screen B	SD-L-W-H+0.5 mM 3-AT	>1000
$Isl1_{\Delta LIM}$ screen A	SD-L-W-H+2.5 mM 3-AT	35
$Isl1_{\Delta LIM}$ screen B	SD-L-W-H+2 mM 3-AT	200

The variation in the number of colonies observed during the initial screening was taken as an indication that the screening process was effective. As would be expected, screening with the known protein-protein interaction domain $Isl1_{LIM}$, returned a much larger number (> tenfold) of potential interactors than screening with the same LIM domains bound to a known binding partner, Ldb1, which should prevent the primary peptide-binding face from making interactions (Section 1.2.2.1, Table 3.2).

Due to the large number of colonies observed in the $Isl1_{LIM}$ screens, a fraction of the total number of yeast colonies were isolated for identification: 180 from Screen A, and 260 from Screen B. The colonies isolated from these screens were chosen based on largest size and most intense colour, as these should represent the strongest bait-prey interactions.

3.4 Prey plasmids were isolated and prey proteins identified

The methodology selected for identifying the interacting prey was to isolate the intact plasmids from the yeast for sequencing, identification, and downstream applications. This required extraction of plasmid DNA from the yeast, and amplification in bacteria before final isolation of the plasmid.

Several different methods were trialled to obtain high concentration, pure plasmid from hits (Section 2.3.4). Compared to *E. coli*, *S. cerevisiae* has at least a tenfold lower copy number for plasmids [224], and plasmid recovery proved technically challenging. The protocol used required either a large amount of yeast culture, or a very high efficiency of plasmid extraction, to maximise plasmid yield from yeast.

The final procedure used was adapted from Singh & Weil, and involved lysing yeast through alkaline lysis methods after enzymatic digestion of the outer cell wall of the yeast, followed by clarification of the lysate [225]. This fraction was then transformed into freshly prepared competent HB101 *E. coli*, prepared according to the Inoue method [212]. The combination of these two methods consistently resulted in successful recovery of the prey plasmids for sequencing and identification.

The vast majority of isolated prey plasmids were easily identified through Sanger sequencing, using primers that annealed to sequence upstream and downstream of the prey-coding region. However, there were 32 prey constructs that could not be identified successfully by this approach. This was especially the case for the $Isl1_{\Delta LIM}$ screens, where sequencing of 27 out of 235 prey plasmids resulted in anomalous sequencing data (Figure 3.4). In those cases, the plasmids were discarded, as they most likely represented plasmids with anomalous sequence features, and likely did not contain protein coding sequence.

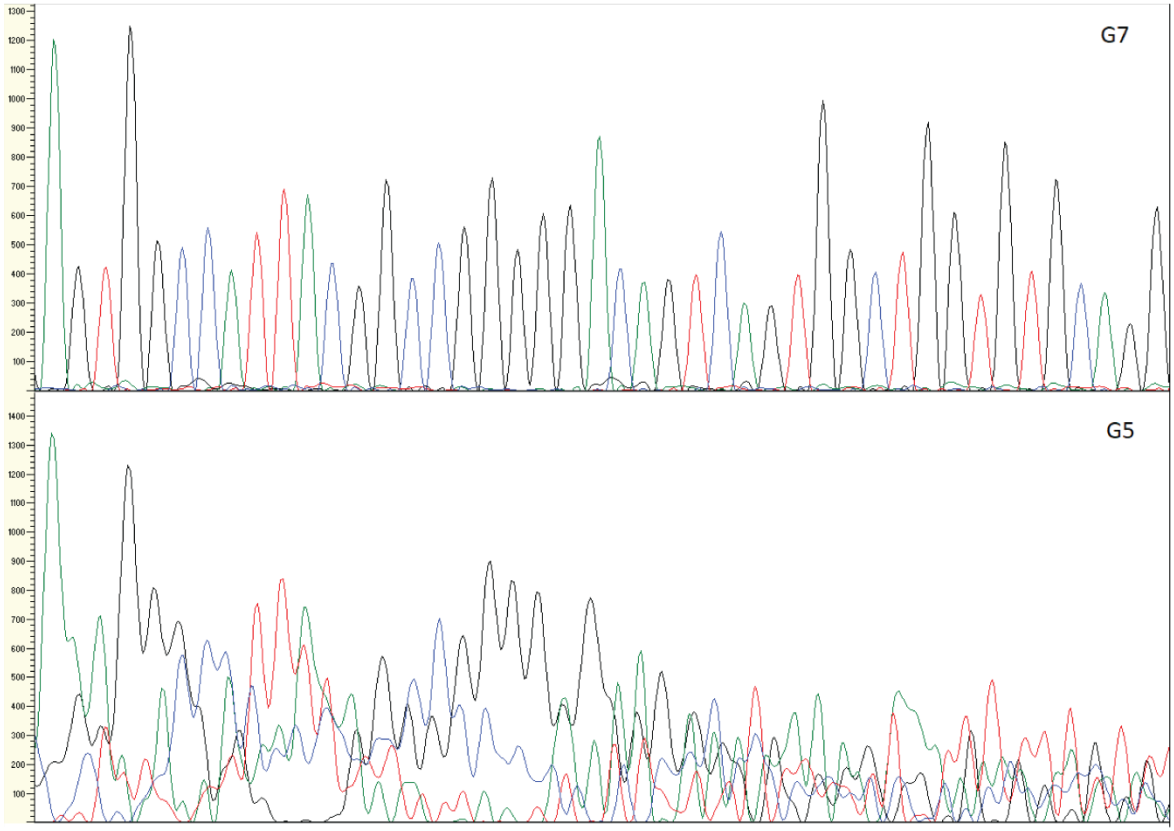


Figure 3.4: Comparison of good quality and anomalous sequencing data of library prey plasmids. Top: sequencing chromatogram from a typical Isl1$_{\Delta LIM}$ prey plasmid (Isl1$_{\Delta LIM}$ hit G7). Bottom: Example of a sequencing chromatogram from a prey plasmid containing an unidentifiable prey gene (Isl1$_{\Delta LIM}$ hit G5), estimated to be from the same region as in the top panel. Note that this region is upstream of the prey-coding region.

3.5 Not all identified prey encoded proteins

Once sequence information was available, each prey construct could be categorised according to sequence type. Through BLAST (Basic Local Search Alignment Tool) searching, it was found that many of the prey-encoded sequences that did not correspond to any known protein (Table 3.3). These prey constructs were most commonly found to contain cDNA corresponding to the 3'-untranslated regions (UTRs) of proteins. This is most likely an artefact from the production of the cDNA library, which used a poly-dT primer to anneal to cellular mRNAs via the 3'-polyadenylation signal (Section 3.3.2). 3'-UTRs are commonly found between the stop codon of an open reading frame and the polyadenylation signal. 5'-degradation of the RNAs during the construction of the library could additionally explain why such a large proportion of identified prey constructs (up to 56% in a given screen) contained only 3'-UTR sequence without any protein-coding region. As these hits did not appear to represent novel binding partners for Isl1, they were not further investigated.

Table 3.3: Proportion of isolated prey encoding proteins across library screens.

Screen	$Ldb1_{LID}/Isl1_{LIM}$	$Isl1_{LIM}$ screen A	$Isl1_{LIM}$ screen B	$Isl1_{\Delta LIM}$ screen A	$Isl1_{\Delta LIM}$ screen B
Prey isolated	74	180*	260*	35	200
Prey identified	74	105	258	31	177
Prey encoding proteins in frame	27 (36%)	73 (70%)	150 (58%)	5 (16%)	23 (12%)
Prey encoding frameshifted proteins	8 (11%)	14 (13%)	33 (13%)	8 (25%)	57 (29%)
Prey not encoding proteins	38 (51%)	18 (17%)	75 (29%)	18 (56%)	94 (47%)
Non-redundant in-frame proteins	11	25	46	4	21

* Screen grew >1000 colonies. Strongest interactions were isolated for identification.

3.5.1 Investigation of frameshifted proteins

A significant proportion of hits (30% overall) encoded protein-coding sequences that were out of frame from the upstream GAL4 protein sequence (Table 3.3). Documentation from Clontech indicated that yeast are tolerant to ribosomal frameshifts, meaning that during translation the mRNA can shift by one base in either direction with respect to the ribosome [211, 226]. This phenomenon would make it possible for these frameshifted sequences to be expressed correctly, even though they are encoded in a different reading frame to the upstream GAL4 sequence. However, it was important to establish whether this was the case for this set of sequences.

The pGADT7-RecAB plasmid encodes a hemagglutinin (HA) tag immediately downstream of the $GAL4_{AD}$, upstream of the multiple cloning site (MCS) in which the prey protein coding sequence is inserted, allowing detection of the expressed protein in an anti-HA Western blot. Several plasmids were chosen that contained protein-coding sequence that would result in a significant size difference of the expressed protein, depending on whether the protein sequence expressed was in frame with the GAL4 sequence or not (Table 3.4).

Table 3.4: Expected size of potentially frameshifted prey. Predicted size of the protein product unique to each clone is given for each reading frame, in kDa. For each prey, the frame that corresponds to a known protein is underlined. Note that these sizes include the size of the upstream sequence, which includes the HA tag, $GAL4_{AD}$, and SV40 NLS (simian virus 40 nuclear localisation sequence) (~22 kDa).

Prey ID	Frame 1 size (kDa)	Frame 2 size (kDa) (GAL4 frame)	Frame 3 size (kDa)
E4	<u>24.4</u>	29.6	23.5
E7	22.9	24.6	<u>44.9</u>
M7	25.8	23.8	<u>34.6</u>
P10	23.5	28.2	<u>33.3</u>
T8	22	24.1	<u>39.2</u>
U5	22.5	29.2	<u>27</u>

The plasmids selected were isolated as part of the $Isl1_{LIM}$ screening process and showed evidence of strong interactions with the $Isl1_{LIM}$ construct. The plasmids were co-transformed into yeast along with Isl_{LIM} and grown in either media selecting for the presence of plasmids (SD-L-W), or media selecting for a strong interaction between bait and prey (SD-L-W-H-A), to determine if inducing protein expression selected for protein produced in a particular frame. The protein was then extracted (Section 2.3.6) and subjected to anti-HA Western blot analysis, to observe the size of the expressed prey protein (Figure 3.5).

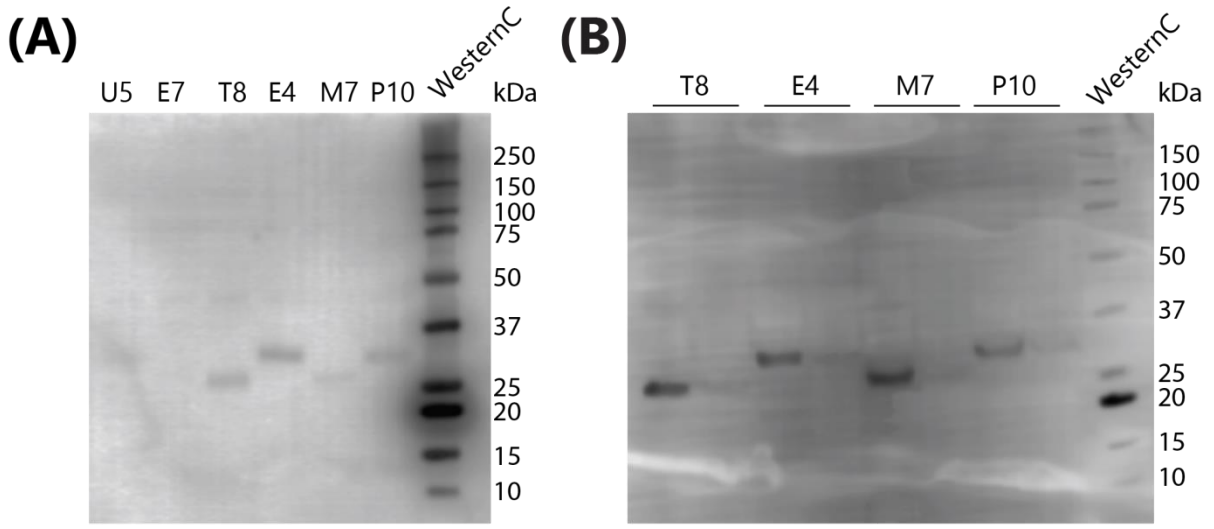


Figure 3.5: Anti-HA Western blots of protein extracts from yeast. Yeast were co-transformed with pGBT9-Isl1$_{LIM}$ and a library pGADT7-RecAB (E4, E7, M7, P10, or U5), and grown to saturation in selective media, before protein was extracted. Total protein extracts were subjected to anti-HA Western blot analysis. (A) Blot of extracts of yeast grown in SD-L-W media, selecting for the presence of plasmids. (B) Blot of extracts of yeast grown in media selecting for plasmid presence (SD-L-W; right lane for each sample) or for a strong interaction (SD-L-W-H-A; left lane for each sample).

Samples U5 and E7 did not show levels of expression detectable by Western blot, so conclusions were drawn from the other four samples. All extracts from media selecting for a strong interaction showed a more intense band, signifying increased levels of protein expression (Figure 3.5B). This may occur as the growth conditions would favour yeast that are able to produce more of the *ADE2* gene product, potentially because they have higher levels of expression of the prey protein, inducing more expression of the *ADE2* gene.

Comparing the sizes of the detected proteins between preys, it appears that E4 and P10 expressed proteins of very similar sizes (around 30 kDa), with the proteins being expressed in M7 and T8 being close in size, but slightly smaller (around 25 kDa). Of the three potential coding frames, only frame 2 fulfils all these criteria, with E4 and P10 being 1.4 kDa different in size (29.6 and 28.2 kDa respectively), M7 and T8 being 0.3 kDa different in size (23.8 and 24.1 kDa respectively), and the two pairs being ~5 kDa different in size (Table 3.4). Given that Frame 2 is expected, these data indicate that no ribosomal frameshifts are occurring. As all these protein sequences that interacted with Isl1$_{LIM}$ appear to represent nonsense peptides, and not a protein-protein interaction that occurs *in vivo*, hits found to encode frameshifted proteins were not pursued further.

3.6 Primary validation of putative interactors to check for prey auto-activation

Identifying false positives and non-specific interactions is an essential step of any screen for novel interaction partners. Yeast two-hybrid experiments can have a high false positive rate due in part to the nature of the artificial fusion proteins used in screening. For this reason, validation experiments formed a crucial step in the workflow. Of the 398 protein-encoding hits from the five screens carried out (Table 3.3), 147 were found to be unique, with the other 149 being duplicates. Additional validation steps were carried out with the pool of 147 putative interactors to test the likelihood that these interactions were both specific and relevant to Isl1. Note that these 147 included hits that encoded different truncation constructs of the same protein.

First, yeast two-hybrid analysis was used to check for auto-activation of the putative interactors. Co-transformations, rather than mating, followed by spot test assays, were conducted for this purpose (Section 2.3.1). Co-transformed yeast were inoculated onto a range of different media (Table 3.1), to test both for auto-activation of the prey construct, and also to probe the strength of putative interactions (Figure 3.6).

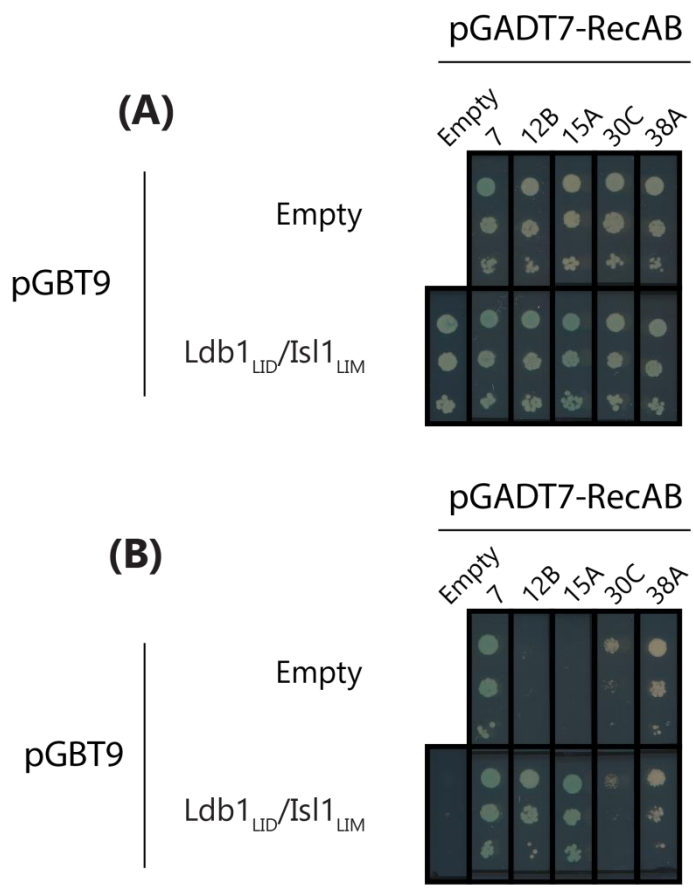


Figure 3.6: An example of yeast two-hybrid assays to assess auto-activation. Yeast co-transformed with one pGBT9 plasmid and one pGAD plasmid were inoculated onto (A) Growth control media (SD-L-W). (B) Selective media (SD-L-W-H+ 0.5 mM 3-AT), and allowed to grow for 3 days. Each set of three spots contains a serial 1:10 dilution of yeast, from OD_{600} 0.2 to 0.002. Prey 12B and 15A were classified as genuine interactors as they displayed no growth when screened against the empty pGBT9; 7, 30C, and 38A were considered false positives.

This process eliminated 12 false positives, as the growth pattern shown with the prey alone for these hits was the same as the growth pattern shown in the presence of both the bait and the prey. It also allowed categorisation of the remaining hits according to strength of interaction (Table 3.5). All hits that encoded different truncation constructs of the same protein behaved consistently. Different trends were observed for each Isl1 construct screened, with the pool of potential LIM domain interactors being the largest, and having the greatest proportion of strong interactions. These trends confirmed observations made during the recovery of hits from initial screening (Section 3.3.3).

Table 3.5 Distribution of interaction strengths according to the Isl1 construct used as bait. Percentages listed are as a proportion of the non-redundant hits for each Isl1 construct.

Construct	Strong	Medium	Weak	False positive
$Isl1_{LIM}$	95% (53)	4% (2)	2% (1)	0
$Isl1_{\Delta LIM}$	4% (1)	36% (9)	24% (6)	36% (9)
$Ldb1_{LID}/Isl1_{LIM}$	0	13% (1)	50% (4)	38% (3)

3.7 Further assessing false positives and non-specific interactions

After eliminating false positives originating from prey auto-activation, the remaining pool of 73 putative interactors may still contain some false positives, and is also likely to contain non-specific interactors. For this reason, further steps were taken to assess the likelihood of each prey protein representing a biologically relevant interaction.

3.7.1 The CRAPome database was used to screen for non-specific binders

The growing number of large-scale protein-protein interaction screens being published in the literature has enabled the creation of resources that identify common hits from screens that are unlikely to represent relevant interactions. CRAPome (Contaminant Repository for Affinity Purification) is one such resource, which predominantly uses mass spectrometry screening data [227]. Hits that occur with a high frequency in the CRAPome dataset are considered to represent non-specific interacting proteins. Of the 73 remaining proteins from screening, there was a relatively small number of proteins that appeared in the CRAPome dataset at a high frequency (Figure 3.7), suggesting that the yeast two-hybrid prey pool does not contain interactions representing common contaminating proteins.

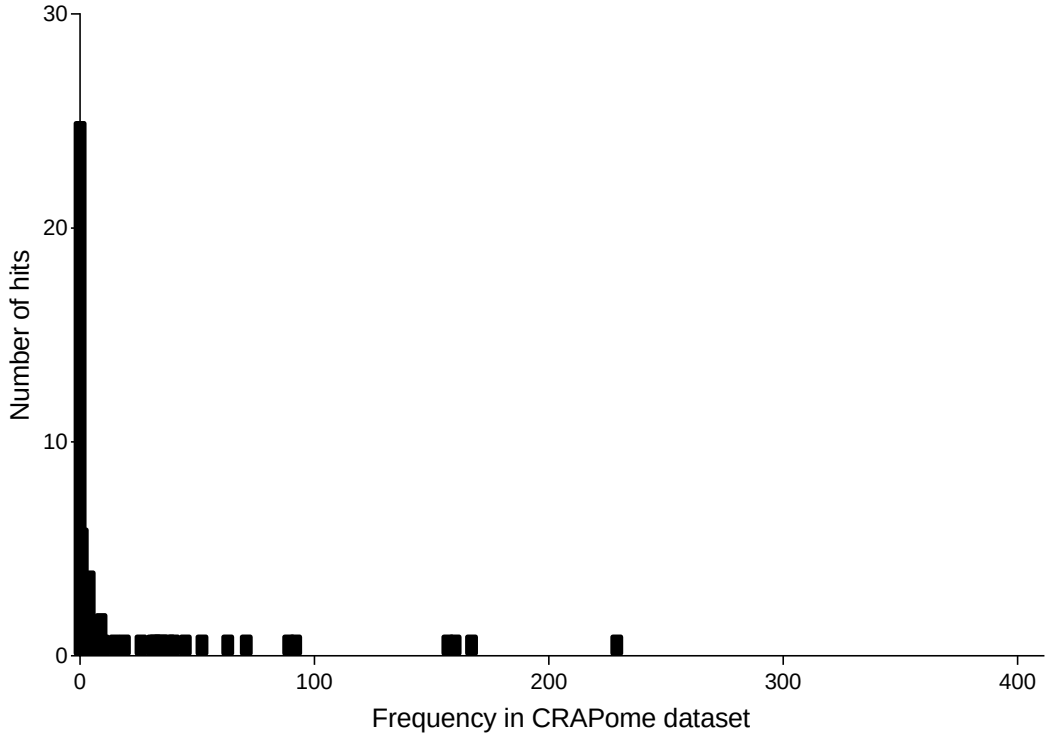


Figure 3.7: Prevalence of putative Isl1-interacting proteins in CRAPome database. 66 out of 73 non-redundant identified putative interactors had entries in CRAPome. The maximum possible score in CRAPome is 411, representing a protein being detected in every screen in the database.

Interestingly, the four highest CRAPome scoring proteins from the pool of potential interactors were all ribosomal proteins. This reflects a trend found across interaction mapping studies in which ribosomal proteins are commonly found as false positives [227]. No prey constructs were discarded at this stage, as very few were present in the CRAPome dataset at high frequencies, and they would likely prove useful comparisons when checking for the specificity of the interaction with Isl1.

3.7.2 Subcellular localisation can be used to assess likelihood of an interaction occurring

Screening hits for subcellular localisation can be a useful indicator for whether it is likely that biologically relevant interaction partners have been detected. Isl1 is primarily localised in the nucleus of cells, meaning the most likely interaction partners will be those also located in the nucleus (Figure 3.8) [228].

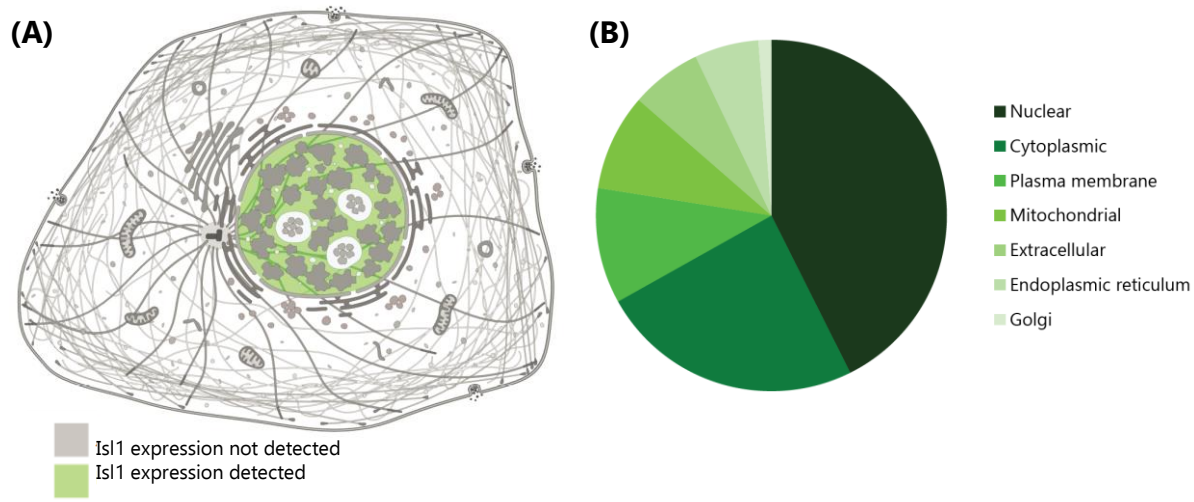


Figure 3.8: Analysing subcellular localisations of Isl1-interacting proteins. (A) Subcellular distribution of Isl1, adapted from Human Protein Atlas (www.proteinatlas.org). (B) Predicted subcellular localisations of potential Isl1-interacting proteins identified across all screens.

Based on either known or predicted localisations, 44% of the total pool of protein hits were proteins that localise in the nucleus (Table 3.6) [229]. Given that nuclear proteins are predicted to comprise 10-20% of the total mammalian proteome [230], this statistic indicates that the screening process enriched for nuclear proteins in the set of putative interactors. This trend was strongest in the $Isl1_{LIM}$ interaction pool. However, given the small sizes of the $Isl1_{\Delta LIM}$ and $Ldb1_{LID}/Isl1_{LIM}$ interaction pools, it is difficult to say whether this trend is significant.

Table 3.6: Subcellular localisations of non-redundant protein hits. Note that several proteins occurred in more than one screen; these are only counted once in the overall pool.

Isl1 bait construct	Nuclear proteins	Non-nuclear proteins	% nuclear proteins
$Isl1_{LIM}$	29	26	53%
$Isl1_{\Delta LIM}$	3	14	18%
$Ldb1_{LID}/Isl1_{LIM}$	1	5	17%
Overall	32	41	44%

It should be noted that in many cases the localisations assigned to putative interacting proteins are predictions only (no experimental data available). Thus, potential interacting proteins were not eliminated from the pool of samples based solely on subcellular localisation.

3.7.3 Using yeast two-hybrid to check for specific binders reveals many non-specific interactors

A straightforward method of determining if a prey protein interacts specifically with Isl1 is to screen the prey against other potential binding partners. Yeast two-hybrid assays were used for screening all potential interactors against Isl2, the close homolog of Isl1, and additional other proteins as explained below.

3.7.3.1 Non-specific interactions in the Isl1$_{LIM}$ prey pool

A set of 54 prey constructs representing each Isl1$_{LIM}$-interacting protein identified were tested for specificity of binding to Isl1$_{LIM}$ by screening against LIM domains from other proteins, including Isl2, LIM domain only protein 4 (Lmo4, another LIM domain transcription factor), and LIM kinase 2 (Limk2, a cytoplasmic LIM protein). Hits that interacted with Lmo4$_{LIM}$ may represent biologically relevant, but less specific, binding partners. Hits that interact with Limk2$_{LIM}$ are more likely to represent non-specific interactions that are not biologically relevant to the function of Isl1.

Of the pool tested, 42 Isl1$_{LIM}$-interacting prey constructs showed strong interactions with Limk2$_{LIM}$, with many also showing strong interactions with Lmo4$_{LIM}$ (Table 3.7). All prey constructs except two showed a strong interaction with Isl2$_{LIM}$. Those were Meprin A subunit beat (Mep1b) and Muskelin (Mkln1), which only showed evidence of weak interactions with any of the tested non-Isl1 LIM domains, highlighting them as potentially specific Isl1$_{LIM}$ interaction partners.

Table 3.7: Screening putative interactors for interaction specificity. Breakdown of putative interactors by Isl1 construct, showing strengths of interactions with Isl1-like proteins.

		Limk2 interaction strength			
Lmo4 interaction strength		Strong	Moderate	Weak	No interaction
	Strong	15	0	0	3
	Moderate	5	0	0	2
	Weak	20	0	0	7
	No interaction	2	0	0	0

Overall, nine of the $Isl1_{LIM}$-interacting pool showed strongest binding to $Isl1_{LIM}$, with minimal binding to the other LIM domains screened. These were: Ddx20, Dfna5, Lace1, Mep1b, Mkln1, Nup50, Rps18, Sparcl1, and Zfand1.

3.7.3.2 Checking the interaction interface of $Ldb1_{LID}$/$Isl1_{LIM}$-interacting prey

A set of six hits from the original $Ldb1_{LID}$/$Isl1_{LIM}$ screen were tested for specificity: Isoaspartyl peptidase (Asrgl1), Coatomer subunit beta (Copb1), COP9 signalosome complex subunit 5 (Cops5), Mkln1, Transient receptor potential channel 1 (Trpc1), and Ubiquitin D (Ubd). Of these, Asrgl1, Cops5, and Mkln1 had also been isolated in the $Isl1_{LIM}$ screening. This was taken as an indication that these proteins may not be interacting with the $Ldb1_{LID}$/$Isl1_{LIM}$ complex as a whole, but might compete with $Ldb1_{LID}$ for binding of $Isl1_{LIM}$. Accordingly, $Ldb1_{LID}$/$Isl1_{LIM}$-interacting hits were screened against $Isl1_{LIM}$, $Isl2_{LID}$, and $Lmo4_{LIM}$, to check both for specificity of binding to $Ldb1_{LID}$/$Isl1_{LIM}$, and for which part of the $Ldb1_{LID}$/$Isl1_{LIM}$ construct was facilitating the interaction.

Asrgl1, Cops5, and Trpc1 were found to interact strongly with $Isl1_{LIM}$ and $Lmo4_{LIM}$, indicating that they most likely interact with LIM domains, but not specifically with Isl1 (Figure 3.9). Therefore, they were not pursued further. Copb1, Mkln1, and Ubd were found to interact weakly with Lmo4, but showed varied interaction strengths with $Isl1_{LIM}$. Copb1 showed weak binding, Ubd showed moderate binding, and Mkln1 showed strong binding to $Isl1_{LIM}$.

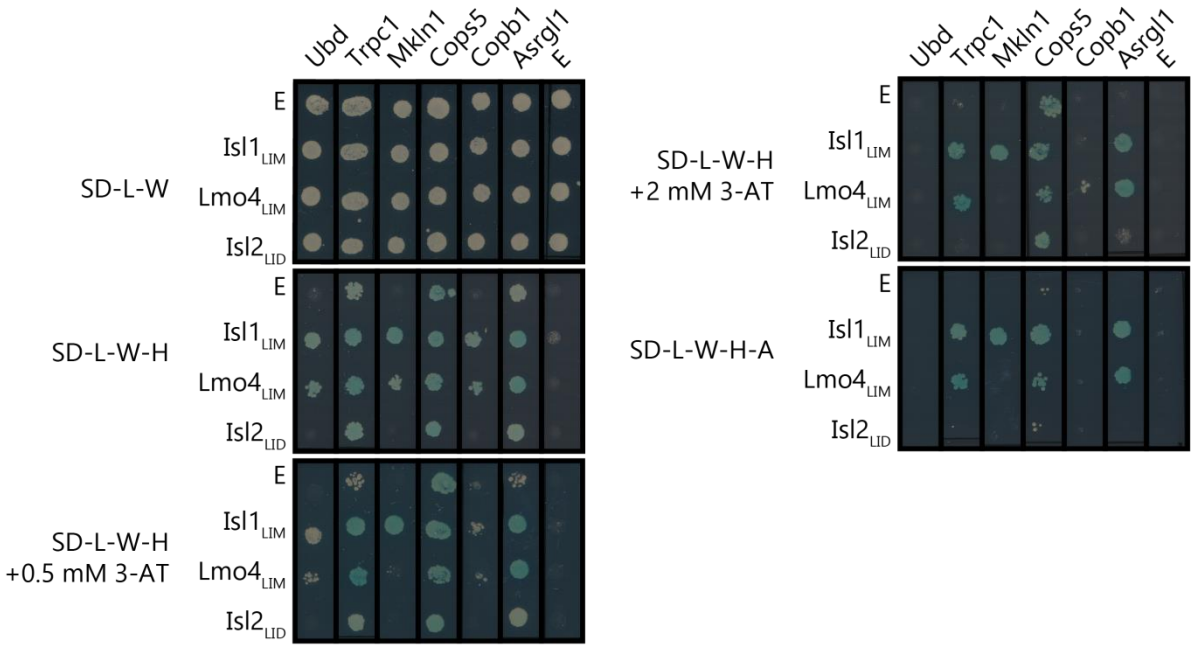


Figure 3.9: Yeast two-hybrid spot test validations for specificity of $Ldb1_{LID}/Isl1_{LIM}$ interacting proteins. Yeast were co-transformed with one pGBT9 plasmid and one pGAD plasmid and grown on a range of selective media to screen for interactions. pGBT9 plasmids used were: empty pGBT9 (E), $Isl1_{LIM}$, $Lmo4_{LIM}$, and $Isl2_{LID}$. pGAD plasmids used were: empty pGAD10 (E), and pGADT7-RecAB plasmids encoding the prey proteins Ubd, Trpc1, Mkln1, Cops5, Copb1, and Asrgl1.

Whereas Mkln1 did appear in $Isl1_{LIM}$ screening, Copb1 and Ubd did not, although they showed an interaction with $Isl1_{LIM}$ in the above specificity validations. This absence may be explained by the fact that the interactions of these two proteins with $Isl1_{LIM}$ are relatively weak, and so would not have been isolated in the $Isl1_{LIM}$ screening process, which focussed on the strongest interactions detected. As Copb1 interacted equally strongly with $Ldb1_{LID}/Isl1_{LIM,}$ $Isl1_{LIM}$, and $Lmo4_{LIM}$, it was not pursued further. This leaves Mkln1 and Ubd as the remaining likely $Ldb1_{LID}/Isl1_{LIM}$-interacting proteins, although it appears likely that they are binding only to $Isl1_{LIM}$.

3.7.3.3 Determining $Isl1_{\Delta LIM}$ prey interaction specificity

$Isl1_{\Delta LIM}$-interacting hits were screened against full length Isl2 and $Isl2_{LID}$ (Figure 3.10). As the C-terminus of Isl1(other than the LID) has an unknown domain structure, further screening against homologous domains could not be conducted. None of the 17 proteins tested showed an interaction with $Isl2_{LID}$. Most of the $Isl1_{\Delta LIM}$-interacting proteins showed

similarly strong interactions with full length Isl1 and Isl2, indicating no preference for interacting with a particular Islet protein.

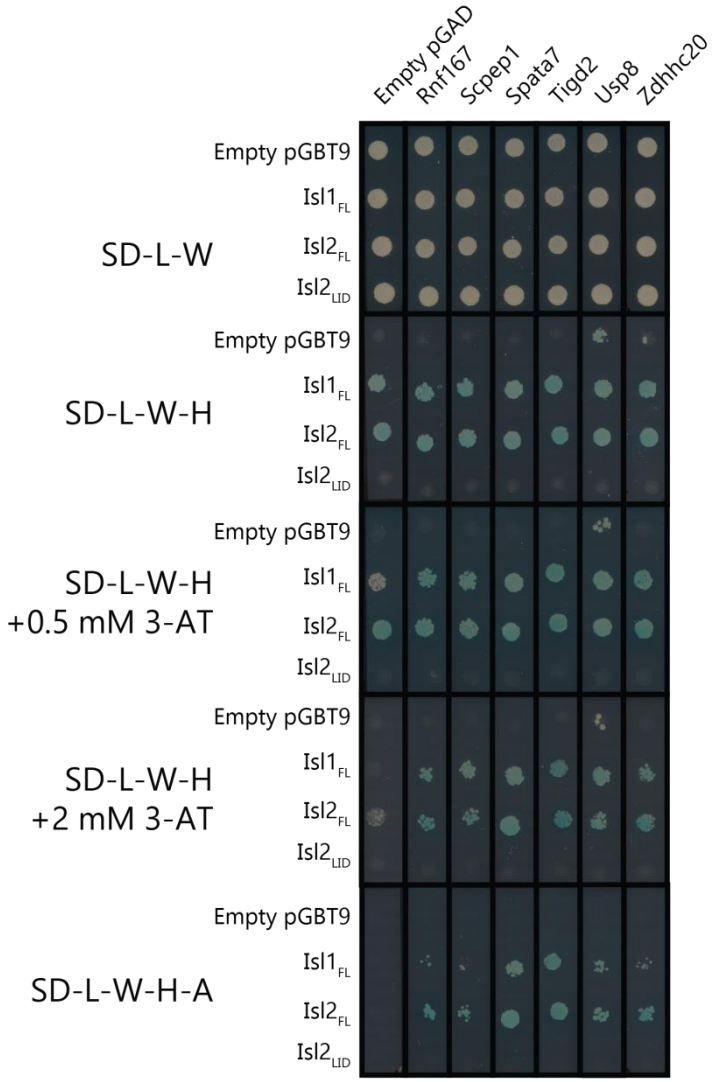


Figure 3.10: Representative yeast two-hybrid spot test validations for specificity of $Ldb1_{LID}$/$Isl1_{LIM}$ interacting proteins. Yeast were co-transformed with one pGBT9 plasmid and one pGAD plasmid, and grown on a range of selective media to screen for interactions. pGBT9 plasmids used were: empty pGBT9, $Isl1_{FL}$, $Isl2_{FL}$, and $Isl2_{LID}$. pGAD plasmids used were: empty pGAD10, and pGADT7-RecAB plasmids encoding the prey proteins Rnf167, Scpep1, Spata7, Tigd2, Usp8, and Zdhhc20.

3.7.3.4 Fhl1

Screening with $Isl1_{LIM}$ resulted in a total of 51 hits encoding 3 different truncations of the protein Four and a half LIM domains protein 1 (Fhl1), making it the most represented protein. In fact, Fhl1 represented 14% of the total pool of $Isl1_{LIM}$ hits, and 23% of the total protein-encoding pool of $Isl1_{LIM}$ hits. Being represented in such high numbers in the sample

pool may indicate a particularly strong interaction, or simply that Fhl1 is an abundant transcript in the library.

Unlike other protein hits, the different Fhl1 fragments identified varied by less than 5 residues (or 15 bp). The shortest construct of Fhl1 found was 19 amino acids in length, and the longest was 23, both containing the very C-terminal portion of the Fhl1 protein. Specificity validation experiments showed that Fhl1 interacted very strongly with both $Isl2_{LIM}$ and $Limk2_{LIM}$, suggesting that it is a non-specific LIM-binding sequence (further discussed in Section 3.8.2).

3.8 Discussion

The data presented in this chapter identified several proteins that represent potentially biologically relevant interaction partners for Isl1. Chapter 4 will further address which of these interactions are likely to be biologically relevant. This discussion will focus on the methodology used to generate this pool of potential Isl1-interactors. Listed in Table 3.8 are the prey constructs that were seen to interact with Isl1 most strongly out of the proteins that they were tested against for specificity (Section 3.7.3).

Table 3.8: Potential Isl1-interacting proteins from yeast two-hybrid screening.

Screen	$Isl1_{\Delta LIM}$	$Isl1_{LIM}$	$Ldb1_{LID}/Isl1_{LIM}$
Protein	Art3	Ddx20	Mkln1
	BC035947	Dfna5	Ubd
	Cops5	Lace1	
	Cyc1	Mep1b	
	Kctd9	Mkln1	
	Lrrc51	Nup50	
	Ly6c1	Rps18	
	Ms4a5	Sparcl1	
	Nectin3	Zfand1	
	Nkiras1		
	Oscp1		
	Rnf167		
	Scpep1		
	Spata7		
	Tigd2		
	Usp8		
	Zdhhc20		

3.8.1 Analysing the methodology used

Overall, the methodology employed here was successful in identifying potential new binding partners for different regions of Isl1. Yeast two-hybrid assays are an established methodology used for screening protein-protein interactions, and so rigorous screening protocols and controls are known.

False positives are well-known to occur in yeast two-hybrid screens. However, they can be minimised in library screening through several strategies. One strategy is minimisation of the auto-activation of the bait by, for example, choosing an appropriate stringency of selective media, and selecting the orientation of bait and prey with respect to the activation and DNA binding domains of GAL4. Generally, the construct fused to the DNA binding domain will show higher auto-activation than the construct fused to the activation domain [231, 232]. This phenomenon is especially true for transcription factors [233], and for this reason Isl1 bait

constructs used were fused to the DNA binding domain, so that auto-activation could be assessed and then minimised in the library screening. This approach is reflected in the primary validations performed herein, where only 12 hits out of 147 were found to be auto-activation false positives (Section 3.6).

Many modern large-scale yeast two-hybrid interaction screening studies do not conduct validation experiments by the same methodology because hits are identified not by isolation of plasmids and sequencing as was carried out here, but by using PCR with primers that flank the prey insert region [234, 235]. Amplicons can thus be sequenced directly to identify the prey constructs. However, although this identification strategy is more efficient than plasmid extraction and sequencing, it does not provide material that can be used in yeast two-hybrid validation experiments. Note that validation with an orthogonal method is a strong indication of a genuine interaction (see Section 4.3.2 for an example). However, in the context of large-scale library screening, it is difficult to find an orthogonal technique that can be conducted efficiently, especially when dealing with large pools of putative interactors [235].

3.8.2 Intrinsic flaws in the initial library used

More than half of the total sample pool were found to not contain protein sequence in frame with the GAL4 domain upstream (364/674 samples). These samples appear to be the result of methodological flaws in the production of the yeast library used for screening. As described earlier, the cDNA library used for screening was produced from mRNA extracted from mouse tissue samples (see Section 3.3.2 and Figure 3.3). This methodology introduced two types of product into the library that led to false positives: non-protein coding RNA sequences, and protein sequences that were inserted into the bait vector in the wrong frame to that of the upstream GAL4 protein sequence (see Section 3.5). Each of these product types can result in nonsense proteins in the context of GAL4-fusion constructs.

Non-coding RNA sequences isolated through the screening process primarily encoded 3'-untranslated regions of genes. These cDNAs may have arisen from partially degraded mRNAs, where the 5' end of the mRNAs had degraded to a point such that all protein-coding sequence was removed, leaving only the 3' untranslated regions, which could still successfully be incorporated into the library as the 3' polyadenylation signals used to generate cDNA were still intact. They may also have arisen during the production of cDNA,

through off-target binding of the primers. Although the presence of 3' UTR-encoding prey is not ideal, the same artefact introduces truncations of protein-encoding mRNAs into the library, which can help to determine the minimal binding domain of an interacting protein. This will be further explored in Chapter 4 (Section 4.3).

A particular potential hit that was likely to have been affected by truncations of protein-encoding mRNAs was Fhl1 (Section 3.7.3.4). The consistently short length of Fhl1 encoded by hits could indicate a very precisely defined minimal binding domain, but it may also indicate that a larger portion of Fhl1 disrupts the observed interaction. Investigation of the domain structure of Fhl1 revealed that the last LIM domain ends in the middle of the prey Fhl1 peptide (Figure 3.11). It seems likely that the Fhl1 peptides found in library screens represent a conformation that would not normally exist *in vivo*, as it would natively form part of a LIM fold. This interaction with $Isl1_{LIM}$ was consequently considered as not biologically relevant.

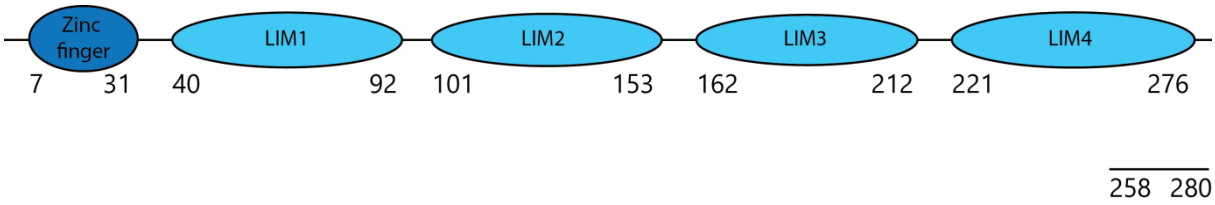


Figure 3.11: Domain structure of Fhl1. Longest portion found during screening is aligned below.

There is no method of selecting the reading frame of the cDNA as it is inserted into the pGADT7-RecAB vector, which means there is only a 1-in-3 chance that the coding sequence being inserted will be in the correct reading frame. Theoretically, yeast are tolerant to ribosomal frameshifts, meaning that out of frame sequences could still be able to produce the correct protein product [226, 236]. However, work here shows no evidence for ribosomal frameshifts (Section 3.5.1). Consequently, at least 2/3 of the sequences in the prey library would not represent an actual protein sequence, and with the possibility of nonsense protein sequences giving rise to false positives (e.g., through expression of short peptides), this statistic represents a significant inefficiency of cDNA library design. This problem was exemplified in this work by the ~30% of hits encoding out of frame sequences (Section 3.5), highlighting the need for appropriate validation procedures. One possible benefit of this

inefficiency is that although these hits clearly do not represent biologically relevant binding partner, they could prove useful in the development of peptide inhibitors to the bait.

3.8.3 Considering the pool of prey

Most hits found during screening were isolated from screening with the $Isl1_{LIM}$ construct, which is consistent with the presence of a well characterised protein-protein interaction domain in that construct [45]. This set of interactions has an established pipeline for characterisation [53, 55, 57, 58]. However, the interactions with $Isl1_{\Delta LIM}$ and $Ldb1_{LID}$/$Isl1_{LIM}$ may represent interesting interactions with potentially novel mechanisms of interaction. As the C-terminus of Isl1 remains undescribed in terms of function, it would be of great interest to investigate interactions occurring in this region.

Many of the prey constructs that interacted strongly with $Isl1_{LIM}$ were discovered to interact strongly with $Limk2_{LIM}$ and/or $Lmo4_{LIM}$ (Section 3.7.3). It is possible that those hits represent several novel LIDs that target a broad array of LIM domains, in a similar way to the primary LIM cofactor Ldb1, which interacts with LIM domains from all LIM-HD and Lmo proteins [54]. Given the low levels of conservation in LID sequences, it is entirely possible that these proteins, though having very low levels of sequence similarity, bind to LIM domains via the same mechanism. However, these interactions could be artefacts of the screening process. In particular, the hydrophobic nature of unbound LIM domains may have enriched the sample pool for non-specific interaction partners. As the yeast two-hybrid system takes many proteins out of their native environment, many prey constructs can contain exposed hydrophobic regions (especially if the construct comprises part of a normally folded domain) that in these conditions could interact strongly with hydrophobic $Isl1_{LIM}$ [54]. This kind of false positive is unlikely to be detected using the primary validation with empty pGBT9, as described above, as the $GAL4_{DBD}$ is a well-folded domain with few exposed hydrophobic regions (Section 3.6) [233, 237].

The $Isl1_{\Delta LIM}$ screening pool had a particularly large proportion of hits that did not encode proteins. The reason for this remains unclear. The two characterised features of the $Isl1_{\Delta LIM}$ construct are the homeodomain and the LID. While the homeodomain is positively charged, it is more likely to interact with DNA than protein sequence, and the LID is intrinsically disordered, and only forms a stable structure when bound to specific LIM binding partners.

Neither of these structures should be prone to non-specific protein-protein interactions, so the presence of so many nonsense peptide hits is anomalous.

The remaining pool of prey were further considered to assess whether they represent a likely biologically relevant interaction with Isl1. Those investigations form the basis of Chapter 4.

4 Assessing potential Isl1 interactors

4.1 Introduction

After eliminating false positives, and screening for interaction specificity, 27 candidates for specific interaction with Isl1 remained. These are listed in Table 4.1. In this Chapter their likelihood of being biologically relevant interactions was assessed by considering what was known about those proteins in the literature, and the physical properties of the constructs that were identified as hits. Following this assessment a strong candidate was further assessed experimentally through attempts to define a minimal binding domain and validate the interaction by an orthogonal method.

Table 4.1: Isl1-interacting proteins remaining after validations. If a protein was isolated in screens with different Isl1 constructs, the total number of clones isolated is given in brackets.

Protein	Isl1-interacting construct	Number of clones isolated	Length of protein isolated
Ecto-ADP-ribosyltransferase 3 (Art3)	$Isl1_{\Delta LIM}$	1	Full protein
BC035947	$Isl1_{\Delta LIM}$	1	Unknown
COP9 signalosome complex subunit 5 (Cops5)	$Isl1_{\Delta LIM}$	3 (11)	38-334
Cytochrome C1 (Cyc1)	$Isl1_{\Delta LIM}$	1 (3)	296-335
BTB/POZ domain-containing protein KCTD9 (Kctd9)	$Isl1_{\Delta LIM}$	1	309-339
Leucine-rich repeat-containing protein 51 (Lrrc51)	$Isl1_{\Delta LIM}$	1	Full protein
Lymphocyte antigen 6C1 (Ly6c1)	$Isl1_{\Delta LIM}$	1	74-131
Membrane-spanning 4-domains, subfamily A, member 5 (Ms4a5)	$Isl1_{\Delta LIM}$	1	Full protein
Nectin-3 (Nectin3)	$Isl1_{\Delta LIM}$	1	54-549
NF-kappa-B inhibitor-interacting Ras-like protein 1 (Nkiras1)	$Isl1_{\Delta LIM}$	1	Full protein

Oxidored-nitro domain-containing protein 1 (Oscp1)	$Isl1_{\Delta LIM}$	1	Full protein
E3 ubiquitin-protein ligase RNF167 (Rnf167)	$Isl1_{\Delta LIM}$	1	Full protein
Retinoid-inducible serine carboxypeptidase (Scpep1)	$Isl1_{\Delta LIM}$	1	Full protein
Spermatogenesis-associated protein 7 homolog (Spata7)	$Isl1_{\Delta LIM}$	2	200-582
Tigger transposable element-derived protein 2 (Tigd2)	$Isl1_{\Delta LIM}$	1	25-525
Ubiquitin carboxyl-terminal hydrolase 8 (Usp8)	$Isl1_{\Delta LIM}$	1	941-1080
Palmitoyltransferase ZDHHC20 (Zdhhc20)	$Isl1_{\Delta LIM}$	1	64-380
Probable ATP-dependent RNA helicase DDX20 (Ddx20)	$Isl1_{LIM}$	1	458-825
Gasdermin-E (Dfna5)	$Isl1_{LIM}$	2	238-512
AFG1-like ATPase (Lace1)	$Isl1_{LIM}$	1	428-480
Meprin A subunit beta (Mep1b)	$Isl1_{LIM}$	1	398-704
Nuclear pore complex protein Nup50 (Nup50)	$Isl1_{LIM}$	1	39-464
40S ribosomal protein S18 (Rps18)	$Isl1_{LIM}$	5	63-152
SPARC-like protein 1 (Sparcl1)	$Isl1_{LIM}$	14	335-650
AN1-type zinc finger protein 1 (Zfand1)	$Isl1_{LIM}$	4	189-268
Muskelin (Mkln1)	$Isl1_{LIM}$, $Ldb1_{LID}/Isl1_{LIM}$	1 4	113-735
Ubiquitin D (Ubd)	$Ldb1_{LID}/Isl1_{LIM}$	1	Full protein

4.2 Assessing the literature

The candidates under consideration are broken into groups according to which screen or screens they were identified as hits: $Isl1_{\Delta LIM}$ only, $Isl1_{LIM}$ only, $Ldb1_{LID}/Isl1_{LIM}$ only, and both $Isl1_{LIM}$ and $Ldb1_{LID}/Isl1_{LIM}$ (Mkln1).

4.2.1 $Isl1_{\Delta LIM}$ interactors

Of the 17 proteins found as potential binding partners to $Isl1_{\Delta LIM}$, nine immediately seem unlikely, due to their notated functions and localisations in the cell. Art3 is extracellular, Ly6c1, Ms4a5, Nectin3, Oscp1, Rnf167, and Zdhhc20 all function in the plasma membrane, Scpep1, Rnf167, and Zdhhc20 are localised to the lysosome, and Spata7 localises to the microtubule network in cilia [238-246]. It seems unlikely that Isl1 could interact with these proteins as part of its role in regulating gene expression. Interactions with proteins localised to the lysosome may still be biologically relevant, as they could be involved in the degradation of Isl1. However, the lysosome is primarily involved in the degradation of larger assemblies, and it is more likely that Isl1 degradation would occur through the ubiquitin degradation pathway [247, 248]. Indeed, the ubiquitin ligase Rlim/Rnf12 has been identified as catalysing the ubiquitination of LIM-HD proteins that bind to the common LIM-HD/Lmo-binding protein Ldb1, although not specifically Isl1 [249, 250].

Several proteins found in the pool of putative $Isl1_{\Delta LIM}$ interactors, such as Lrrc51, BC035947, and Tigd2 have not been characterised at all, making it difficult to comment on whether these proteins could be of further interest or not. The remaining five proteins (Kctd9, Usp8, Cops5, Cyc1, and Nkiras1) are discussed below individually.

4.2.1.1 Kctd9

Although Kctd9 was predicted to be an extracellular protein, is has been characterised as an E3 ubiquitin ligase [251]. It forms homopentamers through its BTB/POZ domain, and can form a larger complex alongside cullin proteins, potentially acting as the substrate recognition subunit of the E3 complex [251]. An interaction between Kctd9 and Isl1 would be of interest, as it may reveal a specific proteasomal degradation pathway for Isl1. However, it should be noted that the construct isolated during screening consists of only the C-terminal 30 residues, and 17 of these residues are predicted to contribute to one of the ordered pentapeptide repeats (Figure 4.1). Consequently, the interaction with Isl1 should first be

tested with a larger portion of the protein. Whether this interaction can occur with Kctd9 in a pentameric state should also be investigated, to establish if this interaction could occur *in vivo*.

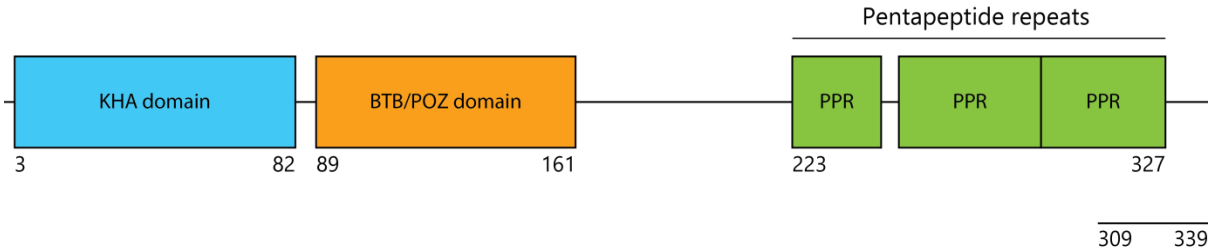


Figure 4.1: Domain structure of Kctd9. Fragment of Kctd9 isolated from yeast two-hybrid screening is shown beneath.

4.2.1.2 Usp8

Usp8 is also involved in the ubiquitination system, but has been shown to cleave conjugated ubiquitin from proteins, rather than tagging proteins for proteasomal degradation [252]. The clone of Usp8 isolated in the screening process encodes a truncated product of Usp8, containing only the last 140 amino acids (residues 941-1080) (Figure 4.2). The characterised structure of Usp8 shows that this truncation would likely eliminate two beta sheets and a zinc coordination site (Figure 4.2B), as well as disrupting the hydrophobic core of the fold in that region (Figure 4.2C) [253]. The truncation would likely leave the remaining protein only partially folded, with exposed hydrophobic surfaces, so is likely to represent an interaction artefact rather than a native interaction between Usp8 and Isl1.

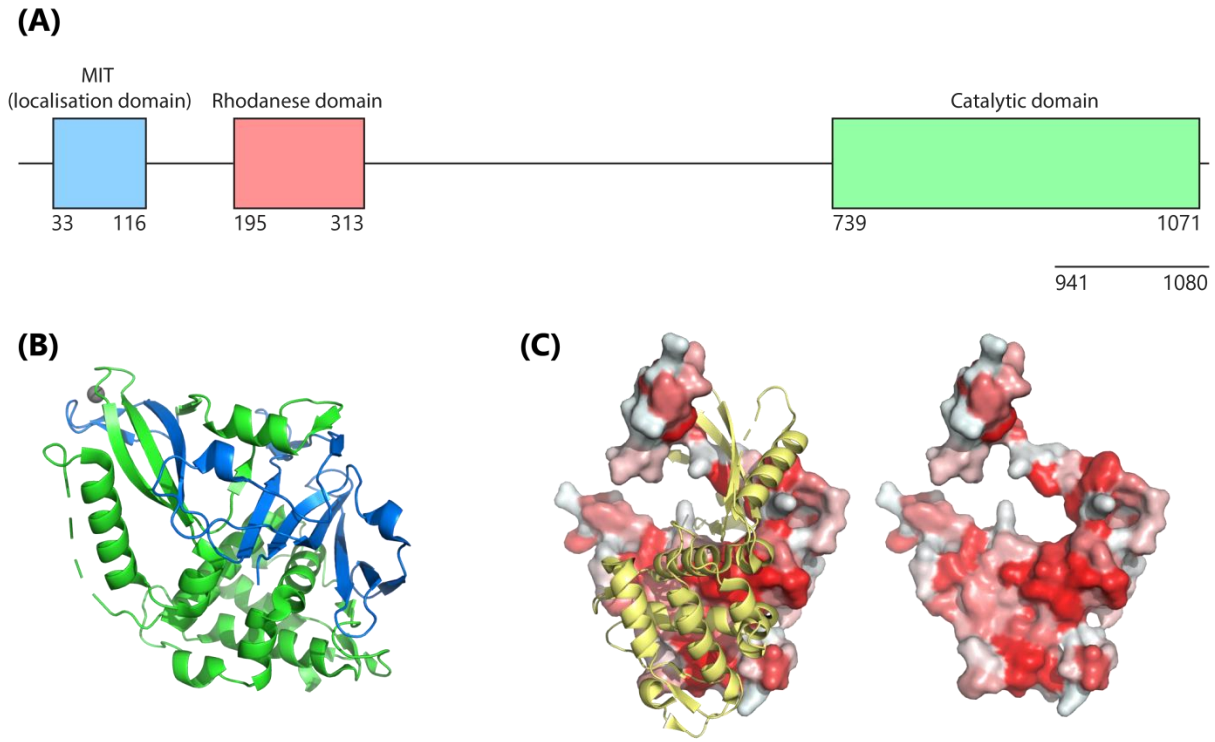


Figure 4.2: Structure of Usp8. (A) Domain structure of Usp8, with the fragment of Usp8 isolated from yeast two-hybrid screening shown beneath. (B) and (C) show the solved structure of the C-terminal catalytic domain (PDB:2GFO). (B) Residues found in the $Isl1_{\Delta LIM}$-interacting construct are shown in blue, remaining residues are shown in green. (C) Surface representation of $Isl1_{\Delta LIM}$-interacting Usp8, where red represents hydrophobic surfaces, with (left) and without (right) non-$Isl1_{\Delta LIM}$-interacting residues (yellow).

4.2.1.3 Cops5 and Cyc1

Cops5 and Cyc1 were both isolated in multiple screens, but only $Isl1_{\Delta LIM}$ binding appears to be real based on the following data. Cops5 was isolated in screens with all three Isl1 constructs used, but it was discarded in the $Isl1_{LIM}$ and $Ldb1_{LID}$/$Isl1_{LIM}$ pools because of high auto-activation levels (Figure 4.3A). Repeated testing of auto-activation levels of Cops5 gave variable results. Although some experiments showed low levels of auto-activation, most showed high levels of auto-activation. This suggests that the binding seen with Cops5 may represent an artefact, and not a genuine interaction with $Isl1_{\Delta LIM}$.

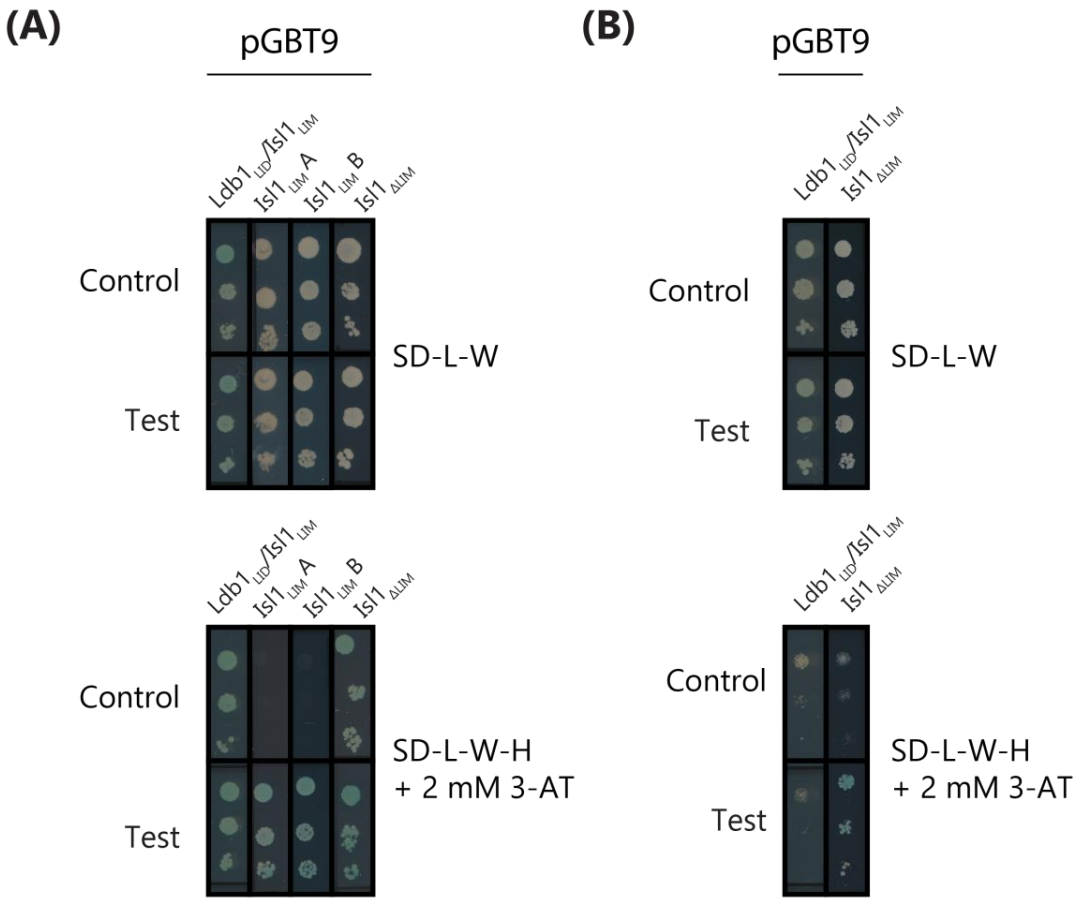


Figure 4.3: Cops5 and Cyc1 yeast two-hybrid spot tests. . Each set of three spots contains a serial 1:10 dilution of yeast, from OD_{600} 0.2 to 0.002. (A) Validation spot test assays of Cops5 against $Ldb1_{LID}/Isl1_{LIM}$, $Isl1_{LIM}$ and $Isl1_{\Delta LIM}$, with two iterations of $Isl1_{LIM}$ testing. (B) Validation spot test assays of Cyc1 against $Ldb1_{LID}/Isl1_{LIM}$ and $Isl1_{\Delta LIM}$.

Cyc1 was isolated in the $Ldb1_{LID}/Isl1_{LIM}$ screen, but upon further testing did not show an interaction with $Ldb1_{LID}/Isl1_{LIM}$. Yeast two-hybrid spot test assays show a moderately strong interaction with $Isl1_{\Delta LIM}$ (Figure 4.3B). However, given the well-established role of Cyc1 in the mitochondrial electron transport chain, it also seems unlikely as an biologically relevant interaction partner for Isl1 [254].

4.2.1.4 Nkiras1

Nkiras1 plays a role in regulating NF-κB signalling, by preventing the degradation of NF-kappa-B inhibitor beta [255, 256]. This interaction could be of interest, as NF-κB signalling is a major signalling pathway involved in development. Of note, the clone isolated during screening encodes the full length Nkiras1 protein, making the observed interaction more likely to indicate a true binding event.

4.2.2 Isl1$_{LIM}$ interactors

The Isl1$_{LIM}$ pool of interactors represents the interactions most likely to be biologically relevant, as the LIM domains are known protein interaction domains, and the proteins remaining in this pool all showed evidence of strong and specific interactions with Isl1.

4.2.2.1 Sparcl1/Ecm2/Sc1

Secreted protein, acidic and rich in cysteine-like 1 (Sparcl1) is a glycoprotein named for its similarity to an earlier discovered protein Sparc (Secreted protein, acidic and rich in cysteine) [257, 258]. Due to its similarity to Sparc, Sparcl1 has been speculated to play a role in regulating the extracellular matrix, potentially through acting as a collagen chaperone [259]. Sparcl1 contributes to proliferation and survival of cells, although the mechanism by which this occurs is unknown [260]. In line with this role, Sparcl1 expression was correlated with several kinds of cancer, including prostate and colorectal cancers [261, 262].

That Sparcl1 was recovered 14 times from the Isl1$_{LIM}$ screen highlights it as a protein of interest (Table 4.1). However, although Sparcl1 shows a strong interaction with Isl1$_{LIM}$ in yeast two-hybrid assays, it seems unlikely that this interaction would occur *in vivo*. Sparcl1 is exported from the cell in order to regulate the extracellular matrix. This makes it unlikely as a biologically relevant interaction partner candidate for Isl1, as they likely would never co-localise. Like many extracellular proteins, Sparcl1 also contains many disulfide bonds and is heavily post-translationally modified [263]. These modifications should not be present in the intracellular environment of the yeast two-hybrid assay. These reasons make it probable that the interaction observed between Isl1 and Sparcl1 is an artefact of the experimental system used.

4.2.2.2 Ddx20/Dp103/Gemin3

DEAD-box helicase DDX20 (Ddx20) is a protein with several roles in the cell [264]. It was first discovered as part of a complex involved in spinal muscular atrophy [265]. This complex interacts with a broad range of RNAs and is involved in the assembly of small nuclear ribonucleoproteins (snRNPs) [265-267]. The DEAD-box motif in Ddx20 provides the RNA binding function required for this complex [268]. Beyond this, Ddx20 has also been shown to play a role as a transcriptional regulator, through binding to Egr (Early growth response) protein family members, to be important in ovarian development, and to play a role in cell

signalling pathways involving NF-κB and p53 [269-272]. Deletion of Ddx20 is lethal, with fertilised eggs failing to progress past the two-cell stage (blastocoel), indicating that it plays a major role in early embryonic development [272].

Both Isl1 and Ddx20 localise to the nucleus and are involved in transcriptional regulation. Deletion of either of these proteins results in lethality at an early stage of embryonic development. These common attributes suggest that an interaction between Isl1 and Ddx20 may be biologically relevant, and is worth further investigation.

4.2.2.3 Zfand1

AN1-type zinc finger protein 1 (Zfand1) is a largely uncharacterised protein. It is named for the presence of two AN1-type zinc fingers near its N-terminus. The AN1-type zinc finger is found in proteins associated with stress responses, across animals and plants [273]. The structure of the AN1-type zinc finger from the yeast protein Cuz1 (Cdc48-associated UBL/zinc finger protein-1) was recently solved [274]. Cuz1 is a homolog of Zfand1, and is involved in targeted protein degradation through the ubiquitin conjugation pathway [275]. The Cuz1-proteasome association was found especially in cases of exposure to metalloids, indicating a role in stress response.

Recently, Zfand1 was shown to also regulate proteasomal protein degradation in response to stress, specifically from the metalloid arsenite [276]. This group of chemicals, which contain oxidised arsenic, have been shown to be carcinogenic, as a result of inhibiting DNA repair [277].

As Zfand1 is predicted to localise to the nucleus, the interaction between Isl1 and Zfand1 could be of further interest. However, the interaction between Isl1 and a larger portion of Zfand1 must first be tested, as the clones isolated from screening all encode truncation products of Zfand1 containing only the C-terminal 78 amino acids (Figure 4.4). This truncation begins in the middle of the ordered C-terminal region, which is predicted to adopt a ubiquitin-like fold [276]. It is difficult to assess whether this truncation product reflects how the protein would behave and fold *in vivo* without further structural data.

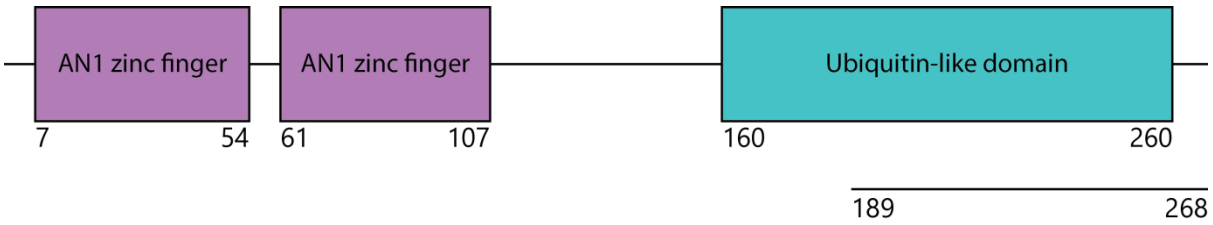


Figure 4.4: Domain structure of Zfand1. Fragment of Zfand1 isolated from yeast two-hybrid screening is shown beneath.

4.2.2.4 Dfna5/Gsdme/Dfna5h

Non-syndromic hearing impairment protein 5 (Dfna5), also known as Gasdermin-E (Gsdme), was initially identified in a mutant form as a gene responsible for causing hereditary deafness [278, 279]. Dfna5 has since been found to play a role in the DNA damage response, and its expression is downregulated through promoter methylation in several types of cancer [280-282]. These observations suggest that Dfna5 plays a role as a tumour suppressor. In support of this idea, further investigation has linked Dfna5 to pro-apoptotic pathways [283, 284]. Dfna5 is cleaved by caspase 3 in the apoptotic cascade, into N- and C-terminal fragments (Figure 4.5). The N-terminal fragment then translocates to the plasma membrane through the action of a membrane targeting sequence, where it may potentially oligomerise to form a pore that promotes cell lysis [284].

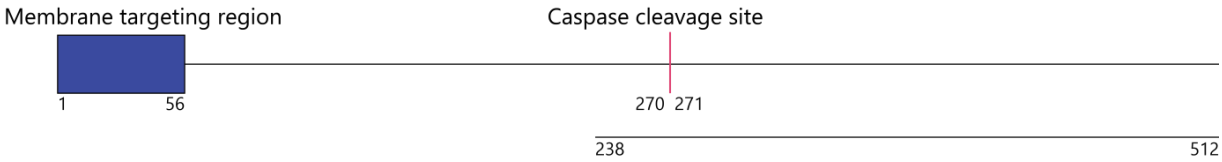


Figure 4.5: Domain structure of Dfna5. Fragment of Dfna5isolated from yeast two-hybrid screening is shown beneath.

Whether or not an interaction between Isl1 and Dfna5 is biologically relevant is difficult to predict, considering that Dfna5 appears to be a multi-functional protein. A Dfna5:Isl1 interaction could form part of a cascade to regulate gene expression in response to either DNA damage signals or pro-apoptotic signals. Whereas the N-terminal fragment of Dfna5 has been investigated, the role of the C-terminal fragment has not been explored - this is especially relevant as the construct of Dfna5 isolated interacting with $Isl1_{LIM}$ lacks the first 237 residues (approximately half) of the protein.

4.2.2.5 Lace1/Afg1l

AFG1-like ATPase (Afg1l), also known as Lactation elevated protein 1 (Lace1), was originally discovered as a protein that was highly upregulated in expression in lactating mouse breast tissue [285]. Lace1 is a homolog of the yeast Afg1 protein, a mitochondrial protein involved in proteolysis of electron transport chain proteins [286]. Lace1 was recently shown to play a similar role, promoting degradation of the subunits in complex IV of the electron transport chain [287]. There is also evidence that Lace1 promotes apoptosis through association with the tumour suppressor protein p53, and sequestering it to the mitochondria [288].

Given that Lace1 functions primarily in the mitochondria, an interaction with Isl1 appears unlikely in cells, as there is no evidence for the presence or function of Isl1 in mitochondria. However, this protein is still poorly characterised; if it can sequester p53 to the mitochondria, it is unlikely that it is always localised to the mitochondria, and so may at some stage co-localise with Isl1.

4.2.2.6 Mep1b

Meprin A subunit beta (Mep1b) is a membrane protein with a large extracellular region. It normally functions as a proteolytic homo-oligomeric enzyme, breaking down target proteins [289]. This raises several impediments to a biologically relevant interaction with Isl1 as its localisation in membranes, along with a large extracellular component, reduces the likelihood of co-localisation with Isl1. That Mep1b normally exists as a membrane bound homo-oligomer means that the context of the yeast two-hybrid assay may affect the structure of Mep1b. In yeast two-hybrid experiments, Mep1b should not be localised to a membrane, meaning regions of the protein that would normally be buried may be exposed, permitting the formation of non-native interactions. Overall, it seems unlikely that Isl1 and Mep1b would interact *in vivo.*

4.2.2.7 Rps18 and the ribosomal interaction

Ribosomal proteins have previously been identified as commonly occurring false positives in yeast two-hybrid screens [231]. The five ribosomal proteins identified from the $Isl1_{LIM}$ screen as potential interactors were Rpl9 from the large ribosomal subunit, and Rps18, Rps23, Rps26, and Rps29 from the small ribosomal subunit. Whereas only Rps18 was classed as a specific interactor from yeast two-hybrid experiments (Section 3.7.3), Isl1 would be near all

of these proteins during translation. Figure 4.6 shows these proteins highlighted in a recent structure of the intact ribosome, derived from electron microscopy [290].

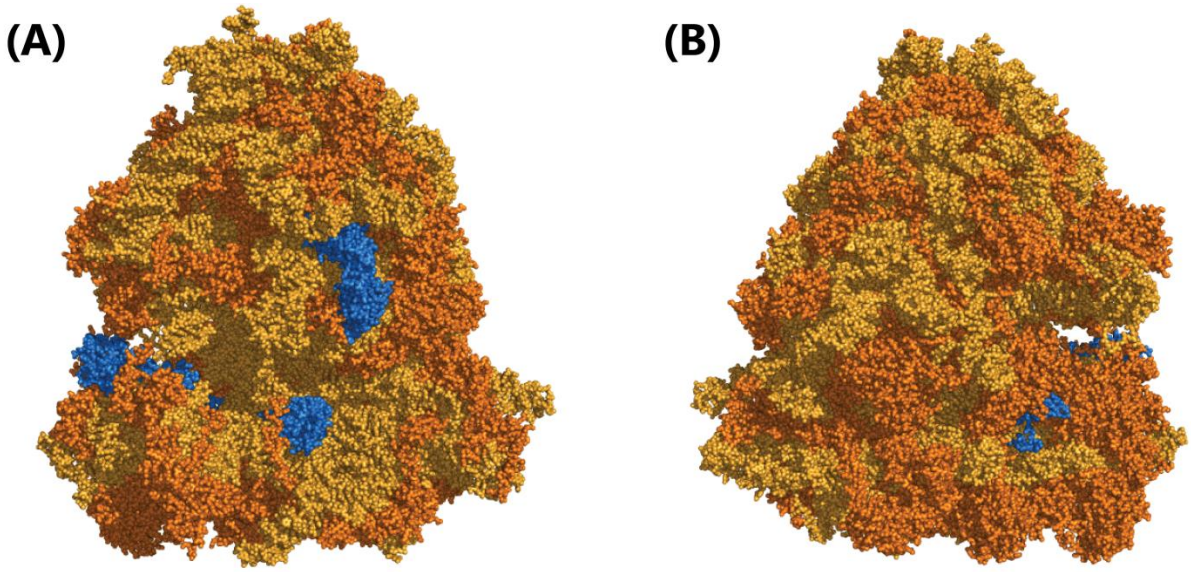


Figure 4.6: Electron microscopy model of an intact mouse ribosome. Shown are two orientations: (A) front and (B) rotated 180°. RNA is shown in light orange, proteins from the screen are shown in blue, and other ribosomal proteins are shown in dark orange (PDB: 5LKS).

Although the putative Isl1-interacting proteins are somewhat clustered, other ribosomal proteins in the vicinity have not been identified as Isl1 interactors. If these interactions occur *in vivo*, they are most likely to be transient interactions, which are only in effect during the translation of Isl1. These interactions could potentially aid in the folding of Isl1 as it is translated, especially considering that these ribosomal interactions are seen with the LIM domains of Isl1, which are located at the N-terminus of the protein. In this way, transient interactions with the ribosome could prevent off-target binding of the nascent Isl1 LIM domains, until the C-terminal LID is translated and can displace the ribosomal proteins. Although speculative, there is evidence that interactions of this type can form between nascent translating proteins and the ribosome, and that these interactions can be important in ensuring correct folding of the translating protein [291, 292]. However, none of the ribosomal proteins identified as putative interactors are close to the exit tunnel of the ribosome. It would be difficult for a translating protein to interact with any of these specific ribosomal proteins because they would not be sufficiently close. Additionally, the yeast two-hybrid environment removes all of the contextual ribosomal interactions that would be occurring in the cell. In

particular, Rps18 normally contacts the 16S rRNA, making it likely that the presence of this rRNA would block any interaction with $Isl1_{LIM}$, or influence the folding of Rps18 in a way that disrupts the Rps18:Isl1 interaction [290]. Because of this, it seems probable that the interaction between Isl1 and Rps18 would not occur *in vivo*.

4.2.2.8 Nup50 and the nuclear pore interaction

The protein nucleoporin 50 (Nup50) forms part of the nuclear pore complex and is directly involved in protein import into the nucleus [293, 294]. An interaction between Isl1 and Nup50 could be central to the import of Isl1 into the nucleus. The canonical nuclear import pathway relies on importin proteins binding target proteins via a nuclear localisation sequence (NLS), before travelling through the nuclear pore [295]. This process primarily involves transient interactions between the importin protein and the nuclear pore complex, with no interactions occurring between the nuclear pore and the cargo protein. Interestingly, Nup50 may be an exception to this general rule, as Nup50 was shown to displace NLSs from importin proteins [293, 296]. This type of activity is generally considered a means by which the nuclear pore can help to detach imported proteins from the nuclear import chaperone. A Nup50:Isl1 interaction would then most likely involve an Isl1 NLS.

To date no NLS has been formally identified in Isl1. There is evidence that helix 3 of the canonical homeodomain can act as an NLS (Section 5.10.4.2), but the interaction with Nup50 involved $Isl1_{LIM}$, not a construct containing the homeodomain [297]. Residues 7-14 in Isl1 (PPKKKRLI) resemble a common NLS motif [295]. However, the $Isl1_{LIM}$ construct used for yeast two-hybrid screening is missing the first four amino acids of this potential NLS, so it is likely that the interaction detected here between Isl1 and Nup50 involves additional regions of $Isl1_{LIM}$. Overall, an interaction between Nup50 and Isl1 could be of further interest, as it may reveal a mechanism for Isl1 transport into the nucleus.

4.2.3 $Ldb1_{LID}$/$Isl1_{LIM}$ interactors

Compared to the other two Isl1 constructs screened, there were very few $Ldb1_{LID}$/$Isl1_{LIM}$-interacting proteins isolated and identified. This suggests that the LIM domains of Isl1 are normally bound by one partner at a time.

The only protein found to interact solely with $Ldb1_{LID}/Isl1_{LIM}$ was Ubiquitin D/Fat10, a small protein that is structurally similar to diubiquitin [298]. Like canonical ubiquitin, Fat10 can be conjugated to target proteins, promoting their degradation by the proteasome [299, 300]. To date, one E1 ubiquitin ligase has been identified that can bind both canonical ubiquitin and Fat10 [301]. Fat10 has been functionally linked to the immune response, with overexpression of Fat10 promoting apoptosis of the cell [298, 300, 302, 303].

An interaction between Isl1, Ldb1, and Fat10 could be possible, as Fat10 has been shown to localise to the nucleus [300]. However, neither Isl1 nor Ldb1 has been demonstrated to act as a ubiquitin conjugation enzyme, or as part of the ubiquitin degradation pathway. In the absence of other data, it is likely that this interaction is not biologically relevant.

4.2.4 Mkln1

Muskelin (Mkln1) was identified in the screens for both $Isl1_{LIM}$ and $Ldb1_{LID}/Isl1_{LIM}$. Interestingly, yeast two-hybrid experiments testing for specificity showed that there was no interaction between Mkln1 and $Ldb1_{LID}$, and Mkln1 showed an equally strong interaction with $Isl1_{LIM}$ and $Ldb1_{LID}/Isl1_{LIM}$. This raises the possibility that Mkln1 in these experiments competes with $Ldb1_{LID}$ for binding to $Isl1_{LIM}$, or binds a different surface on the LIM domains.

Mkln1 was first identified as playing a role in maintaining the structure of the cytoskeleton, and in cell adhesion [304]. However, it has also been well characterised as part of a complex containing Ran-binding protein M (RanBPM), which is involved in transcriptional regulation [305, 306]. Further investigation has revealed that the localisation of Mkln1 depends on its protein binding partners [307, 308]. Mkln1 is predicted to contain an N-terminal discoidin domain, a central LisH (lissencephaly-1 homology) domain immediately followed by a CTLH (C-terminal to LisH) domain, and a C-terminal kelch repeat β-propeller domain (Figure 4.7) [304, 306, 309]. Both the discoidin domain and the LisH domain have been implicated as protein dimerisation domains. It has been shown that Mkln1 can exist in the cell as a monomer, dimer, or tetramer, with these different states associated with differing subcellular localisations [309-311].

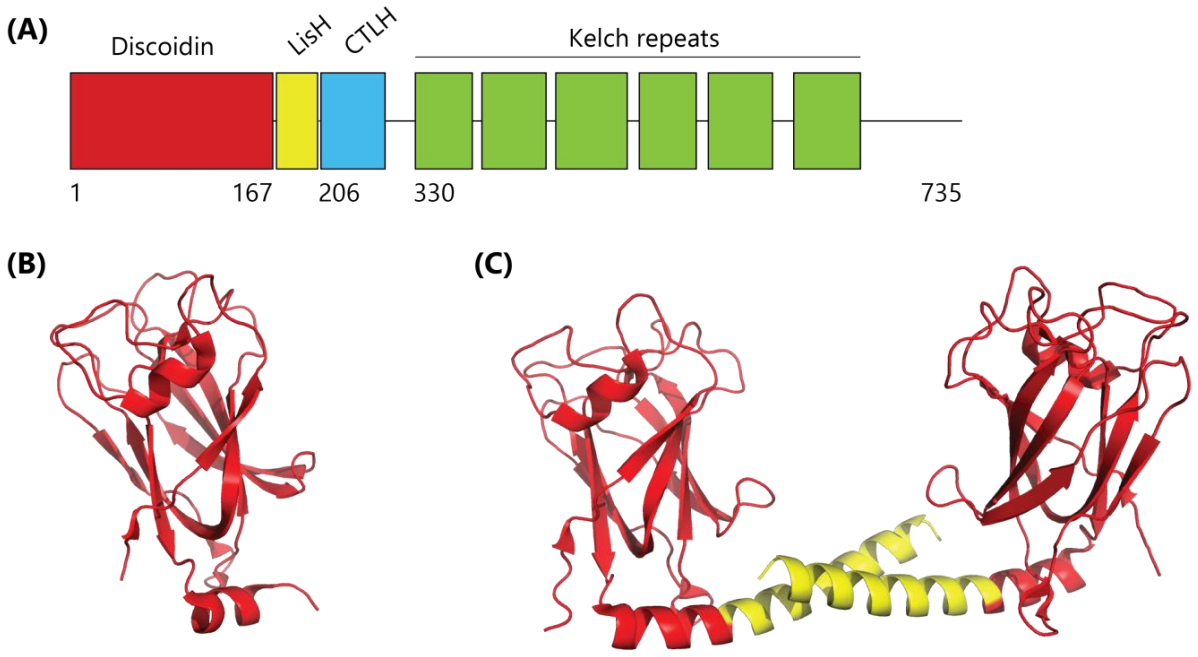


Figure 4.7: Domain structure of Mkln1. (A) Schematic of the domain organisation of Mkln1. (B) Structure of the Mkln1 discoidin domain. (PDB: 4PQQ) (C) Structure of the discoidin domain (red) and a portion of the LisH domain (yellow), showing LisH dimerisation (PDB: 4OYU).

The structures of the isolated discoidin domain (Figure 4.7B) and the discoidin domain with a portion of the LisH domain have been determined (Figure 4.7C) [311, 312]. There is limited structural information available about the CTLH domain, but it is consistently found in proteins downstream of a LisH domain, suggesting that it is required for correct folding or function.

The interaction between Isl1 and Mkln1 was selected for further validation, both because of its identification in both $Isl1_{LIM}$ and $Ldb1_{LID}/Isl1_{LIM}$ screens, and because the literature indicates that an interaction with Isl1 is plausible, in terms of both function and localisation.

4.3 Validating the interaction with Muskelin

Mkln1 was isolated 5 times over the course of yeast two-hybrid screening, with each clone encoding a different portion of the Mkln1 protein (Figure 4.8).

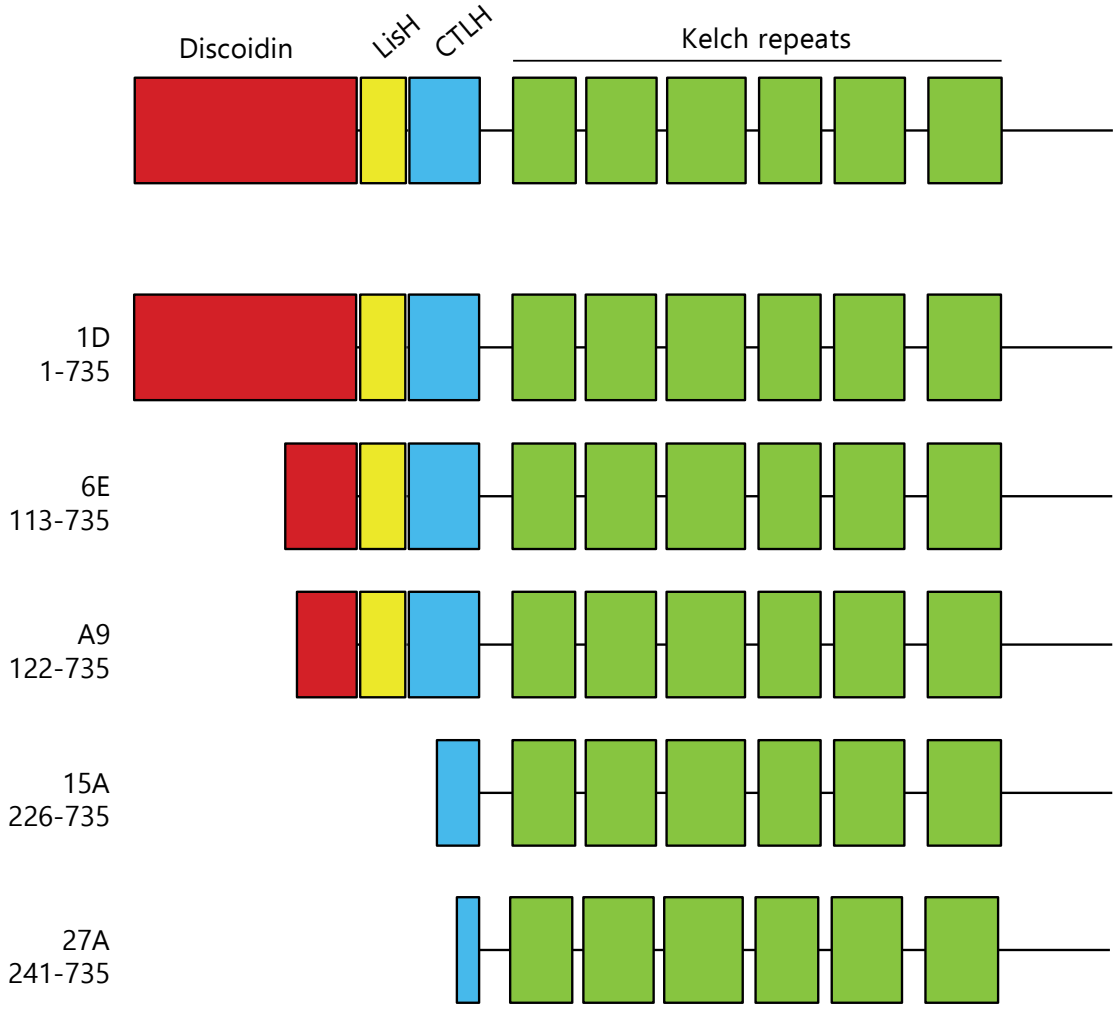


Figure 4.8: Schematic of Mkln1 truncations identified throughout yeast two-hybrid screening. Clones 1D, 6E, 15A, and 27A were identified in the $Ldb1_{LID}$/$Isl1_{LIM}$ screen; A9 was identified in $Isl1_{LIM}$ screening.

These constructs all contained the C-terminal CTLH, and kelch repeat domains, indicating that the region of interaction with Isl1 was likely to lie in this region. Constructs were designed to test this and to further narrow down the minimal Isl1-binding region of Mkln1 (Figure 4.9). Constructs were designed around domain boundaries, in the hopes of preserving correct folding. As the kelch repeats collectively form a β-propeller structure, they were treated as one domain. Full amino acid sequences of both Mkln1 and Isl1 constructs used can be found in Appendix F.

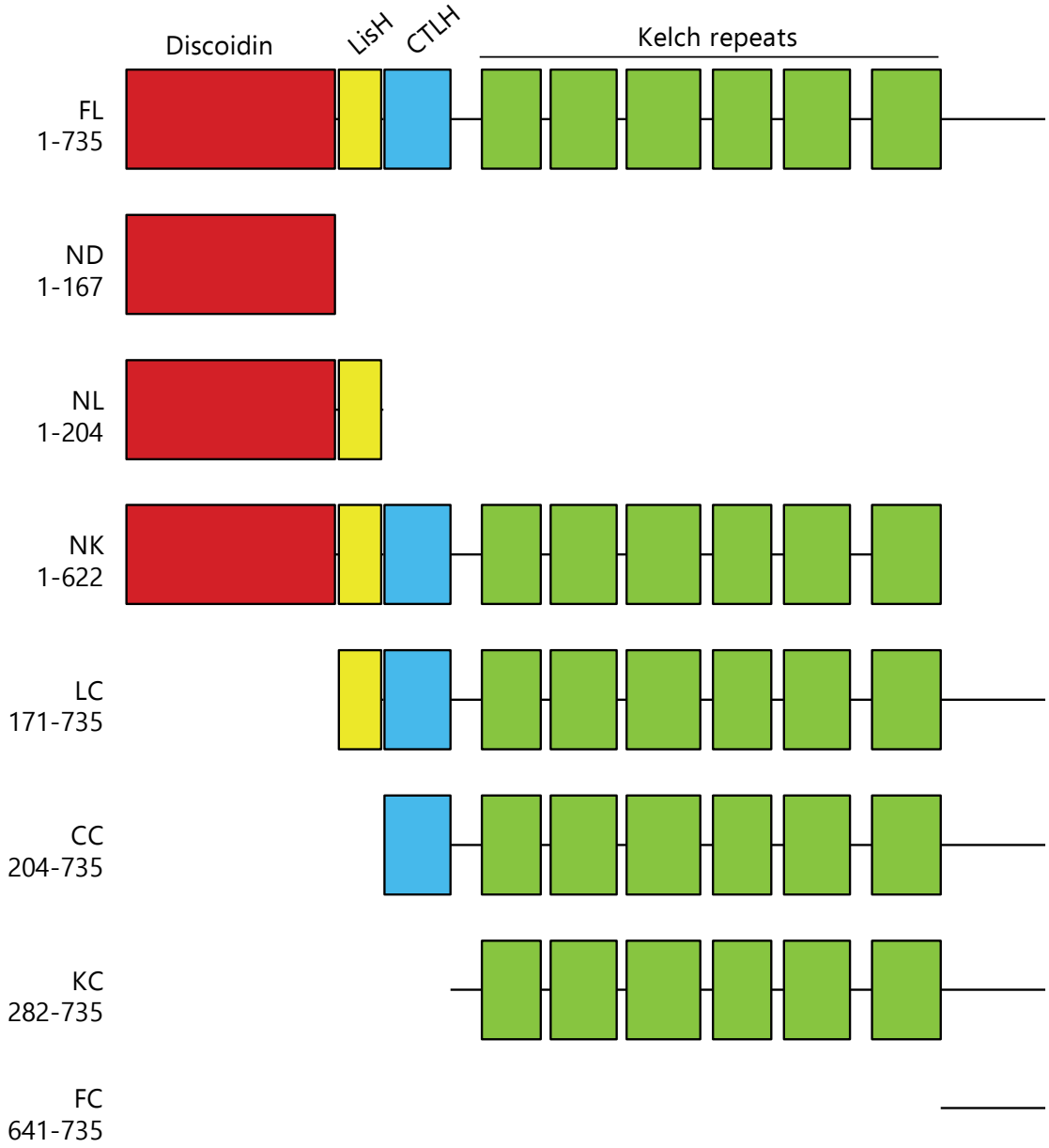


Figure 4.9: Truncation constructs of Mkln1 produced to validate interaction with Isl1.

These truncation constructs were screened for interactions with Isl1 in yeast two-hybrid experiments and in co-immunoprecipitation experiments using protein produced in mammalian cells, to probe the interaction in conditions that more closely represent the native environments for the two proteins *in vivo*.

4.3.1 Yeast two-hybrid shows an interaction between Isl1 and the CTLH region of Mkln1

Yeast two-hybrid analysis was used to test interactions between $Isl1_{LIM}$, $Ldb1_{LID}$/$Isl1_{LIM}$, and the various Mkln1 constructs produced. Full length Isl1 was not tested due to the high levels of auto-activation produced by the full length Isl1 construct. These experiments showed evidence of strong interactions between Isl1 constructs and the Mkln1 constructs NK, LC, and CC, and a weaker interaction with the original 1D clone (Figure 4.10). Other constructs showed some growth under the weakest selection conditions for interactions with $Isl1_{LIM}$, but this was at similar level to the negative control and was not evident under more stringent selection conditions.

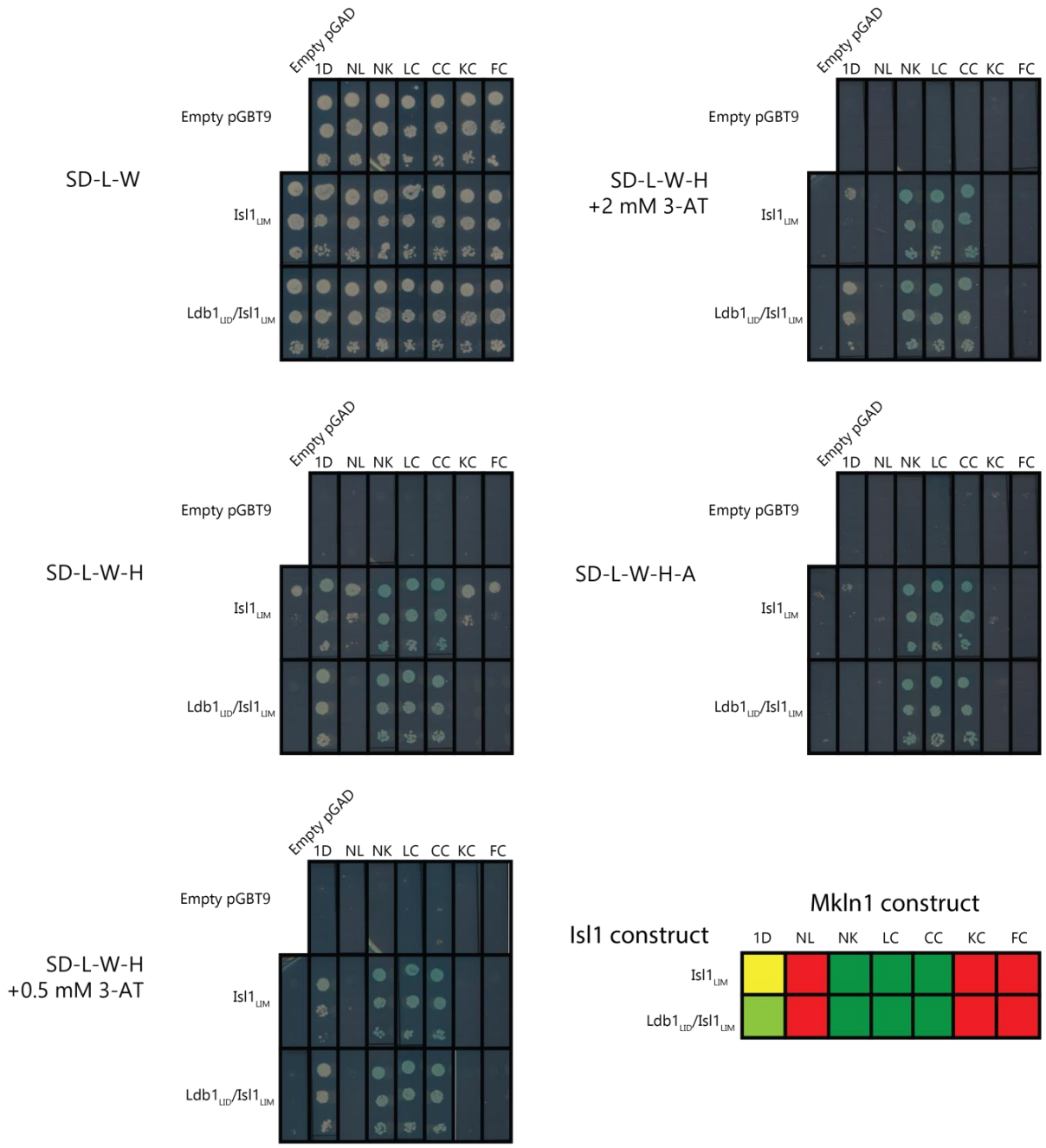


Figure 4.10: Spot tests of Mkln1 constructs against Isl1 constructs. Yeast were co-transformed with one pGBT9 plasmid and one pGAD plasmid, and grown on a range of selective media to screen for interactions. . Each set of three spots contains a serial 1:10 dilution of yeast, from OD_{600} 0.2 to 0.002. pGBT9 plasmids used were: empty pGBT9, pGBT9-$Isl1_{LIM}$, and pGBT9-$Ldb1_{LID}$/$Isl1_{LIM}$. pGAD plasmids used were: empty pGAD10, pGADT7-RecAB containing the library screen clone 1D, and pGADT7-RecAB containing the Mkln1 truncation constructs NL, NK, LC, CC, KC, and FC. Lower right presents a scoring table for interactions, with red representing no interaction, and darker shades of green representing stronger interactions.

The interactions of $Mkln1_{NK}$, $Mkln1_{LC}$, and $Mkln1_{CC}$ with $Isl1_{LIM}$ and $Ldb1_{LID}/Isl1_{LIM}$ cannot be classified in terms of strength of binding from this experiment, as all combinations showed growth under the most stringent selection conditions. The full length Mkln1 construct (1D), which showed a weaker interaction than any of the truncations, was found to contain a frameshift in the sequence at the very beginning of the Mkln1 coding region, meaning the protein produced is most likely not full length Mkln1 (as discussed in Section 3.5.1).

Overall, the yeast two-hybrid experiments indicated that the CTLH domain region of Mkln1 is necessary for the interaction with Isl1. It remains unclear whether this region interacts with the Isl1 LIM domains on a different face to that of the conventional LID binding face.

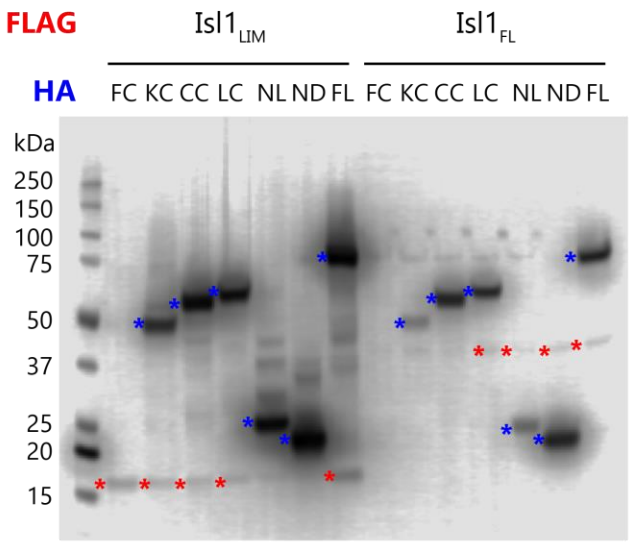
4.3.2 Co-immunoprecipitation results do not clearly indicate a minimal binding region of Mkln1

An orthogonal system was utilised to further validate the interaction between Isl1 and Mkln1. Isl1 and Mkln1 constructs were subcloned into two versions of pcDNA3.1: one encoding an N-terminal FLAG tag and the other encoding an N-terminal HA tag. Gibson cloning was used for this purpose and to simultaneously correct the frameshift found in the construct 1D, hereafter referred to as $Mkln1_{FL}$.

After constructs were subcloned, HEK Expi293F™ cells were co-transfected with combinations of two plasmids for protein expression. Each pair of plasmids transfected consisted of one encoding a FLAG-tagged bait and one encoding an HA-tagged prey. Both orientations of Isl1 and Mkln1 as bait and prey were used.

Three days after co-transfection, cells were harvested. Western blotting was used to confirm the expression of both bait and prey proteins, through the simultaneous use of anti-FLAG and anti-HA antibodies (Figure 4.11). The majority of constructs showed expression, with the consistent exception of $Mkln1_{FC}$. It is possible that $Mkln1_{FC}$ was expressed, but was not detectable through Western blotting in the conditions used as it was too small for the transfer conditions used.

(A)



(B)

Construct	Expected size (kDa)
FLAG-Isl1$_{FL}$	39.2
FLAG-Isl1$_{LIM}$	16.1
HA-Mkln1$_{FL}$	87.2
HA-Mkln1$_{ND}$	21.5
HA-Mkln1$_{NL}$	26.1
HA-Mkln1$_{LC}$	67.7
HA-Mkln1$_{CC}$	63.7
HA-Mkln1$_{KC}$	54.8
HA-Mkln1$_{FC}$	13.2

Figure 4.11: Expression of FLAG- and HA-tagged proteins from HEK293 cells. (A) Representative anti-FLAG/anti-HA Western blot, showing expression of 14 co-transfections. (B) Expected sizes of FLAG- and HA-tagged constructs. HA-tagged Mkln1 constructs and FLAG-tagged Isl1 constructs are marked with blue and red asterisks, respectively.

Immunoprecipitation was performed using an anti-FLAG antibody immobilised on beads, with immunoprecipitates being eluted from the beads by adding 3× FLAG peptide to compete off the FLAG-tagged protein and any binding partners. Samples of co-immunoprecipitation input and eluate were subjected to anti-FLAG/anti-HA Western blotting to detect both bait and prey, and evidence for any interactions between the two (Figure 4.12). A representative experiment is shown in Figure 4.12A, with a summary of all the experiments in Figure 4.12C.

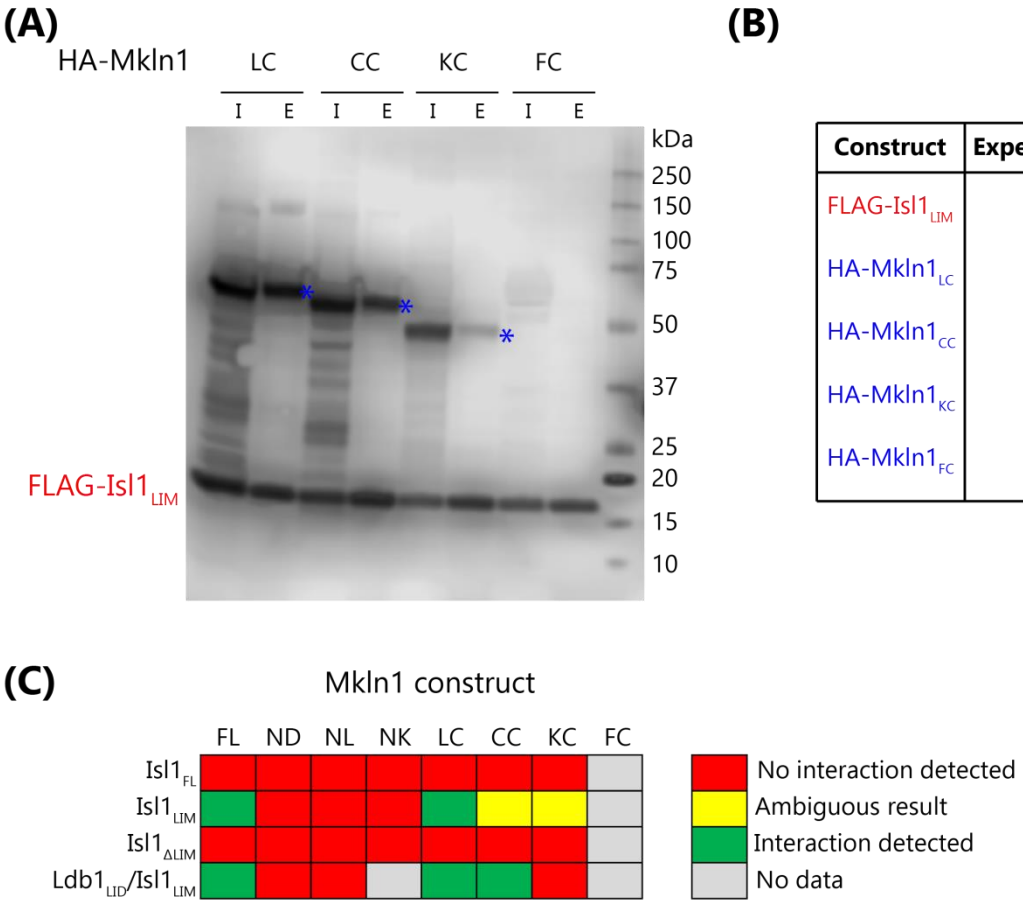


Figure 4.12: Detecting the interaction between Isl1 and Mkln1 using co-immunoprecipitation. (A) Representative Western blot of co-immunoprecipitation experiments. I: input; E: elution. HA-tagged Mkln1 constructs are marked with blue asterisks. (B) Expected size of FLAG bait and HA prey. (C) Summary of co-immunoprecipitation results. Ambiguous result indicates either a weak band was detected (For example $Isl1_{LIM}$ against $Mkln1_{KC}$ in (A)), or a band was not consistently detected in replicated experiments/both orientations. No data indicates expression of one or both partners was not observed, meaning that the interaction could not be tested.

Most Mkln1:Isl1 combinations showed no interaction by co-immunoprecipitation (Figure 4.12C). Consistent with the yeast two-hybrid data, the $Isl1_{\Delta LIM}$ construct did not interact with any Mkln1 construct, and the Mkln1 constructs ND and NL did not interact with any Isl1 construct. In further agreement, $Isl1_{LIM}$ and $Ldb1_{LID}/Isl1_{LIM}$ both interacted with $Mkln1_{FL}$, $Mkln1_{LC}$, and $Mkln1_{CC}$, although the interaction between $Isl1_{LIM}$ and $Mkln1_{CC}$ was only observed with FLAG-Isl1 and HA-Mkln1, not the reverse orientation (i.e., with the tags swapped). In contrast to the yeast two-hybrid results, however, $Mkln1_{NK}$ did not interact with either $Isl1_{LIM}$ or $Ldb1_{LID}/Isl1_{LIM}$, although no data could be obtained for $Mkln1_{NK}$ against $Ldb1_{LID}/Isl1_{LIM}$. Additionally, HA-$Mkln1_{KC}$ was shown to weakly bind to FLAG-$Isl1_{LIM}$ by

co-immunoprecipitation, where no interaction was observed using yeast two-hybrid analysis (Figures 4.10, 4.12A and C).

An unexpected result from the co-immunoprecipitation data was the lack of any interaction detected between any Mkln1 construct and $Isl1_{FL}$. Given that expression of $Isl1_{FL}$ was observed consistently, and that using FLAG-$Isl1_{FL}$ effectively enriched for $Isl1_{FL}$, it seems unlikely that this was due to technical issues such as poor expression levels. It is possible that the intramolecular $Isl1_{LIM}$:$Isl1_{LID}$ interaction may occlude the Mkln1 binding site (Section 1.3). This would seem unlikely given that $Isl1_{LID}$ binds across the same interface as $Ldb1_{LID}$, which does not obstruct the Mkln1 interaction, but it is possible, as $Isl1_{LID}$ and $Ldb1_{LID}$ do have different kinetics of binding [313].

Overall the co-immunoprecipitation results, while in partial agreement with the yeast two-hybrid results, did not clearly define a minimal binding region for the Isl1:Mkln1 interaction.

4.3.3 Classifying the Isl1/Mkln1 interaction

Taking the yeast two-hybrid and co-immunoprecipitation results together, it is likely that there is a genuine interaction between the two proteins, as there were interactions detected between Isl1 and Mkln1 in both systems. This interaction is likely to involve the LisH and CTLH domains, and may also require the Kelch repeats. More work is required to determine which regions are involved in binding, and which are required for correct folding, as there were no characterisation experiments performed to confirm whether the Mkln1 constructs used were correctly folded.

It is conceivable that the experimental conditions found in the co-immunoprecipitation scenario impacted the interaction, preventing it from forming between certain combinations of constructs [314]. This could be due to steric constraints introduced by the presence of tags, or by the immobilisation of proteins using affinity beads. It is also possible that endogenous proteins from the HEK293 cells bound to either of the interaction partners in preference to the interaction being tested and prevented that interaction from taking place, as there were no purification steps performed to separate the proteins of interest from the rest of the soluble cell lysate.

Another factor to consider is the tendency for Mkln1 to form dimers and tetramers: it may be that the interaction with Isl1 can only occur with one oligomeric form of Mkln1, and that the truncation constructs used prevent Mkln1 from adopting this state [311].

Further experimentation to determine the nature of this interaction, as well as its biological relevance, should be conducted to confidently show a bona fide interaction between these two proteins. However, this was not possible within the time frame of this thesis.

4.4 Discussion

The data presented in this chapter identified several proteins that represent potential biologically relevant interaction partners for Isl1. What follows is a summary of the likely biological relevance of these identified interactions, followed by an evaluation of the success of the screening process.

4.4.1 Assessing the pool of remaining prey

Of the 27 prey proteins that were found to represent specific Isl1-interacting proteins in yeast two-hybrid analysis, 12 are plausible interactors based on known functional information about these proteins (Section 4.2, Table 4.2).

Table 4.2: Classification of identified proteins according to likelihood of biological relevance of the interaction.

Isl1 bait	Unlikely to interact	Insufficient information to comment on	Likely to interact
$Isl1_{LIM}$	Sparcl1	Zfand1	Mkln1
	Mep1b	Lace1	Ddx20
	Rps18		Dfna5
			Nup50
$Isl1_{\Delta LIM}$	Art3	Lrrc51	Kctd9
	Ly6c1	BC035947	Nkiras1
	Ms4a5	Tigd2	
	Nectin3		
	Oscp1		
	Rnf167		
	Zdhhc20		
	Scpep1		
	Spata7		
	Usp8		
	Cops5		
	Cyc1		
$Ldb1_{LID}/Isl1_{LIM}$		Ubd	Mkln1

Of these 12 proteins, only Lrrc51, Ubd, Nkiras, and Mkln1 have been tested in their full-length form. The others were recovered from the screens as truncated constructs. The first step in further validating these interactions should be obtaining the full-length protein and re-testing the interaction with Isl1, both in yeast two-hybrid assays, and in an orthogonal system such as co-immunoprecipitation.

Mkln1 is a putative interactor of interest, being the only validated prey protein that could interact with more than one Isl1 construct. Further, the mammalian co-immunoprecipitation data did indicate an interaction between Isl1 and Mkln1. However, the conflicting data for the minimal binding region of Mkln1 needs to be resolved, and additional *in vivo* data is needed to confirm the biological relevance of the interaction.

Confirmation of the co-localisation of Isl1 with these potential interaction partners in cells is an important future experiment. Evidence for co-localisation would build confidence in the biological relevance of these interactions. Once minimal binding domains have been established, and binding sites characterised, it would also be possible to test the functionality of these interactions by mutating the binding sites and then observing resulting phenotypes in cell lines or animal models.

4.4.2 Assessing the success of library screening for identifying new binding partners

There are several ways of assessing whether the screening performed here was an overall success. One is to assess the number of novel binding partners identified compared to the total number of hits analysed. Another is to compare the interactors identified from these screens with known binding partners for Isl1 in the literature. Both assessments are discussed below.

4.4.2.1 Assessing the number of novel binding partners identified

Several proteins were identified that could represent novel, specific, and biologically relevant binding partners for Isl1. However, only 6% of the total sample pool obtained from yeast two-hybrid library screening was found to represent specific interactions with Isl1 (Figure 4.13). If this number is expanded to include less specific interactions, including interactions with the close homolog Isl2 or other nuclear LIM proteins, which could still be biologically relevant, it reaches 22%. This means the majority of hits from screening did not represent potential binding partners for Isl1. This inefficiency draws attention to the problems associated with library screening, and for this particular system used, as was discussed in the previous chapter (Section 3.8.1).

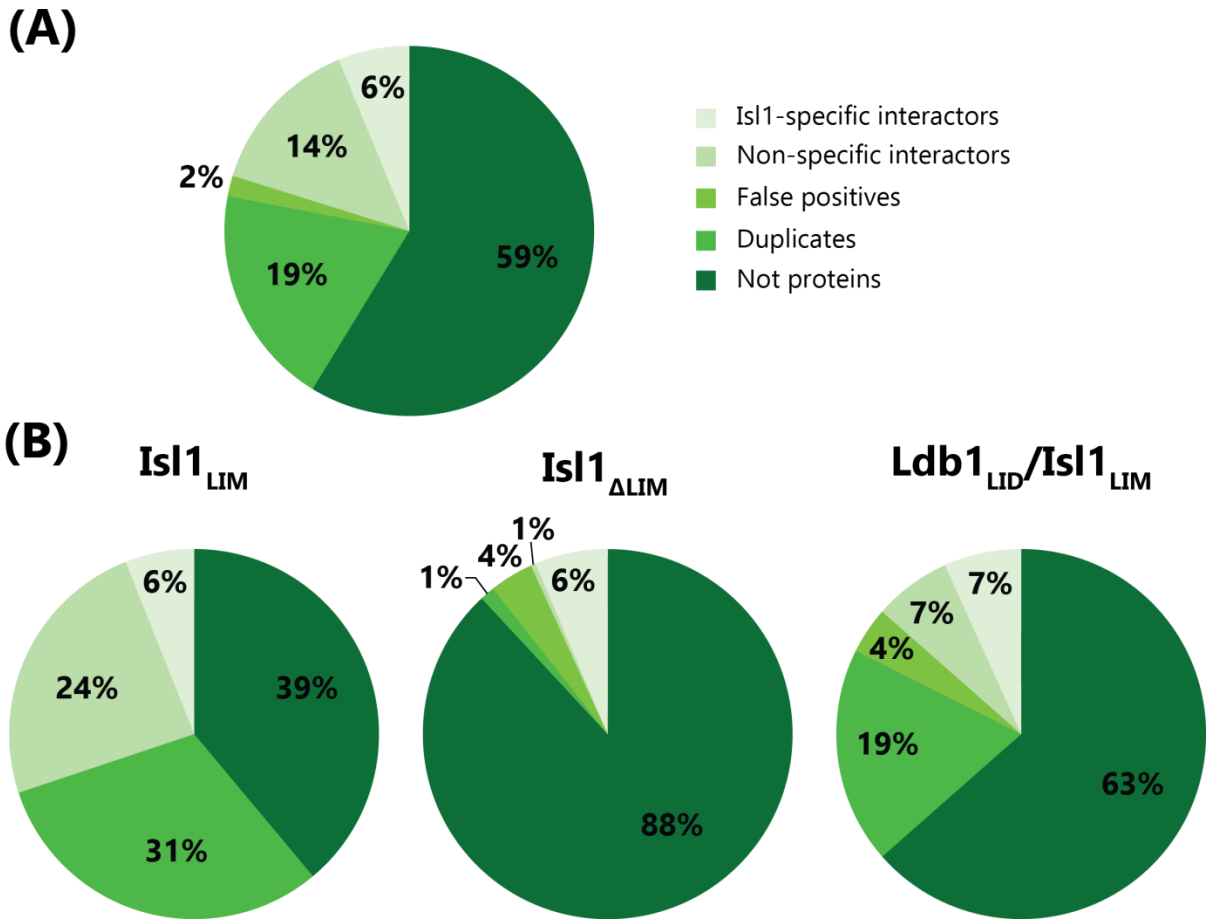


Figure 4.13: Breakdown of the categorisation of hits isolated in yeast two-hybrid library screening. (A) All screens combined. (B) Hits according to Isl1 bait.

4.4.2.2 Known Isl1-interacting proteins in the pool of detected interactors

The identification of known Isl1-interacting proteins in the screens should function as positive controls, indicating that the screening was thorough and that the baits were behaving as they would *in vivo*.

One glaring absence from the pool of interacting proteins is Ldb1, the major cofactor of all LIM-HD proteins, including Isl1 (Section 1.2.2.1). Ldb1 was expected to be present in the pool of $Isl1_{LIM}$-interacting proteins. The interaction between $Ldb1_{LID}$ and $Isl1_{LIM}$ is well characterised, and has been previously observed in yeast two-hybrid assays in the Matthews laboratory [51, 54]. Ldb2 was identified from $Isl1_{LIM}$ screening, which should offer some reassurance. However, Ldb2 was among the prey that had protein sequence encoded out of frame with the upstream GAL4 sequence, and M7 was an Ldb2-encoding hit that was screened to check for whether a frameshift was occurring during protein expression (Section 3.5.1). As no frameshift was observed, Ldb2 cannot be considered as a positive control.

Other LIM-HD proteins like Lhx3 and Lhx4 should also have been present in the $Isl1_{\Delta LIM}$-interacting pool. While the LID in Isl1 does not interact with as many proteins as does $Ldb1_{LID}$, $Isl1_{LID}$ is known to interact with Lhx3 and Lhx4 with reasonable affinity [54, 57, 313]. Neither of these proteins was identified in the $Isl1_{\Delta LIM}$ screens.

Given the known strength of these interactions, the absence of Ldb1, Lhx3, and Lhx4 in the prey pool suggests that they were not screened against Isl1. As these proteins are present at low abundance in cells, it is possible that they were not successfully subcloned into the pGADT7-RecAB yeast library, or that the yeast containing these plasmids did not successfully mate with the bait strain. If this were the case, repeated screening should allow the identification of these proteins, and may also reveal more novel binding partners.

An alternative explanation is that transcripts for Ldb1, Lhx3, and Lhx4 were not present in the mRNA pool used to generate the library used for screening. While the cDNA library used is marketed as universal, the transcripts are derived from adult mice [315]. As the LIM-HD proteins play many roles in embryonic development, it is possible that they are not expressed at sufficiently high levels in adult mice to ensure that they were present in the cDNA library. Certainly, Lhx3 and Lhx4 are expressed at very low levels in the adult mouse. However, Ldb1 is expressed at levels comparable to those of Zfand1, which was identified four times during screening [316].

The use of an adult-derived cDNA library reveals a larger issue with the work presented here. While many LIM-HD proteins are expressed both during development and in the fully developed adult mouse, there are many key developmental proteins that are only expressed during embryonic development. These would be ideal targets to test for interaction with Isl1, as Isl1 is implicated in the embryonic development of numerous tissues. However, these proteins are not present in the cDNA library used. This represents a significant flaw in the approach used here. This flaw was not noticed until the absence of Ldb1 was confirmed, leaving insufficient time to repeat the experiments with an embryonic mouse cDNA library, and highlights the issue that the output of any library screen is dependent on the composition of the library.

4.4.3 Future work to map the interaction partners of Isl1

Although yeast two-hybrid screening may not have been successful in generating a comprehensive map of Isl1 interacting proteins, it still provided useful avenues for further study. Many of the proteins identified as putative binding partners need further validation, through the use of mammalian expression and co-immunoprecipitation. Ddx20, Dfna5, Nup50, Kctd9, and Nkiras1 would be good targets for this, as they are the most likely candidates to represent biologically relevant interactions (Table 4.2).

Assuming the Isl1:Mkln1 interaction is validated through co-localisation studies, it would be interesting to determine the structure of the Isl1:Mkn1 complex. Once minimal binding domains have been identified, the relevant domain constructs could be produced. Recombinant Mkln1 was previously expressed in bacteria, in sufficient yields for crystallography studies [311]. Isolated $Isl1_{LIM}$ would prove difficult to obtain, given the propensity for unbound LIM domains to aggregate in solution, but co-expression and purification of the two proteins could provide soluble complex. $Ldb1_{LID}$/$Isl1_{LIM}$ can be produced in sufficient yields using bacterial expression [54], suggesting this approach would likely be successful. One key question that would be answered by structure determination of the complex is whether the Isl1 LIM domains can simultaneously bind multiple proteins (Mkln1 and $Ldb1_{LID}$), or if Mkln1 competes off $Ldb1_{LID}$. This information would give valuable insights into the complexity of transcriptional regulation by Isl1.

There are also several proteins already proposed in the literature as Isl1 interactors that need more thorough examination, including Pou4f2, Phox2a, Neurod1, and Stat3 (Section 1.5) [113, 187, 317]. None of these complexes have been confirmed to feature a direct interaction with Isl1. Stat3 is of particular interest, as it has been shown to cooperate with the motor neuron complex, in which Isl1 is bound to Ldb1 and Lhx3 via its LIM and LID regions respectively (Section 1.4.1, Table 1.3). If a direct interaction was detected between Isl1 and Stat3, it could reveal a new binding interface on Isl1.

Additional interaction partners for Isl1 could be discovered through alternate techniques, which in combination with the yeast two-hybrid screens presented here would help improve confidence of their biological relevance. Recent interaction mapping studies have shown the potential of co-immunoprecipitation coupled with mass spectrometry (co-IP/MS) analysis in

detecting interactions in a high throughput manner [318-321]. While yeast two-hybrid methodology is well established in the Matthews laboratory, co-IP/MS analysis would have been a good alternate technique to identify new Isl1-interacting proteins. Compared to yeast two-hybrid library screening, the co-IP methodology can be easily applied to a range of different samples, such as different cell lines or mouse tissues, as there is no need to procure or produce cDNA libraries. There are also many software suites available for analysing the results of such experiments in a high throughput manner. However, the main barrier to a co-immunoprecipitation approach is the requirement for a suitable antibody against the target protein. Although many antibodies against Isl1 are available, these antibodies are primarily used for immunofluorescence or Western blotting, and require testing to confirm that they are suitable for co-immunoprecipitation before a large scale interaction screen could be conducted. Preliminary experiments to this end were attempted for this project, and demonstrated the need for validation of the antibody. None of the three antibodies tested was sufficiently specific for co-immunoprecipitation (data not shown).

4.4.4 Conclusion

Overall, the aim of identifying novel binding partners to Isl1 was achieved, but use of more efficient methodology would have allowed for further investigation of relevant proteins, and potentially the identification of more binding partners. As it stands, there are several novel putative binding partners of Isl1 identified in this work. Further work could address the nature of these interactions, whether they are biologically relevant, and their consequences for development.

Part II

Protein-DNA interactions

5 Investigating the mechanisms behind the action of the motor neuron complex

5.1 Introduction

This chapter describes efforts to understand the interactions between homeodomains and DNA that are associated with the motor neuron complex, as described in Section 1.4.1. All protein constructs were cloned prior to the commencement of this thesis, unless otherwise specified (Section 5.2). Expression and purification protocols had been established, but were further optimised during the course of this thesis to enable production of sufficient amounts of high purity protein for downstream characterisation and structural studies (Section 5.3).

5.2 Construct design

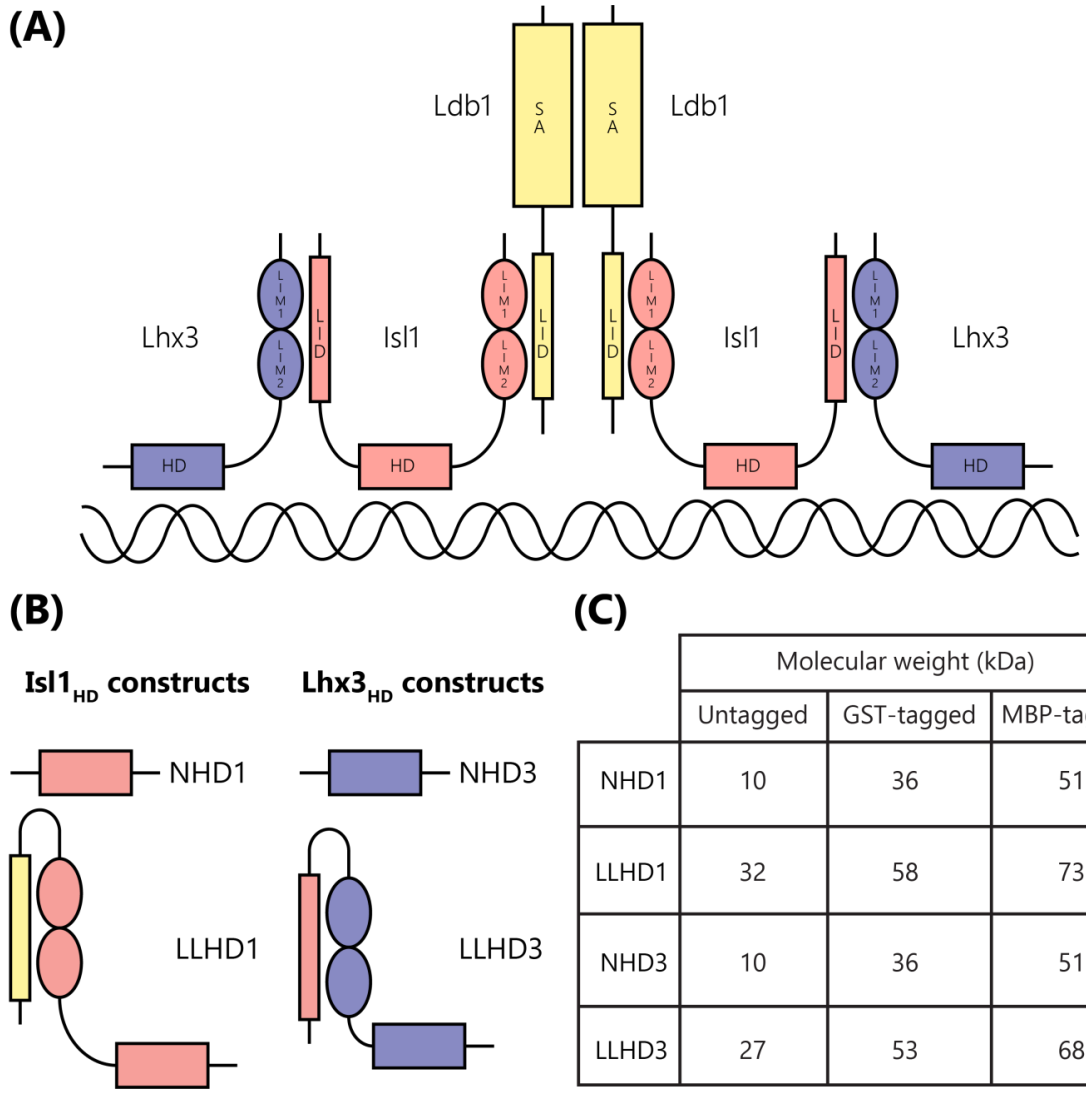
Various constructs of Isl1, Lhx3, and Ldb1 were generated for this project, including truncation mutants of single proteins as well as fusion constructs comprising domains of two proteins. These constructs were designed with the aim of understanding which regions are important for imparting affinity and specificity to the protein-DNA interactions that occur *in vivo*. Full construct sequences can be found in Appendix G.

5.2.1 Single homeodomain constructs

Constructs encoding the isolated homeodomains from Isl1 and Lhx3, named NHD1 and NHD3 respectively, had already previously been cloned (Tom Drury, University of Sydney). The canonical homeodomain is 60 residues long, including the N-terminal arm that is important for interacting with DNA. However, difficulties were encountered observing DNA binding using 60 residue homeodomain constructs (Tom Drury, unpublished data), so the constructs used throughout this thesis were 80 residues long, with extended N- and C- termini [214].

Additional single homeodomain constructs, containing upstream protein-protein interaction domains, were also produced to better mimic the motor neuron complex *in vivo* (Figure 5.1A). These constructs were named LLHD3, which encoded $Isl1_{LID}$ fused to $Lhx3_{LIM+HD}$,

and LLHD1, which encoded $Ldb1_{LID}$ fused to $Isl1_{LIM+HD}$ (Figure 5.1B). LLHD3 was produced during my Honours work in 2014, and LLHD1 was produced for this thesis.

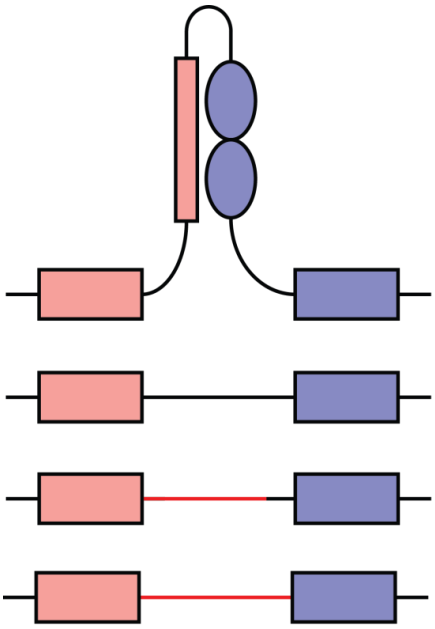


	Molecular weight (kDa)		
	Untagged	GST-tagged	MBP-tagged
NHD1	10	36	51
LLHD1	32	58	73
NHD3	10	36	51
LLHD3	27	53	68

Figure 5.1: Single homeodomain constructs used in the project. (A) Schematic model of the motor neuron complex found in development. (B) Single homeodomain constructs used throughout the project. (C) Molecular weights of single homeodomain constructs with various protein fusion tags.

5.2.2 Fusion homeodomain constructs

Four homeodomain fusion constructs were used over the course of the work described in this thesis. All homeodomain fusion constructs used contained $Isl1_{HD}$ at the N-terminus, followed by a varied region, with $Lhx3_{HD}$ at the C-terminus (Figure 5.2). The construct 2HDLL contains native sequence that would naturally occur between the two homeodomains when in complex — the C-terminal LID of Isl1, and the N-terminal LIM domains of Lhx3.



	Molecular weight (kDa)		
	Untagged	GST-tagged	MBP-tagged
2HDLL	37	64	78
2HDN	21	48	62
2HD17	20	47	61
2HD23	20	47	61

Figure 5.2: Homeodomain fusion constructs used throughout the project. Isl1 is shown in pink and Lhx3 in purple; red line indicates glycine/serine linker, in place of native protein sequence.

The constructs 2HDN, 2HD17, and 2HD23 were designed to contain 23-residue disordered linkers between the two homeodomains, in place of the native protein-protein interaction domains. In 2HDN, the linker contains a portion of sequence from Isl1 (the sequence between the end of the homeodomain and the beginning of the LID), and a portion of sequence from Lhx3 (the sequence between the end of the LIM domains and the beginning of the homeodomain). In 2HD23, all of this sequence is replaced with a glycine/serine linker. In 2HD17, the 6 residues N-terminal to $Lhx3_{HD}$ are derived from native Lhx3 sequence in case those residues play a role in DNA binding or folding. 2HDLL was cloned by Tom Drury (University of Sydney). The remaining three constructs were produced during my Honours work in 2014 [209].

These constructs were generated to determine if any of the protein sequence between the homeodomains and their adjacent LIM or LID domains has an effect on DNA binding. In addition, 2HD17 and 2HD23 were produced in case 2HDN proved difficult to work with. For example, if the native linker in 2HDN was prone to cleavage by proteases, the glycine/serine linkers in 2HD17 and 2HD23 should be less prone to proteolysis.

5.3 Optimisation of homeodomain expression and purification

Initially, production of the homeodomain constructs gave poor protein yield, with low reproducibility (Figure 5.3). Some optimisation of protein expression was conducted during my Honours work, but more was needed to develop protocols that reliably yielded sufficient protein for a variety of experiments.

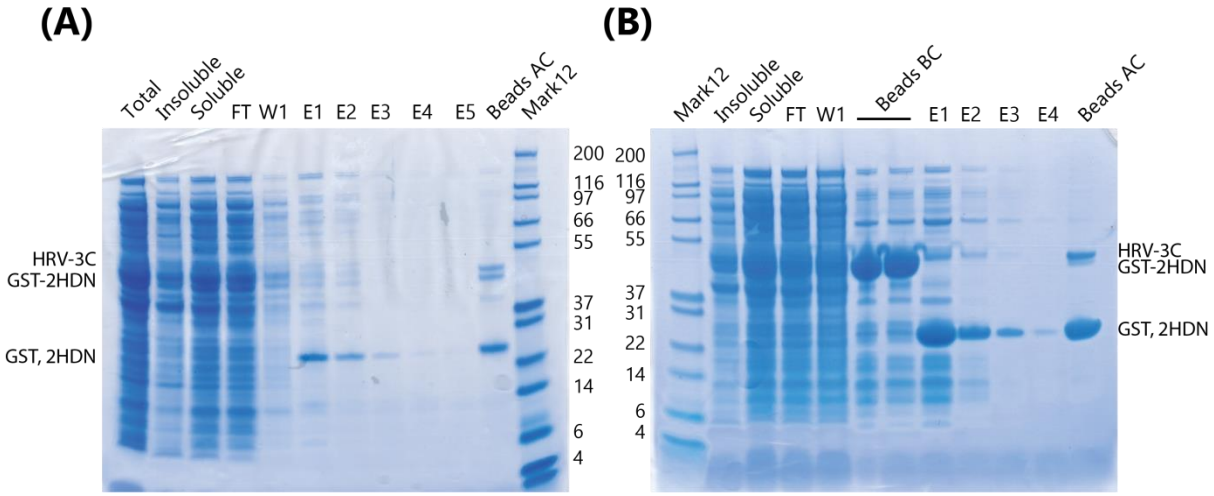


Figure 5.3: Variability of 2HDN purification. Two examples of glutathione affinity purification of 2HDN using non-optimised protocols. FT: flow-through; W1: first wash; BC: before cleavage; AC: after cleavage; E1-E5: elution fractions 1-5. (A) Example of purification from 2015 with low yield. (B) Example of a purification from 2014 with high yield. Note that GST (glutathione S-transferase) and 2HDN cannot be resolved using SDS-PAGE in the conditions used.

Initial expression conditions used homeodomain constructs subcloned into vectors encoding either an N-terminal GST tag (2HDN, 2HD17, 2HD23, LLHD3) or maltose binding protein (MBP) tag (2HDLL, NHD1, NHD3), under the control of a *lacUV* promoter. The use of an MBP tag for 2HDLL, NHD1, and NHD3 was to improve protein solubility and overall yield, so the remaining constructs (2HDN, 2HD17, 2HD23 and LLHD3) were subcloned into the same vector for this thesis in the hope of improved expression levels. However, this modification did not lead to higher yields (Figure 5.4).

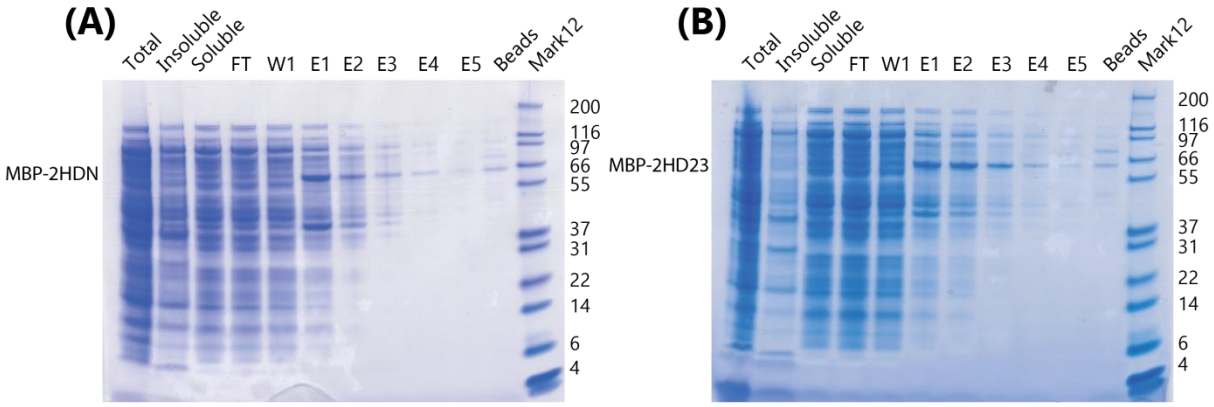


Figure 5.4: Amylose affinity purifications of homeodomain constructs with an N-terminal MBP tag. (A) Purification of MBP-2HDN; (B) Purification of MBP-2HD23.

Inconsistencies in culture growth were observed with all constructs. For example, cultures would often not grow, or take more than four hours to reach a density suitable for induction of protein expression. These inconsistencies could indicate toxicity resulting from leaky expression of the protein constructs, as small amounts of the DNA-binding homeodomains could bind to bacterial DNA, interfering with cell growth or metabolism.

Constructs were subsequently subcloned (for this thesis) into a pET-DUET expression vector encoding an N-terminal GST fusion under the control of a T7 promoter, to reduce possible leaky expression (Section 2.1.4). The possible effects of leaky expression and protein toxicity were tested by comparing the time taken for test cultures, each containing the same plasmid in a different *E. coli* strain, to reach an OD_{600} (optical density at 600 nm) of 0.7 from the same starting density (Table 5.1). Previous experiments used ampicillin to maintain the plasmid, but degradation of that antibiotic can lead to inconsistent bacterial growth and loss of the plasmid. Accordingly, carbenicillin was used instead as it is more resistant to degradation [322, 323].

Table 5.1: Testing for leaky expression and toxicity. Three strains of *E. coli* (BL21 (DE3), BL21 Gold (DE3) (pLysS), and Rosetta 2 (pLysS)) were transformed with an expression vector encoding GST-2HDLL, under the control of a T7 promoter. Cultures (10 mL) supplemented with 100 µg/mL carbenicillin (and 34 µg/mL chloramphenicol for pLysS strains) were inoculated with transformed bacteria with an initial OD_{600} of 0.05, and incubated with shaking until reaching an OD_{600} of 0.7 ± 0.05. The time taken to reach this OD_{600} is reported in minutes (N = 2).

Temperature (°C)	**BL21 (DE3)**	**BL21 Gold (DE3) (PLysS)**	**Rosetta 2 (PLysS)**
25	360	280	295
37	205	160	160

pLysS strains contain a plasmid that expresses T7 lysozyme, which inhibits T7 polymerase produced from the DE3 prophage, reducing leaky expression [324]. Both strains that contained a pLysS plasmid reached an OD_{600} of 0.7 in similar time. The slower growth of the BL21 (DE3) cultures suggests that leaky expression is occurring in that strain and impacting the growth of the bacteria.

Final expression protocols for all homeodomain constructs consisted of the following: transformed BL21 Gold (DE3) (PLysS) *E. coli* were grown at 37 °C in LB supplemented with 100 µg/mL carbenicillin, 34 µg/mL chloramphenicol, and 0.2% w/v glucose, until reaching an OD_{600} of 0.7, at which point protein expression was induced by the addition of 0.4 mM IPTG (Isopropyl β-D-thiogalactopyranoside) (Section 2.5.2). Cultures expressing proteins containing LIM and LID domains were incubated at 25 °C for 16-20 hours post-induction, rather than at 37 °C for three hours post-induction for HD-only constructs.

Purification protocols were adapted from already established protocols [214]. However, nucleic acid contamination was observed in partially purified protein during this thesis work, so a PEI (polyethylenimine) precipitation step was introduced to precipitate nucleic acids prior to affinity chromatography (Figure 5.5). This new step achieved the goal of removing nucleic acid contamination without compromising protein yield. It was also observed that elution fractions from glutathione resin contained fewer contaminants when the lysate had been subjected to PEI precipitation prior to GSH-affinity chromatography.

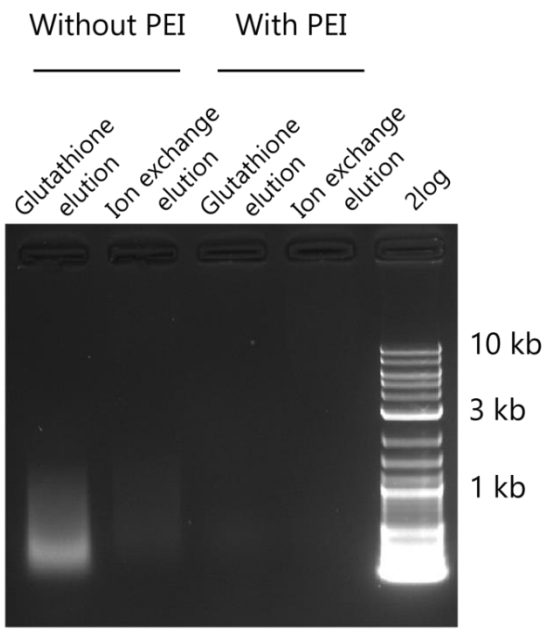


Figure 5.5: Removal of nucleic acid contamination by PEI precipitation. Elution fractions from the purification process of 2HDLL were analysed by agarose electrophoresis and visualised by staining with HydraGreen™.

The final optimised protocol consisted of cell lysis by French press, nucleic acid precipitation by PEI, GSH-affinity chromatography, and cation exchange chromatography (Section 2.5.3). All homeodomain-containing constructs could be purified to a high standard using this set of protocols (Figure 5.6, Appendix H).

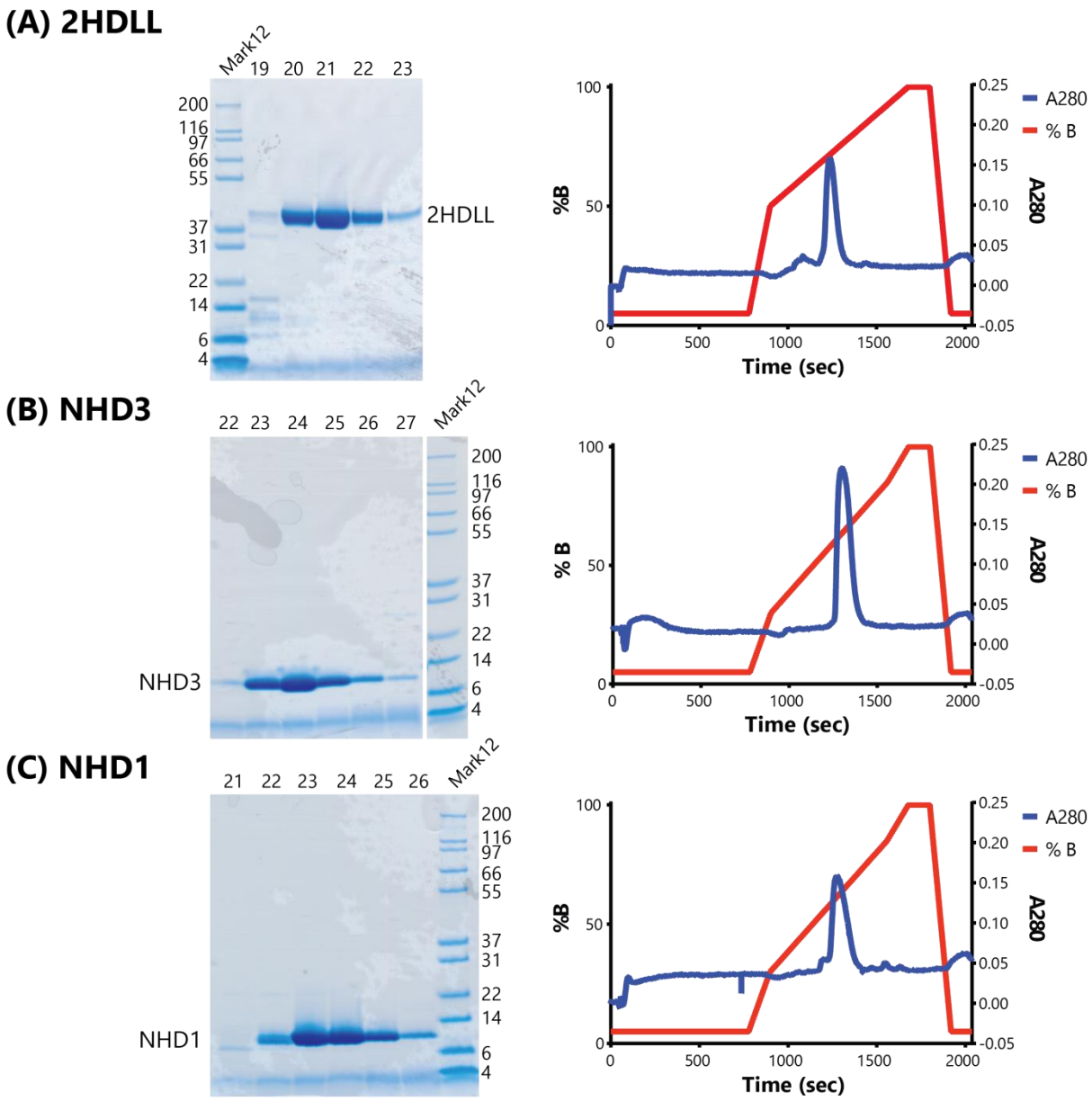


Figure 5.6: Examples of purifications of homeodomain constructs. Further examples can be found in Appendix H. (A) 2HDLL purification, showing cation exchange elution fractions 19-23 on SDS-PAGE, and elution profile. (B) NHD3 purification, showing cation exchange elution fractions 22-27 on SDS-PAGE, and elution profile. (C) NHD1 purification, showing cation exchange elution fractions 21-26 on SDS-PAGE, and elution profile.

5.4 Isolated homeodomains are folded in solution

In order to confirm that the homeodomain constructs produced were in a form that could bind to DNA, far-UV circular dichroism (CD) and nuclear magnetic resonance (NMR) were used to investigate the folded state of the proteins.

The far-UV CD spectra observed are typical of folded proteins with maxima below 200 nm (Figure 5.7). The double minima observed (around 208 and 222 nm) are characteristic of alpha-helices and are consistent with the 3-helix fold of the canonical homeodomain, with additional disordered residues. There was no appreciable difference between the signals recorded for NHD1 and NHD3, suggesting that they are in very similar states in terms of secondary structure.

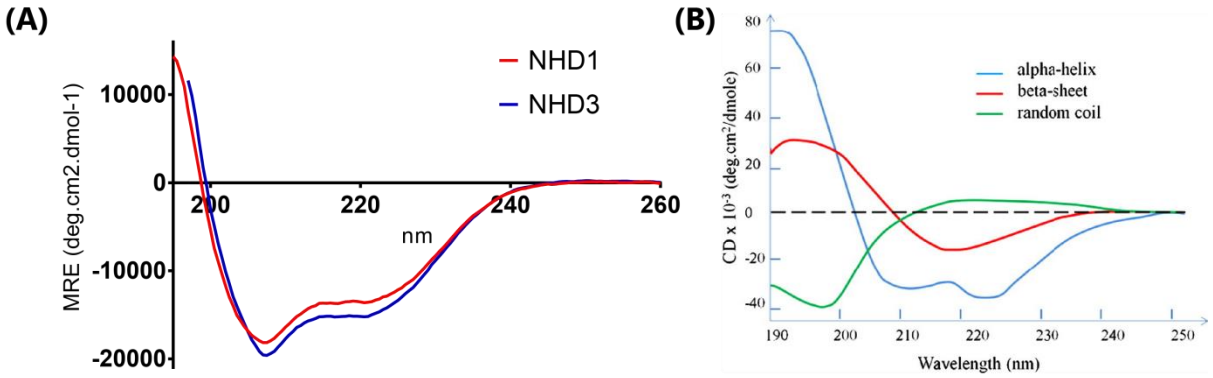


Figure 5.7: Far-UV CD spectra of single homeodomains. (A) Spectra for NHD1 (red) and NHD3 (blue), collected at 20 °C. (B) Typical spectra characteristic of alpha helices (green), beta sheets (blue), and random coil (red) Figure taken from [325].

^{1}H 1D NMR spectra taken for each of NHD1 and NHD3 showed sharp dispersed peaks characteristic of folded protein (Figure 5.8). In particular, peaks below 1 ppm and around 7 ppm are good indicators of ‘foldedness’. Once more, the spectra are very similar, indicating NHD1 and NHD3 are likely to have similar tertiary structure.

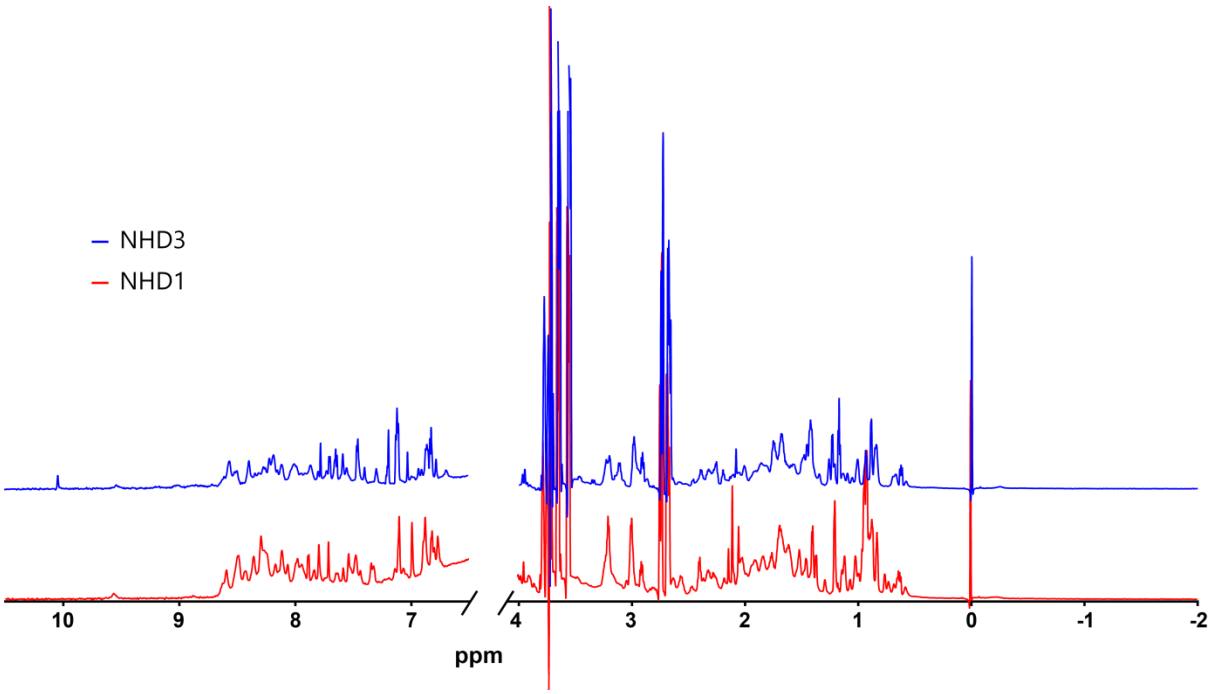


Figure 5.8: ^{1}H 1D NMR spectra. NHD1 is shown in red and NHD3 is shown in blue. Peaks in the 7-10 ppm and 1-3 ppm range are characteristic of folded protein.

5.4.1 Investigation of protein fold stability of individual homeodomains

^{1}H 1D NMR was used to assess the stability of the fold of both NHD1 and NHD3. Spectra were recorded for each protein at both 25 °C and 37 °C. The spectra at the higher temperature for both proteins showed a disappearance and broadening of peaks characteristic of partial unfolding (Figure 5.9). This was particularly noticeable in the aromatic region between 8 and 9 ppm.

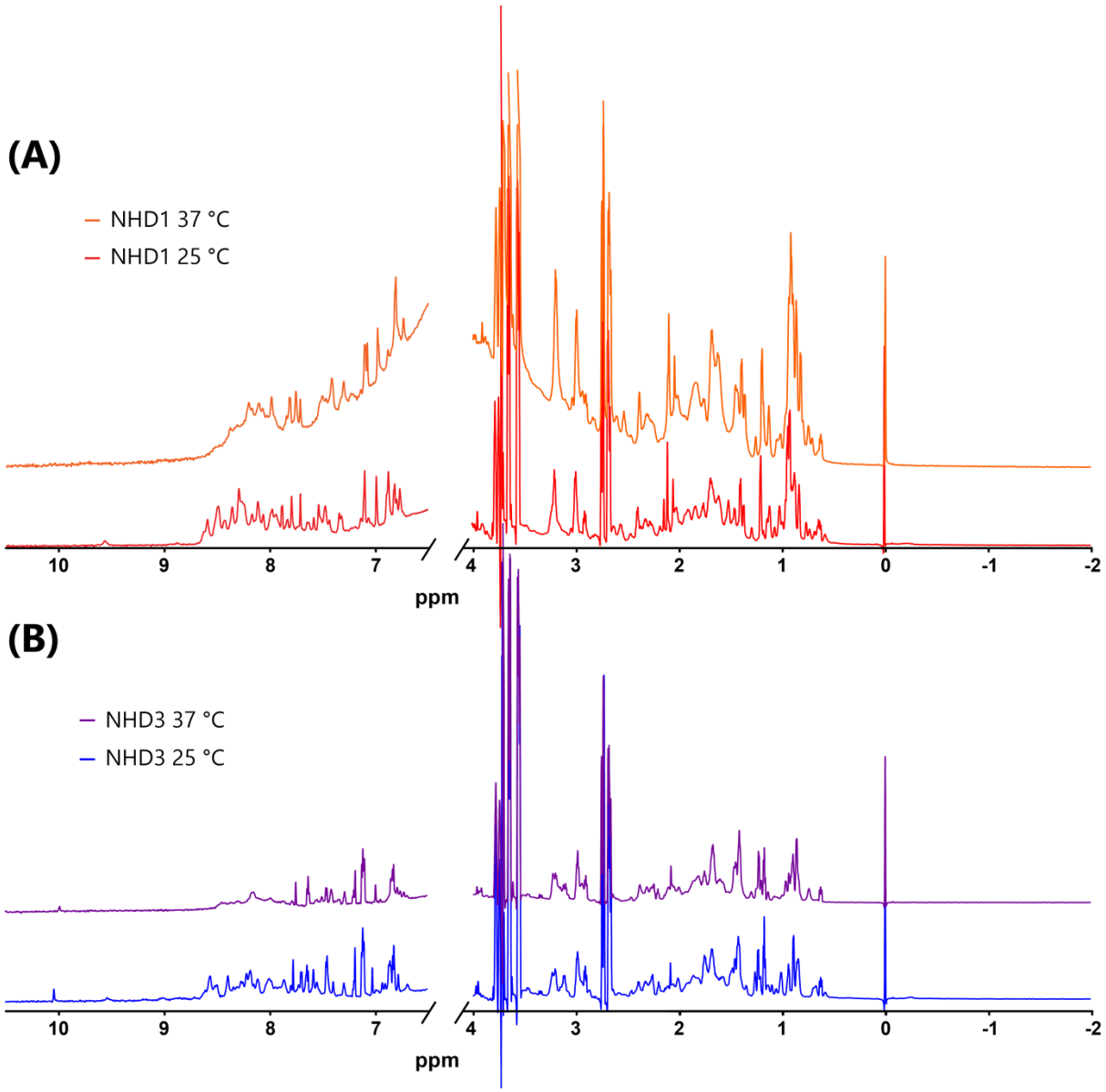


Figure 5.9 Comparison of NMR spectra taken at 25 °C and 37 °C. (A) NHD1 spectra; (B) NHD3 spectra.

The apparent unfolding with higher temperature was subsequently monitored using far-UV CD over the temperature range 10 °C to 90 °C. Both homeodomains appeared folded at ≤ 25 °C. The proteins exhibited a similar thermal denaturation profile, with an apparent T_m (melting temperature) at ~42 °C (Figure 5.10), as indicated by the disappearance of the signals at 208 and 222 nm, and a shift to typical unfolded spectra (see random coil, Figure 5.7B). Collectively these data indicate that $Lhx3_{HD}$ and $Isl1_{HD}$ have similar properties in terms of protein folding and stability.

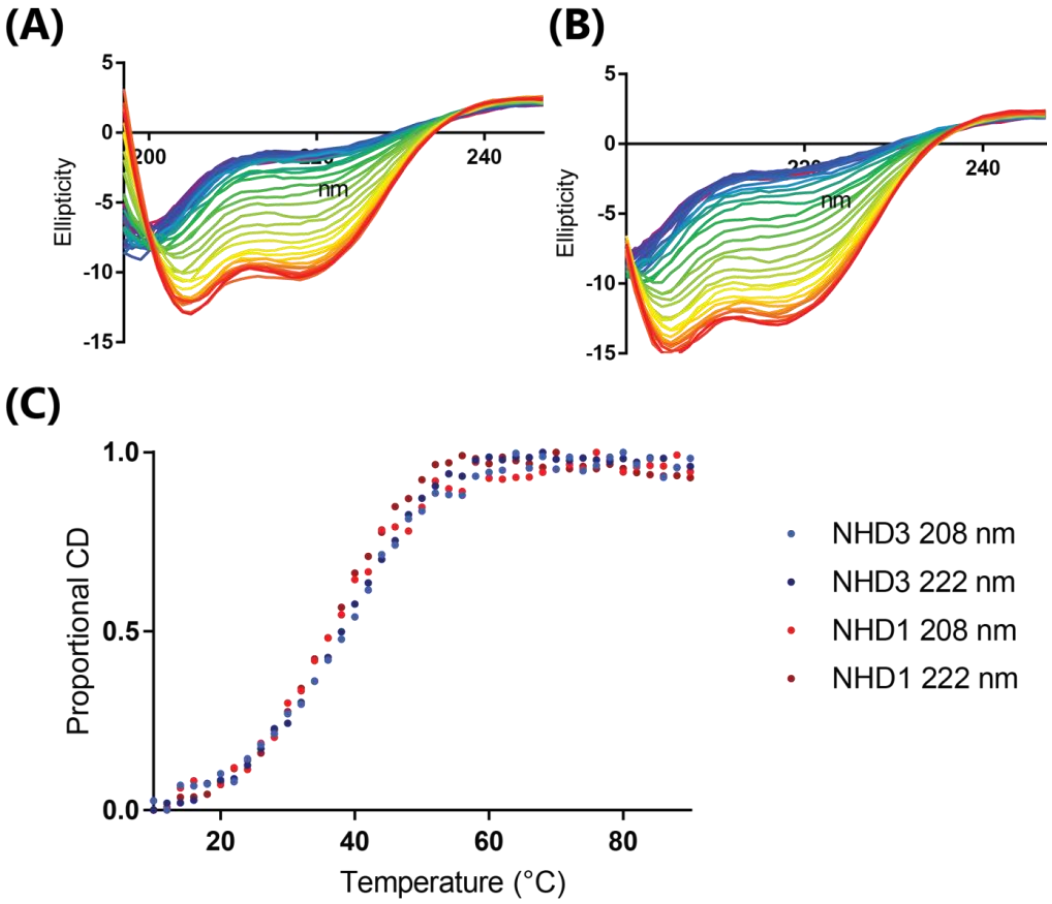


Figure 5.10: Far-UV CD melting profiles of single homeodomains. Lowest temperatures are shown in red; higher temperatures progress through a rainbow, with highest temperatures in purple. (A) Spectra of NHD1. (B) Spectra of NHD3 taken at 2 °C increments from 10 °C to 90 °C (according to sample temperature), heating at 2 °C/min. (C) Plot of temperature against proportional CD signal at 208 nm and 222 nm for both NHD1 and NHD3.

5.5 Using electrophoretic shift assays to probe *in vivo* DNA binding

Electrophoretic mobility shift assays (EMSAs) were initially used to monitor DNA binding of $Isl1_{HD}$ and $Lhx3_{HD}$. EMSAs can be used to observe protein-DNA interactions over a range of affinities, as well as to observe multiple binding events. Additionally, the appearance of a shifted band in a gel can give indications of binding kinetics [326]. However, the EMSA is not a true equilibrium experiment, and so cannot be used for precise measurements of thermodynamics or kinetics. All homeodomain constructs were tested against sequences sourced from promoters that are associated with the appropriate developmental complex *in vivo*: *GSU*, *GA*, and *Hb9* (Table 5.2).

Table 5.2: Binding sequences used for EMSA binding studies. A fluorescein moiety was present at the 5' end of each sequence. Note that all oligonucleotides used were double-stranded. Putative homeodomain binding site is shown in bold.

Name	Gene	Oligonucleotide sequence	Bound by
Isl1GA	*GA* (proglucagon)	ACCGCG**TAATAT**CTG	Isl1[327]
Lhx3GSU	*GSU* (glycoprotein hormone subunit α)	ACTTAGC**TAATTA**ATGTG	Lhx3[102, 328]
M100	*Hb9* (homeobox protein 9)	CGGC**CATTAGCCAAATTA**CGGC	Isl1/Lhx3 in complex [99]

5.5.1 $Lhx3_{HD}$ binds with specificity, but $Isl1_{HD}$ does not

NHD3 was tested for its ability to bind M100, Isl1GA, and Lhx3GSU (Figure 5.11). Of those three potential interactions, NHD3 was seen to bind with high affinity only to the Lhx3GSU oligonucleotide, as shown by the clear shifted band present in the EMSA (Figure 5.11C). There is no clear gel-shift for the other two interactions (Figure 5.11A and B). However, disappearance of the free DNA indicates some binding between NHD3 and Isl1GA and M100. The concentration range at which this occurs, between 500 nM and 5 μM, suggests non-specific binding (discussed further in Section 5.10.1) [329].

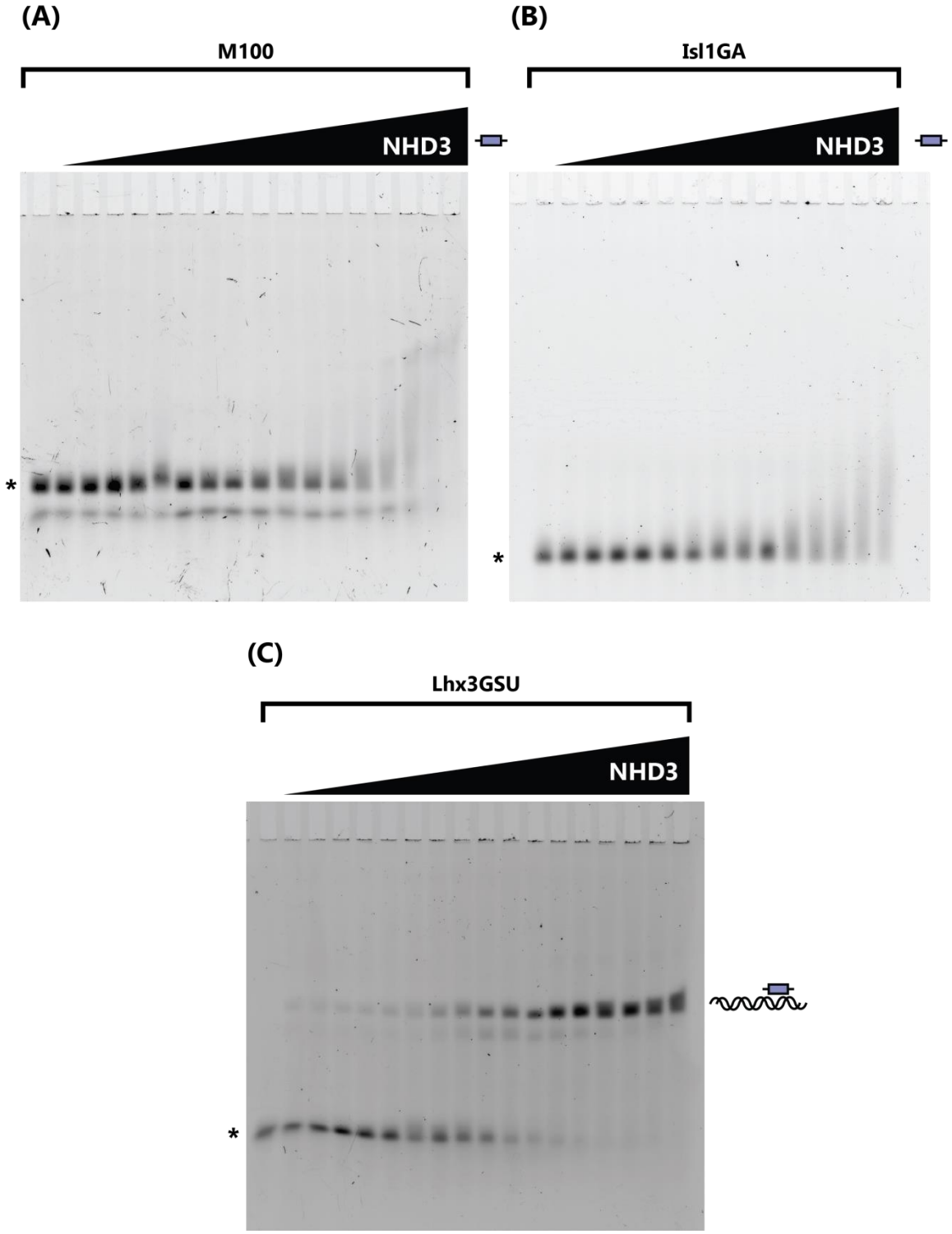


Figure 5.11: Fluorescent EMSA titrations of NHD3. * denotes the unbound oligonucleotide band. Concentrations range from 80 pM-5 µM NHD3, proceeding in a twofold concentration series. The leftmost lane of each set contains no protein. (A) M100. (B) Isl1GA. (C) Lhx3GSU.

Under the same electrophoresis conditions, NHD1 did not show a clear shifted band with any sequence tested (Figure 5.12). While the disappearance of the free oligonucleotide at higher concentrations of protein similarly suggests non-specific DNA binding, it is markedly weaker than the binding seen with NHD3. The behaviour of both proteins binding to DNA was reproducible, indicating a difference in the typical DNA binding behaviour of NHD1 and NHD3.

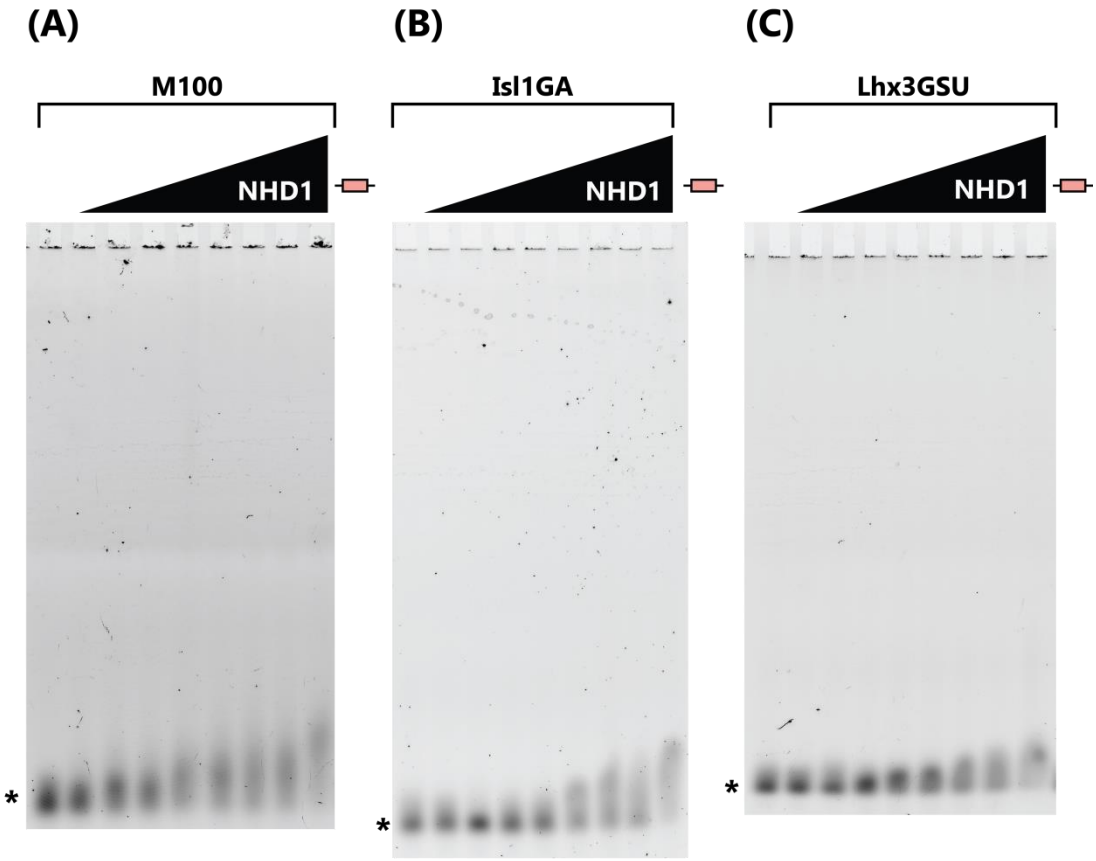


Figure 5.12: Fluorescent EMSA titrations of NHD1. * denotes the unbound oligonucleotide band. Concentrations range from 40 nM-5 µM NHD1, proceeding in a twofold concentration series. The leftmost lane of each set contains no protein. (A) M100. (B) Isl1GA. (C) Lhx3GSU.

5.5.2 $Lhx3_{HD}$ and $Isl1_{HD}$ do not bind co-operatively when separated

There exist in the literature several examples of homeodomains binding cooperatively in transcriptional complexes [330, 331]. To investigate whether this was the case for Isl1 and Lhx3, the purified NHD1 and NHD3 were tested for binding to M100 by EMSA in isolation and in combination (Figure 5.13).

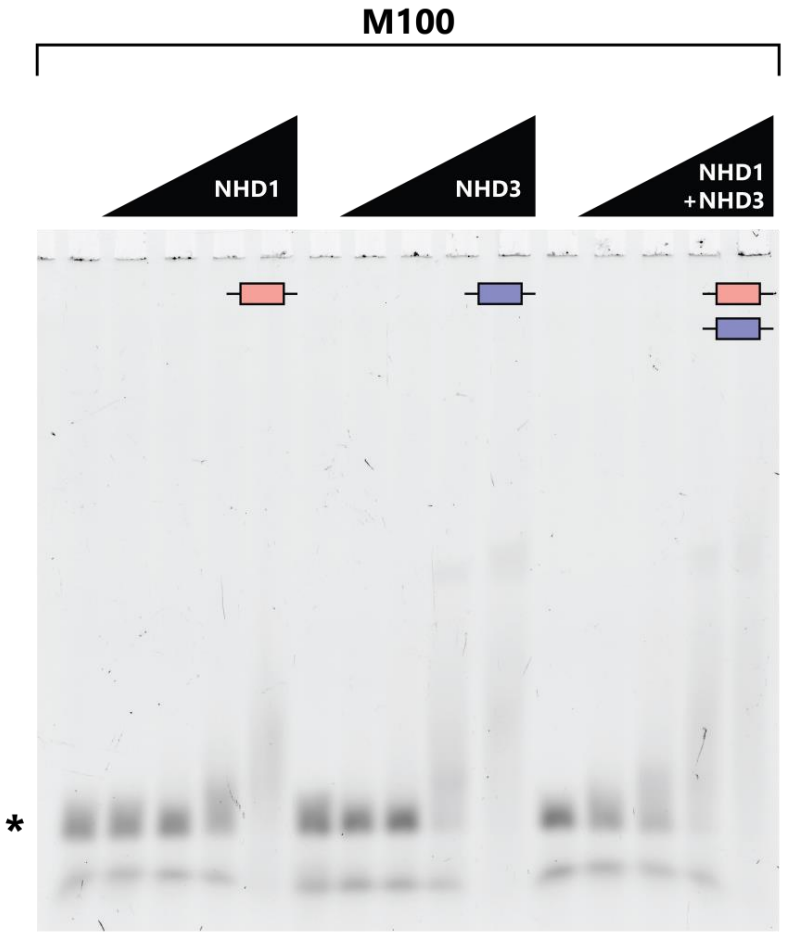


Figure 5.13: Assessing cooperative binding between $Isl1_{HD}$ and $Lhx3_{HD}$ to the M100 sequence. * denotes the unbound oligonucleotide band. The leftmost lane of each set contains no protein. Concentration range proceeds in a tenfold concentration series. From left to right: 0-5 µM NHD1, 0-5 µM NHD3, and 0-5 µM NHD1 and equimolar NHD3 (total 0-10 µM protein).

There was slightly more binding seen in the combined NHD1+NHD3 series, due to the overall higher number of DNA-binding protein molecules in solution, but as this minor increase was not more than the sum of the single homeodomain binding experiments, there was no evidence for cooperative binding in this experiment. Cooperative binding would manifest as a significantly earlier gel-shift of the DNA, and a higher shifted band being visible (representing two homeodomains bound to the same DNA oligonucleotide). This result suggests that $Isl1_{HD}$ and $Lhx3_{HD}$ do not directly influence each other's DNA binding behaviour.

5.6 Seeking an explanation for the observed binding behaviour of $Isl1_{HD}$

In contrast with the above data, published literature has shown Isl1 binding to DNA using EMSAs (further discussed in Section 5.10.2) [104, 327]. However, many of these studies observed binding in the presence of whole nuclear extracts, containing potential binding partners for Isl1 that could confer tighter binding than would be observed using the isolated homeodomain [126, 130]. Other studies have used Isl1 fused to other protein motifs such as TrpE, HA (hemagglutinin), and GST [44, 332], but to date no studies have shown the isolated homeodomain of Isl1 binding directly to DNA. Given that many of these studies used larger portions of Isl1 than just the homeodomain, the binding behaviour of $Isl1_{HD}$ constructs containing additional protein domains was investigated. The binding behaviour of $Lhx3_{HD}$ was also assessed for comparison.

5.6.1 The presence of upstream LIM:LID interaction regions has no direct effect on binding of $Isl1_{HD}$ or $Lhx3_{HD}$

In order to more closely mimic the biological context of these proteins within the ternary Lhx3/Isl1/Ldb1 complex, constructs were produced containing native sequence of Isl1 and Lhx3, from the LIM domains to the homeodomains (Figure 5.2). As these proteins are normally found in complexes, and the LIM domains tend to aggregate in the absence of a binding partner (Section 1.2.2.1), an N-terminal tethered LID sequence from a binding partner was also included, to form LLHD1 and LLHD3 (see Section 5.2.1 for more details). EMSAs were used to assess the binding of LLHD3 and LLHD1 to the same oligonucleotides as were tested with NHD3 and NHD1 (Figure 5.14).

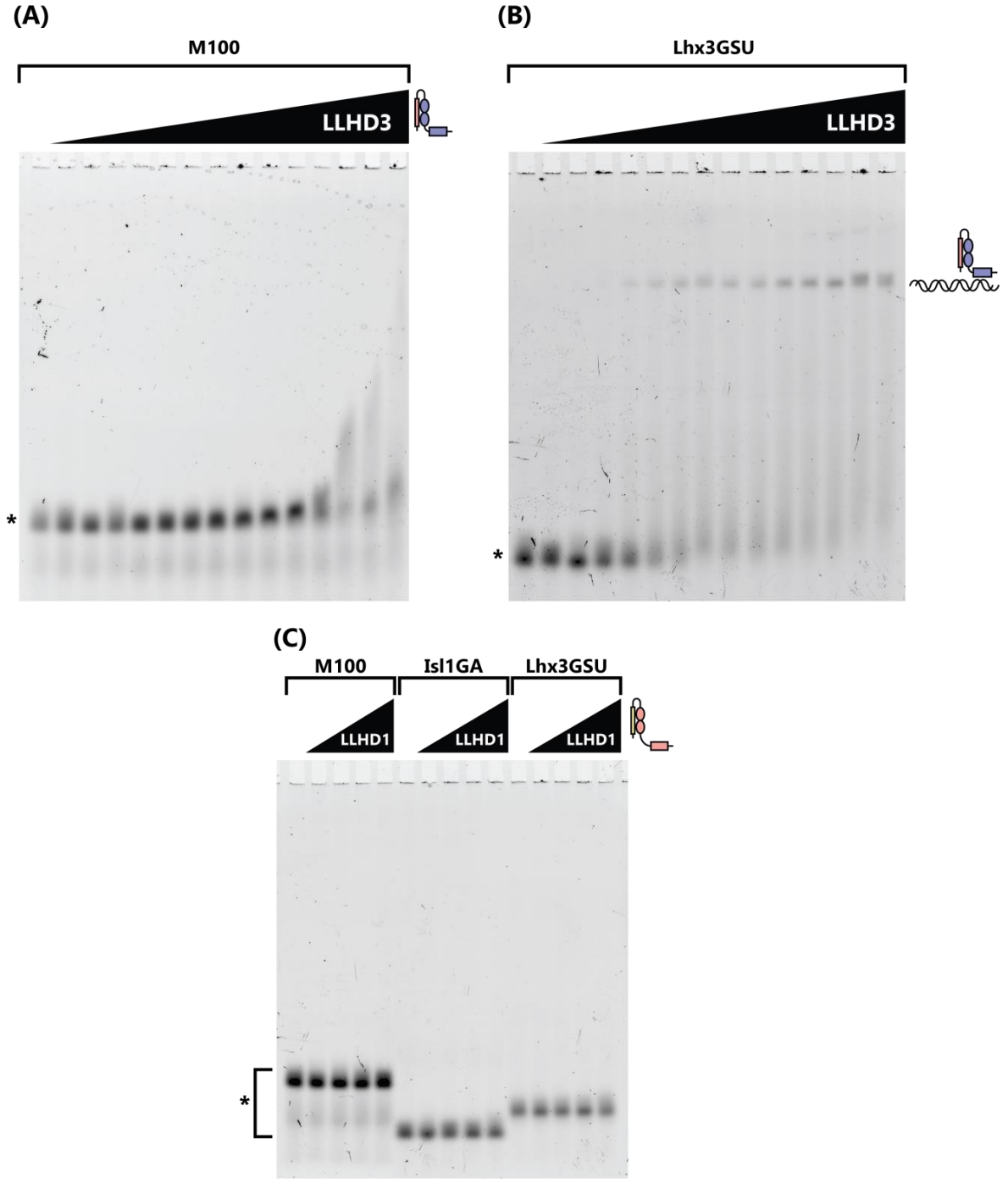


Figure 5.14: EMSA titrations of homeodomain constructs with N-terminal LIM domains. * denotes the unbound oligonucleotide band. The leftmost lane of each set contains no protein. (A) LLHD3 against M100. LLHD3 concentrations range from 80 pM-5 µM, in a twofold concentration series. (B) LLHD3 against Lhx3GSU. LLHD3 concentrations range from 80 pM-5 µM, in a twofold concentration series. (C) LLHD1 against M100, Isl1GA, and Lhx3GSU. LLHD1 concentrations range from 5 nM-5 µM, in a tenfold concentration series.

LLHD3 showed clear binding to Lhx3GSU, and weaker binding to M100, but LLHD1 showed no binding to any oligonucleotide tested. These binding profiles are essentially identical to those seen for NHD3 (Figure 5.11) and similar for NHD1, but with less apparently non-specific binding (Figure 5.12). Thus, the presence of the LIM:LID interaction region does not affect binding of $Lhx3_{HD}$, and if it does affect the binding of $Isl1_{HD}$, it appears to be in an inhibitory manner.

5.6.2 The presence of a dimerising domain can affect DNA binding preferences of homeodomains

The influence of an N-terminal dimerisation domain was also investigated to mimic some of the previous studies that reported DNA binding activity of Isl1 (Section 5.6). This was achieved by omitting the HRV-3C cleavage step during purification of the homeodomain constructs (Section 2.5.3), to obtain GST-HD constructs with the ability to dimerise through GST, which has a dimerisation affinity with a dissociation constant (K_d) of ~ 1 µM [333]. GST-fused NHD1 and NHD3 were again tested against the same three oligonucleotides in EMSAs (Figure 5.15).

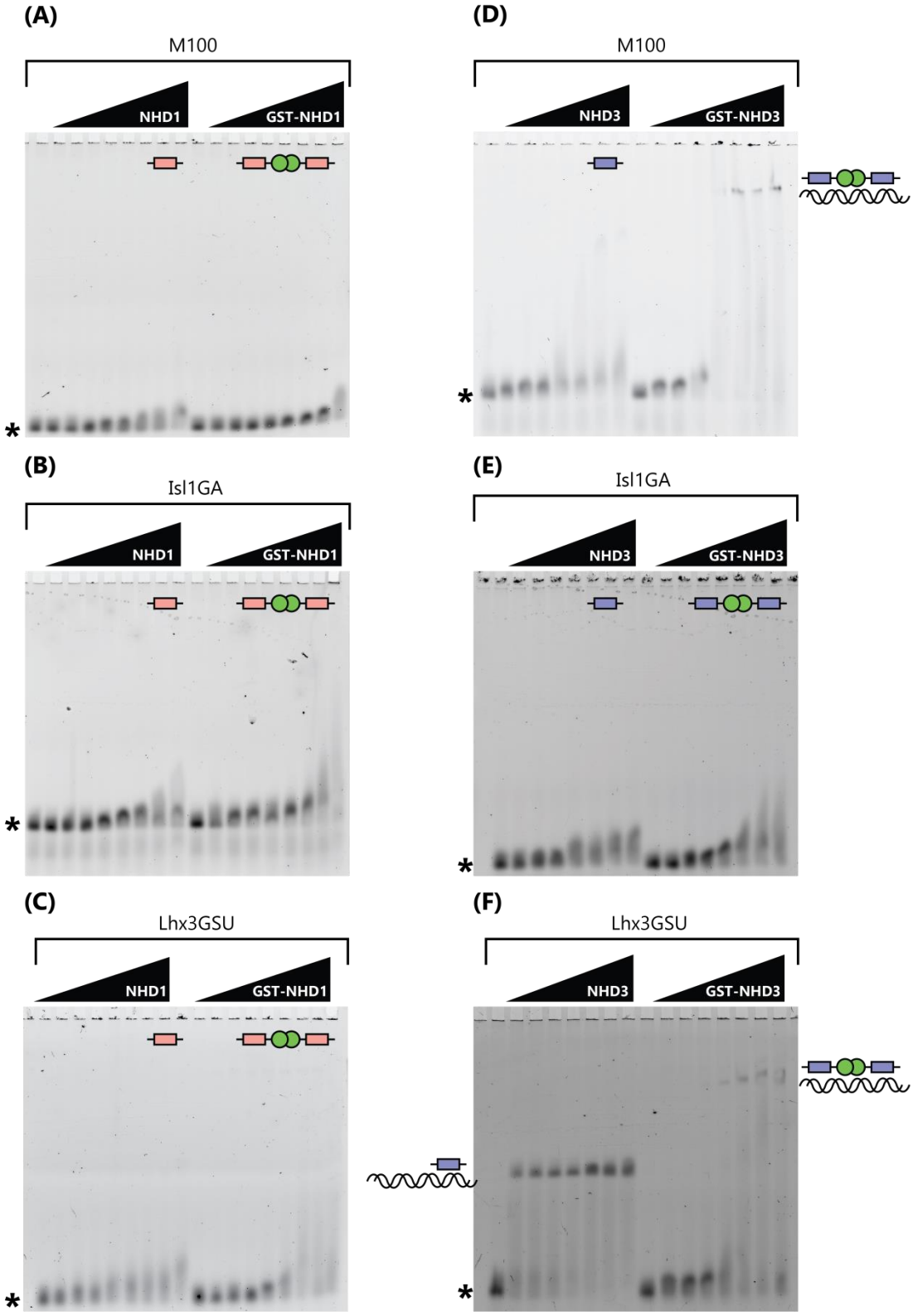


Figure 5.15: Behaviour of GST-fusion NHD1 and NHD3. * denotes the unbound oligonucleotide band. The leftmost lane of each set contains no protein. Protein concentrations range from 40 nM-5 µM, proceeding in a twofold dilution series. (A) NHD1/GST-NHD1 against M100. (B) NHD1/GST-NHD1 against Isl1GA. (C) NHD1/GST-NHD1 against Lhx3GSU. (D) NHD3/GST-NHD3 against M100. (E) NHD3/GST-NHD3 against Isl1GA. (F) NHD3/GST-NHD3 against Lhx3GSU.

DNA binding of GST-NHD1 is slightly tighter than NHD1, but not by a substantial amount (Figure 5.15A-C). In contrast, binding of GST-NHD3 to DNA shows a very different pattern of behaviour to NHD3 (Figure 5.15D-F). Binding to Lhx3GSU was severely reduced, while binding to M100 was increased. Increased binding to M100 could be explained by the oligonucleotides ability to bind two homeodomains simultaneously, whereas Lhx3GSU can only bind one. However, the reduced binding to Lhx3GSU is not readily reconciled.

5.7 Fusion constructs of Isl1 and Lhx3 behave differently to the individual homeodomains

2HDN and 2HDLL were used to further probe the binding behaviour of the homeodomains from Isl1 and Lhx3 when in combination. 2HDN and 2HDLL were both seen to bind to the M100 oligonucleotide with clear shifted bands (Figure 5.16). EMSAs containing 2HDLL presented with a more intense shifted band than those containing 2HDN, but the shifts were evident at very similar protein concentrations.

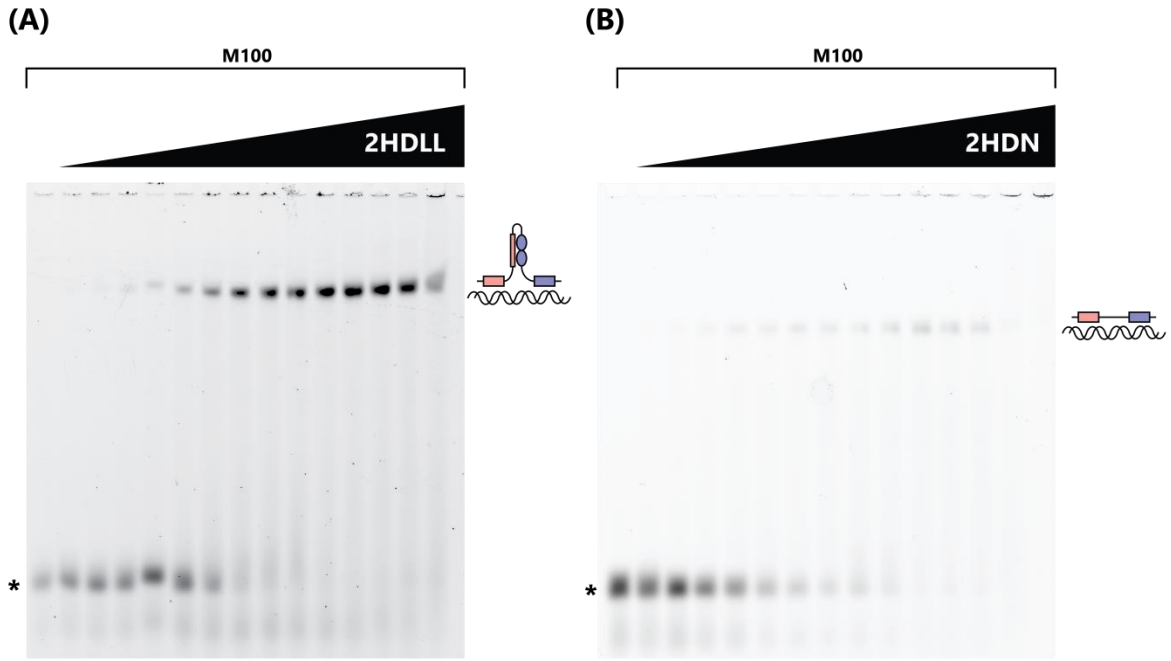


Figure 5.16: Examples of M100 binding by 2HDLL and 2HDN. * denotes the unbound oligonucleotide band. Concentrations range from 80 pM-5 µM protein, proceeding in a twofold concentration series. The leftmost lane of each set contains no protein. (A) 2HDLL titrated against M100. (B) 2HDN titrated against M100.

Some binding was observed for 2HDN and 2HDLL to the Isl1GA and Lhx3GSU sequences (Figure 5.17). However, this binding manifested as a disappearance of unbound oligonucleotide at higher concentrations, indicating weaker, possibly non-specific, binding. 2HD23 bound in the same manner as 2HDN (Figure 5.18). As such, 2HD17 was not tested, as the information it could provide was considered redundant.

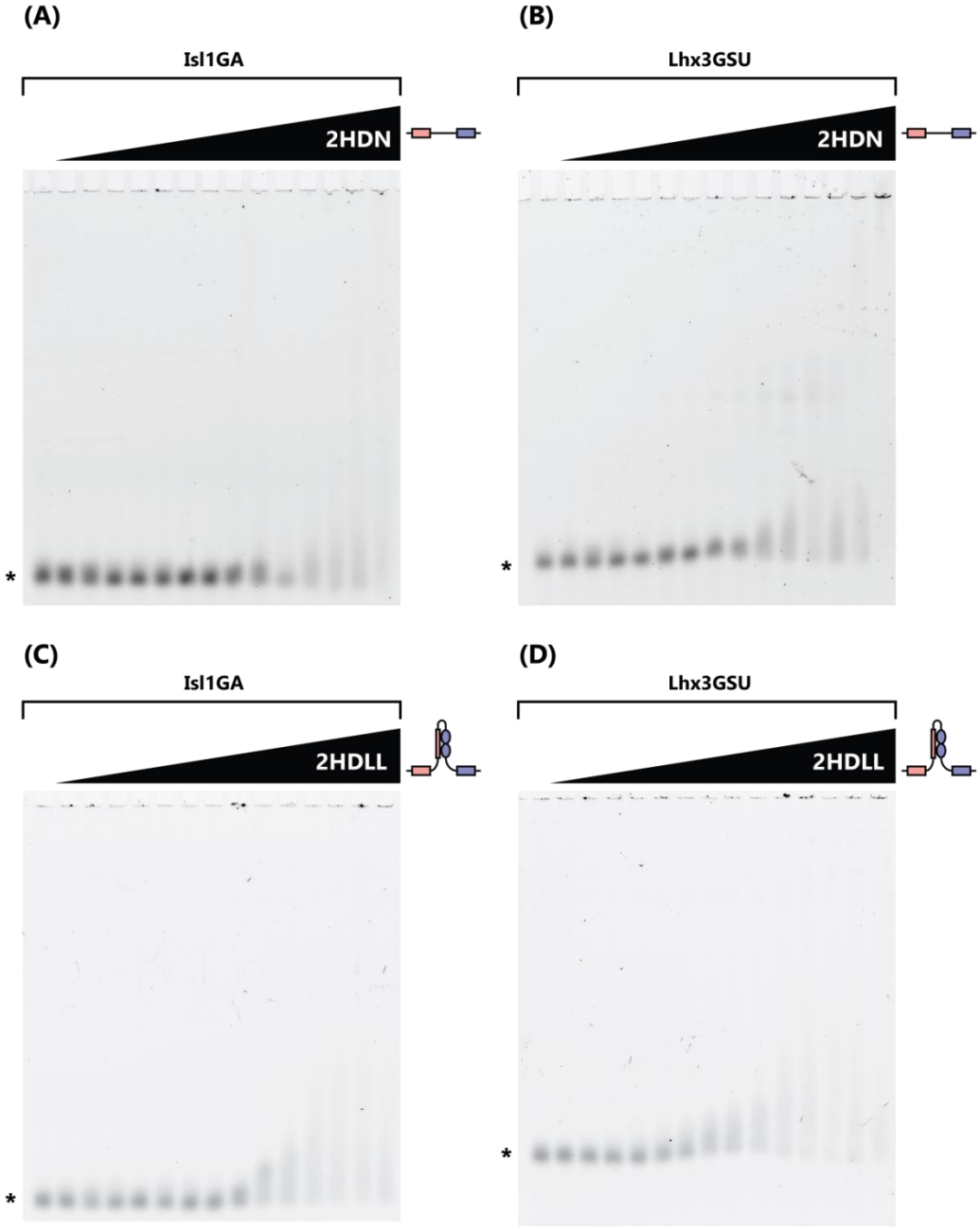


Figure 5.17: Binding to single homeodomain binding site oligonucleotides by 2HDN and 2HDLL. * denotes the unbound oligonucleotide band. Concentrations range from 80 pM-5 µM protein, proceeding in a twofold concentration series. The leftmost lane of each set contains no protein. (A) 2HDN against Isl1GA. (B) 2HDN against Lhx3GSU. (C) 2HDLL against Isl1GA. (D) 2HDLL against Lhx3GSU.

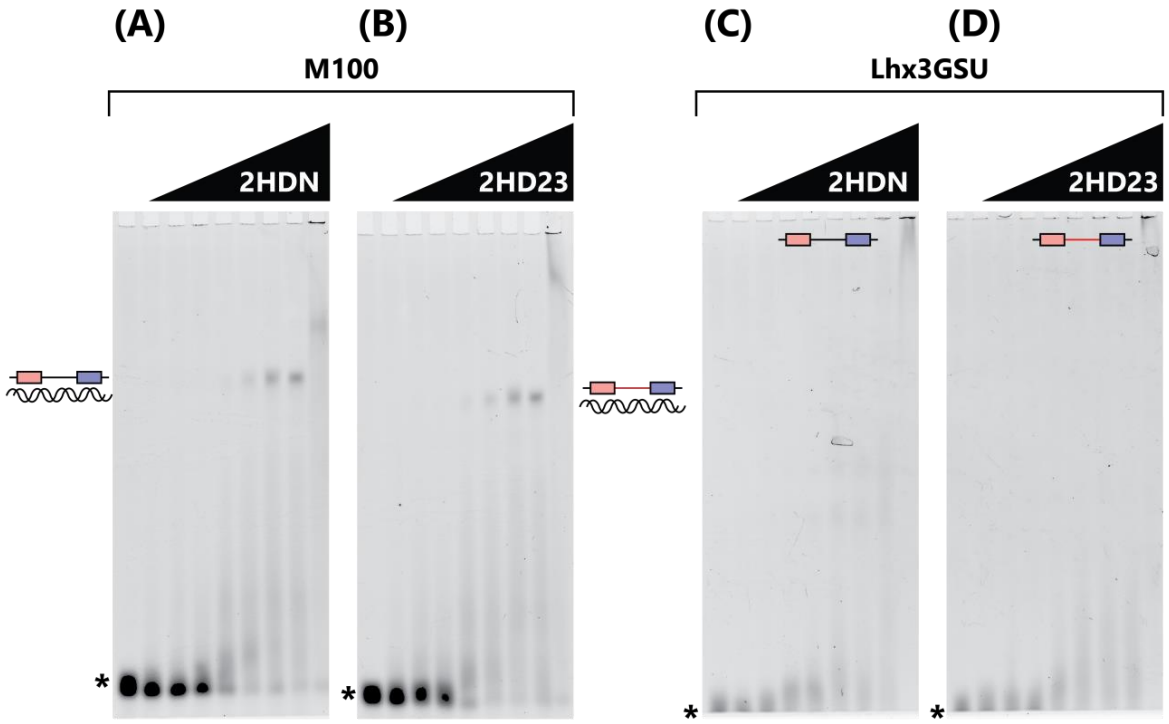


Figure 5.18: Comparison of DNA binding behaviour for 2HDN and 2HD23. * denotes the unbound oligonucleotide band. The leftmost lane of each set contains no protein. Protein concentrations range from 40 nM-5 μM, proceeding in a twofold dilution series. (A) 2HDN against M100. (B) 2HD23 against M100. (C) 2HDN against Lhx3GSU. (D) 2HD23 against Lhx3GSU.

Together, these results show that tethering the two homeodomains results in altered specificity of DNA binding compared to the individual homeodomains. That all fusion homeodomain/homeodomain constructs show the same changes in specificity suggests that the native LIM:LID interaction that brings the two homeodomains into proximity in the context of the ternary complex does not play a direct role in determining the overall specificity of DNA-binding by these homeodomains. However, the slightly tighter binding of 2HDLL compared to the fusions with synthetic linkers does suggest that the LIM:LID regions could have an indirect influence on binding, perhaps by restricting the conformations of the homeodomains to allow the adoption of a structure that facilitates tighter DNA binding.

5.8 Quantitation of DNA-binding affinities

Densitometry was used to obtain an estimate of the dissociation constants for the protein-DNA complexes as described in Section 5.5, Section 5.6, and Section 5.7. Although there are limitations that prevent the technique from being truly quantitative, it can be useful in identifying broad trends. Table 5.3 lists K_ds derived from EMSAs, in the combinations described above.

Table 5.3: K_ds (M) of homeodomain:DNA complexes, as determined from EMSAs (n = 2-3). * denotes combinations that were only measured once and are indicative only. N/A indicates combinations for which dissociation constants could not be calculated.

Protein construct	M100	Lhx3GSU	Isl1GA
NHD1	$3.0 \pm 2 \times 10^{-6}$	$2.0 \pm 0.9 \times 10^{-6}$	$5.9 \pm 0.3 \times 10^{-6}$
LLHD1	N/A	N/A	N/A
GST-NHD1	2×10^{-6} *	$4 \pm 4 \times 10^{-6}$	$6 \pm 5 \times 10^{-6}$
NHD3	$4 \pm 3 \times 10^{-6}$	$9 \pm 3 \times 10^{-9}$	$2 \pm 1 \times 10^{-6}$
LLHD3	1×10^{-6} *	1.5×10^{-8} *	N/A
GST-NHD3	$5 \pm 5 \times 10^{-7}$	$5 \pm 3 \times 10^{-6}$	2×10^{-6} *
2HDLL	$2.4 \pm 0.3 \times 10^{-8}$	$4 \pm 2 \times 10^{-6}$	$7 \pm 5 \times 10^{-7}$
2HDN	$1.6 \pm 0.4 \times 10^{-7}$	$4.4 \pm 0.9 \times 10^{-7}$	$5 \pm 2 \times 10^{-7}$

The majority of binding events for which K_ds could be estimated by this method have a K_d in the micromolar range (Table 5.3). These values are around the limit of measurement by this method and in the range for non-specific protein-DNA interactions [334]. In contrast, the dissociation constants in the nM range were considered to be specific interactions. These high affinity interactions include NHD3 against Lhx3GSU, 2HDLL and 2HDN against M100, and GST-NHD3 against M100. Interestingly, all of these strong interactions include $Lhx3_{HD}$, suggesting that it drives DNA binding to either its own target (Lhx3GSU) as a single protein or to motor neuron genes (M100) as part of Isl1 and Lhx3-containing complexes. It is notable that two copies of $Lhx3_{HD}$ within a dimer (GST-NHD3) showed approximately the same strength of binding to M100 as 2HDLL, the tightest binding fusion construct mimic of the ternary complex. These data also indicate that $Isl1_{HD}$ does not strongly bind DNA in an independent manner, but may contribute to the binding in multiprotein DNA-binding complexes.

5.9 Further investigations into characterising the binding of homeodomain-DNA complexes

Given that data from EMSAs is semi-quantitative, and several interactions appear to lie at or below the limit of detection, attempts were made to obtain additional quantitative binding data.

5.9.1 Trialling the use of circular dichroism to observe changes in the stability of homeodomains upon DNA binding

Far-UV CD can be used to monitor changes in protein stability or folding that result from binding to DNA. As shown in Section 5.4, NHD1 and NHD3 both exhibit spectra consistent with an alpha-helical structure. However, before observing the behaviour of protein-DNA complexes in CD, the behaviour of DNA by itself was assessed, to ensure the signal from the DNA would not interfere with the protein signals, or vice versa.

5.9.1.1 Observing the melting behaviour of DNA in circular dichroism

Spectra recorded of the M100 oligonucleotide from 10-90 °C did not result in significant signal changes in the range of 200-225 nm, meaning the DNA should not interfere with the homeodomain spectra in the same region (Figure 5.19). However, the DNA signal at 247 nm did shift over the temperature gradient, similar to the disappearance of signal seen upon protein unfolding at shorter wavelengths (Section 5.4.1, Figure 5.10). This shift could be exploited as a way to estimate the melting temperature of the DNA in CD - that is, the temperature at which the two strands of DNA dissociate.

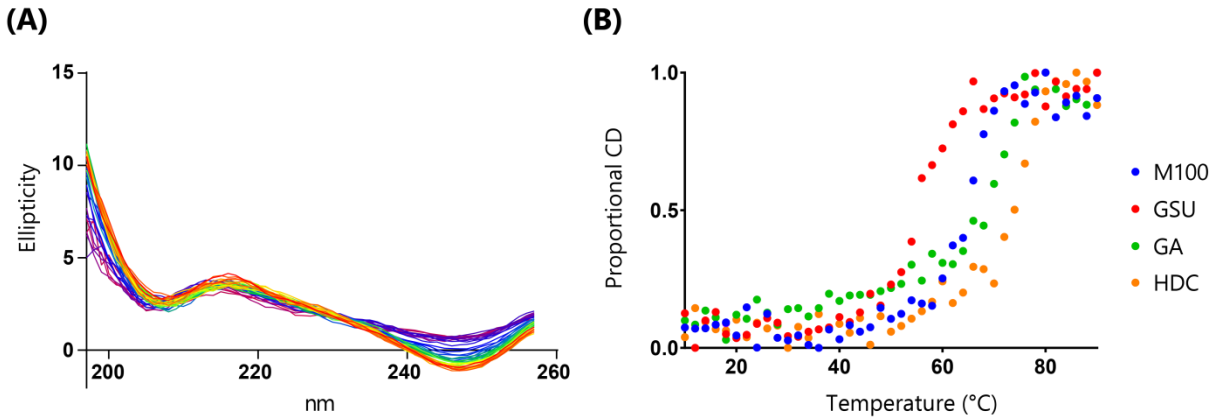


Figure 5.19: Far-UV CD signals of oligonucleotides. (A) Far-UV CD spectra of M100 oligonucleotide, taken every 2 degrees from 10 °C to 90 °C, with lowest temperatures shown in red, and higher temperatures progressing through a rainbow, with highest temperatures in purple. (B) Melt curves of oligonucleotides, using the signal at 247 nm.

The melting temperatures of the oligonucleotides used were estimated, and compared with the theoretical melting temperatures of the sequences used (Table 5.4). A negative control oligonucleotide, HDC, was included. The sequence of HDC was derived from the sequence of Isl1GA, removing the TAAT core homeodomain binding site. The estimated melting temperatures from recorded spectra were considered close enough to the theoretical values to demonstrate that the change in signal at 247 nm does report dissociation of dsDNA.

Table 5.4: Observed and theoretical melting temperatures of DNA sequences used in far-UV CD. Theoretical melting temperatures reported are from the manufacturer of the oligonucleotides (Integrated DNA Technologies; Coralville, IA).

Oligonucleotide name	Sequence	Theoretical melting temperature (°C)	Observed melting temperature (°C)
M100	GCGCATTAGCCAAATTACG CGCGTAATCGGTTTAATGC	64	63
Isl1GA	GCACCGCGTAATATCTGCG CGTGGCGCATTATAGACGC	68	68
Lhx3GSU	ACTTAGCTAATTAAATGTG TGAATCGATTAATTTACAC	54	55
HDC	CACGTGCCGTCAGCGGTAC GTGCACGGCAGTCGCCATG	69	74

5.9.1.2 Observing the foldedness of homeodomains in the presence of DNA

It has been established in the literature that upon binding of DNA, homeodomains become more helical in structure [335, 336]. This was explored with NHD1 and NHD3 (Figure 5.20).

Both curves shift to the right, with a more intense signal at ~208 nm, indicating a more folded conformation.

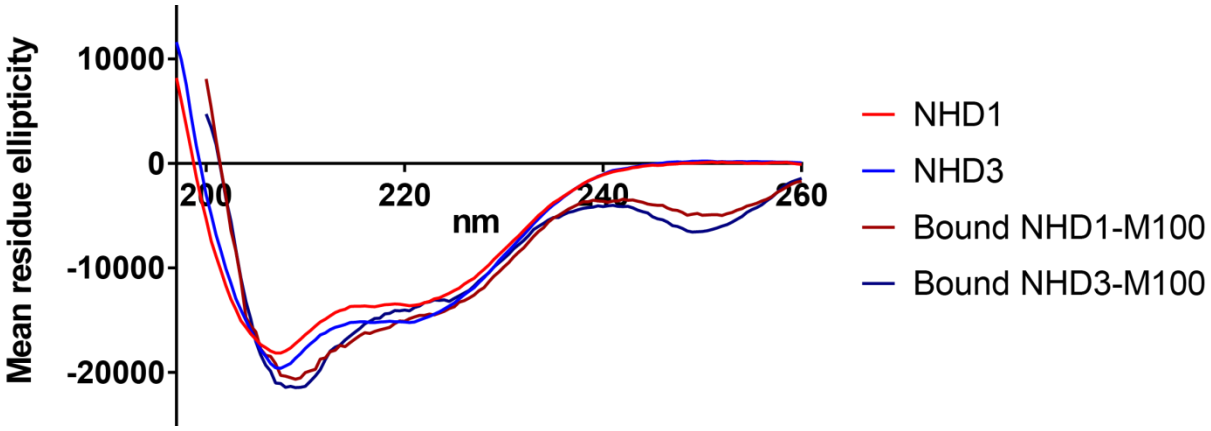


Figure 5.20 CD spectra of homeodomain with and without DNA (M100). Bound spectra were generated by subtracting the spectrum of M100 alone from the spectra of each homeodomain protein in the presence of M100.

5.9.1.3 Homeodomains melt at a higher temperature in the presence of DNA

Combinations of NHD1 and NHD3 with different oligonucleotides were subjected to thermal melts analysed by far-UV CD. For all oligonucleotides tested the homeodomains melted at a higher temperature in the presence of an oligonucleotide (Table 5.5). Notably, this increase in melting temperatures included samples with the negative control oligonucleotide, HDC. Binding to this oligonucleotide should be weak and non-specific in nature.

Table 5.5: Comparison of homeodomain melting temperatures (°C) with and without DNA present. Melting temperatures were estimated as the point of 50% intensity of proportional CD for each sample, measured at 222 nm.

	Oligonucleotide				
	None	M100	Isl1GA	Lhx3GSU	HDC
NHD1	43	68	73	67	73
NHD3	43	68	74	65	66

The melting temperatures of the oligonucleotides was also assessed (Table 5.6). Although the DNA melting temperatures generally increase in the presence of the homeodomains, the changes are generally less than for the proteins upon addition of oligonucleotide.

Table 5.6: Comparison of DNA melting temperatures (°C) with and without homeodomain present. Melting temperatures were estimated as the point of 50% intensity of proportional CD for each sample, measured at 247 nm.

Oligonucleotide	Alone	with NHD1	with NHD3
M100	63	68	68
Isl1GA	68	73	74
Lhx3GSU	55	66	67
HDC	74	78	71

It is notable that the melting temperatures of the oligonucleotide and protein in any combination are extremely similar, except in the case of the HDC oligonucleotide. The similar melting temperatures observed are consistent with stabilisation of both entities by complex formation, and dissociation of the protein-DNA complex leading to unfolding. For HDC, interaction with NHD1 appears to be stabilising for both the protein and the DNA, whereas with NHD3 the interaction is stabilising for the protein but apparently destabilising for the DNA. The mechanisms for such binding and stabilising behaviours are not known.

5.9.2 Quantification of binding affinities

Microscale thermophoresis (MST) and isothermal titration calorimetry (ITC) were attempted to acquire more quantitative information about the binding of homeodomain:DNA interactions.

5.9.2.1 MST

Thermophoresis is a phenomenon in which different particles move differently within a temperature gradient. In microscale thermophoresis experiments, application of a small thermal pulse sets up a temperature gradient, and molecules will move into or out of the hot spot depending on their properties. This results in changes of local concentration which can be measured, as one partner is fluorescently tagged and so can be monitored. By setting up a concentration series with the partner protein, binding affinities can be determined, because the thermophoretic behaviour of the labelled molecule changes according to whether or not it is bound by the titrated partner [337]. For the experiments described here, the DNA was fluorescently labelled (Section 2.6.4, Appendix A).

Unfortunately, this approach did not work well for these homeodomain-DNA interactions. Some sample data is shown for 2HDLL and M100 experiments, but similar problems were experienced with other pairs. For this interaction, which appears to be a high affinity interaction by EMSAs (Section 5.7), the expected outcome was a standard monophasic or biphasic hyperbolic binding curve indicating high affinity specific binding, with or without non-specific binding at higher protein concentrations. Instead any changes in fluorescence were very small and the data were very noisy at higher concentrations (Figure 5.21), with similar results seen when the MST power was increased twofold.

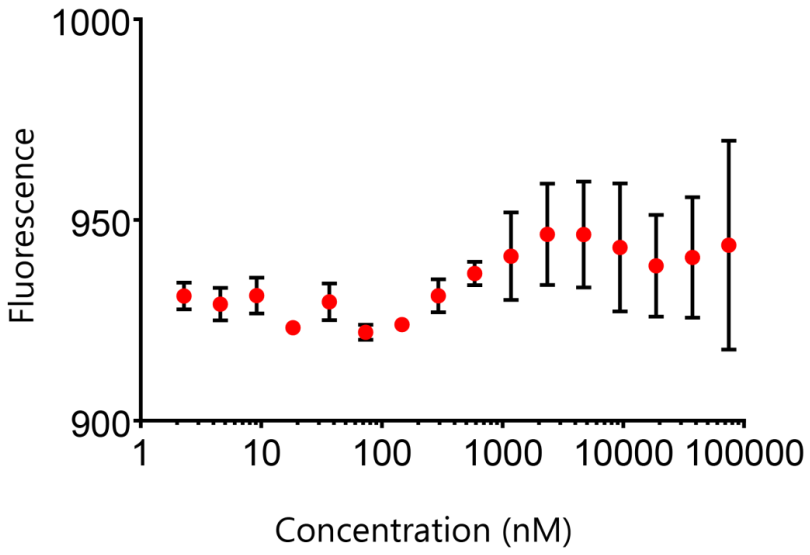


Figure 5.21: Microscale thermophoresis data for the interaction between 2HDLL and M100. Data were taken at 20% MST power. Error bars represent standard deviation, with 5 technical replicates.

One reason for these problems could be the experimental set up. The M100 oligonucleotide was used at a concentration of 50 nM to enable detectable fluorescence. As the 2HDLL:M100 interaction appears to be of high affinity (~50 nM according to EMSA data), this oligonucleotide concentration may be too close to the K_d for valid measurement of the affinity of the interaction. However, similar results were also seen for apparently lower affinity interactions.

One quality control check for this technique is to record the fluorescence of all the samples before thermophoresis. As all samples contain the same amount of DNA (and so fluorescence label), they should have the same intensity of fluorescence, but here there was significant variation observed (Figure 5.22). The same effect was noted for sample capillaries with

different coatings (which can resolve this problem for some molecules [338]). In general, there was lower fluorescence at the higher concentrations, so it is possible that binding of 2HDLL to the M100 oligonucleotide could have quenched the fluorescence of the oligonucleotide, influencing the thermophoresis measurements taken. This was unexpected, as these oligonucleotides are identical to those used for fluorescent EMSAs and the fluorescent tag is not immediately adjacent to the binding site. However, due to these difficulties, it was concluded that MST was not a suitable technique for studying this system.

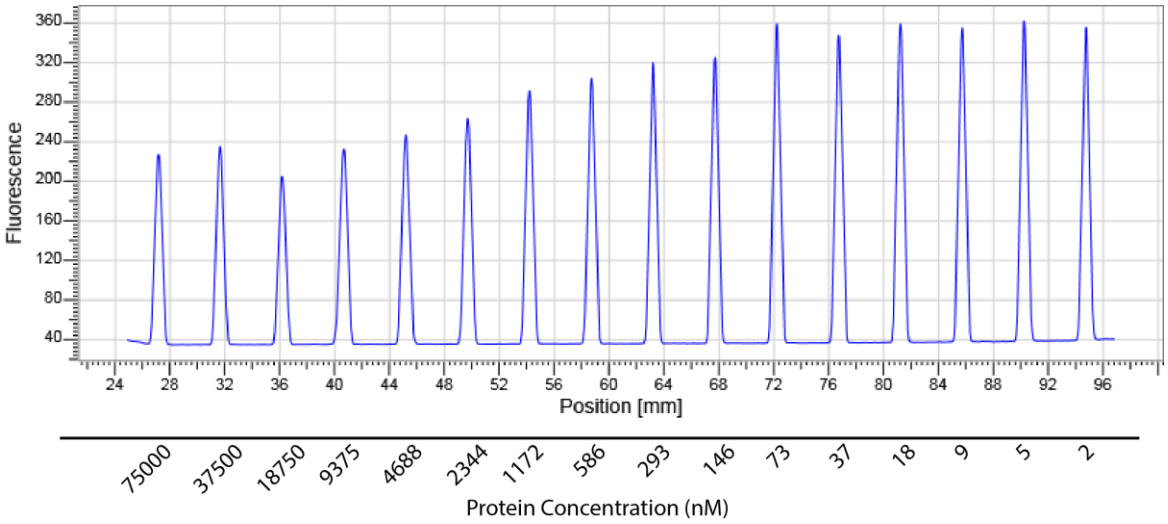


Figure 5.22: Capillary fluorescence scan of 2HDLL:M100 samples prior to thermophoresis measurements. Concentration range proceeds in a twofold dilution series from 75 µM (leftmost) to 2.3 nM (rightmost). 50 nM M100 oligonucleotide was present in each sample.

5.9.2.2 ITC

ITC measures the heat released (or absorbed) from binding events. The data gathered from this technique can give precise thermodynamic information including K_ds, binding stoichiometry, and enthalpies of binding. ITC titrations were performed using the homeodomains from each of Lhx3 (NHD3) and Isl1 (NHD1) against their respective target binding sequences, as well as against M100 and the negative control oligonucleotide HDC (sequences listed in Table 5.4). Note that these oligonucleotides were not fluorescently labelled.

These experiments showed some promise. However, these data could not be fitted to reasonable titration curves, as the fits obtained reported aberrant stoichiometries of binding. Optimisation of experimental conditions is required in order to obtain data of sufficient

quality for quantitative analysis. In particular, as this technique is very sensitive to concentration, more careful estimates of concentrations for both protein and DNA could help to resolve the observed inconsistencies associated with stoichiometry. Additionally, higher concentrations of materials are required for more complete titrations for better estimations of binding affinities (Figure 5.23). These optimisation experiments could not be carried out during the timeframe of this thesis. Nevertheless, qualitative assessment of these preliminary data based on the shapes of the binding curves indicates higher levels of binding by $Lhx3_{HD}$ than $Isl1_{HD.}$

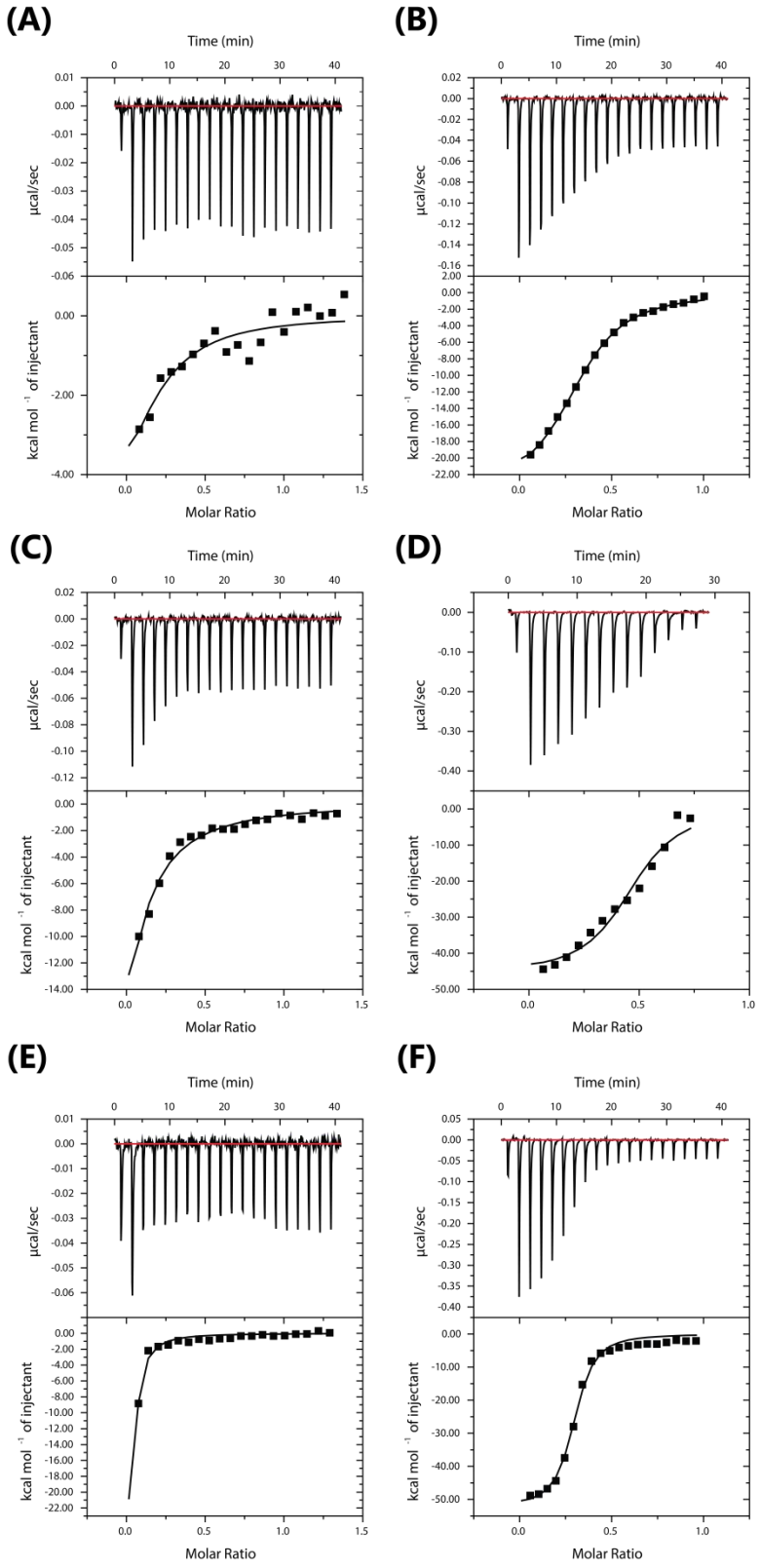


Figure 5.23: ITC titrations of homeodomains against binding sequences. (A) NHD1 against the negative control HDC; (B) NHD3 against the negative control HDC; (C) NHD1 against Isl1GA; (D) NHD3 against Lhx3GSU; (E) NHD1 against M100; (F) NHD3 against M100. Note that these curves are not scaled to each other.

5.10 Discussion

The work described here shows two very similar DNA-binding domains, with markedly different DNA-binding behaviours. Whereas $Isl1_{HD}$ and $Lhx3_{HD}$ show similar properties in terms of folding and stability, they exhibit very different DNA-binding properties. Isl1 was seen to bind DNA only in a weak, probably non-specific manner, but Lhx3 bound with much higher affinity to a specific sequence (Lhx3GSU). The DNA-binding behaviour of Isl1 is unusual for a homeodomain, and bears closer examination.

5.10.1 Comparing the observed binding behaviour to the reported behaviour in the literature

The DNA-binding specificity of homeodomains has been extensively studied, both in individual cases and in a high throughput manner [46, 63, 69, 339-342]. These studies have been successful in identifying a plethora of consensus sequences for specific homeodomains, as well as providing quantitative data on the homeodomain-DNA interaction. Specific binding to DNA is normally observed with affinities in the nanomolar range, with non-specific binding being with approximately micromolar affinity [343, 344]. These measurements support the EMSA data presented in Sections 5.5-5.8, in which NHD3 and 2HDLL were seen to bind to one target sequence with nanomolar affinity, and all other binding combinations were observed with affinities in the micromolar range.

Although cooperative binding was not observed between NHD1 and NHD3, many instances of cooperative binding between different homeodomain proteins have been observed, including some that resulted in altered specificity [345-349]. Many of these examples involve proteins from the Pre-B-cell leukaemia transcription factor (Pbx) or Meis family [350-354]. These proteins contain homeodomains, but are generally involved in DNA-binding only in the presence of an additional homeodomain [349, 355, 356]. Similar to the behaviour seen with Isl1 and Lhx3, the combination of a homeodomain protein and a Pbx/Meis protein in complex shows different DNA-binding preferences to the individual proteins [350]. There is limited data to show if Pbx or Meis proteins can bind to DNA in isolation. However, the mechanism of action of Pbx/Meis-homeodomain complexes has been explored, and a short YPWM motif in the partner protein was shown to interact with the Pbx/Meis homeodomain, forming the basis for complex formation and the change in binding specificity [357, 358]. No

such motif can be found in the sequence of either Isl1 or Lhx3, meaning that though the behaviour of these proteins may be similar, the mechanisms are probably different.

Another instance of a homeodomain with markedly different DNA-binding behaviour is the transcriptional regulator homeodomain-only protein (Hopx) [359]. Hopx is a small protein that consists almost entirely of a homeodomain fold [360]. This homeodomain has not been shown to bind to DNA, but was shown to act to regulate expression of genes in many contexts, including during development [360-364]. The inability of Hopx to bind DNA arises from its sequence, which has a very low level of conservation compared to other mammalian homeodomains [360]. Specifically, Hopx is missing key DNA-binding residues at positions 2, 4, 51, 53, and 55 of the homeodomain fold. Whereas Isl1 may behave like Hopx in terms of DNA-binding, it does not share these sequence discrepancies (discussed further in Section 5.10.3).

5.10.2 Isl1 binding in the literature

The data presented here are in contrast with published literature on Isl1. Several papers report *in vitro* specific binding of Isl1 to DNA and showed evidence of this binding in EMSAs [44, 104, 126, 130, 136, 332]. It should be noted that the work presented here is the only study that used the isolated homeodomain of Isl1. All other studies used fusion proteins containing Isl1, and/or used un-purified full length Isl1 in complex mixtures of proteins such as cell lysates. In those environments, there is the potential for Isl1 to interact with other proteins. Any resulting DNA-binding does not necessarily reflect the properties of Isl1 alone, as Isl1 forms many protein-protein interactions with other DNA-binding proteins as part of its function (Section 1.5). This is especially the case with experiments performed using full length Isl1 produced in reticulocyte lysate, as reticulocytes are known to express Ldb1, a binding partner of Isl1 that could bridge interactions between Isl1 and other DNA-binding proteins [365, 366].

Nevertheless, other indirect evidence does support the data in this chapter that suggests that isolated $Isl1_{HD}$ is unable to bind DNA strongly. Many transcription factor binding site databases now exist, providing a valuable resource for investigating transcription factor protein-DNA interactions. These databases either lack, or have poor quality, binding data for Isl1, but have plentiful data for Lhx3 and most other LIM-homeodomain proteins [3, 367-

370]. Although those studies do not address negative data, a consistent absence of Isl1 in the databases implies that weak or no binding is the norm for Isl1 in the absence of binding partners.

5.10.3 Checking $Isl1_{HD}$ sequence against overall homeodomain conservation

Homeodomain-DNA interactions have in general been well characterised [68, 340, 342, 371]. It is possible to compare the amino acid sequence of Isl1 with that of other homeodomains to try and detect sequence differences that may explain its unusually poor DNA-binding behaviour.

Many residue positions in the canonical 60 residue homeodomain fold are highly conserved (Figure 5.24). This is especially seen in the third alpha-helix, comprising residues 41-58, which binds the major groove of DNA (Section 1.2.2.3). Nearly half (28 from ~60) of the positions in the homeodomain sequence have one amino acid identity that occurs in 50% or more sequences. These residues can be classified as contributing to DNA-binding, or to protein fold stability (Table 5.7).

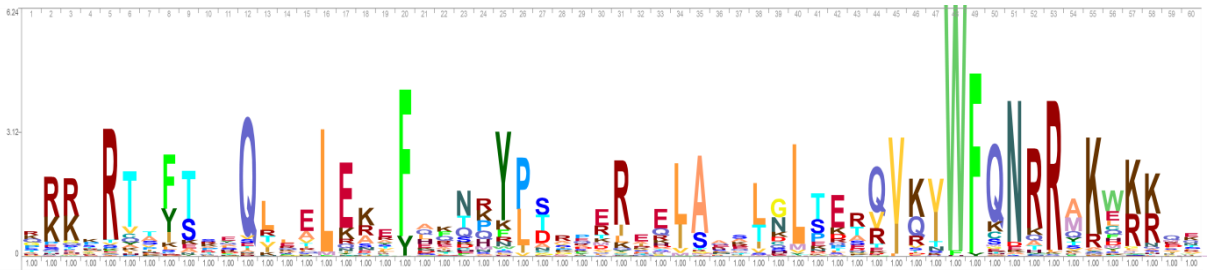


Figure 5.24: Sequence conservation across 213 human homeodomains. This figure was generated using Skylign, inputting 213 human homeodomain sequences from the Homeodomain Resource Databank that were aligned using Clustal Omega [67, 372, 373].

Table 5.7: Conserved amino acids in the homeodomain fold. Prevalence was calculated from a set of 213 60-residue homeodomain sequences [69, 374].

Position	Isl1 amino acid	Conserved amino acid	Prevalence	Position	Role
49	F	F	98%	Helix 3	Hydrophobic core
48	W	W	98%	Helix 3	Hydrophobic core
16	L	L	96%	Helix 1	Hydrophobic core
53	R	R	95%	Helix 3	DNA interacting
51	N	N	90%	Helix 3	DNA interacting
40	L	L	88%	Flexible loop	Hydrophobic core
20	Y	F	87%	Helix 1	Hydrophobic core
5	R	R	86%	N-terminal tail	DNA interacting
12	Q	Q	84%	Helix 1	Unclear
55	K	K	81%	Helix 3	DNA interacting
35	V	A	76%	Helix 2	Unclear
52	K	R	74%	Helix 3	Salt bridge
45	I	V	73%	Helix 3	Hydrophobic core
25	R	Y	70%	Flexible loop	DNA interacting
17	R	E	69%	Helix 1	Salt bridge
50	Q	Q	68%	Helix 3	DNA interacting
31	K	R	66%	Helix 2	DNA interacting
34	L	L	66%	Helix 2	Hydrophobic core
6	T	T	63%	N-terminal tail	Unclear
38	T	L	56%	Helix 2	Hydrophobic core
9	N	T	56%	Helix 1	Unclear
57	K	K	55%	Helix 3	DNA interacting
8	L	F	53%	N-terminal tail	Hydrophobic core
26	P	P	52%	Flexible loop	Introduces turn
42	P	E	52%	Helix 3	Salt bridge
13	L	L	51%	Helix 1	Unclear
44	V	Q	50%	Helix 3	Unclear
2	T	R	50%	N-terminal tail	DNA interacting

Of the 28 conserved residues listed, Isl1 differs from the conserved residue primarily in regions that influence fold stability [374]. Notably, residue 25 in Isl1 is a tyrosine, whereas it is arginine in the majority of other homeodomains. This position is classified as lying in a DNA-interacting loop, so could influence DNA binding in Isl1. However, it cannot fully explain the DNA-binding behaviour of Isl1, as Lhx3 also has a tyrosine residue at this position. Overall, there are no residues where the sequence in Isl1 varies from the canonical or highly conserved sequences that would account for weakened DNA binding.

5.10.4 Potential non-canonical roles for $Isl1_{HD}$

Alternative roles for $Isl1_{HD}$ are possible. Homeodomains were recently found to be involved in several cellular functions beyond canonical DNA binding. These include binding of methylated DNA and acting as nuclear localisation sequences, as detailed below.

5.10.4.1 Methylated DNA binding

A recent paper highlighted the potential of homeodomains to bind methylated DNA, offering a potential novel role for Isl1 [103]. However, upon closer observation of the data presented this is thrown into question. In the publication, several residues were identified as being crucial for this interaction, specifically those in positions 43, 47, and 54 of the canonical homeodomain fold. For binding methylated DNA the following residues are favoured: arginine at position 43, a valine or isoleucine at position 47, and a valine or alanine at position 54. These are all relatively common residues: of the 213 sequences analysed above, 30% of homeodomains contained R43, 80% contained a valine or isoleucine at position 47, and 40% contained either a valine or an alanine at position 54. Of the sequences analysed 10% fulfilled all three of these requirements, with this figure expanding to 53% for sequences containing two or more requirements.

These numbers suggest that many homeodomains may have the potential to bind methylated DNA (assuming the identified residues do direct such binding), in direct contradiction of the findings of the study, which showed that most homeodomains cannot bind methylated DNA. It is worth noting that even with structural data providing a very clear picture of sequence preferences for homeodomain-DNA binding, it is still not possible to pinpoint the interactions responsible for dictating the specificity of these interactions.

5.10.4.2 Homeodomains as nuclear localisation sequences

Homeodomains can act to direct cell localisation, in addition to their DNA binding capability. Examples have been found of homeodomains containing both nuclear export sequences and nuclear localisation sequences [297, 375, 376]. Helix 3 of the canonical homeodomain contains a sequence (RRMKWKK) that could function as a nuclear localisation sequence [297]. Given that this region is highly conserved across all homeodomain containing proteins, it is possible that the homeodomain of Isl1 can target the protein to the nucleus. However, the full Isl1 protein is comparatively small, and should be able to diffuse into the nucleus without

the aid of nuclear transport chaperones [377]. Additionally, the sequence at the beginning of the LIM domains of Isl1 contains a canonical nuclear localisation sequence, which is more likely to play a role in the localisation of Isl1 (See Section 4.2.2.8).

5.10.5 Consequences of a weak DNA-binding Isl1

It is possible that isolated Isl1 does not bind DNA with an affinity sufficient for transcriptional regulation. Rather, it may be necessary for Isl1 to interact with other DNA-binding proteins, through its LIM and LID regions, to target DNA. Interestingly, of the LIM-HD proteins, only Isl1 and its close homolog Isl2 contain identified LIDs. Their role in the ternary complex may be to modify the properties of Lhx protein, with small contributions from the homeodomain in the Isl protein contributing to DNA binding. This aspect of ternary complex formation and DNA binding is explored in the following Chapter.

6 Structural studies on the 2HDLL:M100 complex

Although the qualitative DNA binding data from Chapter 5 show there are differences in DNA binding behaviour between the individual homeodomains and the homeodomains in ternary complex mimics, they do not fully explain the differences. In this chapter, attempts were made to gather more detailed structural information about the 2HDLL:M100 complex, to better understand how Lhx3 and Isl1 bind their target DNA sites.

The whole 2HDLL:M100 complex was chosen for study, as structures are already available for $Isl1_{HD}$ and $Isl1_{LID}$:$Lhx3_{LIM}$. Additionally, DNA-binding data indicate that the binding behaviour of the Lhx3 and Isl1 homeodomains alters when the two are in close proximity, such as in a protein fusion like 2HDLL. Attempts were not made to crystallise isolated 2HDLL due to expected flexibility of the linker region.

6.1 The 2HDLL:M100 complex is monodisperse in solution

Size exchange chromatography coupled with multi-angle laser light scattering analysis (SEC-MALLS) was used to establish the stoichiometry of the 2HDLL:M100 complex. This technique also provides information about sample heterogeneity, which could hamper structure determination efforts.

2HDLL and M100 were mixed in at 100 µM in a 1:1 ratio and subjected to SEC-MALLS (Figure 6.1). The elution profile shows a major peak at ~23 mins and two smaller peaks at ~34 and 38 min. Only the first peak was of a high enough intensity to register scattering or UV absorbance (Figure 6.1B). It showed strong absorbance at both 280 nm and 260 nm, indicating that both protein and DNA are present. The difference in signal intensity between the peaks indicates that the majority of the protein and DNA in solution are in complex.

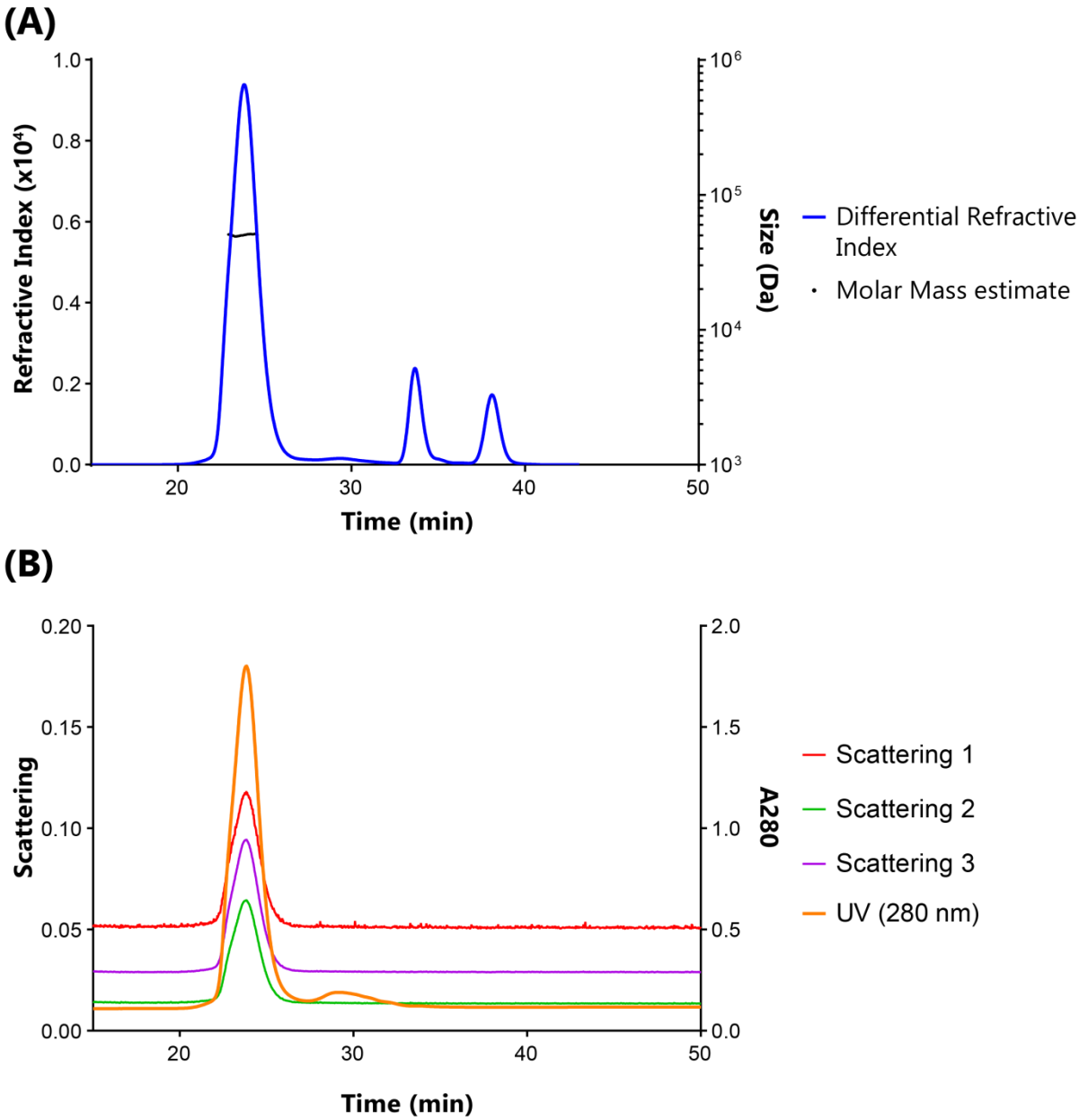


Figure 6.1: Elution profiles of 2HDLL:M100 from SEC-MALLS. SEC-MALLS was conducted at 1 mL/min, room temperature, in 20 mM sodium phosphate monobasic/dibasic pH 7.4, 100 mM NaCl, 1 mM DTT. (A) Measurements of differential refractive index, as well as molar mass estimate of major peak. (B) UV absorbance and scattering readings from 3 different angles.

Using Astra (Wyatt Technology) analysis to estimate the mass of the major species eluting from SEC-MALLS yields a value of 50-51 kDa, which is in good agreement the expected size of 49 kDa for the 2HDLL:M100 complex with 1:1 stoichiometry, as the error of estimating solution weight average molecular weight by this approach is ~10% [378]. The smaller peaks may represent uncomplexed protein or DNA, but due to the low signal intensity this could not be investigated. The lack of higher molecular weight peaks indicates

that the complex itself is not forming higher molecular weight species (higher order oligomers or aggregates). These data indicate that the protein-DNA complex forms a monodisperse complex with at least moderate affinity (greater than K_d ~10^{-7} M) under near-physiological buffer conditions, making it a suitable candidate for structural studies [379].

6.2 Attempting to solve the structure of 2HDLL:M100 through crystallography

Several different oligonucleotides were used in attempts to crystallise the 2HDLL:M100 complex (Table 6.1). Each contained the M100 sequence with a one-base overhang, with the exception of M100c20b, which was blunt-ended. Crystallisation screens were set up with one well for each buffer condition containing 2HDLL:M100, and a second containing only M100, for ready identification of crystals containing only DNA.

Table 6.1: Oligonucleotide sequences used for crystallography trials.

Name	Sequence (5'-3')
M100c21	ACGCCATTAGCCAAATTACGC GCGGTAATCGGTTTAATGCGT
M100c20	ACGCATTAGCCAAATTACGC GCGTAATCGGTTTAATGCGT
M100c20b	CCGCATTAGCCAAATTACGC GGCGTAATCGGTTTAATGCG
M100c14	TCATTAGCCAAATTA GTAATCGGTTTAATT

6.2.1 Initial crystallisation condition screening

Initially screening was conducted using M100c21, as this oligonucleotide most closely mimicked the sequence used in DNA binding studies. The one-base overhang in the design can be useful in helping DNA to form extended structures that encourage regular crystal formation [334]. Unfortunately, while nucleation (as evidenced by the formation of microcrystals or spherulites) was observed in protein-DNA wells in several conditions, these species dissolved upon contact, or were too small for recovery (Figure 6.2). In all cases where nucleation was observed in protein-DNA wells, it was also observed in the DNA-only condition, suggesting that DNA was the entity undergoing nucleation. The species observed for the protein-DNA samples were highly birefringent, which is consistent with nucleic acid crystallisation [380, 381]; however, no evidence could be obtained to determine whether or not protein was present.

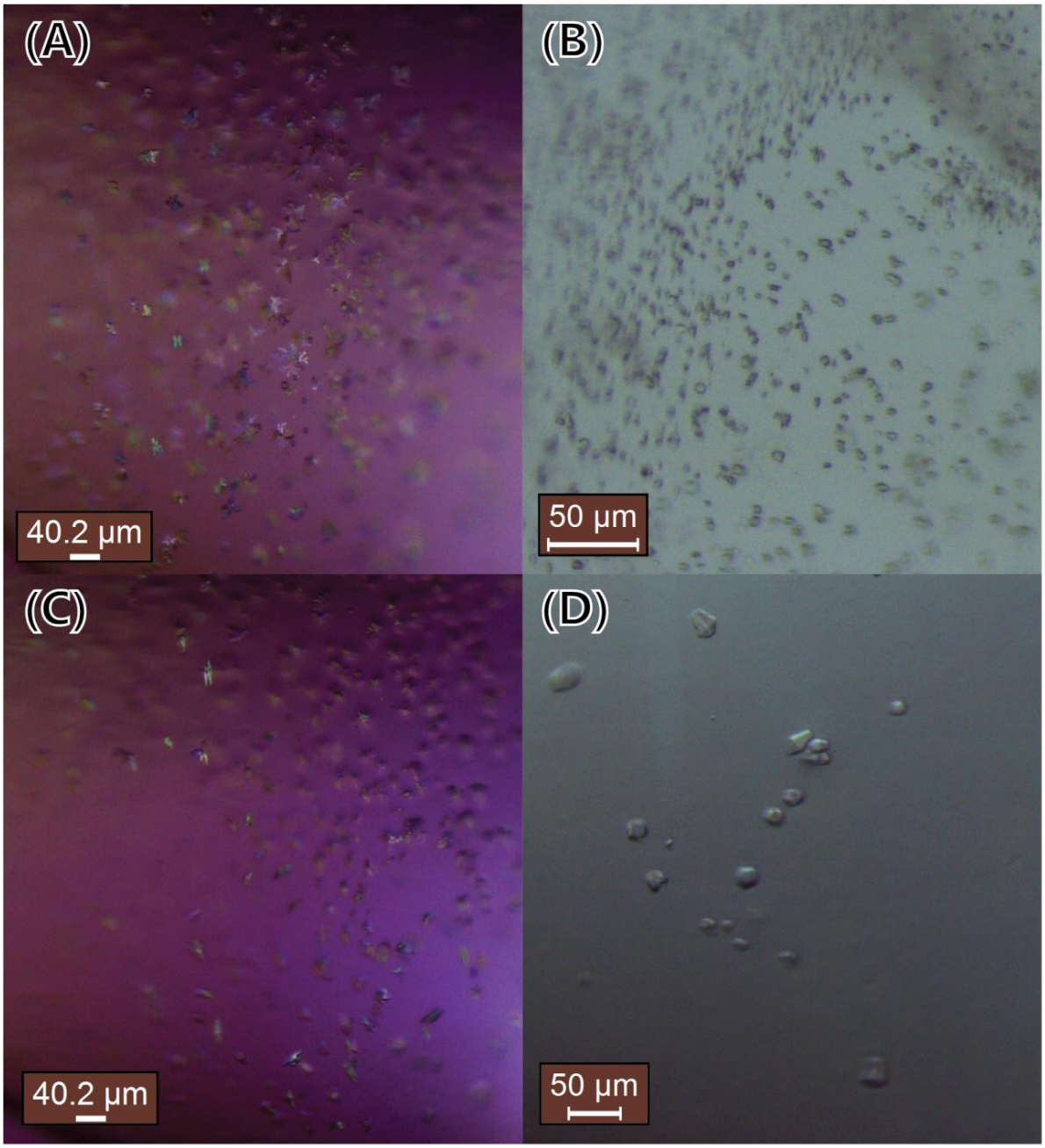


Figure 6.2: Examples of nucleation in 2HDLL:M100c21 wells. (A) and (B) from Morpheus screen; (C) from Index HT; (D) from JCSG+ HT.

Different versions of the M100 oligonucleotide were then designed - the original M100 oligonucleotide was shortened into one version 20 bases long (M100c20), and one version 14 bases long (M100c14), based on oligonucleotide lengths that have previously been successful in giving homeodomain-DNA structures by X-ray crystallography (Table 6.1) [65, 330, 382, 383]. A one-base overhang was retained in both oligonucleotides.

Both 2HDLL:M100c14 and 2HDLL:M100c20 crystal screens resulted in conditions with evidence of protein-DNA nucleation (Figure 6.3). However, screening with M100c14 did not result in any 2HDLL:M100c14 crystals. Overall there were relatively few conditions that showed evidence of nucleation, and those that did were consistent with DNA-only crystals, as crystals observed in 2HDLL:M100c14 drops looked the same as those in M100c14-only drops.

The initial crystallisation trials for 2HDLL:M100c20 were more promising. Many conditions showed protein-DNA nucleation, but while the majority of these were consistent with DNA-only crystals in the corresponding DNA-only wells, several conditions gave rise to likely protein-DNA crystals (Figure 6.3). While there was still nucleation or crystallisation observed in the corresponding DNA-only drops for these conditions, in all cases there were differences in morphology observed between the two wells. Before pursuing optimisation, tests were carried out to establish whether protein was present in these crystals.

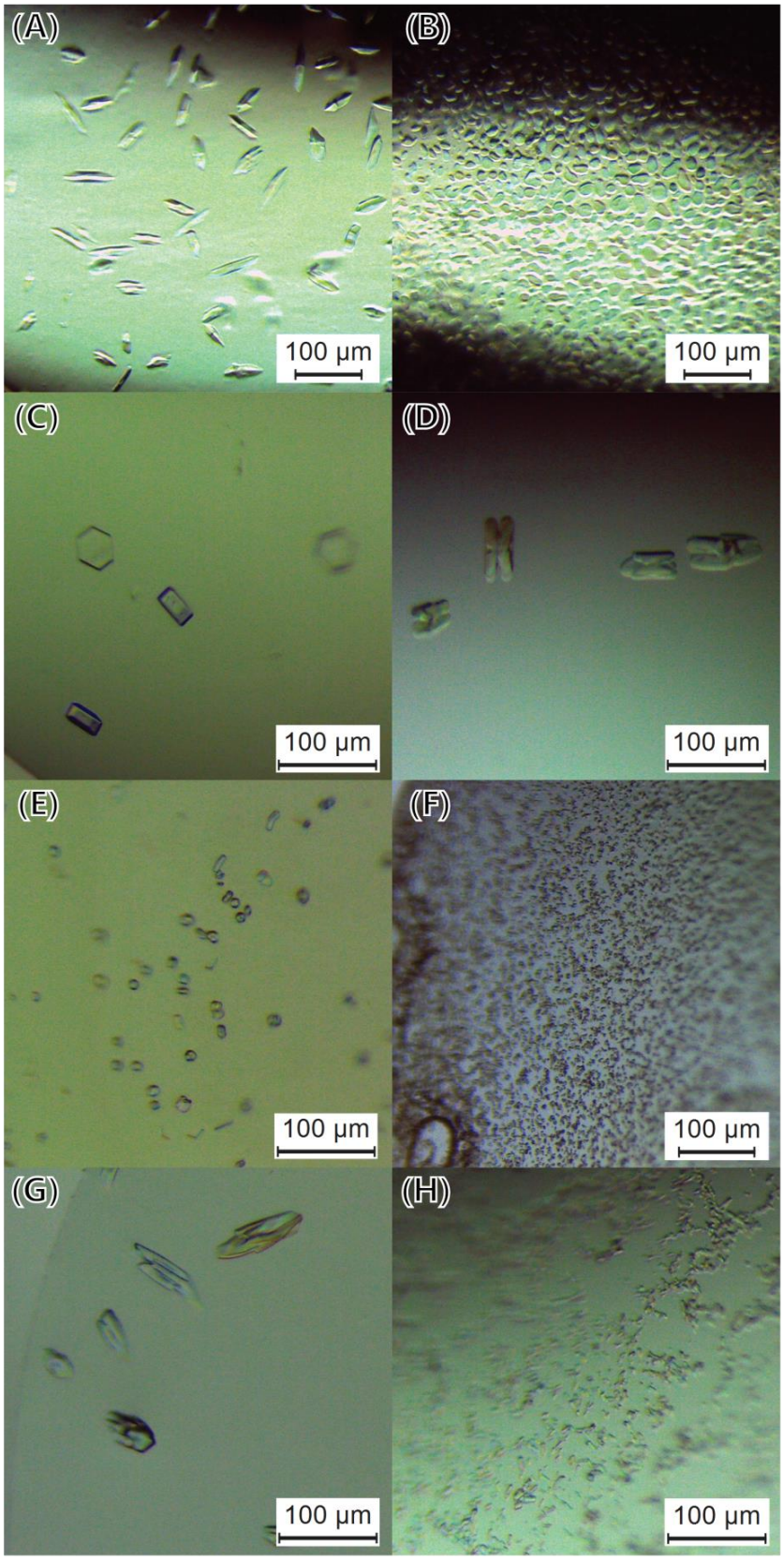
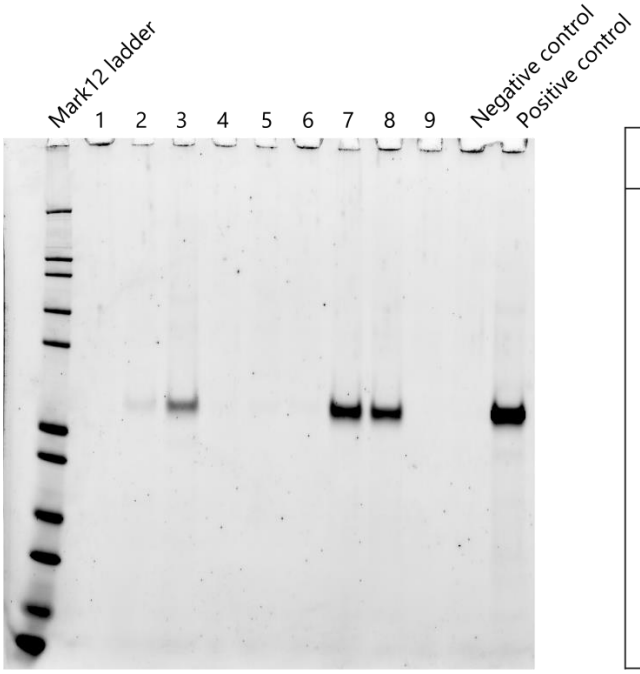


Figure 6.3: Conditions showing 2HDLL:M100c20 crystallisation. (A), (C), (E), and (G) are conditions containing protein and DNA; (B), (D), (F), and (H) contain only DNA. (A) and (B) from PEG Ion, condition A7; (C) and (D) from Crystal Suite, condition H3; (E) and (F) from JCSG+ HT, condition H4; (G) and (H) from Natrix, condition F5.

The putative protein-DNA crystals observed in the 2HDLL:M100c20 trays were small but could be recovered from the crystal trays. However, their small size meant that mounting crystals and obtaining diffraction signals was difficult. For this reason, a gel-based method was used to check for the presence of protein. Crystals were recovered from several conditions and dissolved, before being analysed by SDS-PAGE with visualisation using SYPRO Ruby staining (Figure 6.4, Section 2.6.7.2). SYPRO Ruby can be used to visualise as little as ~5 ng of protein, making it suitable for detecting the small amounts of protein contained in recovered crystals [384-386].



Lane	Oligo	Crystallography solution
1	M100c20	PACT A11
2	M100c14	Natrix F9
3	M100c20	JCSG+ H11
4	M100c20	Natrix E2
5	M100c20	Natrix F3
6	M100c20	JCSG H4
7	M100c20	Crystal Screen H3
8	M100c20	Natrix F4
9	M100c20	PACT A11

Figure 6.4: SYPRO Ruby stained SDS-PAGE of putative 2HDLL:M100 crystals. Mark12 ladder was included for size estimation. Lanes 1-9 contain dissolved crystal. The negative control lane contains sample prepared by mimicking recovering a crystal in a clear drop. The positive control lane contains the entire contents of a crystallography drop, taken from a drop with no signs of nucleation.

Crystal Screen condition H3 (lane 7, Figure 6.4) and Natrix condition F4 (lane 8, Figure 6.4) both showed strong bands at the correct molecular weight for 2HDLL, indicating that these crystals contain protein. JCSG condition H11 (lane 3, Figure 6.4) also showed a moderately strong band, but while recovering these crystals a ‘skin’ was also recovered from the drop. This skin may have contained protein, so whether there was protein in the crystals was not confirmed here. The components found in Crystal Screen H3 and Natrix F4 are listed in Table 6.2.

Table 6.2: Initial crystallisation conditions that grew putative protein-DNA crystals.

Screen ID	Components
H3, Crystal Screen	0.2 M Magnesium chloride hexahydrate 0.1 M Tris, pH 8.5 3.4 M 1,6-Hexanediol
F4, Natrix	0.08 M Strontium chloride hexahydrate 0.04 M Sodium cacodylate, pH 6 35% (+/−)-2-Methyl-2,4-pentanediol (MPD) 12 mM spermine tetrahydrochloride

6.2.2 Optimising 2HDLL:M100c20 crystals

Both crystallisation conditions contained a divalent cation and chloride-containing salt, as well as a diol-based precipitant, and neither condition was extremely acidic or basic. With this in mind, gradient screens were designed to optimise crystal growth. Two four-gradient screens were designed around these initial crystallisation conditions, varying the concentration of precipitant, the divalent cation, and the pH. These screens successfully replicated the original crystallisation condition, and some larger crystals were observed in conditions with lower concentrations of precipitant.

Larger crystals were seen in the gradient screen containing hexanediol (Figure 6.5). However, while they still looked morphologically different from the corresponding M100c20 crystals (Figure 6.5C), they more closely resembled the DNA-only crystals from initial screening (Figure 6.5D).

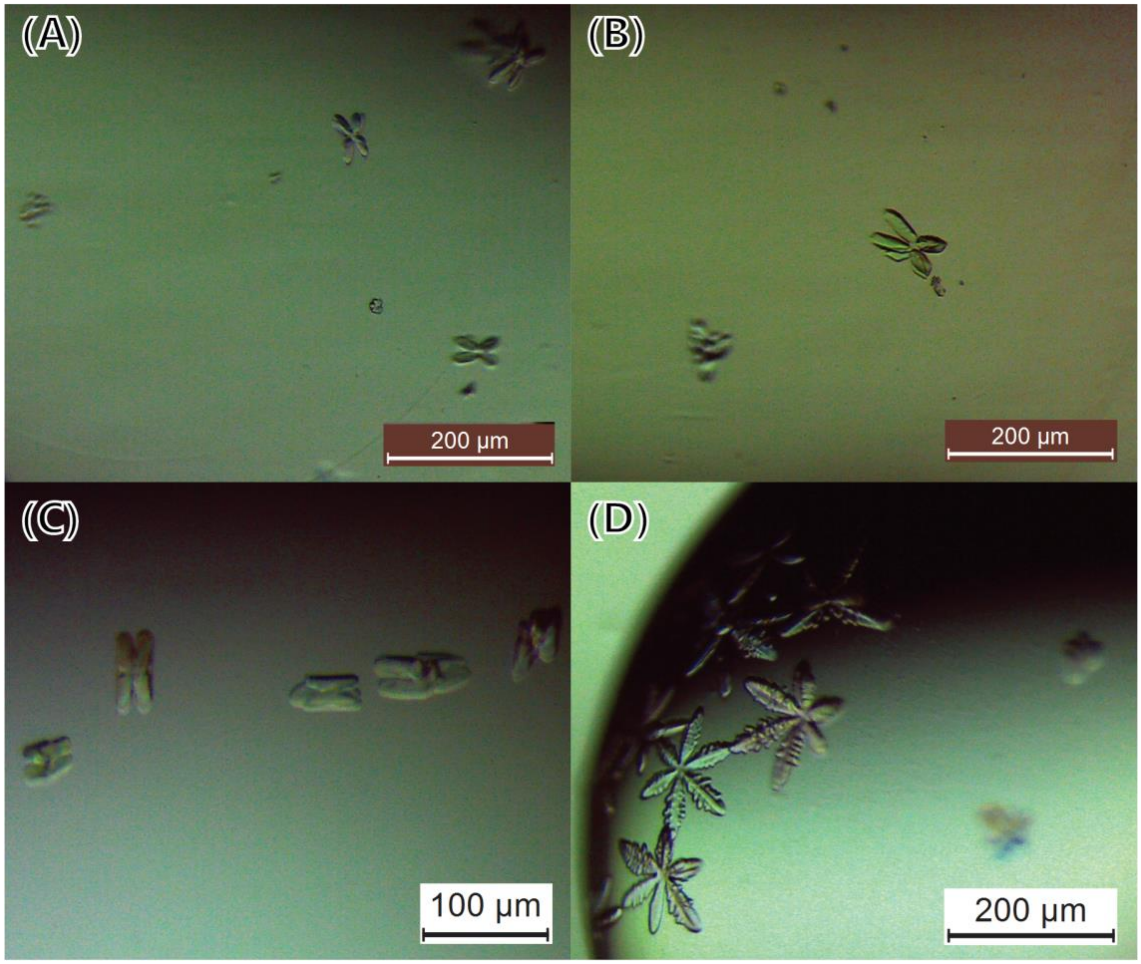


Figure 6.5: Optimised crystals grown in the hexanediol gradient screen. (A) Protein-DNA well, condition F5; (B) protein-DNA well, condition E5; (C) DNA only crystals from initial screening condition (JSCG+ H4); (D) DNA only crystals grown in optimisation condition E5.

Crystals grown in the MPD gradient screen were more morphologically consistent with the initial crystals observed (Figure 6.6). These crystals were larger than those observed in the initial screen, but precipitation was present in the optimisation screen that was not seen in the initial screen. Crystals obtained from both sets of optimisation trays were taken to the Australian Synchrotron for diffraction screening, but no diffraction was observed.

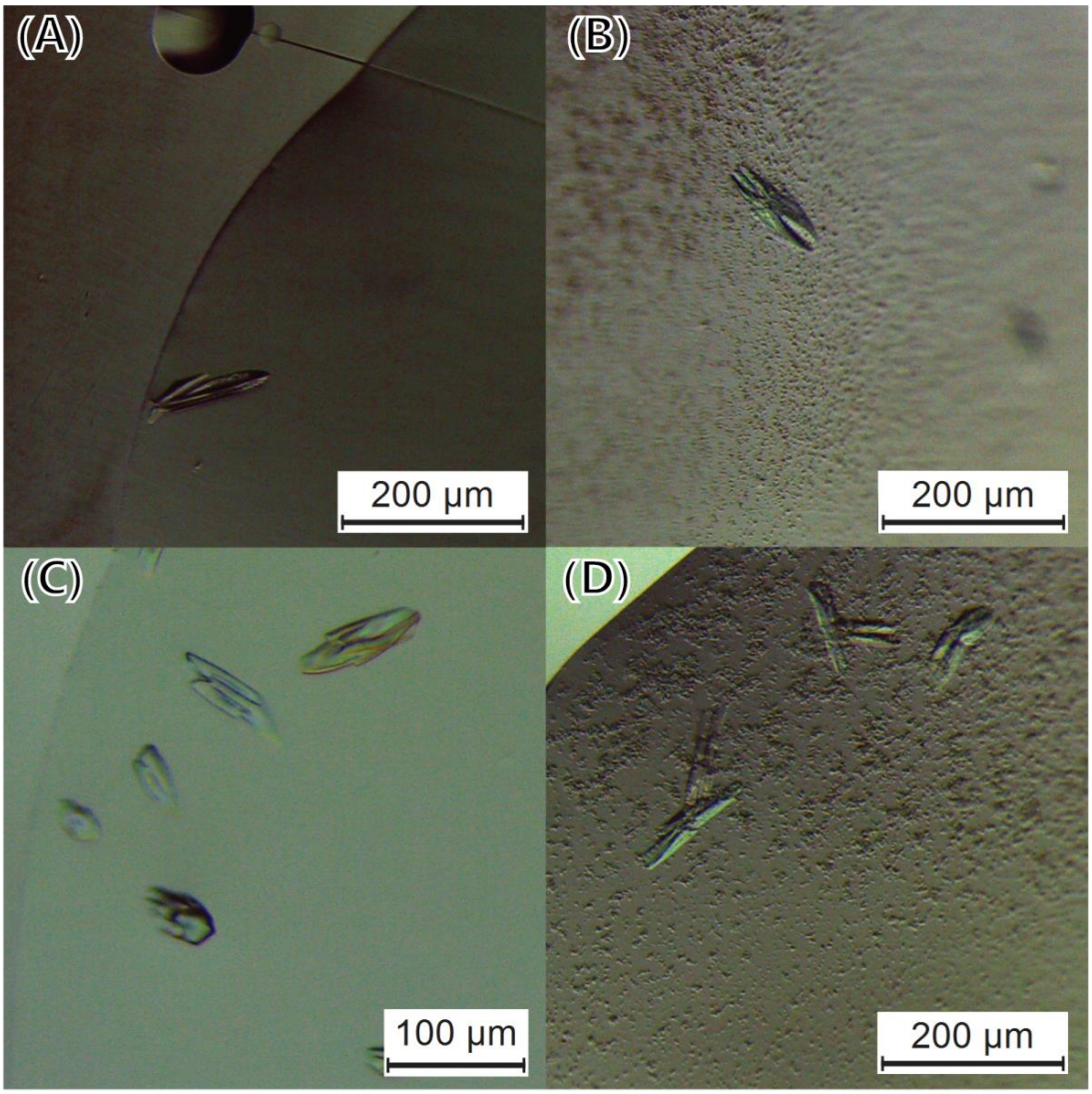


Figure 6.6: Optimised crystals grown in the MPD gradient screen. (A) Protein-DNA well from MPD optimisation screen condition A3; (B) Protein-DNA well from MPD optimisation screen condition B9; (C) Protein-DNA well from initial screening condition crystals (Natrix F5); (D) Protein-DNA well from MPD optimisation screen condition C9.

As the M100c20 oligonucleotide showed the most promise from the crystallisation trials described above, a variant of this oligonucleotide, but without the 1-base overhang (M100c20b; Table 6.1), was tested. The removal of the overhang no longer predisposes the DNA to form long strands through the crystals, which may reduce the overall chance of nucleation, but may also allow the protein-DNA complex to pack in an overall more favourable way. However, this strategy only resulted in DNA-only crystals. No protein-DNA screening conditions were found to have crystals likely to contain both the protein and the DNA, as all observed crystals closely resembled those seen in the corresponding DNA-only condition.

As the attempts to crystallise the 2HDLL:M100 complex did not result in diffracting crystals, alternative strategies were pursued to obtain structural information about the complex.

6.3 Investigating low resolution structure with small angle X-ray scattering

Small angle X-ray scattering (SAXS) was pursued as an alternative strategy to crystallography, as a technique that could provide information about the shape and component disposition of the complex in solution.

6.3.1 SAXS background

SAXS is based on the principle that particles exposed to a focussed X-ray beam will cause scattering of the X-rays (reviewed in [387]). This scattering is recorded at a detector, across varying angles of diffraction (Figure 6.7). The intensity and distribution of scattering reflects properties of the particle being measured [388]. Fourier transforms of scattering profiles result in a distribution of atom-pair distances, weighted by the product of their scattering factors. Different samples have different scattering factors. For example, DNA scatters more than protein, so a more intense signal will be observed from two atoms within a DNA oligonucleotide, compared to two atoms within a protein of the same spacing.

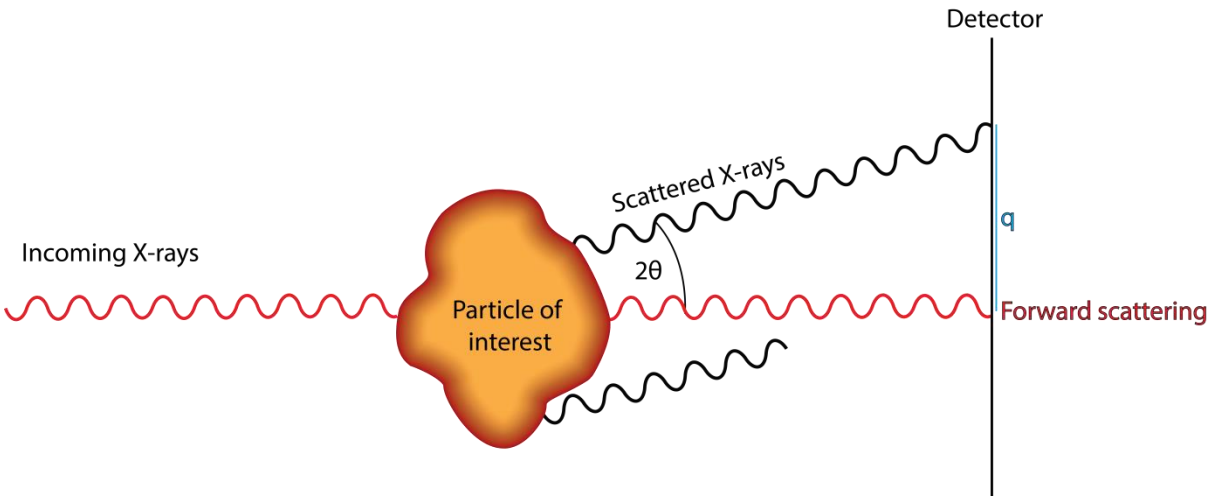


Figure 6.7: Schematic of SAXS, showing the diffraction and detection of X-rays. Figure adapted from [387].

SAXS data can be used to make inferences about various structural characteristics, such as overall solution conformation and conformational flexibility of the particle [389]. However, SAXS can only be used to obtain structural information when data is recorded from pure, homogenous samples. SAXS analysis relies on aggregated data, so any impurities or heterogeneity within the sample will skew downstream analyses [390]. This thesis will not go into depth on the mathematical background of SAXS analysis, but several key parameters are explained below.

6.3.1.1 Experimental terms

SAXS data is commonly presented in the form of X-Y plots of *q* against *I(q)*, with varying manipulations applied (discussed in [387, 391, 392]). *q* is the length of the scattering vector (Figure 6.7) and is related to the wavelength of the applied radiation, through the formula:

$$q = \frac{4\pi}{\lambda} \cdot \sin(\theta) \qquad \text{Equation 1}$$

In Equation 1, θ is half the angle of the scattering, with reference to the forward scattering (Figure 6.7). λ is the wavelength of the incident beam. *q* can be expressed in nm^{-1} or $Å^{-1}$. In some studies, variables called *s* or *h* are used in the place of *q*. Definitions of these terms can vary, so all data presented in this thesis will use *q*.

The function *I(q)* is a measure of the intensity of the scattering, in absolute units (cm^{-1}). *I(q)* values are the product of the volume of the particle (*V*), the electron density of the particle (*ρ*), the forward scattering (I_0), and the form factor *P(q)*, which is a function that reflects the shape of the particle emitting scattering:

$$I_1(q) = I_0 \cdot \rho_1^2 \cdot V_1^2 \cdot P(q) \qquad \text{Equation 2}$$

Plotting *q* against *I(q)* then reflects the intensity of scattering across a range of *q* angles. Commonly used plots include the logarithmic plot, Guinier plot, and Kratky plot, each of which provides different assessments about the sample quality and or particle properties.

6.3.1.2 Plots

The logarithmic plot is commonly used to confirm the dispersity of the sample [392]. The logarithmic plot shows both q and $I(q)$ on a logarithmic scale. For a homogenous and monodisperse sample, the logarithmic plot will be a flat line at low q. An upturn as q approaches 0 is an indication of aggregation, and a downturn is an indication of interparticle interference, a phenomenon whereby molecules in solution repel each other [393]. Both phenomena can skew downstream analyses.

The Guinier plot, q^2 against $\ln[I(q)]$, is used to calculate R_g and to estimate $I(0)$ by extrapolation [394-396]. $I(0)$ cannot be directly measured because of the beamstop placement in experimental setups. Because of its similarity to the logarithmic plot, this plot can also reveal information about the dispersity of the sample. The Guinier plot is generally only used to present data at low q (where qR_g is less than 1.3), as this region is the most linear; this is referred to as the Guinier region [395, 397].

The Kratky plot displays scattering data in a form that provides an indication of flexibility in the sample (Figure 6.8) [389, 398]. The dimensionless Kratky plot removes any influence of concentration differences, which allows for direct comparisons between different data sets [397, 399]. In both forms of the plot, the height of the peak at low s, in comparison to the increasing linear trend at higher s, is indicative of flexibility.

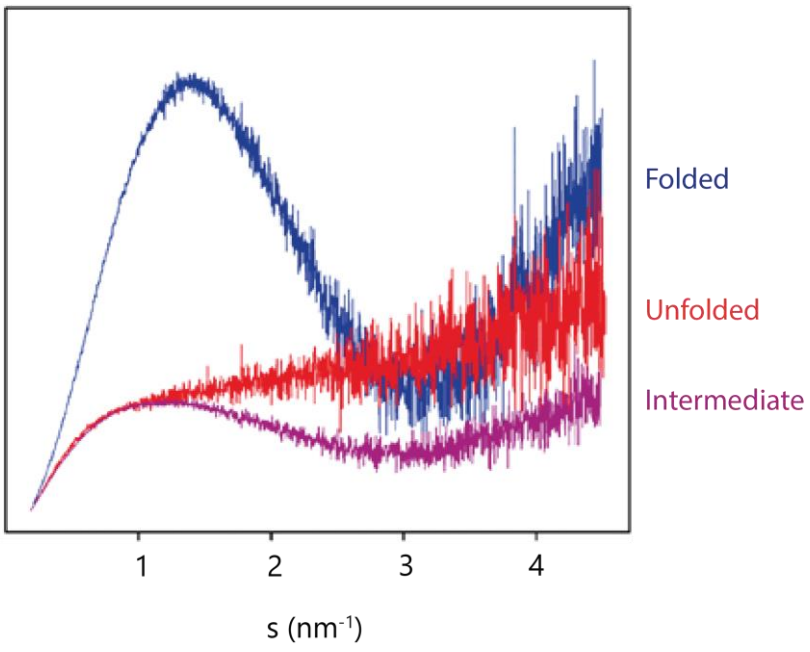


Figure 6.8: Examples of Kratky plots. Figure adapted from [389].

6.3.1.3 Calculated terms

One important calculated parameter is the radius of gyration (R_g), which is an indicator of the size and shape of a particle [395]. It is the mean root square of distances from the centre of the particle being analysed. The R_g is used for modelling calculations and can be a useful indicator of data quality. R_g is calculated from the inverse Fourier transform of *I(q)*, but can also be visualised as the gradient of the curve of the Guinier plot.

The molecular weight of the particle being observed can also be calculated from SAXS data [390, 400]:

$$Mw = \frac{I(0) \,.\, N_A}{C(\Delta\rho \,.\, v)^2} \qquad \text{Equation 3}$$

Equation 3 uses Avogadro's number (N_A), along with the properties of the particle in the chosen solvent, namely electron density (ρ) and partial specific volume (v), as well as the concentration of the particle (C) and $I(0)$. This metric allows confirmation of the size of the particle and can reveal polydispersity present in the sample. Due to the relationship with concentration, inaccurate concentration measurements can affect the calculated molecular weight, making it less accurate.

6.3.2 Trialling SAXS with the 2HDLL:M100 complex

Data were collected using 2HDLL and the M100c20 oligonucleotide, using the SAXS/WAXS beamline at the Australian Synchrotron. Initially, two experimental set ups were used: static SAXS, where each sample is exposed to X-rays in a capillary, and SEC-SAXS, where the sample is injected into a SEC column, in-line with a SAXS apparatus such that eluting material from the SEC column is then exposed to X-rays. Both experimental set ups were utilised to determine the optimal conditions for observation of the 2HDLL:M100 complex. It was important to ensure that small amounts of potentially unbound protein and DNA seen in the SEC-MALLS data (Figure 6.1) would not influence the SAXS measurements and downstream analyses (Section 6.1). Extensive standards have been established for reporting of SAXS data [390]. Appendix E contains complete reporting of sample details, data collection, and analysis.

6.3.3 SEC-SAXS indicates the 2HDLL:M100 complex is homogenous

SEC-SAXS was performed using the same column as previously used in SEC-MALLS experiments, allowing for easy identification of the complex (Section 6.3.1). The elution profile looked very similar to that seen during SEC-MALLS, with the majority of the signal detected belonging to the 2HDLL:M100 complex peak (Figure 6.9).

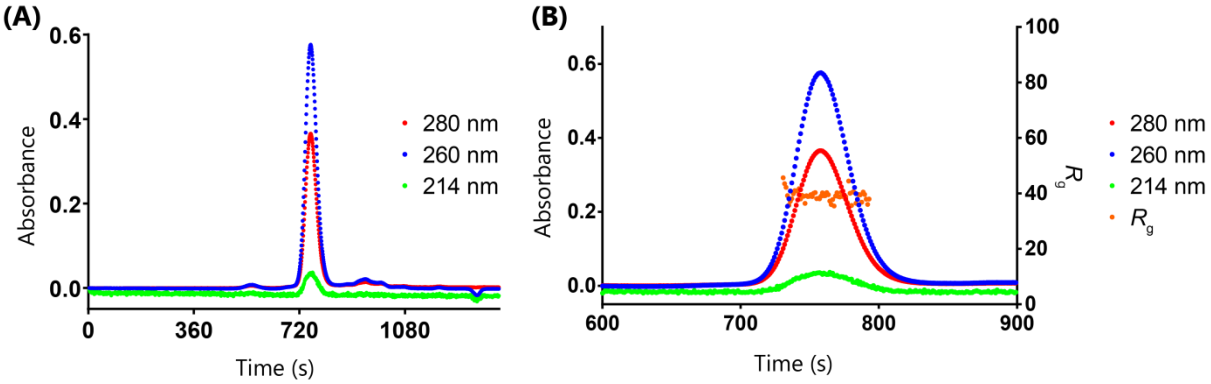


Figure 6.9: Elution profile of 2HDLL:M100 on SEC-SAXS, showing absorbance at three wavelengths. (A) Full elution profile. (B) 2HDLL:M100 elution peak, showing absorbance and R_g values.

R_g analysis of each frame measured show an overall stable R_g across the elution peak, of approximately 39 Å (Figure 6.9B). This provides additional evidence that the complex is homogenous, and gives a figure for comparison in static SAXS data analysis. Due to the dilution of the sample as it runs through the SEC apparatus, the signal:noise ratios for SEC-SAXS data were too low for modelling purposes. Static SAXS was conducted at a higher concentration, providing a higher signal:noise ratio that was more suitable for downstream structural analysis.

6.3.4 Analysing static SAXS data

Data for M100 and M100: 2HDLL were collected across a range of different concentration samples (Section 2.6.8). Data for 2HDLL was collected at a single concentration. Complete parameters and statistics from analysis can be found in Appendix E; below is a summary of the features of the data.

Logarithmic plots show that most of the data sets are homogenous, as indicated by the linearity of the data at low values of log(q) (Figure 6.10, Section 6.3.1.2). The datasets for the lowest concentration samples show lower signal:noise ratios than desired, but all samples had sufficient signal intensity for analysis.

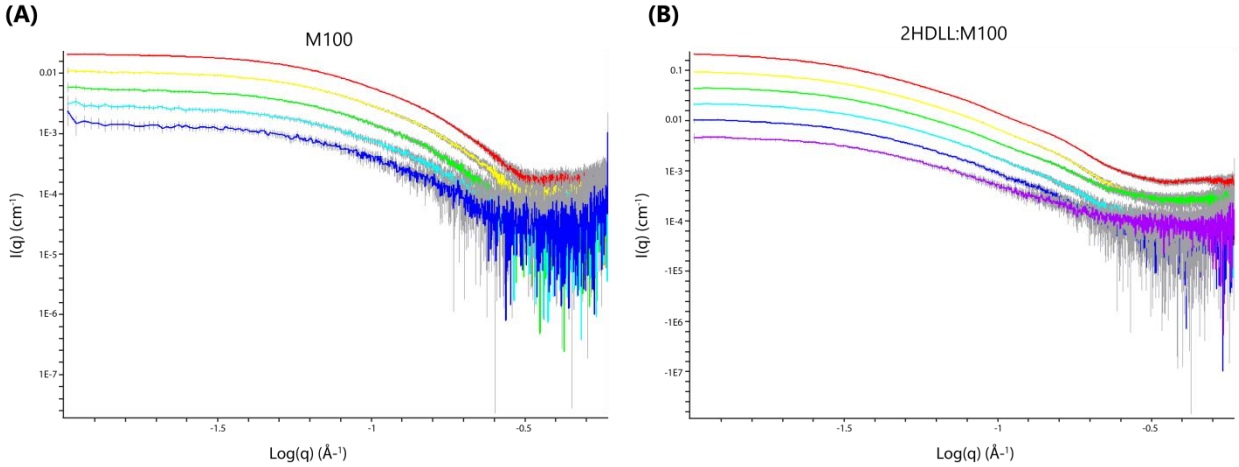


Figure 6.10: Baseline subtracted static SAXS data for concentration series. Logarithmic plots of (A) M100 and (B) 2HDLL:M100. Plots show a concentration series with the highest concentration being shown in red, then progressing through a rainbow.

R_g analysis for M100 samples showed an R_g consistently between 21.44 and 22.38 Å (Table 6.3). However, the reported R_gs for 2HDLL:M100 samples were found to be concentration dependent, with higher concentrations reporting higher R_gs. This was especially noticeable for the highest concentration 2HDLL:M100 sample and can be observed in differing gradients of the datasets in Guinier plots (Figure 6.11). This variation was taken as an indication of potential aggregation, given that the SEC-SAXS data indicates the R_g should be around 39 Å. For this reason 2HDLL:M100 data sets E and F, which were consistent with the SEC-SAXS data, were used for structural analysis.

Table 6.3: Experimental R_gs (Å) for M100 and 2HDLL:M100 SAXS samples.

Sample	M100		2HDLL:M100	
	Rg	Max qRg	Rg	Max qRg
A	21.44	1.26	47.89	0.97
B	21.82	1.25	42.65	1.15
C	22.28	1.29	40.74	1.23
D	22.38	1.26	40.38	1.19
E	21.88	1.3	38.97	1.28
F	21.82	1.3	36.61	1.3

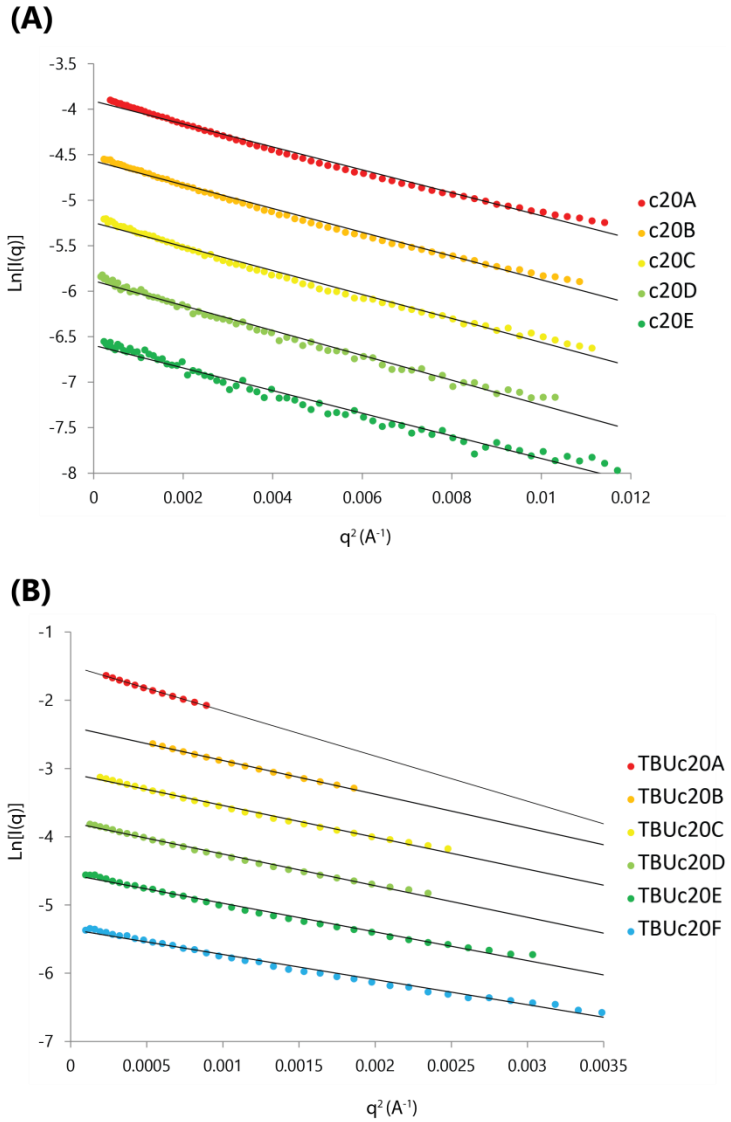


Figure 6.11: Guinier plots of M100 alone and 2HDLL:M100 SAXS datasets. (A) Guinier region of M100 samples, showing the points used for R_g analysis and linear fits; (B) Guinier region of 2HDLL:M100 samples, showing the points used for R_g analysis, and linear fits. qRg maxima used for fitting are listed in Table 6.3.

Kratky plots showed that data for the DNA alone (M100) gave rise to a curve consistent with that of a folded species, whereas 2HDLL alone and 2HDLL:M100 samples gave rise to curves characteristic of a species with both folded and flexible regions (Figure 6.12). The flexible nature of the complex could explain why crystallography efforts failed (Section 6.2).

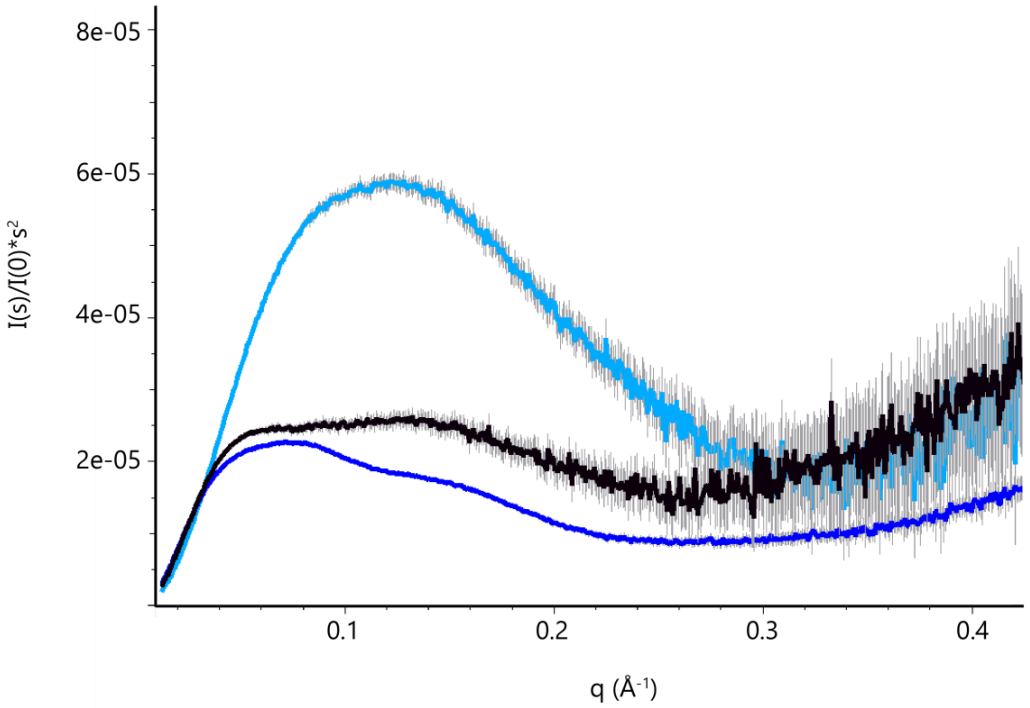


Figure 6.12: Kratky plots of SAXS data. Measured data with M100 alone A (light blue), 2HDLL alone (black), and 2HDLL:M100 A (dark blue).

The dimensionless Kratky plot can be used to more directly compare the flexibility of different particles (Section 6.3.1.2). This plot again suggests that 2HDLL is flexible, by the plateaued peak shape and the dispersion of the data at higher qR_g values (Figure 6.13). The data for M100 suggests an ordered species, and the data for 2HDLL:M100 shows an intermediate state between the two.

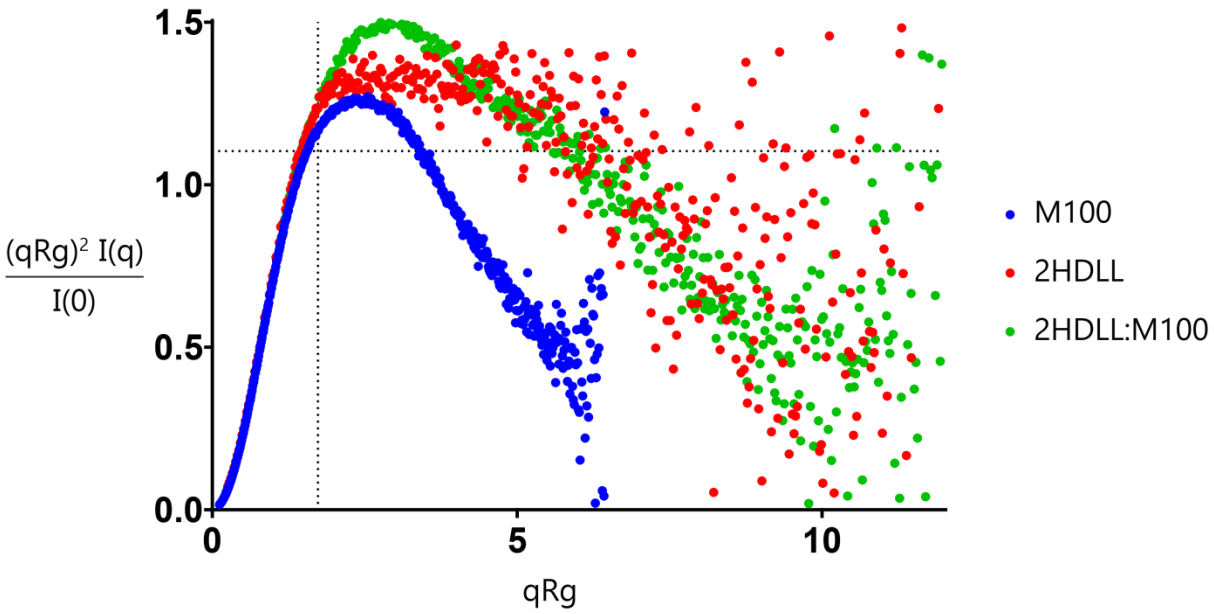


Figure 6.13: Dimensionless Kratky plots of SAXS data. Dotted line indicates the point at which qRg = √3, $(qRg)^2I(q)/I(0) = 1.104$; this point is a local maximum for folded globular proteins.

It is reasonable to assume that if the flexible 2HDLL bound DNA it would have reduced flexibility and therefore a different overall shape. Thus, data from 2HDLL in isolation was not used to model 2HDLL bound to M100. In contrast, based on existing structural data for homeodomain-DNA complexes, the much more rigid M100 is unlikely to be significantly affected by binding (Section 1.2.2.3).

6.3.5 Modelling the 2HDLL:M100 complex

The modelling program MONSA was used to construct potential envelopes of the 2HDLL:M100 complex. MONSA is an *ab initio* modelling algorithm that uses iterative dummy atom modelling to find a solution that most closely matches the inputted SAXS data [401]. Baseline subtracted data, the proportional volumes of each species in solution, and the d_{max} (maximum dimension of the particle being modelled) are the only inputs for required for MONSA. The R_g may also be supplied as an additional modelling constraint.

MONSA can be used to model complexes, by inputting multiple data sets and specifying which complex components are present in each data set. MONSA can also accommodate phases with different scattering densities. In this case, this allows the DNA and protein

components to be distinguished. For modelling the 2HDLL:M100 complex, data sets for both the complex and the DNA alone were input, for the reasons outlined at the end of the previous section.

The inherent flexibility within the complex means that as the data reaches higher q values, it will not tend towards 0 as quickly as would a highly ordered species, which impacts the modelling performance, as MONSA assumes that the data tends towards 0. To prevent this from affecting the models generated, the data was truncated to a q cut off of 0.3. The signals beyond this point are dominated by short range internal density fluctuations, which MONSA does not consider as it uses uniform density modelling for each component. As a result, truncation of the data should not impact the validity of the fits generated.

Due to uncertainty around the d_{max} of the complex, as well as the R_g, these parameters were varied for different iterations of modelling. Each combination of parameters was used to generate three iterations of modelling. The resulting ensemble of models showed varying conformations of the complex. Four representative models will be discussed, as they illustrate the variations and similarities present within the ensemble (Figure 6.14).

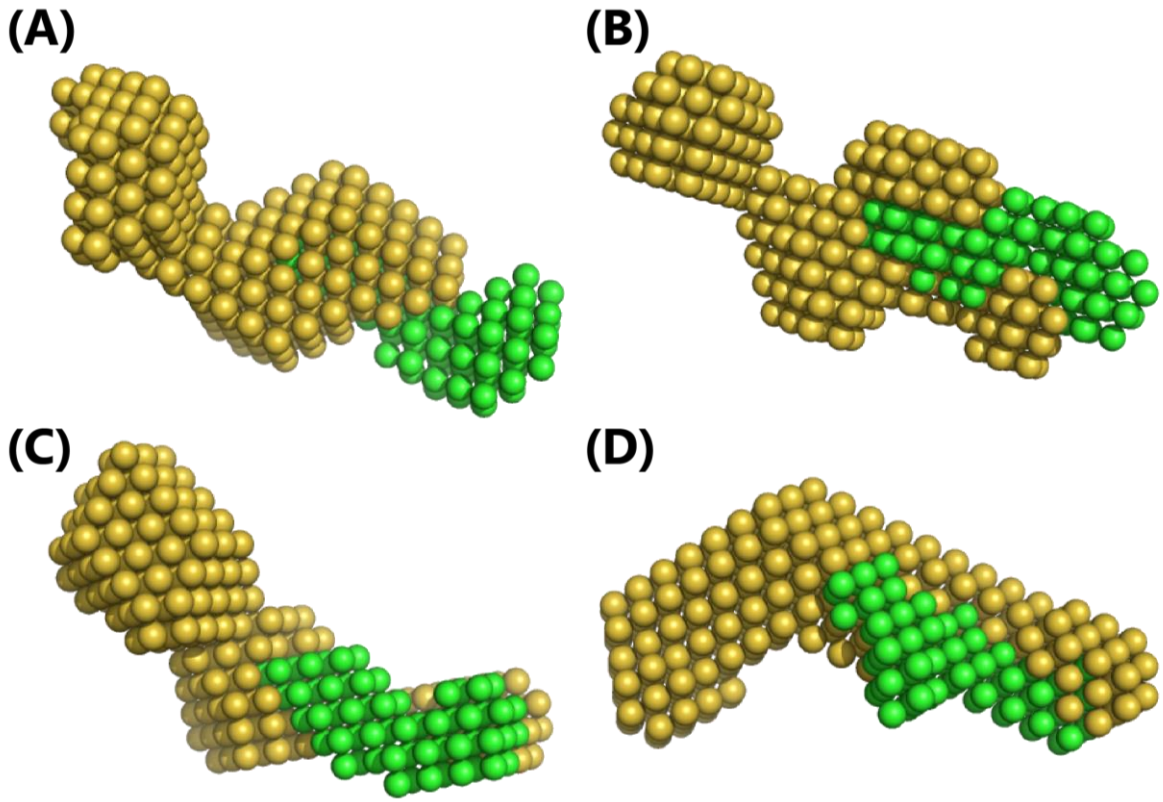


Figure 6.14: Examples of 2HDLL:M100 models generated using MONSA. These models were generated by restricting the R_g of 2HDLL during modelling. (A) R_g of 35; (B) R_g of 37; (C) R_g of 39; (D) R_g of 42. The parts of the model corresponding to DNA are shown in green and the protein component is shown in gold.

In all of these models the DNA appears to have been modelled effectively as a roughly rod-like or elongated volume. The models also place the protein in close proximity to the DNA, indicative of an interaction, but also show a substantial volume of protein that does not appear to contact the DNA. This volume potentially represents the LIM:LID part of the construct that lies between the two homeodomains of 2HDLL.

There was significant variation amongst the models in terms of how much of the protein was placed in contact with the DNA. Several models showed an extended interaction interface between the DNA and the protein, suggesting that both homeodomains bind the DNA (Figure 6.14 C and D), but others showed only a small portion of protein interacting with the DNA, which could indicate single homeodomain binding (Figure 6.14 A and B). In those models where the protein is not interacting along the length of the DNA, it is interesting to note that the protein-DNA interface is primarily localised to one end of the DNA, as opposed to the centre. This may reflect a tight interaction between $Lhx3_{HD}$ and the AAATTA site in the M100c20 oligonucleotide. While these models show different conformations, they all show a

similar fit to the data (Figure 6.15). Since all of the models fit the data equally well, it cannot be determined which conformation is more likely to represent the 2HDLL:M100 complex. Indeed, it is possible that all of these conformations are represented in solution.

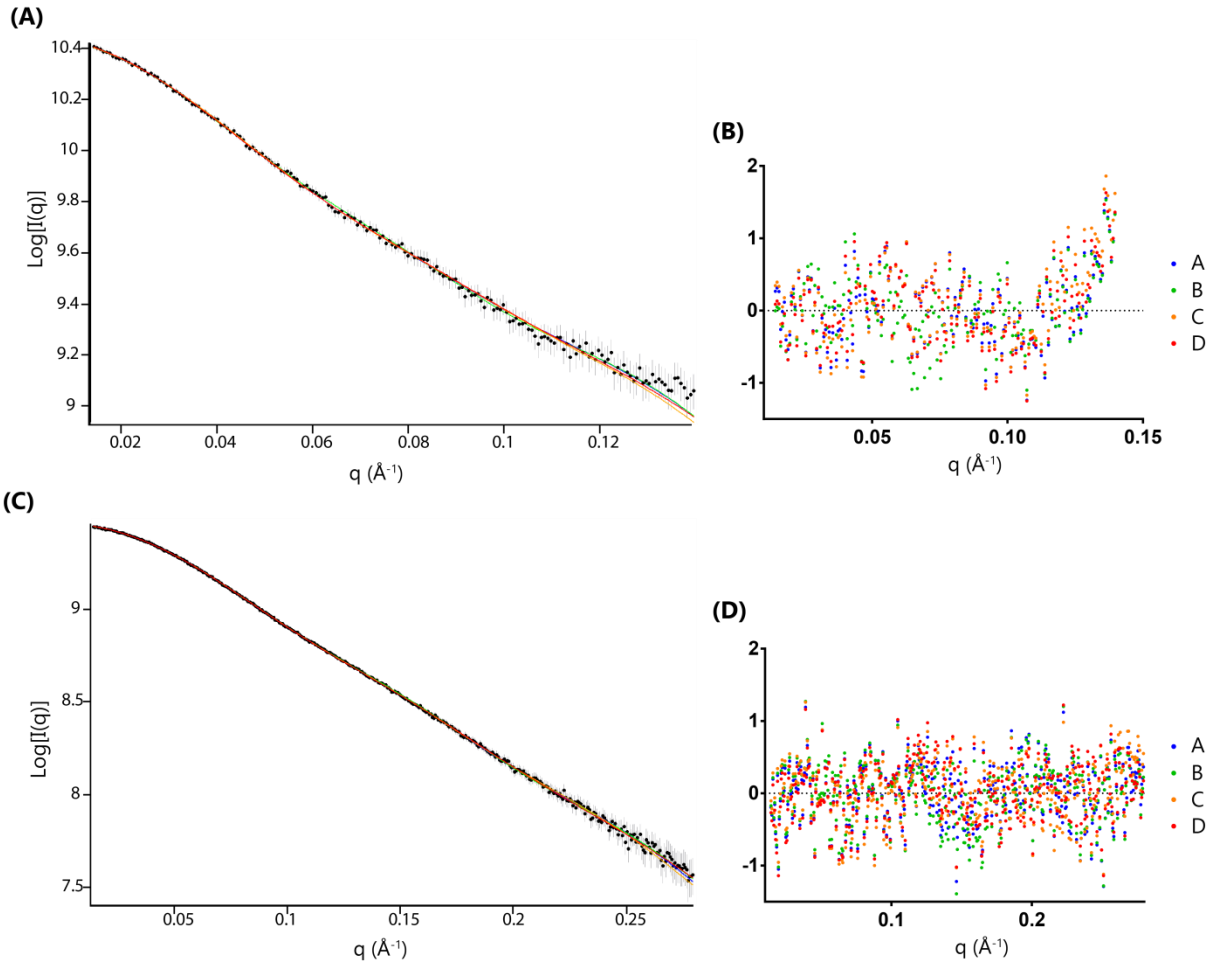


Figure 6.15: Assessing the fits of 2HDLL:M100 models generated. The same models are used as in Figure 6.14. (A) Plot of fits overlayed on experimental data for 2HDLL:M100; (B) Difference plot of fits to 2HDLL:M100; (C) Plot of fits overlayed on experimental data for M100; (B) Difference plot of fits to M100. Difference plots were generated by calculating the difference between the experimental data and the fit, then dividing by the error of the fit.

It is not possible to determine how $Isl1_{HD}$ is interacting with the DNA from the SAXS data presented in this chapter. More data is required to gain further insight into the conformational and structural interplay between the protein and DNA components of this complex. As the 2HDLL:M100 complex has been shown here to be flexible in nature, it may not be possible to obtain one definitive structure of the complex.

6.4 Discussion

This chapter has sought to gain insight into the structure of the Lhx3:Isl1 DNA-binding module. Due to the flexible nature of the 2HDLL:M100 observed in SAXS, it is not yet possible to construct a definitive model of how this complex binds DNA. However, the data presented here have provided hints as to the main features of binding.

Chapter 5 showed that the Isl1 homeodomain in isolation is unable to bind DNA with high affinity or specificity. The additional SAXS data presented in this chapter shows that when brought into close proximity to $Lhx3_{HD}$ within fusion constructs, $Isl1_{HD}$ can potentially bind to DNA. This data gives rise to the following question about the function of Isl1: if the protein only makes a small contribution to DNA-binding in the context of a larger complex, how does this affect its function? Several theories are plausible.

6.4.1 Mechanisms of Isl1 and Lhx3 cooperation

When Isl1 and Lhx3 are in combination they preferentially bind a sequence different to that bound by Lhx3 only (Section 5.7) [99]. Consequently, the two homeodomains must be somehow influencing the binding preferences of each other.

6.4.1.1 The potential for homeodomain heterodimerisation

One simple explanation for this would be a direct interaction occurring between $Isl1_{HD}$ and $Lhx3_{HD}$. Although the involvement of homeodomains in protein-protein interactions is not common, dimerisation of homeodomains has been previously observed [342, 371, 402]. The DNA binding specificities of these heterodimers were shown to be different to those of the individual homeodomains. In many cases, these interactions rely on residues upstream or downstream of one of the homeodomains. However, there are only six residues between the end of the Lhx3 LIM domains (definitions as judged by solved crystal structures (PDB: 2JTN and 2RGT)) and the beginning of the canonical homeodomain. If the residues immediately upstream of the Lhx3 homeodomain were involved in an interaction with $Is1l_{HD}$, there would likely be steric hindrance between the Isl1 homeodomain and the Lhx3 LIM domains. Further, DNA binding experiments using Isl1/Lhx3 fusion constructs in this thesis have demonstrated that the linker sequence between the two is not important in binding specificity of the complex, as binding behaviour is similar whether there is a native linker (2HDN) or a glycine-serine linker (2HD23) (Section 5.7). This would mean that any interaction between

the two homeodomains would occur through the homeodomains themselves, which is unlikely.

6.4.1.2 A kinetic explanation for changing DNA-binding preferences

Another potential explanation is that Isl1 does bind to DNA but has kinetic properties (e.g. rapid off-rates) that make it difficult to observe by the methods used herein. If this were the case, the binding of Lhx3 to DNA, and the tethering of the protein components through LIM-LID complexes could lock the DNA-bound $Isl1_{HD}$ in place. The presence of $Isl1_{HD}$ could alter the specificity of $Lhx3_{HD}$, by making it sterically less favourable to bind a TAATTA site, and more favourable to bind a CATTAGCCAAATTA sequence, which has room for two homeodomains to be bound simultaneously. This type of mechanisms resembles the way that homeodomains are thought to bind non-specifically to DNA, and then move around and along the DNA until they find a binding site of high affinity [403]. This binding mechanism would be especially relevant in an *in vivo* environment. Since Isl1 appears to be unable to stably bind DNA independently, it seems likely that the Lhx3:Isl1 interaction would form first, and then the complex would bind cognate sites in DNA. Having two DNA binding domains linked together through protein complex formation, with one being a tighter binder than the other, has shown to be an effective way for a complex to search for recognition sites on DNA more effectively [404].

6.4.2 Future directions

Efforts to obtain a high resolution crystal structure of the 2HDLL:M100 complex were unsuccessful. However, the SAXS data has provided new insight into the flexibility of the complex. Further investigations into the conformation of the complex with further SAXS experiments may extend our knowledge more. These investigations are currently underway, as more SAXS data was gathered in November 2018. This data focussed on 2HDLL and M100c20b, but also on other protein constructs in combination with M100: 2HDN, LLHD3, and NHD3. Although initial analysis of this additional SAXS data looks promising, thorough analysis and modelling could not be conducted in the timeframe of this thesis. These other protein constructs may be useful in gathering further structural insights. The flexibility seen in the conformation of 2HDLL, both alone and in complex with M100, may potentially be due to movement between the LIM:LID regions and the homeodomains respectively. If this is the case, SAXS data for the 2HDN:M100 complex may not be influenced by the flexibility

that prevents rigorous structural modelling of the 2HDLL:M100 complex. Since 2HDN appears to bind DNA in a very similar manner to 2HDLL (Section 5.7), it may then also be of interest to attempt to solve the structure of 2HDN:M100 through X-ray crystallography.

The binding behaviour of 2HDLL to M100 could also be further characterised to provide more information about how binding is facilitated. This includes the use of surface plasmon resonance (SPR) to probe the kinetic properties of binding, as well as uncovering which specific amino acids are involved in binding through HSQC (heteronuclear single quantum coherence) NMR titrations.

A larger focus on the DNA side of the interaction may also provide new information. Varying the spacing of the CATTAG and AAATTA sites within the M100 oligonucleotide and checking the binding with 2HDLL could help to indicate whether a direct Isl1:Lhx3 interaction influences the DNA binding of the overall complex. It would also be intriguing to check the binding behaviour of GST-NHD3 dimers to such DNA mutants, to see if they still follow the same binding trends as 2HDLL.

6.4.3 Conclusion

Without a clearer structural picture of how the motor neuron complex is binding DNA, it is impossible to precisely define the DNA-binding behaviour of Isl1. At this stage, it appears likely that Isl1 does contribute to DNA-binding when in complex with a binding partner, such as Lhx3. It is possible that this behaviour is functionally relevant, and acts as a mechanism by which Isl1 can act in combination with many different DNA-binding proteins, allowing precise targeting of specific DNA sequences in an efficient manner. This could provide a partial explanation for how Isl1 plays a role in regulating gene expression in such a diverse range of tissues, even though it does not appear to bind DNA when in isolation.

7 Conclusions

This thesis has investigated the protein-DNA and protein-protein interactions that allow Isl1 to function as a transcriptional regulator with diverse roles across many tissues, in the hopes of gaining new insights into how Isl1 acts in a wide variety of cellular contexts.

Chapters 3 and 4 investigated the protein interaction domains of Isl1, searching for novel binding partners through yeast two-hybrid library screening. These screens were successful in identifying putative novel binding partners for Isl1, with Mkln1 being a potential binding partner of interest based on its ability to bind both $Isl1_{LIM}$ and $Ldb1_{LID}/Isl1_{LIM}$ in two orthologous systems. Other potential binding partners identified herein may also be of interest, although close assessment of the literature suggests that some may not represent biologically relevant interactions (Chapter 4). This identification of binding events that may be real but are not biologically relevant highlights the need for good moderation of large screening datasets of this type, and for thorough validation of putative interactions.

Chapters 5 and 6 shifted the focus onto the mechanisms by which Isl1 and Lhx3 interact with DNA as isolated LIM-HD proteins or when in combination. Chapter 5 probed the DNA-binding behaviour of the two homeodomains and reveals that Lhx3 and Isl1 have very different DNA binding behaviours, despite being very similar in terms of sequence, fold and stability. It appears that Isl1 is not able to bind to DNA, including its reported recognition sequences, with high affinity in the absence of a protein binding partner that promotes DNA-binding. In Chapter 6, the behaviour of the Isl1 and Lhx3 homeodomains in complex was further investigated, with the key goal being determination of a structure of the 2HDLL:M100 complex. Although an atomic resolution structure was not produced, SAXS has provided new insights into this apparently dynamic complex (Section 6.3). The combination of EMSA data presented in Chapter 5 and SAXS data presented in Chapter 6 provide biophysical evidence that the Isl1 homeodomain can bind to DNA directly to modulate the stronger binding of Lhx3. Additional structural and biophysical studies, informed by the presented data herein, should provide confirmation of how these two homeodomain proteins can bind different DNA sites as single entities or through combined efforts.

7.1 The role of the Isl1 in a broader context

Aside from its role in the motor neuron complex, the action of the Isl1 homeodomain is also relevant to the other roles of the Isl1 protein. The data presented here using Isl1/Lhx3 fusion constructs provides hints about how Isl1 can play a role in many different contexts.

As Isl1 has been shown to play a role regulating gene expression in many tissues (Section 1.4), it may be that weak DNA binding is essential to its function. Making a smaller contribution to DNA binding affinity could allow Isl1 to participate in a wider range of protein-DNA complexes, targeting a broader range of DNA sequences than would be possible if it bound DNA with a higher affinity or increased specificity. It is also possible that the combination of a weakly DNA-binding Isl1 in complex with another protein that binds DNA more strongly allows more rapid identification and binding of target sequences, which can be crucial in a developmental context, where cell fate decisions are made with precise timing [403, 404]. This ability to use non-specific binding to search for specific gene targets is a known feature of homeodomains [405].

The influence of Isl1 on the DNA-binding of partner proteins could be easily investigated through the production and study of other fusion constructs containing $Isl1_{HD}$ and another DNA-binding domain. Studying the DNA-binding behaviour of such fusion constructs would allow a broader insight into how the Isl1 homeodomain influences the DNA-binding preferences of other proteins. As Isl1 is thought to cooperate with a wide variety of transcription factors *in vivo*, it is likely that it influences the binding specificities of at least some other protein binding partners in a similar manner to that seen in the case of the motor neuron complex.

It is plausible that the primary role of Isl1 is as a protein-protein adaptor, bringing together other transcription factors so that they may bind DNA. In support of this statement, Isl1 and Isl2 are the only LIM-homeodomain proteins to have two known protein-protein interaction interfaces (LIM domains and LIDs), and so are the only proteins in the LIM-HD transcription factor family that can form higher order complexes through those two interfaces. However, all the LIM-HD proteins have uncharacterised C-termini, so dual protein-binding sites may not be exclusive to the Islet proteins. Moreover, the Islet proteins may also contain additional interaction domains that have not yet been identified.

It must be noted that LIM domains themselves can bind to multiple binding partners concurrently. Examples of this include the case of the hematopoietic transcriptional complex that contains Lmo2 [406, 407], and the cytoskeletal complex that contains another LIM-domain containing protein, Testin [408, 409]. In situations such as this, binding partners are in very close proximity, meaning they have the opportunity to interact with each other, as well as with the LIM domain protein. The LIDs appear to be more limited in their ability to interact with other partners. To date they appear to bind only one protein at any instance, and whereas $Ldb1_{LID}$ can interact with multiple partners [51], $Isl1_{LID}$ preferentially binds only Lhx3 and Lhx4, with apparently lower affinity binding to Lmx1b [54]. The LIDs from Isl1 and Isl2 have another proposed role in shielding the Isl1 LIM domains from non-specific or off-target interactions [54]. However, it may be possible for the intramolecular LIM:LID interaction to co-exist with another LIM domain based protein-protein interaction. This may act as a mechanism by which higher order complexes are formed: initial binding of a protein to the $Isl1_{LIM}$ domains, before displacement of the $Isl1_{LID}$ and formation of further protein-protein interactions.

From the data presented here, it is clear that the study of both protein-protein and protein-DNA interactions are important in examining the function of Isl1. Data from Chapter 5 suggests that Isl1 may not be able to function as a DNA-binding transcription factor without the presence of additional protein binding partners. Given that the specificity of the Lhx3:Isl1 complex is different to that of an independent Lhx3, it appears that Isl1 can influence the DNA-binding specificity of its protein binding partners. *In vivo*, this is potentially a mechanism by which Lhx3:Isl1 complexes can target different areas of the genome than Lhx3-only containing complexes. This behaviour extends to other transcriptional complexes containing Isl1. For example, Isl1 has been found to influence the DNA-binding preferences of the transcription factor Phox2a during cranial motor neuron development, in a similar fashion to its behaviour with Lhx3 [105]. Further work may identify other instances of Isl1 modulating the DNA-binding specificity of its protein-protein interaction partners.

It may be that the weak binding of Isl1 to DNA is an advantage in the context of multi-protein complexes binding to DNA. Assuming that this property helps Isl1-containing complexes find their targets more quickly (see above), Isl1 may refine the *in vivo* DNA-targeting properties of any transcriptional complex it is a part of, and allow that complex to find its genomic targets efficiently. As Isl1 can bind multiple partners, it could influence the

DNA-binding properties of many different transcriptional complexes, and thereby direct development in a broad range of contexts.

Further study is needed to determine whether this model of the action of Isl1 is correct. Investigation of the DNA-binding behaviour of Isl1 in the context of different protein binding partners will generate a more complete picture of the role of Isl1 in transcriptional regulation of gene expression, giving new insight into how the interplay of protein-protein and protein-DNA interactions can influence gene regulation.

References

1. Chen, K. and N. Rajewsky, *The evolution of gene regulation by transcription factors and microRNAs.* Nat. Rev. Gen., 2007. **8**: p. 93.
2. Levine, M. and R. Tjian, *Transcription regulation and animal diversity.* Nature, 2003. **424**: p. 147.
3. Lambert, S.A., et al., *The Human Transcription Factors.* Cell, 2018. **172**(4): p. 650-665.
4. Slattery, M., et al., *Absence of a simple code: how transcription factors read the genome.* Trends Biochem. Sci., 2014. **39**(9): p. 381-399.
5. Boyer, L.A., et al., *Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells.* Cell, 2005. **122**(6): p. 947-956.
6. Zardawi, S.J., et al., *Dysregulation of Hedgehog, Wnt and Notch signalling pathways in breast cancer.* Histol Histopathol, 2009. **24**(3): p. 385-398.
7. Futreal, P.A., et al., *A census of human cancer genes.* Nat. Rev. Can., 2004. **4**: p. 177.
8. Srivastava, M., et al., *Early evolution of the LIM homeobox gene family.* BMC Biol., 2010. **8**(1): p. 4.
9. Hunter, C. and S. Rhodes, *LIM-homeodomain genes in mammalian development and human disease.* Mol. Biol. Rep., 2005. **32**(2): p. 67-77.
10. Pfaff, S.L., et al., *Requirement for LIM Homeobox Gene Isl1 in Motor Neuron Generation Reveals a Motor Neuron–Dependent Step in Interneuron Differentiation.* Cell, 1996. **84**(2): p. 309-320.
11. Tsuchida, T., et al., *Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes.* Cell, 1994. **79**(6): p. 957-970.
12. Koulakov, A.A. and D.N. Tsigankov, *A stochastic model for retinocollicular map development.* BMC Neurosci., 2004. **5**(1): p. 30.
13. Thaler, J.P., et al., *LIM Factor Lhx3 Contributes to the Specification of Motor Neuron and Interneuron Identity through Cell-Type-Specific Protein-Protein Interactions.* Cell, 2002. **110**(2): p. 237-249.
14. Caronia-Brown, G., A. Anderegg, and R. Awatramani, *Expression and functional analysis of the Wnt/beta-catenin induced mir-135a-2 locus in embryonic forebrain development.* Neural Dev., 2016. **11**(1): p. 9.
15. Chizhikov, V.V. and K.J. Millen, *Control of Roof Plate Development and Signaling by Lmx1b in the Caudal Vertebrate CNS.* J. Neurosci., 2004. **24**(25): p. 5694-5703.
16. Millonig, J.H., K.J. Millen, and M.E. Hatten, *The mouse Dreher gene Lmx1a controls formation of the roof plate in the vertebrate CNS.* Nature, 2000. **403**: p. 764.
17. Pressman, C.L., H. Chen, and R.L. Johnson, *lmx1b, a LIM homeodomain class transcription factor, is necessary for normal development of multiple tissues in the anterior segment of the murine eye.* genesis, 2000. **26**(1): p. 15-25.
18. Chen, H., et al., *Limb and kidney defects in Lmx1b mutant mice suggest an involvement of LMX1B in human nail patella syndrome.* Nat. Genet., 1998. **19**: p. 51.
19. Vollrath, D., et al., *Loss-of-Function Mutations in the LIM-Homeodomain Gene, LMX1B, in Nail-Patella Syndrome.* Hum. Mol. Genet., 1998. **7**(7): p. 1091-1098.
20. Zhao, Y., et al., *LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum.* Proc. Natl. Acad. Sci., 2007. **104**(32): p. 13182-13186.
21. Costello, I., et al., *Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development.* Genes Dev., 2015. **29**(20): p. 2108-2122.

22. Abellán, A., et al., *Similarities and differences in the forebrain expression of Lhx1 and Lhx5 between chicken and mouse: Insights for understanding telencephalic development and evolution.* J. Comp. Neurol., 2010. **518**(17): p. 3512-3528.
23. Cheah, S.-S., K.M. Kwan, and R.R. Behringer, *Requirement of LIM domains for LIM1 function in mouse head development.* Genesis, 2000. **27**(1): p. 12-21.
24. Porter, F.D., et al., *Lhx2, a LIM homeobox gene, is required for eye, forebrain, and definitive erythrocyte development.* Development, 1997. **124**(15): p. 2935-2944.
25. Hirota, J. and P. Mombaerts, *The LIM-homeodomain protein Lhx2 is required for complete development of mouse olfactory sensory neurons.* Proc. Natl. Acad. Sci., 2004. **101**(23): p. 8751-8755.
26. Iyer, N.R., et al., *Generation of highly enriched V2a interneurons from mouse embryonic stem cells.* Exp. Neurol., 2016. **277**: p. 305-16.
27. Hume, C.R., D.L. Bratt, and E.C. Oesterle, *Expression of LHX3 and SOX2 during mouse inner ear development.* Gene Exp. Patterns, 2007. **7**(7): p. 798-807.
28. Mullen, R.D., et al., *Roles of the LHX3 and LHX4 LIM-homeodomain factors in pituitary development.* Mol. Cell. Endocrinol., 2007. **265–266**(0): p. 190-195.
29. Sheng, H.Z., et al., *Multistep Control of Pituitary Organogenesis.* Science, 1997. **278**(5344): p. 1809-1812.
30. Sharma, K., et al., *LIM Homeodomain Factors Lhx3 and Lhx4 Assign Subtype Identities for Motor Neurons.* Cell, 1998. **95**(6): p. 817-828.
31. Sheng, H.Z., et al., *Specification of Pituitary Cell Lineages by the LIM Homeobox Gene Lhx3.* Science, 1996. **272**(5264): p. 1004.
32. Li, H., et al., *Gsh-4 encodes a LIM-type homeodomain, is expressed in the developing central nervous system and is required for early postnatal survival.* EMBO J., 1994. **13**(12): p. 2876-2885.
33. Paylor, R., et al., *Learning impairments and motor dysfunctions in adult Lhx5-deficient mice displaying hippocampal disorganization.* Physiol. Behav., 2001. **73**(5): p. 781-792.
34. Zhao, Y., et al., *Control of Hippocampal Morphogenesis and Neuronal Differentiation by the LIM Homeobox Gene Lhx5.* Science, 1999. **284**(5417): p. 1155-1158.
35. Grigoriou, M., et al., *Expression and regulation of Lhx6 and Lhx7, a novel subfamily of LIM homeodomain encoding genes, suggests a role in mammalian head development.* Development, 1998. **125**(11): p. 2063-2074.
36. Zhou, C., et al., *Lhx8 mediated Wnt and TGFβ pathways in tooth development and regeneration.* Biomaterials, 2015. **63**: p. 35-46.
37. Zhao, Y., et al., *Isolated cleft palate in mice with a targeted mutation of the LIM homeobox gene Lhx8.* Proc. Natl. Acad. Sci., 1999. **96**(26): p. 15002-15006.
38. Birk, O.S., et al., *The LIM homeobox gene Lhx9 is essential for mouse gonad formation.* Nature, 2000. **403**: p. 909.
39. Bertuzzi, S., et al., *Characterization of Lhx9, a novel LIM/homeobox gene expressed by the pioneer neurons in the mouse cerebral cortex.* Mech. Dev., 1999. **81**(1): p. 193-198.
40. Rétaux, S., et al., *Lhx9: A Novel LIM-Homeodomain Gene Expressed in the Developing Forebrain.* J. Neurosci., 1999. **19**(2): p. 783-793.
41. Failli, V., et al., *Lhx9 and Lhx9α LIM-Homeodomain Factors: Genomic Structure, Expression Patterns, Chromosomal Localization, and Phylogenetic Analysis.* Genomics, 2000. **64**(3): p. 307-317.
42. Sheng, H.Z., et al., *Expression of murine Lhx5 suggests a role in specifying the forebrain.* Dev. Dyn., 1997. **208**(2): p. 266-77.

43. Hobert, O. and H. Westphal, *Functions of LIM-homeobox genes.* Trends Genet., 2000. **16**(2): p. 75-83.
44. Karlsson, O., et al., *Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo-and a Cys-His domain.* Nature, 1990. **344**(6269): p. 879-882.
45. Bach, I., *The LIM domain: regulation by association.* Mech. Dev., 2000. **91**(1–2): p. 5-17.
46. Gehring, W.J., M. Affolter, and T. Burglin, *Homeodomain proteins.* Annu. Rev. Biochem, 1994. **63**: p. 487-526.
47. Krishna, S.S., I. Majumdar, and N.V. Grishin, *Structural classification of zinc fingers: SURVEY AND SUMMARY.* Nucleic Acids Res., 2003. **31**(2): p. 532-550.
48. Dawid, I.B., J.J. Breen, and R. Toyama, *LIM domains: multiple roles as adapters and functional modifiers in protein interactions.* Trends Genet., 1998. **14**(4): p. 156-162.
49. Deane, J.E., et al., *Design, production and characterization of FLIN2 and FLIN4: the engineering of intramolecular ldb1:LMO complexes.* Protein Eng., 2001. **14**(7): p. 493-9.
50. Matthews, J.M. and J.R. Potts, *The tandem β-zipper: Modular binding of tandem domains and linear motifs.* FEBS Lett., 2013. **587**(8): p. 1164-1171.
51. Matthews, J.M. and J.E. Visvader, *LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins.* EMBO Rep., 2003. **4**(12): p. 1132-1137.
52. Bhati, M., et al., *Solution structure of the LIM-homeodomain transcription factor complex Lhx3/Ldb1 and the effects of a pituitary mutation on key Lhx3 interactions.* PLoS One, 2012. **7**(7).
53. Bhati, M., et al., *Crystallization of an Lhx3-Isl1 complex.* Acta Crysallogr. Sect. F, 2008. **64**(4): p. 297-299.
54. Gadd, M.S., et al., *A structural basis for the regulation of the LIM-homeodomain protein islet 1 (Isl1) by intra- and intermolecular interactions.* J. Biol. Chem., 2013. **288**(30): p. 21924-21935.
55. Deane, J.E., et al., *Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex.* EMBO J., 2004. **23**(18): p. 3589-98.
56. Deane, J.E., et al., *Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4.* EMBO J., 2003. **22**(9): p. 2224-33.
57. Gadd, M.S., et al., *Structural basis for partial redundancy in a class of transcription factors, the LIM homeodomain proteins, in neural cell type specification.* J. Biol. Chem., 2011. **286**(50): p. 42971-42980.
58. Bhati, M., et al., *Implementing the LIM code: The structural basis for cell type-specific assembly of LIM-homeodomain complexes.* EMBO J., 2008. **27**(14): p. 2018-2029.
59. Kadrmas, J.L. and M.C. Beckerle, *The LIM domain: from the cytoskeleton to the nucleus.* Nat. Rev. Mol. Cell Biol., 2004. **5**(11): p. 920.
60. Koch, B.J., J.F. Ryan, and A.D. Baxevanis, *The diversification of the LIM superclass at the base of the metazoa increased subcellular complexity and promoted multicellular specialization.* PLoS One, 2012. **7**(3): p. e33261.
61. Banerjee-Basu, S. and A.D. Baxevanis, *Molecular evolution of the homeodomain family of transcription factors.* Nucleic Acids Res., 2001. **29**(15): p. 3258-3269.
62. Derelle, R., et al., *Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes.* Evol. Dev., 2007. **9**(3): p. 212-219.
63. Gehring, W.J., et al., *Homeodomain-DNA recognition.* Cell, 1994. **78**(2): p. 211-223.
64. Clarke, N.D., *Covariation of residues in the homeodomain sequence family.* Protein Sci., 1995. **4**(11): p. 2269-2278.

65. Kissinger, C.R., et al., *Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: A framework for understanding homeodomain-DNA interactions.* Cell, 1990. **63**(3): p. 579-590.
66. Wolberger, C., *Homeodomain interactions.* Curr. Opin. Struct. Biol., 1996. **6**(1): p. 62-8.
67. Banerjee-Basu, S., et al., *The Homeodomain Resource: 2003 update.* Nucleic Acids Res., 2003. **31**(1): p. 304-306.
68. Laughon, A., *DNA binding specificity of homeodomains.* Biochem, 1991. **30**(48): p. 11357-11367.
69. Berger, M.F., et al., *Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences.* Cell, 2008. **133**(7): p. 1266-1276.
70. Dragan, A.I., et al., *Forces Driving the Binding of Homeodomains to DNA.* Biochem, 2006. **45**(1): p. 141-151.
71. Freyd, G., S.K. Kim, and H.R. Horvitz, *Novel cysteine-rich motif and homeodomain in the product of the Caenorhabditis elegans cell lineage gene lin-II.* Nature, 1990. **344**(6269): p. 876-879.
72. Ippel, H., et al., *The solution structure of the homeodomain of the rat insulin-gene enhancer protein Isl-1. Comparison with other homeodomains.* J. Mol. Biol., 1999. **288**(4): p. 689-703.
73. Bhati, M., et al., *Interactions between LHX3- and ISL1-family LIM-homeodomain transcription factors are conserved in Caenorhabditis elegans.* Sci. Rep., 2017. **7**(1): p. 4579.
74. Mann, T., R. Bodmer, and P. Pandur, *The Drosophila homolog of vertebrate Islet1 is a key component in early cardiogenesis.* Development, 2009. **136**(2): p. 317-326.
75. Simossis, V.A. and J. Heringa, *PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information.* Nucleic Acids Res., 2005. **33**(suppl_2): p. W289-W294.
76. Friedrich, F.W., et al., *A novel genetic variant in the transcription factor Islet-1 exerts gain of function on myocyte enhancer factor 2C promoter activity.* Eur. J. Heart Failure, 2013. **15**(3): p. 267-276.
77. Shimomura, H., et al., *Nonsense mutation of islet-1 gene (Q310X) found in a type 2 diabetic patient with a strong family history.* Diabetes, 2000. **49**(9): p. 1597-1600.
78. Thor, S., et al., *The homeodomain LIM protein Isl-1 is expressed in subsets of neurons and endocrine cells in the adult rat.* Neuron, 1991. **7**(6): p. 881-889.
79. Elshatory, Y., et al., *Expression of the LIM-homeodomain protein Isl1 in the developing and mature mouse retina.* J. Comp. Neurol., 2007. **503**(1): p. 182-197.
80. Zhuang, S., et al., *Expression of Isl1 during mouse development.* Gene Exp. Patterns, 2013. **13**(8): p. 407-412.
81. Korzh, V., T. Edlund, and S. Thor, *Zebrafish primary neurons initiate expression of the LIM homeodomain protein Isl-1 at the end of gastrulation.* Development, 1993. **118**: p. 417-425.
82. Zhang, Y., et al., *MicroRNA-31 suppresses the self-renewal capability of α2δ1(+) liver tumor-initiating cells by targeting ISL1.* Oncotarget, 2017. **8**(50): p. 87647-87657.
83. Shi, Q., et al., *ISL1, a novel regulator of CCNB1, CCNB2 and c-MYC genes, promotes gastric cancer cell proliferation and tumor growth.* Oncotarget, 2016.
84. Kondratyeva, L.G., et al., *Dependence of expression of regulatory master genes of embryonic development in pancreatic cancer cells on the intracellular concentration of the master regulator PDX1.* Dokl. Biochem. Biophys., 2017. **475**(1): p. 259-263.

85. Kitchen, M.O., et al., *Methylation of HOXA9 and ISL1 Predicts Patient Outcome in High-Grade Non-Invasive Bladder Cancer.* PLoS One, 2015. **10**(9): p. e0137003.
86. Ericson, J., et al., *Early stages of motor neuron differentiation revealed by expression of homeobox gene Islet-1.* Science, 1992. **256**(5063): p. 1555-1560.
87. Roelink, H., et al., *Floor plate and motor neuron induction by vhh-1, a vertebrate homolog of hedgehog expressed by the notochord.* Cell, 1994. **76**(4): p. 761-775.
88. Maden, M., *Retinoid signalling in the development of the central nervous system.* Nat. Rev. Neurosci., 2002. **3**(11): p. 843-853.
89. Echelard, Y., et al., *Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity.* Cell, 1993. **75**(7): p. 1417-1430.
90. Ericson, J., et al., *Two Critical Periods of Sonic Hedgehog Signaling Required for the Specification of Motor Neuron Identity.* Cell, 1996. **87**(4): p. 661-673.
91. Zhadanov, A.B., et al., *Expression pattern of the murine LIM class homeobox gene Lhx3 in subsets of neural and neuroendocrine tissues.* Dev. Dyn., 1995. **202**(4): p. 354-364.
92. Yamada, T., et al., *Control of cell pattern in the neural tube: motor neuron induction by diffusible factors from notochord and floor plate.* Cell, 1993. **73**(4): p. 673-686.
93. Lee, S., et al., *Fusion protein Isl1–Lhx3 specifies motor neuron fate by inducing motor neuron genes and concomitantly suppressing the interneuron programs.* Proc. Natl. Acad. Sci., 2012. **109**(9): p. 3383-3388.
94. Roy, A., et al., *Onecut transcription factors act upstream of Isl1 to regulate spinal motoneuron diversification.* Development, 2012. **139**.
95. Kim, N., et al., *Functional Diversification of Motor Neuron-specific Isl1 Enhancers during Evolution.* PLoS Genet., 2015. **11**(10): p. e1005560.
96. Nishi, Y., et al., *A direct fate exclusion mechanism by Sonic hedgehog-regulated transcriptional repressors.* Development, 2015. **142**(19): p. 3286-3293.
97. Ericson, J., et al., *Pax6 Controls Progenitor Cell Identity and Neuronal Fate in Response to Graded Shh Signaling.* Cell, 1997. **90**(1): p. 169-180.
98. Lee, S.-K. and S.L. Pfaff, *Synchronization of Neurogenesis and Motor Neuron Specification by Direct Coupling of bHLH and Homeodomain Transcription Factors.* Neuron, 2003. **38**(5): p. 731-745.
99. Lee, S., et al., *A Regulatory Network to Segregate the Identity of Neuronal Subtypes.* Dev. Cell, 2008. **14**(6): p. 877-889.
100. Pfaff, S. and C. Kintner, *Neuronal diversification: development of motor neuron subtypes.* Curr. Opin. Neurobiol., 1998. **8**(1): p. 27-36.
101. Seo, S.Y., B. Lee, and S. Lee, *Critical Roles of the LIM Domains of Lhx3 in Recruiting Coactivators to the Motor Neuron-Specifying Isl1-Lhx3 Complex.* Mol. Cell. Biol., 2015. **35**(20): p. 3579-3589.
102. Roberson, M.S., et al., *Activation of the glycoprotein hormone alpha-subunit promoter by a LIM-homeodomain transcription factor.* Mol. Cell. Biol., 1994. **14**(5): p. 2985-2993.
103. Yin, Y., et al., *Impact of cytosine methylation on DNA binding specificities of human transcription factors.* Science, 2017. **356**(6337).
104. Wang, M. and D.J. Drucker, *The LIM Domain Homeobox Gene isl-1 Is a Positive Regulator of Islet Cell-specific Proglucagon Gene Transcription.* J. Biol. Chem., 1995. **270**(21): p. 12646-12652.
105. Mazzoni, E.O., et al., *Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity.* Nat. Neurosci., 2013. **16**(9): p. 1219-1227.

106. Erb, M., et al., *The Isl1-Lhx3 Complex Promotes Motor Neuron Specification by Activating Transcriptional Pathways that Enhance Its Own Expression and Formation.* eNeuro, 2017. **4**(2).
107. Clovis, Y.M., et al., *Chx10 Consolidates V2a Interneuron Identity through Two Distinct Gene Repression Modes.* Cell Rep., 2016. **16**(6): p. 1642-1652.
108. Arber, S., et al., *Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity.* Neuron, 1999. **23**(4): p. 659-74.
109. Thiebes, K.P., et al., *miR-218 is Essential to Establish Motor Neuron Fate as a Downstream Effector of Isl1-Lhx3.* Nat. Comm., 2015. **6**: p. 7718-7718.
110. Nam, H. and S. Lee, *Identification of STAM1 as a novel effector of ventral projection of spinal motor neurons.* Development, 2016. **143**(13): p. 2334-43.
111. Song, M.-R., et al., *Islet-to-LMO stoichiometries control the function of transcription complexes that specify motor neuron and V2a interneuron identity.* Development, 2009. **136**(17): p. 2923-2932.
112. Thaler, J., et al., *Active Suppression of Interneuron Programs within Developing Motor Neurons Revealed by Analysis of Homeodomain Factor HB9.* Neuron, 1999. **23**(4): p. 675-687.
113. Lee, S., et al., *STAT3 promotes motor neuron differentiation by collaborating with motor neuron-specific LIM complex.* Proc. Natl. Acad. Sci., 2013. **110**.
114. Cai, C.-L., et al., *Isl1 Identifies a Cardiac Progenitor Population that Proliferates Prior to Differentiation and Contributes a Majority of Cells to the Heart.* Dev. Cell, 2003. **5**(6): p. 877-889.
115. Sun, Y., et al., *A central role for Islet1 in sensory neuron development linking sensory and spinal gene regulatory programs.* Nat. Neurosci., 2008. **11**(11): p. 1283-1293.
116. Srivastava, D., *Making or Breaking the Heart: From Lineage Determination to Morphogenesis.* Cell, 2006. **126**(6): p. 1037-1048.
117. Buckingham, M., S. Meilhac, and S. Zaffran, *Building the mammalian heart from two sources of myocardial cells.* Nat. Rev. Gen., 2005. **6**: p. 826.
118. Kelly, R.G., N.A. Brown, and M.E. Buckingham, *The Arterial Pole of the Mouse Heart Forms from Fgf10-Expressing Cells in Pharyngeal Mesoderm.* Dev. Cell, 2001. **1**(3): p. 435-440.
119. Zaffran, S., et al., *Right ventricular myocardium derives from the anterior heart field.* Circul. Res., 2004. **95**(3): p. 261-8.
120. Sun, Y., et al., *Islet 1 is expressed in distinct cardiovascular lineages, including pacemaker and coronary vascular cells.* Dev. Biol., 2007. **304**(1): p. 286-296.
121. Zhou, B., et al., *Nkx2-5- and Isl1-expressing cardiac progenitors contribute to proepicardium.* BBRC, 2008. **375**(3): p. 450-453.
122. Bu, L., et al., *Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages.* Nature, 2009. **460**.
123. Ma, Q., B. Zhou, and W.T. Pu, *Reassessment of Isl1 and Nkx2-5 cardiac fate maps using a Gata4-based reporter of Cre activity.* Dev. Biol., 2008. **323**(1): p. 98-104.
124. Laugwitz, K.-L., et al., *Islet1 cardiovascular progenitors: a single source for heart lineages?* Development, 2008. **135**(2): p. 193-205.
125. Lin, L., et al., *Isl1 is upstream of sonic hedgehog in a pathway required for cardiac morphogenesis.* Dev. Biol., 2006. **295**(2): p. 756-763.
126. Dodou, E., et al., *Mef2c is a direct transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse embryonic development.* Development, 2004. **131**(16): p. 3931-3942.
127. Lin, Q., et al., *Control of Mouse Cardiac Morphogenesis and Myogenesis by Transcription Factor MEF2C.* Science, 1997. **276**(5317): p. 1404-1407.

128. Takeuchi, J.K., et al., *Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development.* Development, 2005. **132**(10): p. 2463-2474.
129. Golzio, C., et al., *ISL1 Directly Regulates FGF10 Transcription during Human Cardiac Outflow Formation.* PLoS One, 2012. **7**(1): p. e30677.
130. Watanabe, Y., et al., *Fibroblast growth factor 10 gene regulation in the second heart field by Tbx1, Nkx2-5, and Islet1 reveals a genetic switch for down-regulation in the myocardium.* Proc. Natl. Acad. Sci., 2012. **109**(45): p. 18273-18280.
131. Wang, Y., et al., *ISL1 and JMJD3 synergistically control cardiac differentiation of embryonic stem cells.* Nucleic Acids Res., 2016. **44**(14): p. 6741-6755.
132. Kang, J., et al., *Isl1 is a direct transcriptional target of Forkhead transcription factors in second heart field-derived mesoderm.* Dev. Biol., 2009. **334**(2): p. 513-522.
133. Liu, Y., et al., *POU homeodomain protein OCT1 modulates islet 1 expression during cardiac differentiation of P19CL6 cells.* Cell Mol. Life Sci., 2011. **68**.
134. Lin, L., et al., *β-Catenin directly regulates Islet1 expression in cardiovascular progenitors and is required for multiple aspects of cardiogenesis.* Proc. Natl. Acad. Sci., 2007. **104**(22): p. 9313-9318.
135. Lu, H., et al., *Wnt-promoted Isl1 expression through a novel TCF/LEF1 binding site and H3K9 acetylation in early stages of cardiomyocyte differentiation of P19CL6 cells.* Mol. Cell. Biochem., 2014: p. 1-10.
136. Pacheco-Leyva, I., et al., *CITED2 Cooperates with ISL1 and Promotes Cardiac Differentiation of Mouse Embryonic Stem Cells.* Stem Cell Rep., 2016. **7**(6): p. 1037-1049.
137. Caputo, L., et al., *The Isl1/Ldb1 Complex Orchestrates Genome-wide Chromatin Organization to Instruct Differentiation of Multipotent Cardiac Progenitors.* Cell Stem Cell, 2015. **17**(3): p. 287-299.
138. Petryszak, R., et al., *Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants.* Nucleic Acids Res., 2015. **44**(D1): p. D746-D752.
139. Witzel, H.R., et al., *The LIM protein Ajuba restricts the second heart field progenitor pool by regulating Isl1 activity.* Dev. Cell, 2012. **23**(1): p. 58-70.
140. Grapin-Botton, A. and D.A. Melton, *Endoderm development: from patterning to organogenesis.* Trends Genet., 2000. **16**(3): p. 124-130.
141. Gittes, G.K., *Developmental biology of the pancreas: A comprehensive review.* Dev. Biol., 2009. **326**(1): p. 4-35.
142. Wilson, M.E., D. Scheel, and M.S. German, *Gene expression cascades in pancreatic development.* Mech. Dev., 2003. **120**(1): p. 65-80.
143. Ahlgren, U., et al., *Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells.* Nature, 1997. **385**: p. 257.
144. Du, A., et al., *Islet-1 is Required for the Maturation, Proliferation, and Survival of the Endocrine Pancreas.* Diabetes, 2009. **58**(9): p. 2059-2069.
145. Yang, Z., et al., *ISL-1 promotes pancreatic islet cell proliferation by forming an ISL-1/Set7/9/PDX-1 complex.* Cell Cyc., 2015. **14**(24): p. 3820-3829.
146. Liu, J., E.R. Walp, and C.L. May, *Elevation of transcription factor Islet-1 levels in vivo increases beta-cell function but not beta-cell mass.* Islets, 2012. **4**.
147. Ediger, B.N., et al., *Islet-1 is essential for pancreatic β-cell function.* Diabetes, 2014.
148. Guo, T., et al., *ISL1 promotes pancreatic islet cell proliferation.* PLoS One, 2011. **6**.
149. diIorio, P.J., et al., *Sonic hedgehog Is Required Early in Pancreatic Islet Development.* Dev. Biol., 2002. **244**(1): p. 75-84.
150. Edlund, H., *Transcribing pancreas.* Diabetes, 1998. **47**(12): p. 1817-1823.

151. Grzeskowiak, R., et al., *Insulin Responsiveness of the Glucagon Gene Conferred by Interactions between Proximal Promoter and More Distal Enhancer-like Elements Involving the Paired-domain Transcription Factor Pax6.* J. Biol. Chem., 2000. **275**(39): p. 30037-30045.
152. Habener, J.F., D.M. Kemp, and M.K. Thomas, *Minireview: Transcriptional Regulation in Pancreatic Development.* Endocrinol., 2005. **146**(3): p. 1025-1034.
153. Zhang, H., et al., *The LIM-Homeodomain Protein ISL1 Activates Insulin Gene Promoter Directly through Synergy with BETA2.* J. Mol. Biol., 2009. **392**(3): p. 566-577.
154. Galloway, J.R., et al., *SSBP3 Interacts With Islet-1 and Ldb1 to Impact Pancreatic β-Cell Target Genes.* Mol. Endocrinol., 2015. **29**(12): p. 1774-1786.
155. Yan, C., et al., *Protein Inhibitor of Activated STAT Y (PIASy) Regulates Insulin Secretion by Interacting with LIM Homeodomain Transcription Factor Isl1.* Sci. Rep., 2016. **6**: p. 39308.
156. Hunter, C.S., et al., *Islet α-, β-, and δ-cell development is controlled by the Ldb1 coregulator, acting primarily with the islet-1 transcription factor.* Diabetes, 2013. **62**(3): p. 875-886.
157. Ediger, B.N., et al., *LIM domain–binding 1 maintains the terminally differentiated state of pancreatic β cells.* J. Clin. Invest., 2017. **127**(1): p. 215-229.
158. Chen, L., et al., *Ssdp proteins interact with the LIM-domain-binding protein Ldb1 to regulate development.* Proc. Natl. Acad. Sci., 2002. **99**(22): p. 14320-14325.
159. van Meyel, D.J., J.B. Thomas, and A.D. Agulnick, *Ssdp proteins bind to LIM-interacting co-factors and regulate the activity of LIM-homeodomain protein complexes in vivo.* Development, 2003. **130**(9): p. 1915-1925.
160. Cai, Y., et al., *Single-Stranded DNA-Binding Proteins (SSBPs) Regulate the Abundance of the LIM-Homeodomain Protein LHX2 and Augment Its Transcriptional Activity.* Blood, 2007. **110**(11): p. 1239-1239.
161. Gong, Z., C.-c. Hui, and C.L. Hew, *Presence of isl-1-related LIM Domain Homeobox Genes in Teleost and Their Similar Patterns of Expression in Brain and Spinal Cord.* J. Biol. Chem., 1995. **270**(7): p. 3335-3345.
162. Puelles, L. and J.L.R. Rubenstein, *Expression patterns of homeobox and other putative regulatory genes in the embryonic mouse forebrain suggest a neuromeric organization.* Trends Neurosci., 1993. **16**(11): p. 472-479.
163. Wu, Y., et al., *LIM Homeodomain Transcription Factor Isl-1 Enhances Follicle Stimulating Hormone-β and Luteinizing Hormone-β Gene Expression and Mediates the Activation of Leptin on Gonadotropin Synthesis.* Endocrinol., 2010. **151**(10): p. 4787-4800.
164. Treier, M., et al., *Hedgehog signaling is required for pituitary gland development.* Development, 2001. **128**(3): p. 377-386.
165. Ericson, J., et al., *Sonic hedgehog induces the differentiation of ventral forebrain neurons: A common signal for ventral patterning within the neural tube.* Cell, 1995. **81**(5): p. 747-756.
166. Ericson, J., et al., *Integrated FGF and BMP signaling controls the progression of progenitor cell differentiation and the emergence of pattern in the embryonic anterior pituitary.* Development, 1998. **125**(6): p. 1005-1015.
167. Dattani, M. and I. Robinson, *The molecular basis for developmental disorders of the pituitary gland in man.* Clin. Genet., 2000. **57**(5): p. 337-346.
168. Wang, H.F. and F.C. Liu, *Developmental restriction of the LIM homeodomain transcription factor Islet-1 expression to cholinergic neurons in the rat striatum.* Neuroscience, 2001. **103**(4): p. 999-1016.

169. Elshatory, Y. and L. Gan, *The LIM-homeobox gene Islet-1 is required for the development of restricted forebrain cholinergic neurons.* J. Neurosci., 2008. **28**(13): p. 3291-7.
170. Ehrman, L.A., et al., *The LIM homeobox gene Isl1 is required for the correct development of the striatonigral pathway in the mouse.* Proc. Natl. Acad. Sci., 2013. **110**(42): p. E4026-E4035.
171. Fragkouli, A., et al., *LIM homeodomain transcription factor-dependent specification of bipotential MGE progenitors into cholinergic and GABAergic striatal interneurons.* Development, 2009. **136**(22): p. 3841-3851.
172. Cho, H.-H., et al., *Isl1 Directly Controls a Cholinergic Neuronal Identity in the Developing Forebrain and Spinal Cord by Forming Cell Type-Specific Complexes.* PLoS Genet., 2014. **10**(4): p. e1004280.
173. Nakagawa, Y. and D.D. O'Leary, *Combinatorial expression patterns of LIM-homeodomain and other regulatory genes parcellate developing thalamus.* J. Neurosci., 2001. **21**(8): p. 2711-2725.
174. Gay, F., et al., *The LIM/Homeodomain Protein Islet-1 Modulates Estrogen Receptor Functions.* Mol. Endocrinol., 2000. **14**(10): p. 1627-1648.
175. Davis, A.M., et al., *Differential Colocalization of Islet-1 and Estrogen Receptor α in the Murine Preoptic Area and Hypothalamus during Development.* Endocrinol., 2004. **145**(1): p. 360-366.
176. Dong, J., S.L. Asa, and D.J. Drucker, *Islet Cell and Extrapancreatic Expression of the LIM Domain Homeobox Gene isl-1.* Mol. Endocrinol., 1991. **5**(11): p. 1633-1641.
177. Kaku, Y., et al., *Islet1 Deletion Causes Kidney Agenesis and Hydroureter Resembling CAKUT.* J. Am. Soc. Nephr., 2013. **24**(8): p. 1242-1249.
178. Zhang, R., et al., *ISL1 is a major susceptibility gene for classic bladder exstrophy and a regulator of urinary tract development.* Sci. Rep., 2017. **7**: p. 42170.
179. Miyazaki, Y., et al., *Bone morphogenetic protein 4 regulates the budding site and elongation of the mouse ureter.* J. Clin. Invest., 2000. **105**(7): p. 863-873.
180. Nishinakamura, R. and M. Sakaguchi, *BMP signaling and its modifiers in kidney development.* Pediatr Nephrol, 2014. **29**(4): p. 681-686.
181. Ching, S.T., et al., *Isl1 mediates mesenchymal expansion in the developing external genitalia via regulation of Bmp4, Fgf10 and Wnt5a.* Hum. Mol. Genet., 2018. **27**(1): p. 107-119.
182. Galli-Resta, L., et al., *Mosaics of islet-1-expressing amacrine cells assembled by short-range cellular interactions.* J. Neurosci., 1997. **17**(20): p. 7831-7838.
183. Bejarano-Escobar, R., et al., *Expression and function of the LIM-homeodomain transcription factor Islet-1 in the developing and mature vertebrate retina.* EEyeR, 2015. **138**(0): p. 22-31.
184. Elshatory, Y., et al., *Islet-1 controls the differentiation of retinal bipolar and cholinergic amacrine cells.* J. Neurosci., 2007. **27**(46): p. 12707-12720.
185. Pan, L., et al., *ISL1 and BRN3B co-regulate the differentiation of murine retinal ganglion cells.* Development, 2008. **135**(11): p. 1981-1990.
186. Mu, X., et al., *Gene-regulation logic in retinal ganglion cell development: Isl1 defines a critical branch distinct from but overlapping with Pou4f2.* Proc. Natl. Acad. Sci., 2008. **105**(19): p. 6942-6947.
187. Li, R., et al., *Isl1 and Pou4f2 Form a Complex to Regulate Target Genes in Developing Retinal Ganglion Cells.* PLoS One, 2014. **9**(3): p. e92105.
188. Radde-Gallwitz, K., et al., *Expression of Islet1 marks the sensory and neuronal lineages in the mammalian inner ear.* J. Comp. Neurol., 2004. **477**(4): p. 412-421.

189. Li, H., et al., *Islet-1 expression in the developing chicken inner ear.* J. Comp. Neurol., 2004. **477**(1): p. 1-10.
190. Deng, M., et al., *Requirement for Lmo4 in the vestibular morphogenesis of mouse inner ear.* Dev. Biol., 2010. **338**(1): p. 38-49.
191. Das, P. and C.L. May, *Expression analysis of the Islet-1 gene in the developing and adult gastrointestinal tract.* Gene Exp. Patterns, 2011. **11**(3): p. 244-254.
192. Terry, N.A., et al., *Impaired enteroendocrine development in intestinal-specific Islet1 mouse mutants causes impaired glucose homeostasis.* Am. J. Physiol. Gastro. Liv. Physiol., 2014. **307**(10): p. G979-G991.
193. Ye, D.Z. and K.H. Kaestner, *Foxa1 and Foxa2 Control the Differentiation of Goblet and Enteroendocrine L- and D-Cells in Mice.* Gastroenterology, 2009. **137**(6): p. 2052-2062.
194. Li, F., et al., *ISLET1-Dependent beta-Catenin/Hedgehog Signaling Is Required for Outgrowth of the Lower Jaw.* Mol. Cell. Biol., 2017. **37**(8).
195. Akiyama, R., et al., *Distinct populations within Isl1 lineages contribute to appendicular and facial skeletogenesis through the β-catenin pathway.* Dev. Biol., 2014. **387**(1): p. 37-48.
196. Mitsiadis, T.A., et al., *Role of Islet1 in the patterning of murine dentition.* Development, 2003. **130**(18): p. 4451-4460.
197. Narkis, G., et al., *Isl1 and Ldb Co-regulators of transcription are essential early determinants of mouse limb development.* Dev. Dyn., 2012. **241**(4): p. 787-791.
198. Nathan, E., et al., *The contribution of Islet1-expressing splanchnic mesoderm cells to distinct branchiomeric muscles reveals significant heterogeneity in head muscle development.* Development, 2008. **135**(4): p. 647-657.
199. Bhattacherjee, V., et al., *Neural crest and mesoderm lineage-dependent gene expression in orofacial development.* Differentiation, 2007. **75**(5): p. 463-477.
200. Hao, A., et al., *The LIM/Homeodomain Protein Islet1 Recruits Janus Tyrosine Kinases and Signal Transducer and Activator of Transcription 3 and Stimulates Their Activities.* Mol. Biol. Cell, 2005. **16**(4): p. 1569-1583.
201. Zhang, Q., et al., *ISL-1 is overexpressed in non-Hodgkin lymphoma and promotes lymphoma cell proliferation by forming a p-STAT3/p-c-Jun/ISL-1 complex.* Mol. Cancer, 2014. **13**(1): p. 1-15.
202. Ando, K., et al., *Isolation and characterization of an alternatively spliced variant of transcription factor Islet-1.* J. Mol. Endocrin., 2003. **31**(3): p. 419-425.
203. Jung, Y., et al., *Isl1β Overexpression With Key β Cell Transcription Factors Enhances Glucose-Responsive Hepatic Insulin Production and Secretion.* Endocrinol., 2018. **159**(2): p. 869-882.
204. Tanabe, Y., C. William, and T.M. Jessell, *Specification of Motor Neuron Identity by the MNR2 Homeodomain Protein.* Cell, 1998. **95**(1): p. 67-80.
205. Thaler, J.P., et al., *A Postmitotic Role for Isl-Class LIM Homeodomain Proteins in the Assignment of Visceral Spinal Motor Neuron Identity.* Neuron, 2004. **41**(3): p. 337-350.
206. Huang, M., et al., *Diverse expression patterns of LIM-homeodomain transcription factors (LIM-HDs) in mammalian inner ear development.* Dev. Dyn., 2008. **237**(11): p. 3305-3312.
207. Edqvist, P.H.D., S.M. Myers, and F. Hallbook, *Early identification of retinal subtypes in the developing, pre-laminated chick retina using the transcription factors Prox1, Lim1, Ap2a, Pax6, Isl1, Isl2, Lim3 and Chx10.* Eur. J. Histochem., 2006. **50**(2): p. 147-54.

208. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat. Methods, 2009. **6**(5): p. 343.
209. Smith, N.C., *Investigating interactions facilitating the function of the developmental transcription factor Isl1*, in *School of Molecular Bioscience*. 2014, University of Sydney: Sydney.
210. Joseph, S., *The LMO4-DEAF1 complex and its biological impacts*, in *School of Molecular Bioscience*. 2013, University of Sydney: Sydney.
211. *Matchmaker Gold Yeast Two-Hybrid System User Manual*. 2013, Clontech Laboratories, Inc.: Mountain View, CA. p. 26.
212. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids.* Gene, 1990. **96**(1): p. 23-28.
213. von der Haar, T., *Optimized Protein Extraction for Quantitative Proteomics of Yeasts.* PLoS One, 2007. **2**(10): p. e1078.
214. Drury, T., *Characterisation of the DNA-binding properties of a LIM-transcriptional complex*, in *School of Molecular Bioscience*. 2011, University of Sydney: Sydney.
215. Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server*, in *The proteomics protocols handbook*. 2005, Springer. p. 571-607.
216. Gorry, P.A., *General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method.* Anal. Chem., 1990. **62**(6): p. 570-573.
217. Fields, S. and O.-k. Song, *A novel genetic system to detect protein-protein interactions.* Nature, 1989. **340**: p. 245.
218. Burstone, M., *Enzyme Histochemistry and its Application in the Study of Neoplasms*. 1962, New York: Academic Press. 621.
219. Julin, D., *Blue-White Selection*, in *Molecular Life Sciences: An Encyclopedic Reference*, E. Bell, Editor. 2014, Springer New York: New York, NY. p. 1-2.
220. Aho, S., et al., *A Novel Reporter Gene MEL1 for the Yeast Two-Hybrid System.* Anal. Biochem., 1997. **253**(2): p. 270-272.
221. Bendixen, C., S. Gangloff, and R. Rothstein, *A yeast mating-selection scheme for detection of protein – protein interactions.* Nucleic Acids Res., 1994. **22**(9): p. 1778-1779.
222. Nisevic, I., *LIM domain-mediated interactions between transcription factors*, in *School of Life and Environmental Sciences*. 2016, University of Sydney: Sydney.
223. Zhulidov, P.A., et al., *Simple cDNA normalization using kamchatka crab duplex-specific nuclease.* Nucleic Acids Res., 2004. **32**(3): p. e37-e37.
224. Gnugge, R. and F. Rudolf, *Saccharomyces cerevisiae Shuttle vectors.* Yeast, 2017. **34**(5): p. 205-221.
225. Singh, M.V. and P. Anthony Weil, *A method for plasmid purification directly from yeast.* Anal. Biochem., 2002. **307**(1): p. 13-17.
226. Dinman, J.D., *Programmed Ribosomal Frameshifting Goes Beyond Viruses: Organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift.* Microbe (Wash. D.C.), 2006. **1**(11): p. 521-527.
227. Mellacheruvu, D., et al., *The CRAPome: a Contaminant Repository for Affinity Purification Mass Spectrometry Data.* Nat. Methods, 2013. **10**(8): p. 730-736.
228. Thul, P.J., et al., *A subcellular map of the human proteome.* Science, 2017. **356**(6340).
229. The UniProt Consortium, *UniProt: the universal protein knowledgebase.* Nucleic Acids Res., 2017. **45**(D1): p. D158-D169.
230. Fink, J.L., et al., *Towards defining the nuclear proteome.* Genome Biol., 2008. **9**(1): p. R15.

231. Golemis, E.A., I. Serebriiskii, and S.F. Law, *The yeast two-hybrid system: criteria for detecting physiologically significant protein-protein interactions.* Curr. Issues Mol. Biol., 1999. **1**(1-2): p. 31-45.
232. Zhong, J., et al., *A Strategy for Constructing Large Protein Interaction Maps Using the Yeast Two-Hybrid System:Regulated Expression Arrays and Two-Phase Mating.* Genome Res., 2003. **13**(12): p. 2691-2699.
233. Van Criekinge, W. and R. Beyaert, *Yeast Two-Hybrid: State of the Art.* Bio. Proc. Online, 1999. **2**: p. 1-38.
234. Ito, T., et al., *Toward a protein–protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins.* Proc. Natl. Acad. Sci., 2000. **97**(3): p. 1143-1147.
235. Brückner, A., et al., *Yeast Two-Hybrid, a Powerful Tool for Systems Biology.* Int. J. Mol. Sci., 2009. **10**(6): p. 2763.
236. Ketteler, R., *On programmed ribosomal frameshifting: the alternative proteomes.* Front. Genet., 2012. **3**: p. 242-242.
237. Marmorstein, R., et al., *DNA recognition by GAL4: structure of a protein-DNA complex.* Nature, 1992. **356**: p. 408.
238. Glowacki, G., et al., *The family of toxin-related ecto-ADP-ribosyltransferases in humans and the mouse.* Protein Sci., 2002. **11**(7): p. 1657-1670.
239. Bothwell, A., P.E. Pace, and K.P. LeClair, *Isolation and expression of an IFN-responsive Ly-6C chromosomal gene.* J. Immunol., 1988. **140**(8): p. 2815-2820.
240. Ishibashi, K., et al., *Identification of a new multigene four-transmembrane family (MS4A) related to CD20, HTm4 and β subunit of the high-affinity IgE receptor.* Gene, 2001. **264**(1): p. 87-93.
241. Satoh-Horikawa, K., et al., *Nectin-3, a new member of immunoglobulin-like cell adhesion molecules that shows homophilic and heterophilic cell-cell adhesion activities.* J. Biol. Chem., 2000. **275**(14): p. 10291-9.
242. Lee, T.-H.D., et al., *Tissue Expression of the Novel Serine Carboxypeptidase Scpep1.* J. Histochem. Cytochem., 2006. **54**(6): p. 701-711.
243. Kobayashi, Y., et al., *Isolation and Characterization of Polyspecific Mouse Organic Solute Carrier Protein 1 (mOscp1).* Drug Metab. Disposition, 2007. **35**(7): p. 1239-1245.
244. Lussier, M.P., et al., *Ubiquitin ligase RNF167 regulates AMPA receptor-mediated synaptic transmission.* Proc. Natl. Acad. Sci., 2012. **109**(47): p. 19426-31.
245. Eblimit, A., et al., *Spata7 is a retinal ciliopathy gene critical for correct RPGRIP1 localization and protein trafficking in the retina.* Hum. Mol. Genet., 2015. **24**(6): p. 1584-601.
246. McMichael, T.M., et al., *The palmitoyltransferase ZDHHC20 enhances interferon-induced transmembrane protein 3 (IFITM3) palmitoylation and antiviral activity.* J. Biol. Chem., 2017.
247. Luzio, J.P., P.R. Pryor, and N.A. Bright, *Lysosomes: fusion and function.* Nat. Rev. Mol. Cell Biol., 2007. **8**: p. 622.
248. Ciechanover, A., *Proteolysis: from the lysosome to ubiquitin and the proteasome.* Nat. Rev. Mol. Cell Biol., 2005. **6**: p. 79.
249. Ostendorff, H.P., et al., *Ubiquitination-dependent cofactor exchange on LIM homeodomain transcription factors.* Nature, 2002. **416**: p. 99.
250. Ostendorff, H.P., et al., *Dynamic expression of LIM cofactors in the developing mouse neural tube.* Dev. Dyn., 2006. **235**(3): p. 786-91.

251. Ji, A.X., et al., *Structural Insights into KCTD Protein Assembly and Cullin3 Recognition.* J. Mol. Biol., 2016. **428**(1): p. 92-107.
252. Wu, X., et al., *Stabilization of the E3 ubiquitin ligase Nrdp1 by the deubiquitinating enzyme USP8.* Mol. Cell. Biol., 2004. **24**(17): p. 7748-57.
253. Avvakumov, G.V., et al., *Amino-terminal Dimerization, NRDP1-Rhodanese Interaction, and Inhibited Catalytic Domain Conformation of the Ubiquitin-specific Protease 8 (USP8).* J. Biol. Chem., 2006. **281**(49): p. 38061-38070.
254. Das, J., *The role of mitochondrial respiration in physiological and evolutionary adaptation.* Bioessays, 2006. **28**(9): p. 890-901.
255. Fenwick, C., et al., *A subclass of Ras proteins that regulate the degradation of IkappaB.* Science, 2000. **287**(5454): p. 869-73.
256. Chen, Y., et al., *Inhibition of NF-κB Activity by IκBβ in Association with κB-Ras.* Mol. Cell. Biol., 2004. **24**(7): p. 3048-3056.
257. Girard, J.-P. and T.A. Springer, *Cloning from purified high endothelial venule cells of hevin, a close relative of the antiadhesive extracellular matrix protein SPARC.* Immunity, 1995. **2**(1): p. 113-123.
258. Nishiu, J., T. Tanaka, and Y. Nakamura, *Identification of a Novel Gene (ECM2) Encoding a Putative Extracellular Matrix Protein Expressed Predominantly in Adipose and Female-Specific Tissues and Its Chromosomal Localization to 9q22.3.* Genomics, 1998. **52**(3): p. 378-381.
259. Martinek, N., et al., *Is SPARC an Evolutionarily Conserved Collagen Chaperone?* J. Dent. Res., 2007. **86**(4): p. 296-305.
260. Oritani, K. and P.W. Kincade, *Lymphopoiesis and Matrix Glycoprotein SC1/ECM2.* Leuk. Lymphoma, 1998. **32**(1-2): p. 1-7.
261. Hurley, P.J., et al., *Secreted protein, acidic and rich in cysteine-like 1 (SPARCL1) is down regulated in aggressive prostate cancers and is prognostic for poor clinical outcome.* Proc. Natl. Acad. Sci., 2012. **109**(37): p. 14977-14982.
262. Yu, S.-J., et al., *SPARCL1, Shp2, MSH2, E-cadherin, p53, ADCY-2 and MAPK are prognosis-related in colorectal cancer.* WJG, 2011. **17**(15): p. 2028-2036.
263. Lundby, A., et al., *Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues.* Nat. Comm., 2012. **3**: p. 876.
264. Curmi, F. and R.J. Cauchi, *The multiple lives of DEAD-box RNA helicase DP103/DDX20/Gemin3.* Biochem. Soc. Trans., 2018. **46**(2): p. 329-341.
265. Charroux, B., et al., *Gemin3: A Novel Dead Box Protein That Interacts with Smn, the Spinal Muscular Atrophy Gene Product, and Is a Component of Gems.* J. Cell Biol., 1999. **147**(6): p. 1181-1194.
266. Cauchi, R.J., K.E. Davies, and J.-L. Liu, *A Motor Function for the DEAD-Box RNA Helicase, Gemin3, in Drosophila.* PLoS Genet., 2008. **4**(11): p. e1000265.
267. Kroiss, M., et al., *Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster.* Proc. Natl. Acad. Sci., 2008. **105**(29): p. 10045-10050.
268. Yan, X., et al., *A novel domain within the DEAD-box protein DP103 is essential for transcriptional repression and helicase activity.* Mol. Cell. Biol., 2003. **23**(1): p. 414-423.
269. Cai, Q., et al., *Epstein-Barr Virus Nuclear Antigen 3C Stabilizes Gemin3 to Block p53-mediated Apoptosis.* PLoS Path., 2011. **7**(12): p. e1002418.
270. Takata, A., et al., *A miRNA machinery component DDX20 controls NF-κB via microRNA-140 function.* BBRC, 2012. **420**(3): p. 564-569.

271. Gillian, A.L. and J. Svaren, *The Ddx20/DP103 Dead Box Protein Represses Transcriptional Activation by Egr2/Krox-20.* J. Biol. Chem., 2004. **279**(10): p. 9056-9063.
272. Mouillet, J.-F.o., et al., *DEAD-Box Protein-103 (DP103, Ddx20) Is Essential for Early Embryonic Development and Modulates Ovarian Morphology and Function.* Endocrinol., 2008. **149**(5): p. 2168-2175.
273. Vij, S. and A.K. Tyagi, *A20/AN1 zinc-finger domain-containing proteins in plants and animals represent common elements in stress response.* Funct. Integr. Genomics, 2008. **8**(3): p. 301-307.
274. Sun, Z.-Y.J., et al., *Solution Structure of the Cuz1 AN1 Zinc Finger Domain: An Exposed LDFLP Motif Defines a Subfamily of AN1 Proteins.* PLoS One, 2016. **11**(9): p. e0163660.
275. Sá-Moura, B., et al., *A conserved protein with AN1 zinc finger and ubiquitin-like domains modulates Cdc48 (p97) function in the ubiquitin-proteasome pathway.* J. Biol. Chem., 2013. **288**(47): p. 33682-33696.
276. Turakhiya, A., et al., *ZFAND1 Recruits p97 and the 26S Proteasome to Promote the Clearance of Arsenite-Induced Stress Granules.* Mol. Cell, 2018. **70**(5): p. 906-919.e7.
277. Lee-Chen, S.F., C.T. Yu, and K.Y. Jan, *Effect of arsenite on the DNA repair of UV-irradiated Chinese hamster ovary cells.* Mutagenesis, 1992. **7**(1): p. 51-55.
278. Camp, G.V., et al., *Localization of a gene for non-syndromic hearing loss (DFNA5) to chromosome 7p15.* Hum. Mol. Genet., 1995. **4**(11): p. 2159-2163.
279. Laer, L.V., et al., *Nonsyndromic hearing impairment is associated with a mutation in DFNA5.* Nat. Genet., 1998. **20**: p. 194.
280. Masuda, Y., et al., *The potential role of DFNA5, a hearing impairment gene, in p53-mediated cellular response to DNA damage.* J. Hum. Genet., 2006. **51**(8): p. 652-664.
281. Kimishige, A., et al., *Identification of DFNA5 as a target of epigenetic inactivation in gastric cancer.* Cancer Sci., 2007. **98**(1): p. 88-95.
282. Kim, M.S., et al., *Aberrant promoter methylation and tumor suppressive activity of the DFNA5 gene in colorectal carcinoma.* Oncogene, 2008. **27**: p. 3624.
283. de Beeck, K.O., et al., *The DFNA5 gene, responsible for hearing loss and involved in cancer, encodes a novel apoptosis-inducing protein.* Eur. J. Hum. Genet., 2011. **19**: p. 965.
284. Rogers, C., et al., *Cleavage of DFNA5 by caspase-3 during apoptosis mediates progression to secondary necrotic/pyroptotic cell death.* Nat. Comm., 2017. **8**: p. 14128.
285. Abrahams, B.S., et al., *Novel Vertebrate Genes and Putative Regulatory Elements Identified at Kidney Disease and NR2E1/fierce Loci.* Genomics, 2002. **80**(1): p. 45-53.
286. Khalimonchuk, O., A. Bird, and D.R. Winge, *Evidence for a Pro-oxidant Intermediate in the Assembly of Cytochrome Oxidase.* J. Biol. Chem., 2007. **282**(24): p. 17442-17449.
287. Cesnekova, J., et al., *The mammalian homologue of yeast Afg1 ATPase (lactation elevated 1) mediates degradation of nuclear-encoded complex IV subunits.* Biochem. J., 2016. **473**(6): p. 797-804.
288. Cesnekova, J., et al., *LACE1 interacts with p53 and mediates its mitochondrial translocation and apoptosis.* Oncotarget, 2016. **7**(30): p. 47687-47698.
289. Bertenshaw, G.P., et al., *Marked Differences between Metalloproteases Meprin A and B in Substrate and Peptide Bond Specificity.* J. Biol. Chem., 2001. **276**(16): p. 13248-13255.

290. Myasnikov, A.G., et al., *Structure–function insights reveal the human ribosome as a cancer target for antibiotics.* Nat. Comm., 2016. **7**: p. ncomms12856.
291. Kramer, G., et al., *The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins.* Nat. Struct. Mol. Biol., 2009. **16**: p. 589.
292. Zhang, G. and Z. Ignatova, *Folding at the birth of the nascent chain: coordinating translation with co-translational folding.* Curr. Opin. Struct. Biol., 2011. **21**(1): p. 25-31.
293. Matsuura, Y. and M. Stewart, *Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling.* EMBO J., 2005. **24**(21): p. 3681-9.
294. Lindsay, M.E., et al., *Npap60/Nup50 Is a Tri-Stable Switch that Stimulates Importin-α:β-Mediated Nuclear Protein Import.* Cell, 2002. **110**(3): p. 349-360.
295. Hicks, G.R. and N.V. Raikhel, *Protein import into the nucleus: an integrated view.* Annu. Rev. Cell Dev. Biol., 1995. **11**(1): p. 155-188.
296. Pumroy, R.A., et al., *Nucleoporin Nup50 stabilizes closed conformation of armadillo repeat 10 in importin alpha5.* J. Biol. Chem., 2012. **287**(3): p. 2022-31.
297. Tilo, M., et al., *Identification of a nuclear localization signal, RRMKWKK, in the homeodomain transcription factor PDX-1.* FEBS Lett., 1999. **461**(3): p. 229-234.
298. Liu, Y.-C., et al., *A MHC-encoded ubiquitin-like protein (FAT10) binds noncovalently to the spindle assembly checkpoint protein MAD2.* Proc. Natl. Acad. Sci., 1999. **96**(8): p. 4313-4318.
299. Hipp, M.S., et al., *FAT10, a ubiquitin-independent signal for proteasomal degradation.* Mol. Cell. Biol., 2005. **25**(9): p. 3483-91.
300. Raasi, S., G. Schmidtke, and M. Groettrup, *The ubiquitin-like protein FAT10 forms covalent conjugates and induces apoptosis.* J. Biol. Chem., 2001. **276**(38): p. 35334-43.
301. Chiu, Y.H., Q. Sun, and Z.J. Chen, *E1-L2 activates both ubiquitin and FAT10.* Mol. Cell, 2007. **27**(6): p. 1014-23.
302. Ross, M.J., et al., *Role of ubiquitin-like protein FAT10 in epithelial apoptosis in renal disease.* J. Am. Soc. Nephr., 2006. **17**(4): p. 996-1004.
303. Canaan, A., et al., *FAT10/diubiquitin-like protein-deficient mice exhibit minimal phenotypic differences.* Mol. Cell. Biol., 2006. **26**(13): p. 5180-5189.
304. Adams, J.C., B. Seed, and J. Lawler, *Muskelin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin-1.* EMBO J., 1998. **17**.
305. Kobayashi, N., et al., *RanBPM, Muskelin, p48EMLP, p44CTLH, and the armadillo-repeat proteins ARMC8α and ARMC8β are components of the CTLH complex.* Gene, 2007. **396**(2): p. 236-247.
306. Umeda, M., H. Nishitani, and T. Nishimoto, *A novel nuclear protein, Twa1, and Muskelin comprise a complex with RanBPM.* Gene, 2003. **303**.
307. Ledee, D.R., et al., *A Specific Interaction between Muskelin and the Cyclin-dependent Kinase 5 Activator p39 Promotes Peripheral Localization of Muskelin.* J. Biol. Chem., 2005. **280**(22): p. 21376-21383.
308. Valiyaveettil, M., et al., *Novel role of the muskelin–RanBP9 complex as a nucleocytoplasmic mediator of cell morphology regulation.* J. Cell Biol., 2008. **182**(4): p. 727-739.
309. Prag, S., Georgina D.M. Collett, and Josephine C. Adams, *Molecular analysis of muskelin identifies a conserved discoidin-like domain that contributes to protein self-association.* Biochem. J., 2004. **381**(2): p. 547-559.
310. Gerlitz, G., et al., *Novel functional features of the Lis-H domain: role in protein dimerization, half-life and cellular localization.* Cell Cyc., 2005. **4**.

311. Delto, Carolyn F., et al., *The LisH Motif of Muskelin Is Crucial for Oligomerization and Governs Intracellular Localization.* Structure, 2015. **23**(2): p. 364-373.
312. Kim, K.H., et al., *Structure of mouse muskelin discoidin domain and biochemical characterization of its self-association.* Acta Crystallogr. Sect. D, 2014. **70**(Pt 11): p. 2863-74.
313. Robertson, N.O., et al., *Disparate binding kinetics by an intrinsically disordered domain enables temporal regulation of transcriptional complex formation.* Proc. Natl. Acad. Sci., 2018. **115**(18): p. 4643-4648.
314. Torrado, M., et al., *Refinement of the subunit interaction network within the nucleosome remodelling and deacetylase (NuRD) complex.* FEBS J., 2017. **284**(24): p. 4216-4232.
315. Miki, T. and S.A. Aaronson, *Efficient directional genetic cloning system*. 1997, US Department of Health and Human Sciences.
316. Papatheodorou, I., et al., *Expression Atlas: gene and protein expression across multiple studies and organisms.* Nucleic Acids Res., 2018. **46**(D1): p. D246-D251.
317. Peng, S.-Y., et al., *ISL1 physically interacts with BETA2 to promote insulin gene transcriptional synergy in non-β cells.* Biochim. Biophys. Acta, 2005. **1731**(3): p. 154-159.
318. Gavin, A.-C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. **415**(6868): p. 141-147.
319. Hein, Marco Y., et al., *A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances.* Cell, 2015. **163**(3): p. 712-723.
320. Malovannaya, A., et al., *Analysis of the Human Endogenous Coregulator Complexome.* Cell, 2011. **145**(5): p. 787-799.
321. Causier, B., *Studying the interactome with the yeast two-hybrid system and mass spectrometry.* Mass Spectrom. Rev., 2004. **23**(5): p. 350-367.
322. Basker, M.J., et al., *Carfecillin: Antibacterial Activity in vitro and in vivo.* Chemotherapy, 1977. **23**(6): p. 424-435.
323. Jacobs, J., et al., *Ampicillin and Carbenicillin Stability in Commonly Used Infusion Solutions.* Drug Intel. & Clin. Pharm., 1970. **4**(8): p. 204-208.
324. Studier, F.W., *Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system.* J. Mol. Biol., 1991. **219**(1): p. 37-44.
325. Wei, Y., A.A. Thyparambil, and R.A. Latour, *Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm.* Biochim. Biophys. Acta, 2014. **1844**(12): p. 2331-2337.
326. Hellman, L.M. and M.G. Fried, *Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions.* Nat. Prot., 2007. **2**: p. 1849.
327. Behravan, G., P.O. Lycksell, and G. Larsson, *Expression, purification and characterization of the homeodomain of rat ISL-1 protein.* Protein Eng., 1997. **10**(11): p. 1327-1331.
328. Meier, B.C., et al., *Characterization of the porcine Lhx3/LIM-3/P-Lim LIM homeodomain transcription factor.* Mol. Cell. Endocrinol., 1999. **147**(1–2): p. 65-74.
329. Lane, D., P. Prentki, and M. Chandler, *Use of gel retardation to analyze protein-nucleic acid interactions.* Microbiol. Rev., 1992. **56**(4): p. 509-528.
330. Wilson, D.S., et al., *High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA.* Cell, 1995. **82**(5): p. 709-719.
331. Mann, R.S., K.M. Lelli, and R. Joshi, *Hox specificity: unique roles for cofactors and collaborators.* Curr. Topics Dev. Biol., 2009. **88**: p. 63-101.
332. Sánchez-García, I., et al., *The cysteine-rich LIM domains inhibit DNA binding by the associated homeodomain in Isl-1.* EMBO J., 1993. **12**(11): p. 4243-4250.

333. Riley, L.G., G.B. Ralston, and A.S. Weiss, *Multimer formation as a consequence of separate homodimerization domains: the human c-Jun leucine zipper is a transplantable dimerization module.* Protein Eng., 1996. **9**(2): p. 223-30.
334. Wilkinson-White, L., et al., *GATA1 directly mediates interactions with closely spaced pseudopalindromic but not distantly spaced double GATA sites on DNA.* Protein Sci., 2015. **24**(10): p. 1649-1659.
335. Carra, J.H. and P.L. Privalov, *Energetics of folding and DNA binding of the MATα2 homeodomain.* Biochem, 1997. **36**(3): p. 526-535.
336. de la Mata, I., et al., *The Impact of R53C Mutation on the Three-Dimensional Structure, Stability, and DNA-Binding Properties of the Human Hesx-1 Homeodomain.* ChemBioChem, 2002. **3**(8): p. 726-740.
337. Wienken, C.J., et al., *Protein-binding assays in biological liquids using microscale thermophoresis.* Nat. Comm., 2010. **1**: p. 100.
338. Seidel, S.A., et al., *Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions.* Methods, 2013. **59**(3): p. 301-315.
339. Jolma, A., et al., *DNA-Binding Specificities of Human Transcription Factors.* Cell, 2013. **152**(1): p. 327-339.
340. Noyes, M.B., et al., *Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites.* Cell, 2008. **133**(7): p. 1277-1289.
341. Desplan, C., J. Theis, and P.H. O'Farrell, *The sequence specificity of homeodomain-DNA interaction.* Cell, 1988. **54**(7): p. 1081-1090.
342. Wilson, D.S. and C. Desplan, *Structural basis of Hox specificity.* Nat. Struct. Biol., 1999. **6**: p. 297.
343. Vukojević, V., et al., *Quantitative study of synthetic Hox transcription factor–DNA interactions in live cells.* Proceedings of the National Academy of Sciences, 2010. **107**(9): p. 4093-4098.
344. Tucker-Kellogg, L., et al., *Engrailed (Gln50→Lys) homeodomain–DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity.* Structure, 1997. **5**(8): p. 1047-1054.
345. Galang, C. and C. Hauser, *Cooperative DNA binding of the highly conserved human Hox 2.1 homeodomain gene product.* The New Biologist, 1992. **4**(5): p. 558-568.
346. Han, K., M.S. Levine, and J.L. Manley, *Synergistic activation and repression of transcription by Drosophila homeobox proteins.* Cell, 1989. **56**(4): p. 573-583.
347. Herzig, S., L. Füzesi, and W. Knepel, *Heterodimeric Pbx-Prep1 Homeodomain Protein Binding to the Glucagon Gene Restricting Transcription in a Cell Type-dependent Manner.* J. Biol. Chem., 2000. **275**(36): p. 27989-27999.
348. Zandvakili, A. and B. Gebelein, *Mechanisms of Specificity for Hox Factor Activity.* Journal of Developmental Biology, 2016. **4**(2): p. 16.
349. Chang, C.P., et al., *Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins.* Genes Dev., 1995. **9**(6): p. 663-674.
350. Slattery, M., et al., *Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins.* Cell, 2011. **147**(6): p. 1270-1282.
351. Jolma, A., et al., *DNA-dependent formation of transcription factor pairs alters their binding specificity.* Nature, 2015. **527**(7578): p. 384-8.
352. Liu, Y., R.J. MacDonald, and G.H. Swift, *DNA binding and transcriptional activation by a PDX1.PBX1b.MEIS2b trimer and cooperation with a pancreas-specific basic helix-loop-helix complex.* J. Biol. Chem., 2001. **276**(21): p. 17985-93.
353. Shen, W.F., et al., *HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells.* Mol. Cell. Biol., 1999. **19**(4): p. 3051-61.

354. Swift, G.H., et al., *An endocrine-exocrine switch in the activity of the pancreatic homeodomain protein PDX1 through formation of a trimeric complex with PBX1b and MRG1 (MEIS2).* Mol. Cell. Biol., 1998. **18**(9): p. 5109-20.
355. Shanmugam, K., et al., *PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins.* Mol. Cell. Biol., 1999. **19**(11): p. 7577-88.
356. Liu, J., et al., *Three-amino-acid-loop-extension homeodomain factor Meis3 regulates cell survival via PDK1.* Proc Natl Acad Sci U S A, 2010. **107**(47): p. 20494-9.
357. Sprules, T., et al., *Lock and key binding of the HOX YPWM peptide to the PBX homeodomain.* J. Biol. Chem., 2003. **278**(2): p. 1053-8.
358. LaRonde-LeBlanc, N.A. and C. Wolberger, *Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior.* Genes Dev., 2003. **17**(16): p. 2060-72.
359. Mariotto, A., et al., *HOPX: The Unusual Homeodomain-Containing Protein.* Journal of Investigative Dermatology, 2016. **136**(5): p. 905-911.
360. Shin, C.H., et al., *Modulation of cardiac growth and development by HOP, an unusual homeodomain protein.* Cell, 2002. **110**(6): p. 725-735.
361. Hawiger, D., et al., *The transcription cofactor Hopx is required for regulatory T cell function in dendritic cell–mediated peripheral T cell unresponsiveness.* Nat. Immunol., 2010. **11**: p. 962.
362. Yamashita, K., H. Katoh, and M. Watanabe, *The Homeobox Only Protein Homeobox (HOPX) and Colorectal Cancer.* Int. J. Mol. Sci., 2013. **14**(12): p. 23231-23243.
363. Trivedi, C.M., et al., *Hopx and Hdac2 Interact to Modulate Gata4 Acetylation and Embryonic Cardiac Myocyte Proliferation.* Dev. Cell, 2010. **19**(3): p. 450-459.
364. Li, D., et al., *Hopx distinguishes hippocampal from lateral ventricle neural stem cells.* Stem Cell Research, 2015. **15**(3): p. 522-529.
365. Song, S.-H., C. Hou, and A. Dean, *A Positive Role for NLI/Ldb1 in Long-Range β-Globin Locus Control Region Function.* Mol. Cell, 2007. **28**(5): p. 810-822.
366. Wadman, I.A., et al., *The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins.* EMBO J., 1997. **16**(11): p. 3145-3157.
367. Matys, V., et al., *TRANSFAC ® : transcriptional regulation, from patterns to profiles.* Nucleic Acids Res., 2003. **31**(1): p. 374-378.
368. Newburger, D.E. and M.L. Bulyk, *UniPROBE: an online database of protein binding microarray data on protein–DNA interactions.* Nucleic Acids Res., 2009. **37**: p. D77-D82.
369. Kulakovskiy, I.V., et al., *HOCOMOCO: a comprehensive collection of human transcription factor binding sites models.* Nucleic Acids Res., 2013. **41**(D1): p. D195-D202.
370. Portales-Casamar, E., et al., *The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences.* Nucleic Acids Res., 2009. **37**(Database issue): p. D54-D60.
371. Joshi, R., et al., *Functional Specificity of a Hox Protein Mediated by the Recognition of Minor Groove Structure.* Cell, 2007. **131**(3): p. 530-543.
372. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol. Syst. Biol., 2011. **7**(1): p. 539.
373. Wheeler, T.J., J. Clements, and R.D. Finn, *Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models.* BMC Bioinformatics, 2014. **15**(1): p. 7.
374. Torrado, M., et al., *Role of Conserved Salt Bridges in Homeodomain Stability and DNA Binding.* J. Biol. Chem., 2009. **284**(35): p. 23765-23779.

375. Hall, M.N., C. Craik, and Y. Hiraoka, *Homeodomain of yeast repressor alpha 2 contains a nuclear localization signal.* Proc. Natl. Acad. Sci., 1990. **87**(18): p. 6954-6958.
376. Maizel, A., et al., *A short region of its homeodomain is necessary for engrailed nuclear export and secretion.* Development, 1999. **126**(14): p. 3183-3190.
377. Paine, P.L., L.C. Moore, and S.B. Horowitz, *Nuclear envelope permeability.* Nature, 1975. **254**(5496): p. 109.
378. Oliva, A., M.a. Llabrés, and J.B. Fariña, *Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering.* J. Pharm. Biomed. Anal., 2001. **25**(5): p. 833-841.
379. Gell, D.A., R.P. Grant, and J.P. Mackay, *The Detection and Quantitation of Protein Oligomerization*, in *Protein Dimerization and Oligomerization in Biology*, J.M. Matthews, Editor. 2012, Springer New York: New York, NY. p. 19-41.
380. Norden, B., M. Kubista, and T. Kurucsev, *Linear dichroism spectroscopy of nucleic acids.* Q. Rev. Biophys., 1992. **25**(1): p. 51-170.
381. Seeds, W. and M. Wilkins, *Ultra-violet microspectrographic studies of nucleoproteins and crystals of biological interest.* Discuss. Faraday Soc., 1950. **9**: p. 417-423.
382. Fraenkel, E., et al., *Engrailed homeodomain-DNA complex at 2.2 å resolution: a detailed view of the interface and comparison with other engrailed structures.* J. Mol. Biol., 1998. **284**(2): p. 351-361.
383. Wolberger, C., et al., *Crystal structure of a MATα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions.* Cell, 1991. **67**(3): p. 517-528.
384. Butt, R.H. and J.R. Coorssen, *Coomassie Blue as a Near-infrared Fluorescent Stain: A Systematic Comparison With Sypro Ruby for In-gel Protein Detection.* Mol. Cell. Proteom., 2013. **12**(12): p. 3834-3850.
385. Steinberg, T.H., et al., *Ultrasensitive fluorescence protein detection in isoelectric focusing gels using a ruthenium metal chelate stain.* Electrophoresis, 2000. **21**(3): p. 486-496.
386. Smejkal, G.B., M.H. Robinson, and A. Lazarev, *Comparison of fluorescent stains: relative photostability and differential staining of proteins in two-dimensional gels.* Electrophoresis, 2004. **25**(15): p. 2511-9.
387. Nadvi, N.A., J.Y.H. Chow, and J. Trewhella, *Small-Angle Scattering of Neutrons and X-Rays*, in *eLS*. 2015.
388. Kratky, O., *X-ray small angle scattering with substances of biological interest in diluted solutions.* Prog. Biophys. Mol. Biol., 1963. **13**: p. 105-IN1.
389. Bernadó, P. and D.I. Svergun, *Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering.* Mol. BioSys., 2012. **8**(1): p. 151-167.
390. Trewhella, J., et al., *2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update.* Acta Crystallogr. Sect. D, 2017. **73**(9): p. 710-728.
391. Feigin, L. and D.I. Svergun, *Structure analysis by small-angle X-ray and neutron scattering*. Vol. 1. 1987: Springer.
392. Jacques, D.A. and J. Trewhella, *Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls.* Protein Sci., 2010. **19**(4): p. 642-657.
393. Chen, S.-H. and D. Bendedouch, *Structure and interactions of proteins in solution studied by small-angle neutron scattering*, in *Methods Enzymol.* 1986, Academic Press. p. 79-116.

394. Putnam, C.D., *Guinier peak analysis for visual and automated inspection of small-angle X-ray scattering data.* J. Appl. Crystallogr., 2016. **49**(Pt 5): p. 1412-1419.
395. Svergun, D.I. and M.H.J. Koch, *Small-angle scattering studies of biological macromolecules in solution.* Rep. Prog. Phys., 2003. **66**(10): p. 1735-1782.
396. Glatter, O. and O. Kratky, *Small angle X-ray scattering*. 1982: Academic press.
397. Rambo, R.P. and J.A. Tainer, *Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology.* Ann. Rev. Biophys., 2013. **42**(1): p. 415-441.
398. Semisotnov, G.V., et al., *Protein globularization during folding. A study by synchrotron small-angle X-ray scattering.* J. Mol. Biol., 1996. **262**(4): p. 559-574.
399. Durand, D., et al., *NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers.* J. Struct. Biol., 2010. **169**(1): p. 45-53.
400. Orthaber, D., A. Bergmann, and O. Glatter, *SAXS experiments on absolute scale with Kratky systems using water as a secondary standard.* J. Appl. Crystallogr., 2000. **33**(2): p. 218-225.
401. Svergun, D., *Ab initio shape determination by simulated annealing using a single phase dummy atom model.* Biophys. J., 1999. **76**: p. 2879-2886.
402. Mann, R.S. and S.-K. Chan, *Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins.* Trends Genet., 1996. **12**(7): p. 258-262.
403. Vuzman, D., A. Azia, and Y. Levy, *Searching DNA via a "Monkey Bar" Mechanism: The Significance of Disordered Tails.* J. Mol. Biol., 2010. **396**(3): p. 674-684.
404. Vuzman, D., M. Polonsky, and Y. Levy, *Facilitated DNA Search by Multidomain Transcription Factors: Cross Talk via a Flexible Linker.* Biophys. J., 2010. **99**(4): p. 1202-1211.
405. Iwahara, J., M. Zweckstetter, and G.M. Clore, *NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA.* Proc. Natl. Acad. Sci., 2006. **103**(41): p. 15062-15067.
406. Wilkinson-White, L., et al., *Structural basis of simultaneous recruitment of the transcriptional regulators LMO2 and FOG1/ZFPM1 by the transcription factor GATA1*. Proc. Natl. Acad. Sci., 2011. **108**(35): p. 14443-14448.
407. El Omari, K., et al., *Structural Basis for LMO2-Driven Recruitment of the SCL:E47bHLH Heterodimer to Hematopoietic-Specific Transcriptional Targets.* Cell Rep., 2013. **4**(1): p. 135-147.
408. Boëda, B., et al., *Tes, a Specific Mena Interacting Partner, Breaks the Rules for EVH1 Binding.* Mol. Cell, 2007. **28**(6): p. 1071-1082.
409. Boëda, B., et al., *Molecular Recognition of the Tes LIM2–3 Domains by the Actin-related Protein Arp7A.* J. Biol. Chem., 2011. **286**(13): p. 11543-11554.

Appendix A - Oligonucleotides used for binding studies

Listed in Table A.1 are all the sequences of oligonucleotides used for binding studies, as well as any modifications.

Oligonucleotide name	Sequence (5'-3')	Modifications	Experiments used for
Isl1GA	ACCGCGTAATATCTG TGGCGCATTATAGAC	5' fluorescein	EMSAs, MST
	GCACCGCGTAATATCTGCG CGTGGCGCATTATAGACGC	None	CD, ITC
Lhx3GSU	ACTTAGCTAATTAAATGTG TGAATCGATTAATTTACAC	5' fluorescein	EMSAs, MST
	ACTTAGCTAATTAAATGTG TGAATCGATTAATTTACAC	None	CD, ITC
M100	CGGCCATTAGCCAAATTACGGC GCCGGTAATCGGTTTAATGCCG	5' fluorescein	EMSAs, MST
	GCGCATTAGCCAAATTACG CGCGTAATCGGTTTAATGC	None	CD, ITC
HDC	CACGTGCCGTCAGCGGTAC GTGCACGGCAGTCGCCATG	None	CD, ITC
M100c21	ACGCCATTAGCCAAATTACGC GCGGTAATCGGTTTAATGCGT	None	Crystallography
M100c20	ACGCATTAGCCAAATTACGC GCGTAATCGGTTTAATGCGT	None	Crystallography
M100c20b	CCGCATTAGCCAAATTACGC GGCGTAATCGGTTTAATGCG	None	Crystallography
M100c14	TCATTAGCCAAATTA GTAATCGGTTTAATT	None	Crystallography

Appendix B - Amino acid sequences of protein fusion tags

Listed in Table B.1 are the sizes and amino acid sequences of protein fusion tags used throughout this thesis.

Table B.1: Fusion protein tags used. Linkers and protease cleavage sites are highlighted in green.

Tag	Size (kDa)	Sequence
FLAG (+ linker)	2.1	MDYKDDDDKGSTRTHNR
HA (+ linker)	2.4	MYPYDVPDYASRGSTRTHNR
MBP (+ linker)	41.0	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTV EHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGL LAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALS LIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQ EPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAK AGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTIN GPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAG INAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVA LKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWY AVRTAVINAASGRQTVDEALKDAQTNSSSGGGGS
GST (+ 3C site)	26.8	MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEG DKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIAD KHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDF ETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDF MLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDK YLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPL GS

Appendix C - Sequencing primers

Various sequencing primers were used throughout this thesis for either confirmation of correct subcloning, or identification of inserts in vectors. Primers and the vectors they were used with are listed in Table C.1.

Table C.1: sequencing primers used. All primers anneal either upstream (forward) or downstream (reverse) of the MCS, unless otherwise specified.

Primer name	Primer sequence (5'-3')	Vector
T7 (forward)	TAATACGACTCACTATAGGG	pcDNA3.1 pET-DUET
BGH (reverse)	CGCCAGGGTTTTCCCAGTCACGAC	pcDNA3.1
malE (forward)	GGTCGTCAGACTGTCGATGAAGCC	pMAL
pMAL3 (reverse)	CGCCAGGGTTTTCCCAGTCACGAC	pMAL
pGAD (forward)	GTATAACGCGTTTGGAATC	pGAD10 pGADT7-RecAB
pGAD (reverse)	CTTAGAGGAGTATAGTTACAT	pGAD10 pGADT7-RecAB
pGBT9 (forward)	TCATCGGAAGAGAGTAG	pGBT9
pGBT9 (reverse)	CGTTTTAAAACCTAAGAGTCAC	pGBT9
pGBT9ab (forward)	AATAATGGTTTCTTAGACGTC	pGBT9 antibiotic resistance gene
pGBT9ab (reverse)	TGCAAGCAGCAGATTAC	pGBT9 antibiotic resistance gene

Appendix D - Library titering experiments

Library titering experiments were performed to confirm that large scale library screening was being conducted correctly. This involved inoculating appropriate selective media with known dilutions of both parental yeast strains, as well as mated yeast (Table D.1). Dilutions were chosen as recommended by the Clontech Matchmaker® Gold Yeast Two-Hybrid System User Manual.

Table D.1: Conditions used for library titering experiments

Yeast strain	Media	Dilutions
Y187 pre-transformed with preys	SD-L	1:100 1:1000 1:10000
Mated yeast	SD-L-W, SD-W, SD-L	1:10 1:100 1:1000 1:10000

After incubating for 3 days (30 °C), colonies were then counted, and calculations performed to determine the viability of the yeast library, the viability of the mated yeast, the number of clones screened, and the mating efficiency (Tables D.2, D.3, D.4, and D.5 respectively). In all cases, the library-containing Y187 yeasts were the limiting partner in determining mating efficiency.

Table D.2: Library viabilities. Viability is calculated by dividing the number of colonies observed by the dilution factor and volume of inoculum. Viability is measured in colony forming units (cfu) per mL.

Library screen	Average number of prey colonies per plate	Viability of library (cfu × 10^6/mL)
$Ldb1_{LID}$/$Isl1_{LIM}$	1056	211
$Isl1_{LIM}$ (2014)	714	143
$Isl1_{LIM}$ (2016)	300	60
$Isl1_{\Delta LIM}$ (2.5 mM 3-AT)	300	60
$Isl1_{\Delta LIM}$ (2 mM 3-AT)	193	39

Table D.3: Viability of mated yeast in interaction screening.

Library screen	Colonies grown of mated yeast				Viability of mated yeast (cfu/mL)
	10^{-2} dilution	10^{-3} dilution	10^{-4} dilution	10^{-5} dilution	
$Ldb1_{LID}/Isl1_{LIM}$	>300	83	20	2	2,000,000
$Isl1_{LIM}$ (2014)	>300	>300	>300	59	59,000,000
$Isl1_{LIM}$ (2016)	>300	>300	255	43	43,000,000
$Isl1_{\Delta LIM}$ (2.5 mM 3-AT)	>300	>300	255	42	42,000,000
$Isl1_{\Delta LIM}$ (2 mM 3-AT)	>300	>300	>300	81	81,000,000

Table D.4: Screening efficiencies of yeast two-hybrid library screens. Number of clones screened is calculated by multiplying the viability of the mated yeast by the total volume of culture.

Library screen	Viability of mated yeast (cfu/mL)	Resuspended total volume of mated yeast (mL)	Number of clones screened (millions)
$Ldb1_{LID}/Isl1_{LIM}$	2,000,000	11.4	23
$Isl1_{LIM}$ (2014)	59,000,000	11.9	702
$Isl1_{LIM}$ (2016)	43,000,000	12.3	529
$Isl1_{\Delta LIM}$ (2.5 mM 3-AT)	42,000,000	12.3	517
$Isl1_{\Delta LIM}$ (2 mM 3-AT)	81,000,000	13.6	1102

Table D.5: Mating efficiencies of yeast two-hybrid library screens. Mating efficiency is calculated by dividing the viability of mated yeast by the viability of the library.

Library screen	Viability of library (cfu × 10^6/mL)	Viability of mated yeast (cfu × 10^6/mL)	Mating efficiency (%)
$Ldb1_{LID}/Isl1_{LIM}$	211	2	0.9
$Isl1_{LIM}$ (2014)	143	59	41
$Isl1_{LIM}$ (2016)	60	43	72
$Isl1_{\Delta LIM}$ (2.5 mM 3-AT)	60	42	70
$Isl1_{\Delta LIM}$ (2 mM 3-AT)	39	81	210

Appendix E - SAXS reporting

Full SAXS experimental details and parameters related to data processing and modelling are listed in Tables E.1 and E.2. Tables E.3 and E.4 report parameters from analyses of static SAXS data sets.

Table E.1: SAXS Experimental details. This includes sample details and data collection parameters.

	2HDLL	**M100c20B**	**2HDLL:M100**
Sample Details			
Organism	*Mus musculus*	*Mus musculus*	*Mus musculus*
Source	Expressed; see Section 5.3	IDT	
Sequence	See Appendix G	Table 6.1	
Extinction coefficient	1.024 (A280, 0.1% w/v)	30933.25 (A260, 0.1% w/v)	Unknown
Partial specific volume (cm^3 g^{-1})	0.731	0.59	0.696
Particle contrast from sequence and solvent constituents (10^{10} cm^{-2})	2.953 (12.404-9.452)	5.344 (14.795-9.452)	3.23 (12.68-9.452)
M_w from chemical composition	45152	12015	57167
SEC-SAXS			
Column	Superose 12 10 x 300 mm		
Loading concentration (mg mL^{-1})	2.55	3	3
Injection volume (µL)	99	100	100
Flow rate (mL min^{-1})	1	1	1
Static SAXS			
Concentration (mg mL^{-1})	0.26	0.75	0.69
Solvent	20 mM sodium phosphate monobasic/dibasic, pH 7.4, 100 mM NaCl, 1 mM DTT		
Data collection parameters			
Instrument	Australian Synchrotron SAXS/WAXS beamline with Dectris PILATUS 1M detector		
Wavelength (Å)	1.07812		
Beam size (µm)	250 × 450		
Camera length (m)	0.9		

q measurement range (A^{-1})	0.006-0.34
Absolute scaling method	Comparison with scattering from 1 mm pure H_2O
Normalisation	To transmitted intensity by beam-stop counter
Monitoring for radiation damage	X-ray dose maintained below 210 Gy
Exposure time	1 sec, 40 exposures per sample
Sample configurations	SEC-SAXS Aspiration of samples into a capillary from a 96-well plate
Sample temperature	27 °C

Table E.2: Data processing details and parameters.

Data reduction, analysis and interpretation			
SAXS data reduction	ScatterBrain 2.82		
Extinction coefficient estimate	ProtParam		
Calculation of volume and contrast	MULCh 1.1		
Basic analyses (Guinier, Kratky, P(r))	PRIMUSqt from ATSAS		
Shape/bead modelling	MONSA		
Three-dimensional graphic model representation	PyMOL 2.0.6		
Structural parameters			
	2HDLL	**M100**	**2HDLL:M100**
Guinier analysis			
I(0) (cm^{-1})	0.0069*	0.024*	0.034*
Rg (Å)	33.77 ± 1.97	19.27 ± 0.10	34.74 ± 0.29
qmin ($Å^{-1}$)	0.00636	0.00636	0.00636
qRg max (qmin = 0.00636)	1.29	1.29	1.34
Coefficient of correlation, R^2	0.75	0.95	0.91
M from I(0) **(ratio to predicted)**	34098.15 (0.92)	18985.65 (1.55)	50730.31 (1.03)
P(r) analysis			
I(0) (cm^{-1})	0.007041 ± 0.000174	0.02373 ± 0.00009	0.03459 ± 0.00032
Rg (Å)	35.56 ± 1.37	20.0 ± 0.12	36.39 ± 0.42
Dmax (Å)	122	67	130
q range ($Å^{-1}$)	0.00860268-0.235868	0.0145871-0.344776	0.0213196-0.238108

X^2	0.8014	0.76	0.7155
M_w from I(0)	34794.94	18772.06	51610
(ratio to predicted)	(0.93515)	(1.53655)	(1.04422)
Porod volume (Å $^{-3}$)	71400	14600	67000
(ratio V_p/calculated M_w)	(1.2792)	(0.7967)	(0.9037)

* Errors in *I*(0) as reported in PRIMUSqt are smaller than the significant figures reported for *I*(0)

Table E.3: Parameters from Guinier analyses of datasets.

Sample concentration (µg/mL)	Rg	I(0)	Pearson's R for fit	q-range	qRg max	Mw from I(0)	Ratio of Mw to expected
M100 alone							
969	21.43±0.23	0.021	0.95	0.01792-0.106842	1.29	12867	1.05
485	21.60±0.27	0.011	0.92	0.015265-0.104188	1.28	13480	1.10
242	22.01±0.82	0.0056	0.76	0.015265-0.105515	1.3	13725	1.12
121	21.89±0.84	0.0029	0.79	0.01261-0.101534	1.29	14215	1.16
61	21.73±1.27	0.0015	0.83	0.015265-0.108169	1.28	14705	1.20
30	21.85±1.58	0.0015	0.86	0.011283-0.108169	1.29	29410	2.40
2HDLL:M100							
3969	47.89±2.28	0.23	0.82	0.015265-0.029866	0.92	60240	1.22
1985	45.85±3.10	0.1	0.89	0.023229-0.043139	0.88	52382	1.06
992	40.72±0.36	0.049	0.98	0.01261-0.049776	1.22	51335	1.04
496	40.72±0.78	0.024	0.97	0.011283-0.048448	1.05	50287	1.02
248	38.69±0.87	0.011	0.92	0.009955-0.055085	1.26	46097	0.93
124	36.91±0.44	0.005	0.97	0.00995-0.060394	1.25	41906	0.85
2HDLL alone							
2550	39.36±1.31	0.042	0.98	0.021902-0.056412	1.18	21199	0.57

Table E.4: Parameters from P(r) analyses of all datasets.

Sample concentration (µg/mL)	Rg	I(0)	Dmax	q range	Q value for fit	Porod volume
M100 alone						
969	22.13±0.08	0.0214	68	0.0206 to 0.3731	0.78	19500
485	22.22±0.13	0.0109	67	0.0206 to 0.3692	0.78	19600
242	22.06±0.22	0.0055	65	0.0219 to 0.3626	0.79	21800
121	22.19±0.41	0.0029	65	0.0179 to 0.3652	0.79	19500
61	21.23±0.50	0.0014	60	0.0126 to 0.3678	0.81	14100
30	21.99±0.63	0.0015	65	0.0126 to 0.362	0.80	13300
2HDLL:M100						
3969	42.64±0.00	0.2127	125	0.0139-0.1665	0.57	107000
1985	46.28±0.51	0.1022	200	0.0126-0.1732	0.65	115000
992	42.85±0.33	0.0493	160	0.0113-0.1957	0.75	125000
496	42.00±0.50	0.0237	145	0.0113-0.1957	0.71	114000
248	39.64±0.40	0.0110	120	0.0139-0.2063	0.79	96000
124	38.21±0.75	0.0050	118	0.0113-0.2156	0.78	56400
2HDLL alone						
2550	40.38±0.26	0.04237	133	0.0113-0.2023	0.77	75600

Appendix F - Amino acid sequences of Mkln1 constructs and Isl1 constructs

The amino acid sequences for Isl1 and Mkln1 constructs used throughout Chapter 4 are listed in Tables F.1 and F.2.

Table F.1: Amino acid sequences of Isl1 constructs used for interaction testing.

Construct	Size (kDa)	Sequence
$Isl1_{FL}$	38.1	GSKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKC AECNQYLDESCTCFVRDGKTYCKRDYIRLYGIKCAKC SIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF ALREDGLFCRADHDVVERASLGAGDPLSPLHPARPLQ MAAEPISARQPALRPHVHKQPEKTTRVRTVLNEKQLH TLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVWFQ NKRCKDKKRSIMMKQLQQQQPNDKTNIQGMTGTPM VAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQS DIDQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTP NSMVASPIEA
$Isl1_{LIM}$	14.9	GSKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKC AECNQYLDESCTCFVRDGKTYCKRDYIRLYGIKCAKC SIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEF ALREDGLFCRADHDVVER
$Isl1_{\Delta LIM}$	24.1	GSHDVVERASLGAGDPLSPLHPARPLQMAAEPISARQP ALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPR PDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSI MMKQLQQQQPNDKTNIQGMTGTPMVAASPERHDGG LQANPVEVQSYQPPWKVLSDFALQSDIDQPAFQQLVN FSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA
$Ldb1_{LID}$/$Isl1_{LIM}$	20.2	GSDVMVVGEPTLMGGEFGDEDERLITRLENTQFDAAN GIDDEGGSGGHMGSGGKRLISLCVGCGNQIHDQYILR VSPDLEWHAACLKCAECNQYLDESCTCFVRDGKTYC KRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECF RCVACSRQLIPGDEFALREDGLFCRADHDVVER

Table F.2: Amino acid sequences of Mkln1 constructs used for interaction testing.

Construct	Size (kDa)	Sequence
Mkln1$_{FL}$	81.6	MAAGGAVAVAPECRLLPYALHKWSSFSSTYLPENILVDKPND QSSRWSSESNYPPQYLILKLERPAIVQNITFGKYEKTHVCNLKK FKVFGGMNEENMTELLSSGLKNDYNKETFTLKHKIDEQMFPC RFIKIVPLLSWGPSFNFSIWYVELSGIDDPDIAQPCLNWYSKYR EQEAIRLCLKHFRQHNYTEAFESLQKKTKIALEHPMLTDMHD KLVLKGDFDACEELIEKAVNDGLXNQYISQQEYKPRWSQIIPK STKGDGETQXLADFWAYSVKENQWTCISRDTEKENGPSARSC HKMCIDIQRRQIYTLGRYLNSSVRNSKSLKSDFYRYDIDTNTW MLLSEDTAADGGPKLVFDHXMCMDSEKHMIYTLGGRILTCNG SVDDSRASEPQFSGLFAFNCQCQTWKLLREDSCNAGPEDIQSRI GHCMLFHSKNRCLYVFGGQRSRTYLNDFFSYDVDSDHVDIISD GTKKDSGMVPMTGFTQRATIDPELNEIHVLSGLSKDKEKREEN VRNSFWIYDIVRNSWSCVYKNDQATKDNLSKSLQEEEPCPRFA HQLVYDELHKVHYLFGGNPGKSCSPKMRLDDFWSLKLCRPSK DYLLRHCKYLIRRHRFEEKAQMDPLSALKYLQNDLYITVDHS DPEETKEFQLLASALFKSGSDFTALGFSDVDHTYAQRTQLFDT LVNFFPDSMTPPKGNLVDLITL
Mkln1$_{ND}$	19.2	MAAGGAVAVAPECRLLPYALHKWSSFSSTYLPENILVDKPND QSSRWSSESNYPPQYLILKLERPAIVQNITFGKYEKTHVCNLKK FKVFGGMNEENMTELLSSGLKNDYNKETFTLKHKIDEQMFPC RFIKIVPLLSWGPSFNFSIWYVELSGIDDPDIVQPCLNW
Mkln1$_{NL}$	23.7	MAAGGAVAVAPECRLLPYALHKWSSFSSTYLPENILVDKPND QSSRWSSESNYPPQYLILKLERPAIVQNITFGKYEKTHVCNLKK FKVFGGMNEENMTELLSSGLKNDYNKETFTLKHKIDEQMFPC RFIKIVPLLSWGPSFNFSIWYVELSGIDDPDIVQPCLNWYSKYR EQEAIRLCLKHFRQHNYTEAFESLQKKTKIAL
Mkln1$_{NK}$	71.8	MAAGGAVAVAPECRLLPYALHKWSSFSSTYLPENILVDKPND QSSRWSSESNYPPQYLILKLERPAIVQNITFGKYEKTHVCNLKK FKVFGGMNEENMTELLSSGLKNDYNKETFTLKHKIDEQMFPC RFIKIVPLLSWGPSFNFSIWYVELSGIDDPDIVQPCLNWYSKYR EQEAIRLCLKHFRQHNYTEAFESLQKKTKIALEHPMLTDMHD KLVLKGDFDACEELIEKAVNDGLFNQYISQQEYKPRWSQIIPKS TKGDGEDNRPGMRGGHQMVIDVQTETVYLFGGWDGTQDLA DFWAYSVKENQWTCISRDTEKENGPSARSCHKMCIDIQRRQIY TLGRYLDSSVRNSKSLKSDFYRYDIDTNTWMLLSEDTAADGG PKLVFDHQMCMDSEKHMIYTFGGRILTCNGSVDDSRASEPQFS

		GLFAFNCQCQTWKLLREDSCNAGPEDIQSRIGHCMLFHSKNRC LYVFGGQRSKTYLNDFFSYDVDSDHVDIISDGTKKDSGMVPM TGFTQRATIDPELNEIHVLSGLSKDKEKREENVRNSFWIYDIVR NSWSCVYKNDQATKDNLSKSLQEEEPCPRFAHQLVYDELHKV HYLFGGNPGKSCSPKMRLDDFWSLK
Mkln1$_{LC}$	62.5	YREQEAIRLCLKHFRQHNYTEAFESLQKKTKIALEHPMLTDMH DKLVLKGDFDACEELIEKAVNDGLFNQYISQQEYKPRWSQIIP KSTKGDGEDNRPGMRGGHQMVIDVQTETVYLFGGWDGTQDL ADFWAYSVKENQWTCISRDTEKENGPSARSCHKMCIDIQRRQI YTLGRYLDSSVRNSKSLKSDFYRYDIDTNTWMLLSEDTAADG GPKLVFDHQMCMDSEKHMIYTFGGRILTCNGSVDDSRASEPQ FSGLFAFNCQCQTWKLLREDSCNAGPEDIQSRIGHDFFSYDVD SDHVDIISDGTKKDSGMVPMTGFTQRATIDPELNEIHVLSGLSK DKEKREENVRNSFWIYDIVRNSWSCVYKNDQATKDNLSKSLQ EEEPCPRFAHQLVYDELHKVHYLFGGNPGKSCSPKMRLDDFW SLKLCRPSKDYLLRHCKYLIRKHRFEEKAQMDPLSALKYLQN DLYITVDHSDPEETKEFQLLASALFKSGSDFTALGFSDVDHTY AQRTQLFDTLVNFFPDSMTPPKGNLVDLITL
Mkln1$_{CC}$	58.4	LEHPMLTDMHDKLVLKGDFDACEELIEKAVNDGLFNQYISQQ EYKPRWSQIIPKSTKGDGEDNRPGMRGGHQMVIDVQTETVYL FGGWDGTQDLADFWAYSVKENQWTCISRDTEKENGPSARSC HKMCIDIQRRQIYTLGRYLDSSVRNSKSLKSDFYRYDIDTNTW MLLSEDTAADGGPKLVFDHQMCMDSEKHMIYTFGGRILTCNG SVDDSRASEPQFSGLFAFNCQCQTWKLLREDSCNAGPEDIQSRI GHDFFSYDVDSDHVDIISDGTKKDSGMVPMTGFTQRATIDPEL NEIHVLSGLSKDKEKREENVRNSFWIYDIVRNSWSCVYKNDQ ATKDNLSKSLQEEEPCPRFAHQLVYDELHKVHYLFGGNPGKS CSPKMRLDDFWSLKLCRPSKDYLLRHCKYLIRKHRFEEKAQM DPLSALKYLQNDLYITVDHSDPEETKEFQLLASALFKSGSDFTA LGFSDVDHTYAQRTQLFDTLVNFFPDSMTPPKGNLVDLITL
Mkln1$_{KC}$	52.4	TETVYLFGGWDGTQDLADFWAYSVKENQWTCISRDTEKENG PSARSCHKMCIDIQRRQIYTLGRYLDSSVRNSKSLKSDFYRYDI DTNTWMLLSEDTAADGGPKLVFDHQMCMDSEKHMIYTFGGR ILTCNGSVDDSRASEPQFSGLFAFNCQCQTWKLLREDSCNAGP EDIQSRIGHCMLFHSKNRCLYVFGGQRSKTYLNDFFSYDVDSD HVDIISDGTKKDSGMVPMTGFTQRATIDPELNEIHVLSGLSKD KEKREENVRNSFWIYDIVRNSWSCVYKNDQATKDNLSKSLQE EEPCPRFAHQLVYDELHKVHYLFGGNPGKSCSPKMRLDDFWS LKLCRPSKDYLLRHCKYLIRKHRFEEKAQMDPLSALKYLQND LYITVDHSDPEETKEFQLLASALFKSGSDFTALGFSDVDHTYA

		QRTQLFDTLVNFFPDSMTPPKGNLVDLITL
Mkln1$_{FC}$	10.8	KHRFEEKAQMDPLSALKYLQNDLYITVDHSDPEETKEFQLLAS ALFKSGSDFTALGFSDVDHTYAQRTQLFDTLVNFFPDSMTPPK GNLVDLITL

Appendix G - Amino acid sequences of homeodomain constructs

The amino acid sequences for all homeodomain constructs used throughout this thesis are listed in Table G.1.

Table G.1: Amino acid sequences of homeodomain containing constructs.

Construct name	Size (kDa)	Sequence
NHD1	9.6	GSARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANP RPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMMK
NHD3	9.6	GSYETAKQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARH VREQLSSETGLDMRVVQVWFQNRRAKEKRLKKDAGRQRW
LLHD3	27.2	GSGTPMVAASPERHDGGLQANPVEVQSYQPPWGGSGGHMGS GGTPEIPMCAGCDQHILDRFILKALDRHWHSKCLKCSDCHVP LAERCFSRGESVYCKDDFFKRFGTKCAACQLGIPPTQVVRRA QDFVYHLHCFACVVCKRQLATGDEFYLMEDSRLVCKADYET AKQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQ LSSETGLDMRVVQVWFQNRRAKEKRLKKDAGRQRW
LLHD1	32.2	GSDVMVVGEPTLMGGEFGDEDERLITRLENTQFDAANGIDDE GGSGGHMGSGGKRLISLCVGCGNQIHDQYILRVSPDLEWHAA CLKCAECNQYLDESCTCFVRDGKTYCKRDYIRLYGIKCAKCSI GFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEFALREDGL FCRADHDVVERASLGAGDPLSPLHPARPLQMAAEPISARQPA LRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPRPDALM KEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMMK
2HDLL	37	GSEKTTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEM TGLSPRVIRVWFQNKRCKDKKRSIMMKQLQQQQPNDKTNIQ GMTGTPMVAASPERHDGGLQANPVEVQSYQPPWGGSGGHM GSGGTPEIPMCAGCDQHILDRFILKALDRHWHSKCLKCSDCH VPLAERCFSRGESVYCKDDFFKRFGTKCAACQLGIPPTQVVRR AQDFVYHLHCFACVVCKRQLATGDEFYLMEDSRLVCKADYE TAKQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVRE QLSSETGLDMRVVQVWFQNRRAKEKRLKKDAGRQRW
2HDN	20.8	GSARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANP RPDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMMKQ LQQQQPNDKTNIQGMTYETAKQREAEATAKRPRTTITAKQLE TLKSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAK

		EKRLKKDAGRQRW
2HD17	20	ARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPRP DALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMMKGGS GGSGGSGGSGGSGGYETAKQREAEATAKRPRTTITAKQLETL KSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEK RLKKDAGRQRW
2HD23	19.7	ARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPRP DALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMMKGGS GGSGGSGGSGGSGGSGGSGGREAEATAKRPRTTITAKQLETL KSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEK RLKKDAGRQRW

Appendix H - Protein Purification

Figure H.1 shows the purity of purified homeodomain constructs not featured in Figure 5.6.

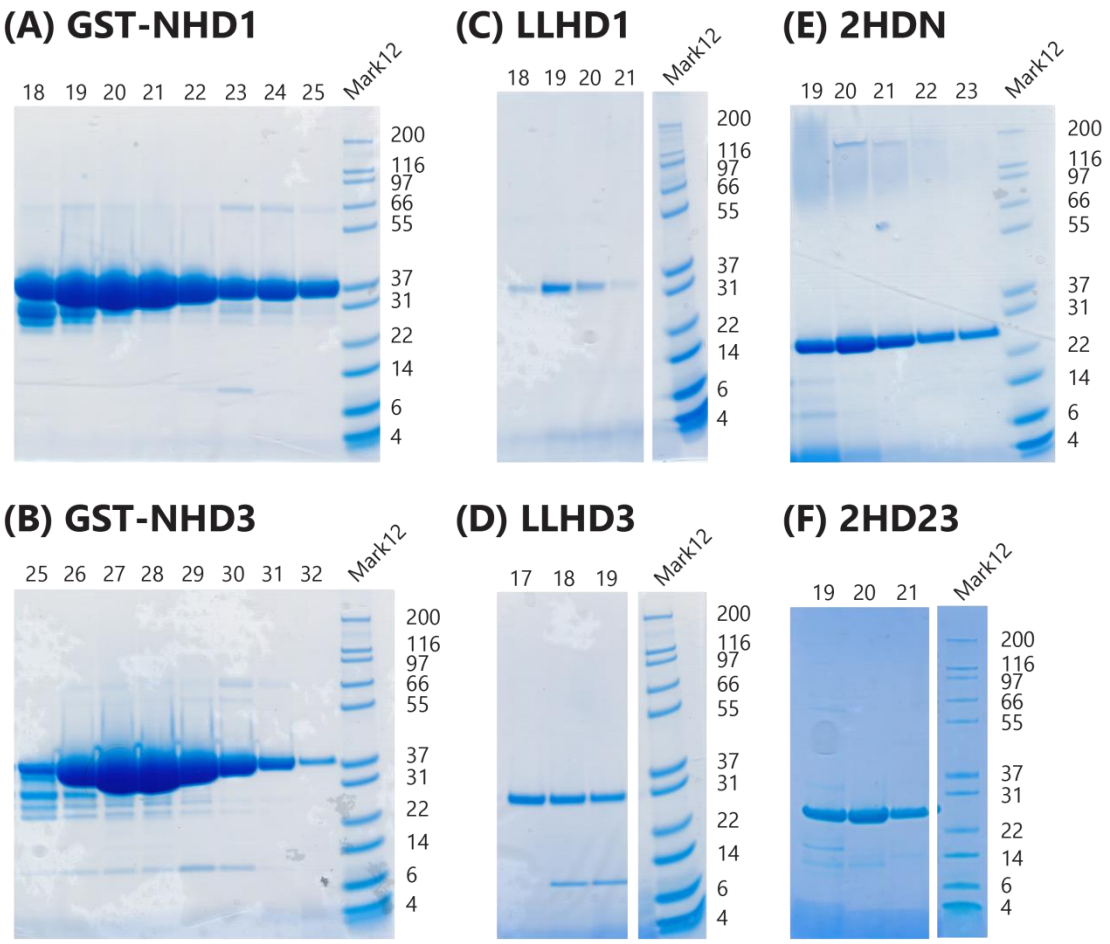


Figure H.1: Further examples of purifications of homeodomain constructs. SDS-PAGE gels showing cation exchange elution fractions for (A) GST-NHD1; (B) GST-NHD3; (C) LLHD1; (D) LLHD3; (E) 2HDN; (F) 2HD23.