DETECTION AND ANALYSIS OF LIM DOMAIN-MEDIATED INTERACTIONS BETWEEN TRANSCRIPTION FACTORS

Ivan Nisevic

A thesis submitted in fulfilment of

the requirements for the degree of

Doctor of Philosophy

School of Life and Environmental Sciences

University of Sydney

Declaration

The experimental work described in this Thesis was performed between March 2008 and May 2013 in the School of Molecular Bioscience at the University of Sydney. I performed the experiments unless otherwise stated. This work has not been submitted, in part or in full, for the purpose of obtaining any other degree.

Abstract

LIM-homeodomain (LIM-HD) proteins are a class of transcription factors involved in tissue specification and cell determination during development and are important in adult gene regulation. Six families of LIM-HD proteins, with two close paralogues in each family, are commonly found in tetrapods. They bind DNA via HDs, whereas their interactions with other proteins are mediated mainly by a pair of closely spaced LIM-domains (LIMs) in each protein. These proteins take part in various transcriptional complexes with Ldb1 and other cofactors that contain LIM-interaction domains (LIDs).

In this thesis, protein-protein interactions of LIM-HD proteins were analysed in order to better understand the molecular mechanisms of transcriptional complex formation.

Based on previous research that showed LIM-LID mediated interactions between Lhx3 and Isl1, yeast two-hybrid mating arrays were used to investigate how widespread protein-protein interactions are amongst the 12 mammalian LIM-HD proteins. Due to high levels of background growth in experiments with full-length proteins in pGBT9 vectors, the mating arrays focused on LIM-domain mediated interactions with full-length LIM-HDs or known LIDs. The arrays revealed a relatively strong interaction between Lhx3 (or Lhx4) and Isl1 (or Isl2), and detected weaker interactions between Lmx1a or Lmx1b and the LIM-binding domain of Isl1.

The contribution of separate LIM-domains to the overall interaction with Ldb1 for each of the proteins was analysed by the same method. In most cases one of the LIM domains in each protein was able to independently interact with the LID domain of Ldb1 by yeast two-hybrid analysis indicating a dominant binder: LIM1 in Isl1 and Isl2, or LIM2 in other proteins. The exceptions were paralogues Lhx1 and Lhx5, for which no separate domain showed interaction with $Ldb1_{LID}$ by this approach. All tandem LIM-domain constructs showed a much stronger interaction with $Ldb1_{LID}$ than any isolated LIM domain supporting the idea that both domains are required for high affinity binding to Ldb1.

Bimolecular Fluorescence Complementation experiments in yeast were designed and conducted as an alternative approach to test interactions between full-length LIM-HD proteins in the hope that a non-transcription based assay would lead to no or less background signal compared to yeast two-hybrid analysis. A plasmid system was developed based on existing yeast

two-hybrid vectors using split green fluorescent proteins in place of domains from the GAL4 transcription factor. The assay was able to detect interactions between different LIMs and their partners but unfortunately interactions between full-length proteins were still difficult to detect due to low fluorescence, self-complementation in the controls and localization effects.

LIM domains from LIM-HD proteins cannot be used in standard bimolecular binding assays because they tend to be insoluble and/or aggregate in the absence of a binding partner. Stable, soluble intramolecular 'tethered complexes' can be generated in which LIMs are tethered to $Ldb1_{LID}$ via a flexible linker. Introduction of a specific protease site into the tether allows the formation of intermolecular cut complexes, which have previously been used in homologous competition ELISA experiments.

In this thesis attempts were made to develop more robust biophysical binding assays that could be used to assess the binding affinities of different LIMs for $Ldb1_{LID}$. Several different labelling approaches were used to generate proteins with fluorescent tags for use in fluorescence anisotropy assays. In one of these approaches expressed protein ligation was applied to generate proteins with an N-terminal fluorescein. Although this labelling strategy was of low efficiency for LIMs-$Ldb1_{LID}$ tethered constructs, some preliminary fluorescence anisotropy experiments were carried out, which indicated that this could be a useful strategy providing a more efficient labelling strategy can be found. GFP-tagged tethered complexes were easier to generate, but could not be used in anisotropy experiments because of the intrinsically high anisotropy of GFP proteins. However, preliminary experiments indicated that these proteins can be used in clear native gel shift competition assays to compare binding affinities of different tandem LIM domains to $Ldb1_{LID}$.

Data presented in this thesis provide valuable insight into protein-protein interactions of LIM-HD transcription factors and the advantages, as well as disadvantages, of applied experimental approaches. The results and their implications are discussed, raising questions that can be resolved in future studies.

Acknowledgements

Most of all, I would like to thank my supervisor Prof. Jacqui Matthews for her support and help in writing this thesis. She has pulled me out of the quicksand many times, invested a lot of effort and showed incredible patience. I am indebted to her. To Prof. Joel Mackay, thank you for your inspirational talks and your passion in science and every sport we played.

I would like to express my gratitude to Mugdha, Morgan, Liza, Sandra, Ann, Roland, Robyn, Eija, Pep, Philippa, Lorna, Arwen, Margie, Kate, Chris, Ingrid, Jason, Chu Wai, Cy, Paula, Sock, Cuong, Mitch, Saad, Marylene, Katschi, Soumya, Nick, David, Krystal, Sarah, Kaavya, Gonzalo, Ana, Angela, Paul and many other wonderful people who showed me how to get around the lab, gave me advice on my experimental work and made my PhD experience more fun. I would especially like to thank Philippa and Neil for proofreading my thesis.

I am grateful to Dr Ben Crossett for his advice on MALDI-TOF, and Dr Brendan Wilkinson for his advice on native chemical ligation and help with reagent purification. Thanks are also due to Dr Tim Newsome for his microscopy-related advice. Thanks to Clare Woodley for everything and all the people from the SMB that I talked to at my progress interviews.

Finally, I need to thank my wife Mirjana, my children and my close family that have stood by me and helped me throughout the years.

I gratefully acknowledge the support and funding of an Australia Postgraduate Award.

Table of contents

List of abbreviations

3-AT	3-amino-1,2,4-triazole
AD	activation domain
ADH1	alcohol dehydrogenase 1
Amp	ampicillin
BiFC	bimolecular fluorescent complementation
BN-PAGE	blue native PAGE
bp	base pairs
BSA	bovine serum albumine
CNS	central nervous system
CN-PAGE	clear native PAGE
CGFP	C-terminal fragment of GFP
$CGFP_N$	CGFP as an N-terminal tag
$CGFP_C$	CGFP as an C-terminal tag
co-IP	co-immunoprecipitation
CRCs	chromatin remodelling complexes
DBD	DNA-binding domain
DTT	dithiothreitol
DSS	2,2-dimethyl-2-silapentane-5-sulfonic acid
ELISA	enzyme linked immuno-sorbent assay
EPL	expressed protein ligation
FA	fluorescence anisotropy
Fl-	fluorescein label
FP	fluorescent protein
GFP	green fluorescent protein
GST	glutathione-S-transferase
HD	homeodomain
His_6	hexahistidine purification tag
IC_{50}	half maximal inhibitory concentration
IPTG	isopropyl β-D-thiogalactopyranoside
K_d	dissociation constant
LB	Luria-Bertani broth

LBD	LIM-binding domain
LIMK1/2	LIM-kinases 1 and 2
LIM-HD	LIM homeodomain protein
LIMs	tandem LIM domains (LIM1+LIM2)
LIMs:LID	complex of LIMs and LID domains
LIMs-LID	tethered construct of LIMs and LID
LMO	LIM only protein
MALDI-TOF	matrix-assisted laser desorption/ionization - time of flight mass spectrometry
MALLS	multi-angle laser light scattering
MBP	maltose-binding protein
MESNA	sodium 2-sulfanylethanesulfonate
MQW	Milli-Q® water
MPAA	4-mercaptophenylacetic acid
NCL	native chemical ligation
NGFP	N-terminal fragment of GFP
NGFPN	NGFP as an N-terminal tag
NGFPC	NGFP as an C-terminal tag
NLS	nuclear localization sequence
NMR	nuclear magnetic resonance
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCA	protein complementation assay
PCR	polymerase chain reaction
pdb (code)	Protein Data Bank accession code
PIC	preinitiation complex
PMSF	phenylmethylsulfonylfluoride
ppm	parts per million
REs	regulatory (response) elements
RNAPII	RNA polymerase II
SD-L-W-H-A	SD media lacking leucine, tryptophane, histidine and adenine
SD media	synthetic dropout media
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography

MALLS	multi-angle laser light scattering
TCEP	tris(2-carboxyethyl)phosphine
tethered construct	two domains connected by a flexible linker with or without a protease site
tethered complex	intramolecular complex formed by domains of a tethered construct
TEV protease (TEV)	tobacco etch virus protease
TFs	transcription factors
TFA	trifluoroacetic acid
tris	tris(hydroxymethyl)aminomethane
X-α-gal	5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside
Y2H	yeast two-hybrid
YPD	yeast peptone dextrose
ZnFs	zinc fingers

Chapter 1 Introduction

1.1 Complex interactions of transcription factors

Regulation of gene expression is a process that has likely been in place for more than 3 billion years, and is still ongoing in every living cell today. It follows inherited genetic programs and epigenetic effects, which are modulated by extracellular signals that are, in turn, brought about through actions of macromolecular complexes.

Transcription factors (TFs) are proteins that are involved in the regulation of gene expression at the level of transcription and take part in the chromatin-bound macromolecular complexes. TF complexes interact with gene-regulatory elements (REs), RNA polymerases, histones, and with each other, to activate or repress transcription [1-4]. These complexes are dynamic and overlapping entities that include the pre-initiation complex (PIC), Mediator complex (or other co-activators), chromatin remodelling complexes (CRCs such as NuRD and BAF), insulators and others. The complexity of TF-interactions is evident in a following model of gene activation.

In order to activate transcription, 'pioneer' transcription factors infiltrate more densely packed chromatin and bind regulatory elements (REs) in the regions that have been primed (or 'bookmarked') by CRCs (Figure 1.1) [5].

Master TFs (e.g., Oct4, Sox2, Nanog, Klf4, etc.) bind primed so-called super-enhancers and are pulled out into more active regions of chromatin while larger complexes are progressively assembled and the local chromatin structure is modified due to CRC activity. The larger complexes potentially prevent the diffusion of regulatory elements back into inactive chromatin. The enhancer-bound complexes recruit components of Mediator (a multiprotein assembly that acts as a co-activator of transcription) or other co-activator complexes, whereas the promoter-bound complexes (which contain a pre-initiation complex at various degrees of assembly) are brought from less active chromatin regions to join them. In some cases, insulator factors such as CTCF rearrange to release a loop that contains the gene that is about to be activated. Chromatin looping caused by interactions between these two types of complexes is

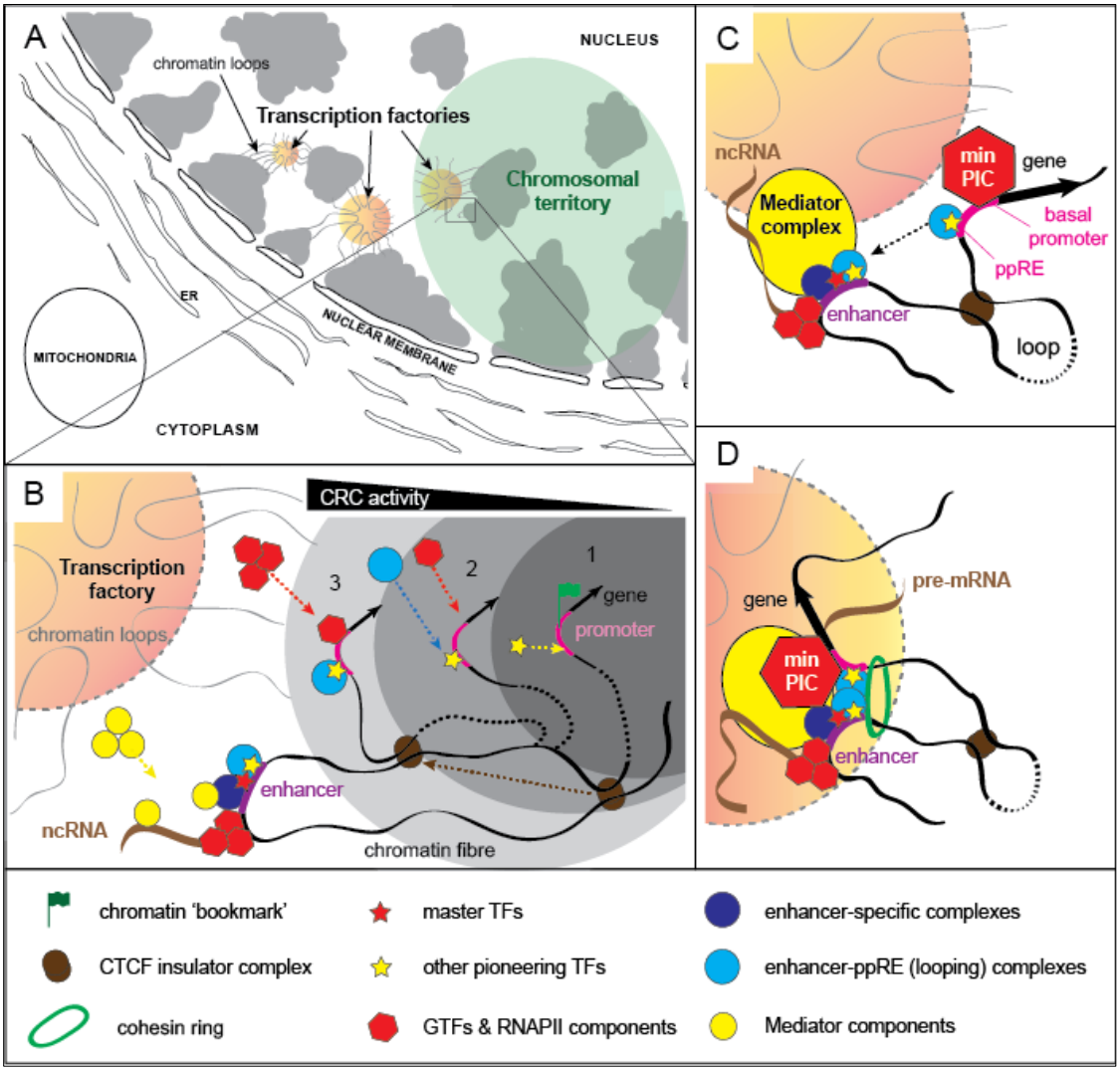


Figure 1.1 A stylized model of transcriptional activation in eukaryotes. (A) So-called transcription factories in the active chromatin are represented by coloured circles and the associated chromatin loops are shown as grey and black lines. A chromosomal territory (area in the nucleus where an interphase chromosome is located) is shown as a green oval. (B) Activation of gene expression. The activity of chromatin remodelling complexes (CRCs; not shown) transforms densely packed inactive chromatin (dark grey) to relaxed active chromatin (white). A typical enhancer in active chromatin is bound by master transcription factors (TFs) and subsequent multiprotein transcriptional complexes. Transcription of non-coding RNA at enhancers attracts RNA-binding components of Mediator to move the assembly towards a local transcription factory. A gene that is about to be activated is 'bookmarked' by CRCs, resulting in a specific 'histone code' that makes the promoter accessible to pioneer TFs (position 1). As additional TFs bind, the promoter is pulled towards the more active chromatin (positions 2 and 3). CTCF insulators release unfavourable loops and create favourable loops that allow activation of a particular gene. Appropriate complexes are progressively assembled at proximal promoter RE (ppRE) and basal promoter (e.g., TATA block) regions. (C) Enhancer-ppRE connecting complexes likely initiate and (D) support the interactions of Mediator (or a similar coactivator) as it covers the basal pre-initiation complex (PIC) with support from cohesin rings. At this point, the promoter is brought into the transcription factory, the RNAPII in PIC is activated and gene transcription starts, generating pre-mRNA. min PIC = minimal pre-initiation complex (complete PolII and all GTFs); ppRE = promoter-proximal response elements; GTF = general transcription factors.

often stabilized by cohesin rings and is characteristic of transcription factories, which are hot-spots of transcription in the nucleus.

During Metazoan development, induced transcriptional programs determine the ultimate fate of stem cells, so specific combinations of TFs are often used as markers of specific cell progenitors and of differentiated cells. Even in terminally differentiated cells, functions and response to stimuli are maintained by actions of different specific TF complexes.

Mutations and agents that affect TF complexes can lead to developmental disorders, cancer and other disease, but conversely, gene regulation shows significant plasticity due to multiple pathways which can compensate for non-functional proteins [6].

1.2 Multidomain nature of TFs

Protein-expressing genes are classified into gene families according to DNA sequence similarities that reflect the common origin and evolution of these sequences. Proteins are often classified according to common structural and/or functional elements (referred to as domains) or by their participation in particular protein complexes or biological processes. As most proteins contain several domains and are multifunctional, these classification systems partially overlap. TFs usually contain DNA-binding domains such as zinc fingers (ZnFs), helix-turn-helix (HTH), basic helix-loop-helix (bHLH), leucine zipper and others, but they generally contain additional domains that can mediate interactions with proteins, RNA and other cofactors. Metazoans tend to have TFs that contain conserved combinations of domains that give rise to complex protein networks that regulate inter-cell communication and create complexes that bind a large repertoire of REs in the non-coding genomic DNA. For example, a few thousand TFs regulate the expression of >20000 genes in the human genome through interactions with over a million known REs [7, 8]. The TF toolkit is enhanced by the existence of alternatively spliced polypeptide variants and modifications.

1.3 Homeodomains (HDs)

Homeodomains (HDs) are ~60 residue DNA-binding HTH domains that were first discovered in proteins coded by homeobox regions of fruitfly homeotic genes. HDs contain an

N-terminal basic 'arm', and two antiparallel shorter α-helices on top of a third, longer, perpendicular helix (Figure 1.2). HDs typically bind the TAAT core DNA sequence as monomers, with lower intrinsic specificity in vitro. Helix 3 is essential for DNA binding and the N-terminal arm that wraps around DNA provides a degree of specificity in some cases (Figure 1.2A). However, HD proteins commonly cooperatively bind as homo- or heterodimers on hybrid or neighbouring DNA sites, which is thought to provide RE-specific binding of complexes in vivo (Figure 1.2B and C) [9-14]. Many HD-proteins contain multiple HD and/or other domains. For example, Zfhx3 contains four HDs and 22 C_2H_2 ZnFs.

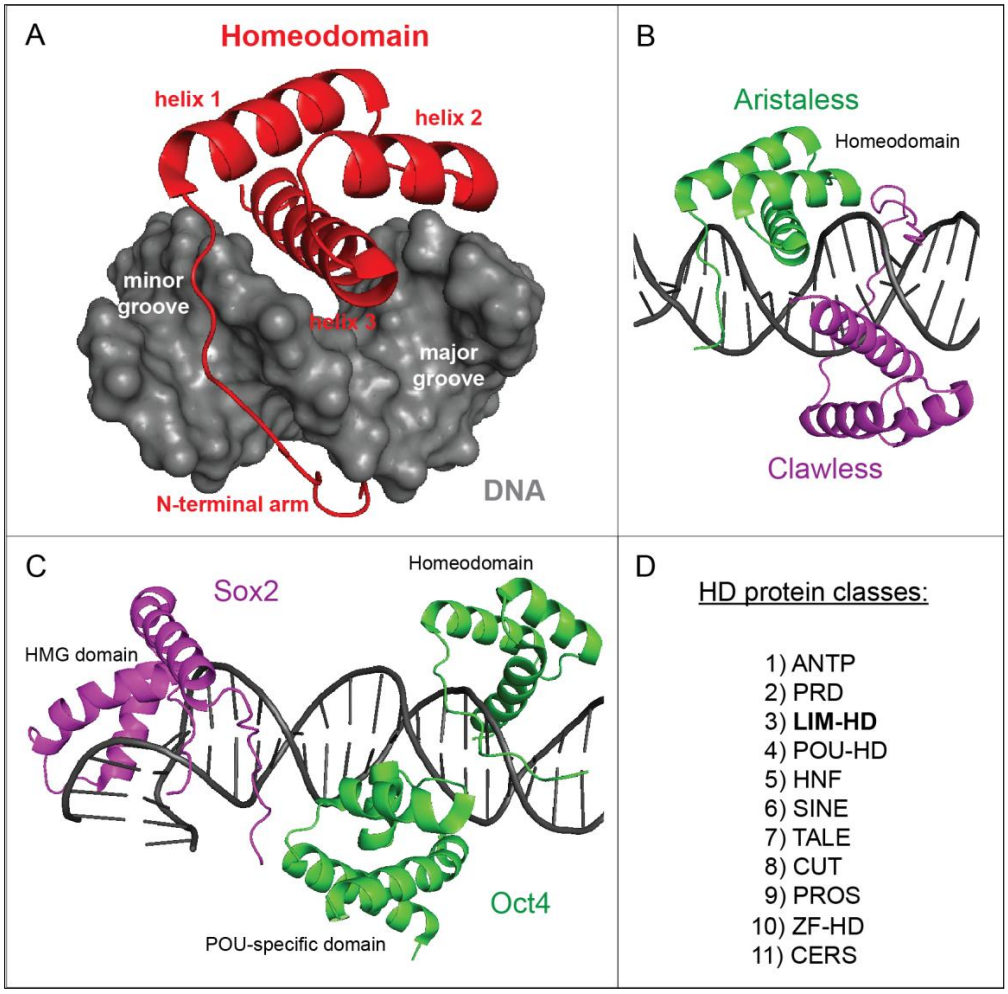


Figure 1.2 Homeodomain (HD) structure and classification. (A) The Clawless HD:DNA complex with DNA in grey surface representation and the HD shown as a red ribbon (pdb code 3A01). (B) HDs from the Drosophila proteins Aristaless and Clawless form a heterodimer on DNA (pdb code 3A01). (C) The Mammalian Oct4 and Sox2 proteins form heterodimeric complexes on DNA (pdb code 1GT0). (D) Classes of HD-containing proteins in humans. DNA-binding specificity is increased by dimerisation on hybrid DNA sites.

Proteins that contain HDs are transcription factors involved in cell specification and the maintenance of cell-specific gene expression in terminally differentiated cells.

Human homeobox genes can be classified into 11 classes (Figure 1.2D), which collectively contain over 100 gene families with a total of 235 genes [15]. One of these classes contains the LIM-HD proteins.

1.4 LIM-homeodomain (LIM-HD) proteins

LIM-HD proteins are a class of related transcription factors that contain two N-terminal, tandemly arrayed, ZnFs known as LIM domains, a central DNA-binding HD and a C-terminal region (Figure 1.3A). Some LIM-HD proteins also have N-terminal regions (95 residues in Lhx6) and/or a LIM-HD connecting region (~80 residues in Lhx2). The C-terminal regions of some LIM-HD proteins were shown to cause transcriptional activation [16, 17] and/or contain additional protein interaction domains [18, 19]. In several cases, alternative splicing can result in variants (isoforms) with alternative N-terminal or C-terminal domains, as well as variants lacking other domains [20-22]. LIM-HD protein variants and etymology are described in Appendix A.

The twelve mammalian LIM-HD genes are grouped into six families, each consisting of a pair of paralogue genes (Figure 1.3B) thought to have formed by genome duplication some time before the last common ancestor of jawed vertebrates. In jawless fish and most other animal groups, one gene per family is present in the genome, whereas additional genome duplication events in teleost fish resulted in more paralogues in each family (Figure 1.3C)[23]. LIM-HD proteins are closely related to LIM-only proteins (LMO1–4), which contain related tandemly arrayed LIM domains but generally little other sequence, and it has been proposed that LMO genes descended from an ancestral LIM-HD gene [24]. Both protein groups were shown to function primarily as transcription factors as detailed below.

LIM-HD proteins can bind DNA as monomers and show a preference for A/T-rich binding sites (Figure 1.3D) [10, 11, 25, 26]. Several promoters and enhancers were shown to contain monomeric, homo- and heterodimeric LIM-HD binding sites [25, 27-38].

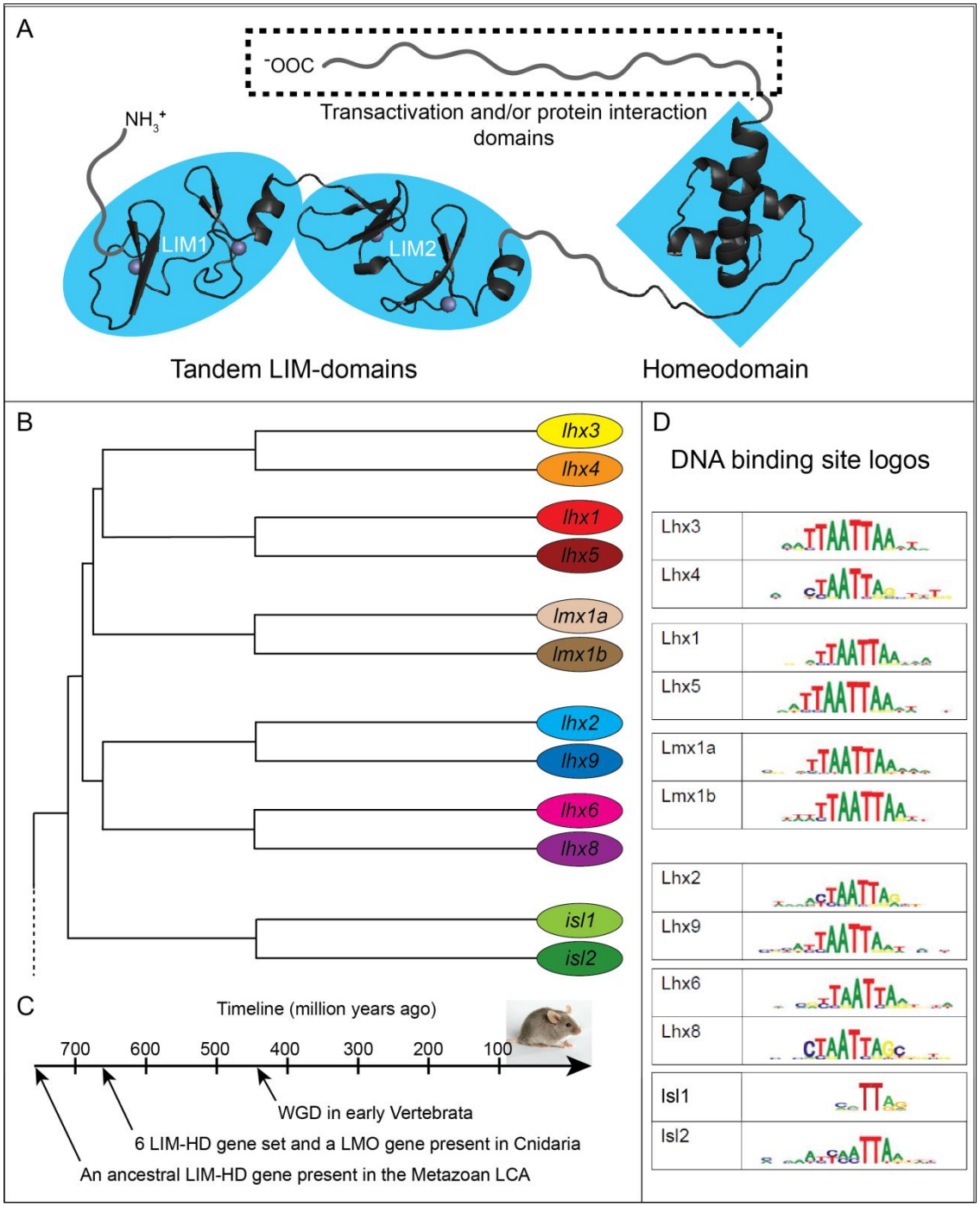


Figure 1.3 LIM-HD proteins. (A) LIM-HD protein structure schematic. Examples of tandemly arrayed LIM domains and HDs that are highly conserved in all LIM-HD proteins are shown; the remaining sequences are generally predicted to be intrinsically disordered. Positions of transactivation and/or protein interaction domains are indicated. The length and sequence of the intrinsically disordered regions varies between LIM-HD proteins. (B) LIM-HD phylogenetic tree [23, 24] with paralogue pairs shown in ovals of similar colours. (C) Evolutionary history of LIM-HD genes. WGD = whole genome duplication.LCA = last common ancestor. (D) Sequence logos of DNA-binding sites for LIM-HD monomers. Taken from Lee et al., 2008 [25](Isl1 logo only) and Berger et al., 2008 [10].

During animal embryonic development, LIM-HD proteins are expressed in cell progenitors and are involved in cell specification. LIM-HD proteins also help to maintain cell identity in specific differentiated cells throughout life. They are especially important for development of neurons in the central nervous system and sensory organs (retina, inner ear and olfactory neurons; Table 1.1). Patterns of expression for vertebrate LIM-HD proteins from the same family are mostly overlapping, but can differ in a few tissues. Combinations of two to three LIM-HD proteins from different families have been detected in the developing and adult pituitary, pancreas, intestine, limbs and reproductive system, whereas a specific LIM-HD protein or paralogue pair can be characteristic for the development of other cells and tissues, including heart, teeth, palate, kidneys, and red blood cells.

Table 1.1 Proposed functions of LIM-HD genes

Gene	Function	Gene	Function
Lhx3	Interneuron and motoneuron specification in the ventral spinal cord; dorso-ventral patterning in the pituitary and for anterior pituitary development [39, 40]; development of inner ear, retina and heart [41, 42].	***Lhx4***	Necessary for pituitary and lung development, axon projections of ventral MNs in spinal cord [41].
Lhx1	Roles in blastoporal organizer activity [17]; important role during gastrulation for axis formation, regulates gastrulation movements [43]; required in both primitive streak-derived tissues and extraembryonic tissues for head formation [44]; controls terminal differentiation and circadian function of the suprachiasmatic nucleus [45]; required for specification of the renal progenitor cell field [46]; directs the trajectories of motor axons in the limb [47].	***Lhx5***	Required for normal development and distribution of Cajal-Retzius cells in mammalian cerebral cortex [48-50]. Both Lhx1/5 are involved in development of the mammalian urogenital system [51], consolidate the inhibitory-neurotransmitter (GABAergic) program of dorsal interneurons in the spinal cord, control Purkinje cell differentiation in cerebellum, activate transcription of Wnt antagonists in the forebrain and specify horizontal cell laminar position in retina [49, 52-55];
Lmx1a	Specification of dopaminergic neurons with a correct midbrain identity; required for the proper formation of the roof plate, cerebellum, and the inner ear [56-58].	***Lmx1b***	Dorso-ventral patterning of limbs and patterning the otic vesicle [59]; important for specification of serotonergic neurons [60, 61] and dopaminergic neurons [58, 62]; regulates anterior eye development [63].

Gene	Function	Gene	Function
Lhx2	Lhx2 controls talamocortical axonal guidance [64]; co-ordinates multiple patterning events for the formation of the optic cup [65]; necessary for differentiation of the neuroretina [66]; involved in trophic hormone gene regulation in the pituitary, development of cerebral cortex (especially hippocampus), limb, and erythropoiesis.	***Lhx9***	Lhx2/9 have roles in patterning the nervous system, wing development, muscle development, axon guidance, neurotransmitter choice [67, 68]; required for development of olfactory sensory neurons. Lhx9 has a role in early gonadal development [69]; important in development of retina and the reproductive system. Lhx2/9 are required for midline axonal crossing in the spinal cord [70].
Lhx6	Roles in mandibular and maxilar paterning; in precursors of cortical interneurons expressing calbindin, parvalbumin, and somatostatin [71]; regulate migration of GABAergic neurons [72].	***Lhx8***	Important in development of basal forebrain cholinergic neurons [73]; important for survival of mesenchymal cells of the tooth germ in development [74]; critical in early folliculogenesis [75]. Lhx6/8 are important for development of striatal interneuron subtype [35, 76, 77].
Isl1	Specification of motoneurons and dorsal interneurons in spinal cord [78]; specification of cholinergic neurons [79, 80]; development of pancreatic mesenchyme and islet cells, development of cardiac progenitor cells, limbs and essential for early pituitary development [32, 40, 81-83]. Isl1 specifies the identity of hypothalamic melanocortin neurons [84].	***Isl2***	Motoneuron specification in the spinal cord and axonal pathfinding; development of retina, inner ear, pancreas; promotes midline crossing in the visual system by gene repression [85].

Homozygous knockout mutants of LIM-HD genes (Table A2) are embryonic or perinatal lethal in mice, except in the following cases. For *Lhx9* pups survive but show defects in gonadal development) [86]. Incomplete penetrance allows survival to weaning for *Lmx1a knockouts* and to adulthood for *Lhx8* (Table A.2, Appendix A). Mutations in LIM-HD genes are associated with a number of syndromes and diseases such as combined pituitary hormone deficiency (*Lhx3* or *Lhx4*), nail-patella syndrome (*Lmx1b*), cleft palate (*Lhx8*), Mayer-Rokitansky-Kuster-Hauser syndrome (*Lhx1*) and others. Aberrant methylation or expression of some LIM-HD genes is associated with some types of cancer (e.g. *Lhx1* in nephroblastoma, *Lhx2* in leukaemia, *Lhx6* in cervical cancer, *Lhx8* in odontoma, *Lhx9* in glioma) [86].

1.5 LIM-HD proteins and the TF-codes for neuron specification

In embryonic development, rostral-caudal (front-back) patterning of the spinal cord is initiated by opposing gradients of caudally-expressed fibroblast growth factor (Fgf) and rostrally produced retinoic acid (RA). These gradients translate into expression patterns of Hox genes that

correspond to their order in the Hox gene cluster. The dorsal-ventral (top-bottom) axis is governed by gradient of Sonic hedgehog protein (Shh; ventral), bone morphogenetic proteins (BMPs; dorsal) and Wnt proteins (dorsal). These gradients trigger transcriptional programs that specify the identities of cells in each progenitor zone, while acting to oppose adjacent transcriptional programs and sharpen boundaries between the neighbouring zones [87].

The combinatorial nature of LIM-HD protein expression in the spinal cord was shown to be important for differentiation of neuron subclasses that segregate into different spinal columns and select specific axonal pathways (Figure 1.4A). This combinatorial expression was particularly evident for motoneurons, and lead to proposition of a motoneuron 'LIM code' [90, 91]. Later, a similar hypothesis was proposed for the definition of boundaries in the developing brain [92, 93].

In the spinal cord, Lhx3 is the first LIM-HD protein that is expressed in mitotic cells from the pMN (progenitor of motoneuron) zone. The expression of the LIM-HD Isl1 and a HD-protein Mnx1(Hb9), is detected early in postmitotic motoneurons, overlapping with the expression of Lhx3 and bHLH factors (Figure 1.4A) [89, 94]. Expression of the LIM-only protein LMO4, and the LIM-HD proteins Isl2, Lhx1, and/or other TFs, is detected in cell groups that define spinal motor columns and later direct axons towards correct targets. Studies performed by Pfaff and colleagues suggest that potential feedback loops, mediated by Isl1, Lhx3, Mnx1, Vsx2 (Chx10 or Hox10), LMO4 and associated TFs, help to maintain motoneuron and V2 interneuron identities (Figure 1.4B) [25, 95].

The term LIM code is an oversimplification because other HD proteins, bHLH, ETS and forkhead TFs are also necessary to define each stage and transition in neuron specification. In vitro experiments show that coexpression of key transfected TFs, Ngn2, Isl1 and Lhx3 (collectively referred to as NIL programming), can rapidly and efficiently program cultivated and induced neural pluripotent stem cells (ipSCs) to acquire cervical motoneuron identity [96-98]. By expressing Phox2a instead of Lhx3 (NIP programming), ipSCs can instead be programmed to acquire cranial motoneuron identity [99].

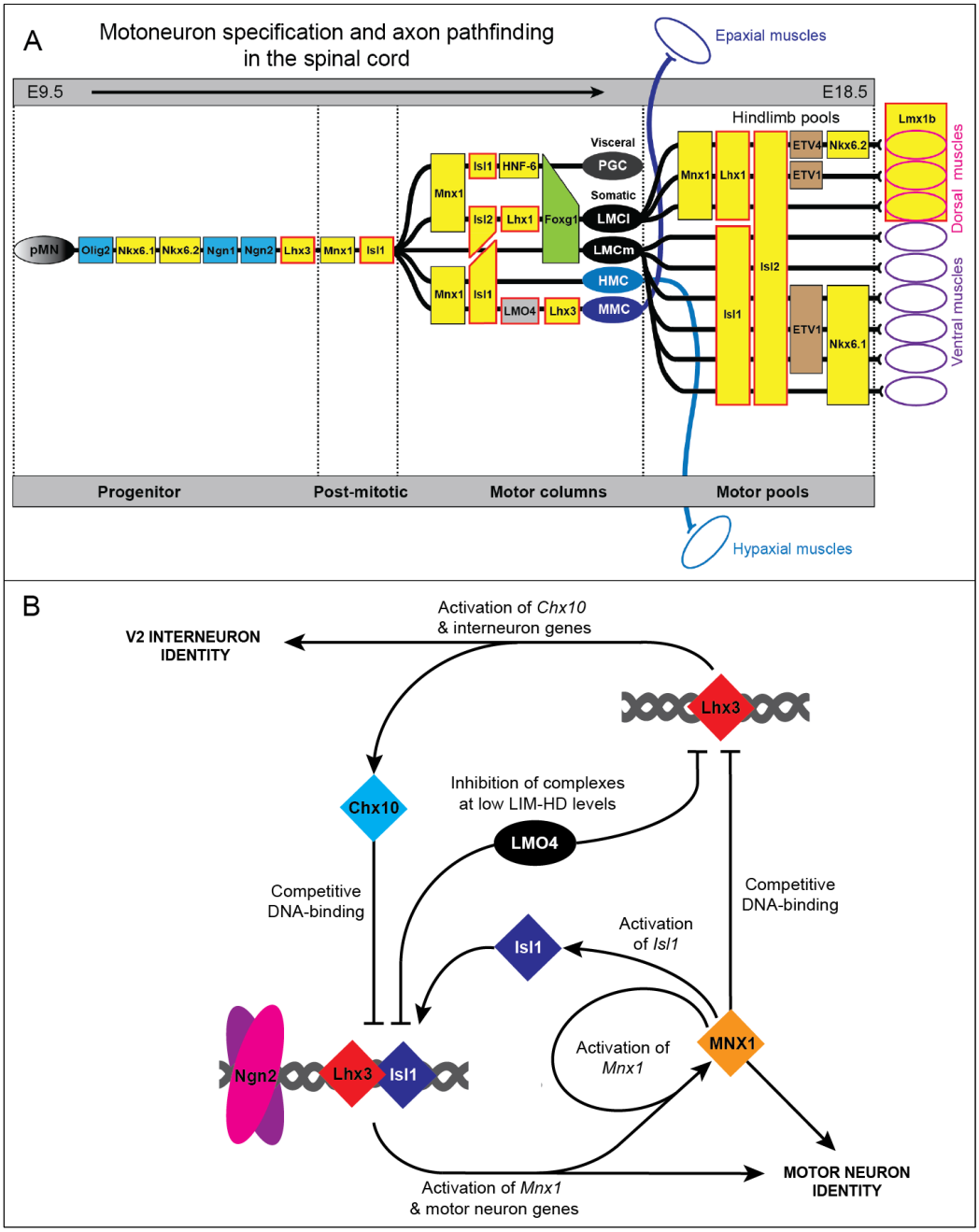


Figure 1.4 Transcription factor code in motoneuron specification and axon pathfinding. A) Proposed model of the TF expression program in spinal motoneuron cell fate determination and axon pathfinding. Ovals represent progenitors from the pMN, motoneuron columns or muscles. Expressed proteins are shown as squares. Sharp angles for Foxg1, Isl1 and Isl2 indicate lower concentrations of these proteins in specific lineages. HD proteins are in yellow and LIM-HD proteins have a red border. LMCl, LMCm, HMC, MMC and PGC refer to lateral portion of lateral motor column, medial portion of lateral motor column, hypaxial motor column, medial motor column and preganglionic column, respectively. This figure is adapted from figures in Alaynick et al., 2011 [88] and Lu et al., 2015 [89]. B) Isl1 and Lhx3 proteins are essential for proposed feedback loops that sharpen the boundary between pMN and p2 progenitor zones in the spinal cord. All HD proteins are shown as squares. LMO4 is thought to inhibit the formation of complexes from low levels of misexpressed LIM-HD proteins.

1.6 LIM-domains

In the early 1990's, a new type of zinc-coordinating domain was discovered in the N-terminal regions of the *C. elegans* HD proteins Lin-11 [100] and Mec3 [101], and rat HD protein Isl1 [102]. The domain was named LIM after the first letters of these proteins and the new protein class was named LIM-HD proteins.

A LIM domain is a ZnF consisting of two sequential zinc-coordinating modules, each of which forms a so-called treble clef fold [103], connected by a common small hydrophobic core (Figure 1.5A). LIM domains share the following consensus sequence:

CX_2**C**X_{16-23}(**H/C**)$X_{2/4}$(**C/H/E**)X_2**C**X_2**C**X_{14-21}(**C/H**)X_{1-3}(**C/H/D/E**),

where the Zn-coordinating residues are shown in bold font. Each of the treble clef folds closely resembles a GATA-type ZnF [104-106], except that in a LIM domain the C-terminal helices of the treble clefs are of variable length, and in the first modules are very short, comprising one turn only (Figure 1.5A). LIM domains contain conserved hydrophobic residues that stabilize the separate treble clefs, form a central LIM domain core and the hydrophobic pockets for protein interactions.

Unlike the GATA ZnFs they closely resemble, most LIM domains do not bind nucleic acids. Rather they bind proteins and often act as adaptors between multiple partner proteins [103, 107-110]. There are, however, some reports that LIM domains from a small number of proteins can bind non-specifically to nucleic acids [111, 112], but at this stage there are no well-characterised studies of DNA binding by LIM domains.

Many LIM-domain proteins are involved in regulation of cytoskeletal dynamics including actin polymerisation, cross-linking and bundling all of which build scaffolding at focal adhesions, adherens junctions and other structures. These molecular processes are a basis for cell spreading, motility, growth, cytokinesis and cell specification. Many LIM-domain proteins seem to shuffle between cytoplasm and the nucleus implying that they are involved in both cell signalling and gene regulation (Figure 1.5B). Given these functions, it is not surprising that LIM domain proteins play important roles in the development of nervous and muscular systems [110].

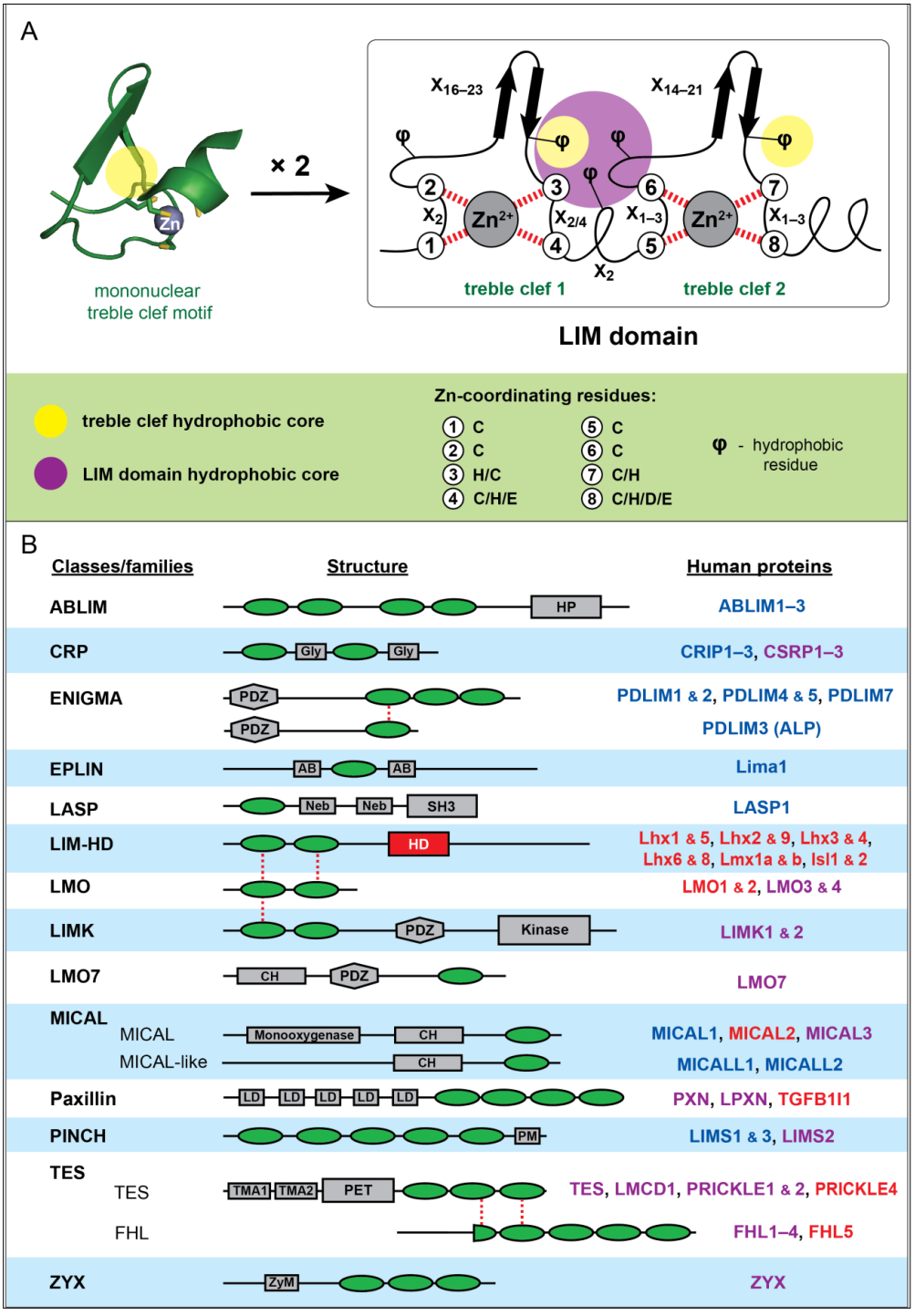


Figure 1.5. LIM domains and LIM domain proteins. A) LIM domain structures contain two fused treble clef motifs, each of which contains a small hydrophobic core (yellow) the first of which forms part of a larger central hydrophobic core (purple). B) Classes of LIM domain proteins, structure, proteins and localization. LIM domains are shown as green ovals, sequence similarity between domains is indicated by dashed red lines. The HD is shown as a red rectangle and other domains are indicated. Proteins in blue font are cytoplasmic, proteins in red font are nuclear and proteins in purple were found both in cytoplasm and nucleus.

1.7 Ldb1/2 cofactors stabilize LIM-HD/LMO proteins and can mediate interactions between relevant chromatin-bound complexes

LIM-domain binding (Ldb) proteins Ldb1 (CHIP/NLI/CLIM2) and Ldb2 (CLIM1) are transcription factors that specifically bind to the LIM-domains of all LIM-HD/LMO proteins, but not other LIM domain proteins [113, 114]. Ldb1 is widely expressed throughout development and in adult tissues, while the expression of Ldb2 is much more restricted [86, 115]. Ldb proteins act as adapters that connect different tissue or stage-specific combinations of TFs (Figure 1.6A) [113, 116].

Ldb proteins contain an N-terminal self-association domain (SA), followed by a Ldb/CHIP conserved domain (LCCD), nuclear localization signal (NLS) and an ~30 residue LIM-interaction domain (LID) (Figure 1.6B).

The ubiquitin-proteasome system (UPS) regulates the turnover of cytoskeletal, signalling and other proteins, including transcription factors. In neural development, the selective removal of TFs by this system is important for migration, axon guidance, neurite outgrowth and synaptic plasticity [117-120]. Rnf12 (RLIM), a E3 ubiquitin-protein ligase, was shown to bind Ldb1, ubiquitinate the protein at K170, and target the protein for proteosomal degradation (Figure 1.6C) [115, 121-123]. In contrast, single-strand DNA binding proteins Ssbp3 (Ssdp1) and Ssbp2 were shown to bind Ldb1, inhibit the interaction with RLIM, prevent RLIM-mediated ubiquitination, and protect Ldb proteins and complexes from proteasomal degradation [124-127].

The Ldb proteins were originally assumed to form dimers [128-139], but in vitro data for the isolated self-association domain, expressed from *E coli.* suggests that it may form higher order complexes [140]. Work is still ongoing to confirm the self-association state of the protein in vitro and in vivo.

Several studies have demonstrated that Ldb1 plays an important role in mediating long range interactions (DNA-looping) between RE-bound haematopoietic complexes. For example, complexes with LMO, GATA, Tal1/E12 and Klf1 TFs in the regulation of β-globin gene transcription (Figure 1.6D) [114, 134, 141-144] or complexes with Isl1 during the activation of genes in cardiac progenitor cells [38] were found to participate in DNA-looping. Ldb-containing complexes that contain Tal1/E12 and LMO proteins are predominantly found to act as

transcriptional activators, but some participation in repression complexes that contain Cbfa2t3 (Eto2) or Fog proteins has been reported [145-147].

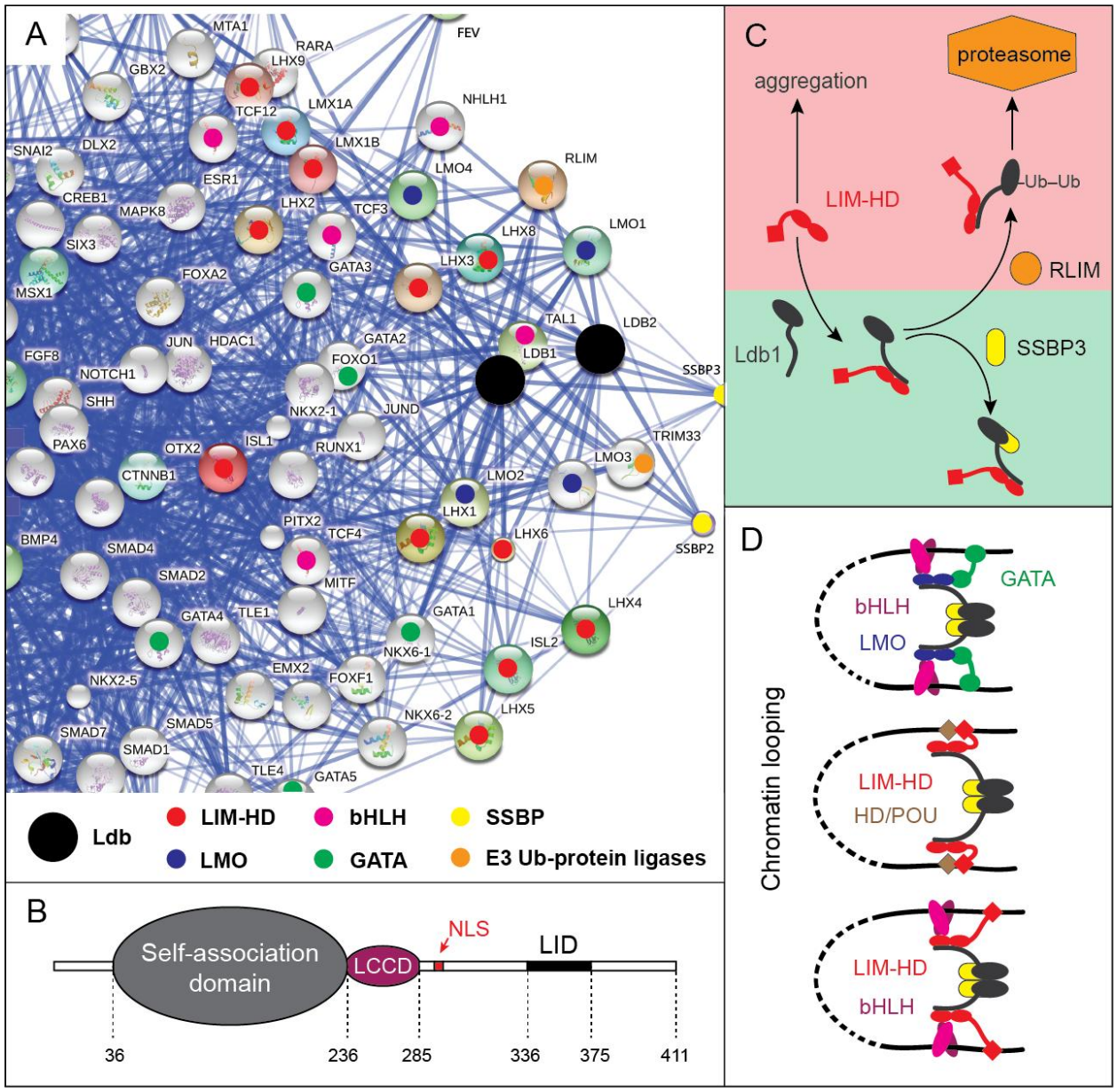


Figure 1.6. LIM-domain binding proteins. A) STRING database interaction map shows genetic and direct interactions, and the relative position of Ldb cofactors and LIM-HDs/LMOs, within the network of cellular interactions from the database. A colour code indicates protein types. B) Domain structure of Ldb proteins. The residue numbers shown correspond to Ldb1. LCCD = Ldb/CHIP conserved domain, NLS = nuclear localization signal and LID = LIM-interaction domain. C) Protein interactions protect LIM-HD proteins from aggregation and Ldb proteins and their partners from proteasomal degradation. Red shading indicates instability and green shading indicates stability. D) Different core Ldb-mediated complexes can loop and mediate contact between distant REs.

1.8 Interactions between Isl1/2 and Lhx3/4 proteins

In the early years of research on LIM domain proteins, various binding studies performed using cell lysates suggested that some LIM domains have the ability to specifically homo- or heterodimerize [148, 149]. Following the discovery of Ldb1 interaction studies suggested that it appeared that Ldb cofactors were likely to mediate all of those interactions *in vivo* [132].

Expression and biophysical characterisation of recombinant LIM-HD/LMO proteins showed that in the absence of Ldb1, the LIM domains of these proteins tend make nonspecific interactions, including forming protein aggregates *in vitro* [150-152]. However, several subsequent studies have demonstrated that some LIM-HD proteins can physically interact with each other and form Ldb1-containing TF complexes that play key roles in the definition of cell identities in the developing spinal cord [78, 132, 153].

Lhx3:Ldb1 complexes are formed through the interaction of the LIM domains from Lhx3 ($Lhx3_{LIMs}$) and the LID from Ldb1 ($Ldb1_{LID}$). These complexes are formed in two neighbouring zones in the developing ventral spinal cord: the p2 zone that produces V2 interneurons, and pMN zone ultimately produces motoneurons. Motoneuron progenitors also express Isl1 allowing the formation of Lhx3:Isl1:Ldb1 complexes, where the LIM domains of Isl1 ($Isl1_{LIMs}$) contact $Ldb1_{LID}$, and $Lhx3_{LIMs}$ interacts with the C-terminal region of Isl1 (Figure 1.7). This complex was shown to be crucial for activation of the *Mnx1* gene and motoneuron specification [25, 78], whereas the Lhx3:Ldb1 complex is required for V2 interneuron specification. The Lhx3:Isl1 complex was later shown to regulate gonadotrope-specific expression of the gonadotrope-releasing hormone receptor (Gnrhr) in the pituitary [32].

Our laboratory further investigated the Isl1:Lhx3 interaction and employed yeast two-hybrid (Y2H) analysis and truncation mutagenesis to define the minimal Lhx3-binding domain in the C-terminal region of Isl1. This ~30-residue domain is termed the LIM-binding domain (LBD) [18]. The laboratory further showed that an LBD is also present in the Isl1 paralogue, Isl2, and that each of Lhx3 and Lhx4 interact with each of $Isl1_{LBD}$ and $Isl2_{LBD}$ [19]. The Lhx3:Isl1 interaction probably exists in all vertebrates and many other animal groups where Isl1 and Lhx3 orthologues are coexpressed. For example, Lim3 and Islet are required for motoneuron

development in fruit flies [154], and the $Ceh14_{LIMs}$ and $Lim7_{LBD}$ domains from *C. elegans* protein orthologues are coexpressed in some neurons and can interact in Y2H assays [155].

An interaction between Lmx1a and Isl1, which was dependent on $Isl1_{LIMs}$, was detected in the same co-immunoprecipitation experiments that identified the Lhx3:Isl1 interaction [132, 154].

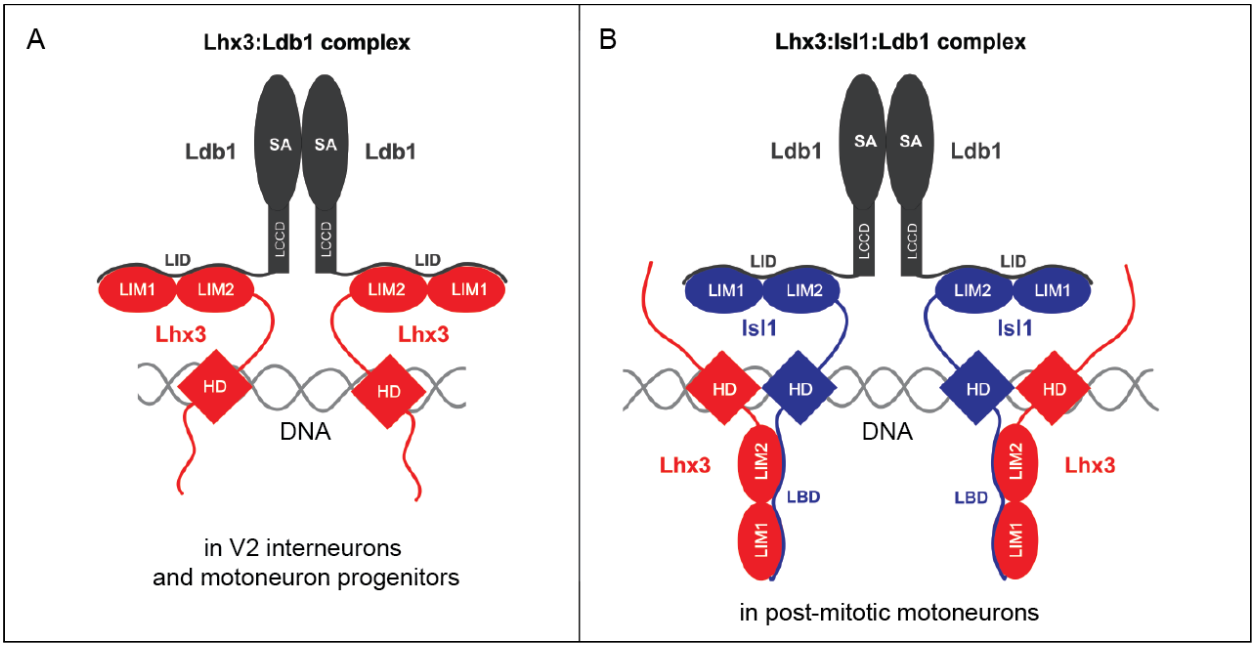


Figure 1.7 Two complexes that differentiate motoneurons from V2 interneurons in the developing ventral spinal cord. A) Lhx3:Ldb1 complex B) Lhx3:Isl1:Ldb1 complex. The stoichiometry of the Ldb mediated complexes is uncertain but for simplicity, is presented as dimeric.

1.9 General characteristics of LIM domain interactions

1.9.1 Tethered LIM-LID 'complexes'

Initial experiments on LIM-only proteins (see section 1.4), in which LMO2 and LMO4 were expressed in bacteria gave rise to very low yields of soluble proteins that were prone to aggregation, making structural studies impossible, as such studies require milligram quantities of soluble protein [150, 151]. Mutagenesis of two non-ligating cysteines to serines in LMO4 only marginally increased solubility, but it was noted that the co-incubation of LIM-constructs with

$Ldb1_{LID}$ peptide significantly reduced aggregation, albeit temporarily. Therefore, LIM domains were expressed as fusion constructs in which the LIM domains were tethered to $Ldb1_{LID}$ by a glycine/serine-linker (see Appendix C). The resulting intramolecular 'complexes' were soluble, folded and stable [150]. The incorporation of a specific protease site in the linker, and comparison of NMR spectra before and after proteolytic cleavage, showed that tethered LIM-LID constructs form native-like complexes [152, 156]. This tethering approach has subsequently allowed structure determination of numerous LIM-partner complexes, and studies of higher order assemblies to be carried out [18, 19, 150, 151, 156-165]. Tethered complexes were used in most in vitro experiments described in this thesis.

1.9.2 LIM domain interfaces and the β-zipper binding mode of linear partner motifs

Currently there are numerous structures that exist of LIM-partner complexes, including LIM domains from the LIM-HD/LMO classes as well as other LIM proteins. These partners can be broken into two main groups: peptides binders (such as LID and LBDs) and globular proteins. Whereas globular proteins can bind various surfaces on LIM domains, including partially exposed hydrophobic treble clef cores (e.g., LMO2:Tal1:E47, pdb code 2YPA; PINCH:ILK, pdb code 2KBX; and TES:MENA, pdb code 2XQN), to date all of the peptide binders bind a common interface, referred to here as the 'front' interface. This front interface is characterised by exposed beta sheets and hydrophobic pockets and surfaces, and is found in LIM-HD/LMO complexes with Ldb1, Isl1/2, CtIP and DEAF1 (pdb codes: 1RUT, 2XJY, 2RGT, 3MMK, 2JTN, 2L4Z, 2MBV), and in the TES:ARP7A complex (pdb code: 2SQN) (Figure 1.8A and B). This class of interaction is discussed in more detail below and an example of binding to other surfaces is shown for LMO2:Tal1:E47 (Figure 1.8C and D).

LIM-binding peptides are generally unstructured prior to binding, and bind the front interface of the LIM domains in an extended conformation in which short β-hairpins in the LIM domains are augmented by short antiparallel β-strands in the peptide, forming a β-zipper [166]. Although peptide binding partners of LIM domains share low sequence identity, hydrophobic residues are conserved in a few positions. These conserved residues, some of which have been

identified as hot-spot residues from mutagenic studies, cover hydrophobic surfaces or fill hydrophobic pockets on the surface of the LIM domain (Figure 1.9A and B). The binding of peptides to tandem LIM domains is modular, where each LIM domain binds to a linear motif in the peptide. Cooperative binding of these modules leads to high affinity and apparently high specificity binding [18, 167]. Interestingly, in some other ZnFs and LIM-domains sequences N- or C-terminal to the ZnF form similar antiparallel β-sheets, but via intramolecular rather than intermolecular β-augmentation.

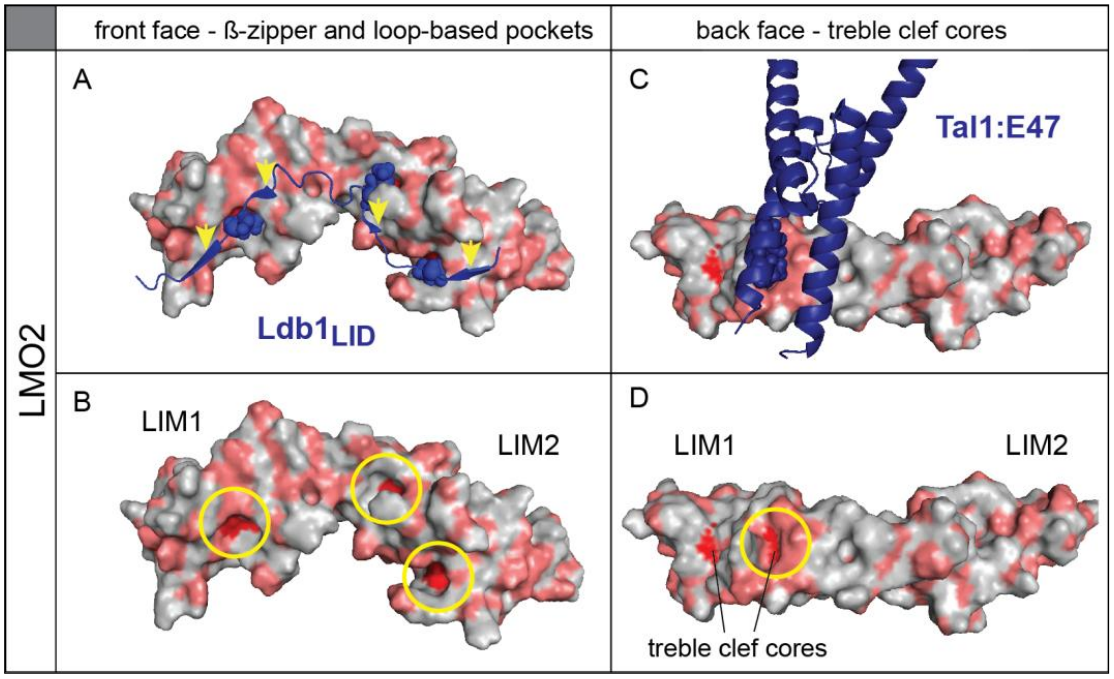


Figure 1.8 Alternate binding modes of tandem LIM domains of LMO2. Hydrophobic atoms are shown as dark salmon with hydrophobic core residues shown in red. A) $Ldb1_{LID}$ binds the front interface of LMO2 in a β-zipper conformation (yellow arrows). Anchoring or 'hot-spot' hydrophobic residues of LID, as defined by mutagenic studies, are shown as blue spheres. B) LMO2 front surface (bound state) shows hydrophobic pockets in which core residues are exposed (yellow circles). C) Tal1:E47 dimer binds to LMO2 back interface and covers a pocket with exposed core of treble clef. Hydrophobic residues of the bHLH dimer are shown as blue spheres. D) Back surface of LMO2 (bound state) with the interacting hydrophobic pocket (yellow circle).

The LIM domains from LIM-HD and LMO proteins share a number of conserved hydrophobic, charged and glycine residues that are not generally conserved in all LIM domain proteins, and preferences for specific residues or residue types exist at many positions (Figure 1.10). These residues are generally not found in the LIM domains from LIM kinases (LIMK1/2), which contain closely related tandem LIM domains, but which do not bind Ldb1, suggesting that

those residues are responsible for the specific binding of Ldb1 to LIM-HD and LMO but not other LIM domain proteins [152]. Mutations of several of these LIM-HD/LMO specific residues in $LMO4_{LIM1}$ to corresponding residues from LIMK1 was shown to inhibit binding to $Ldb1_{LID}$ [152].

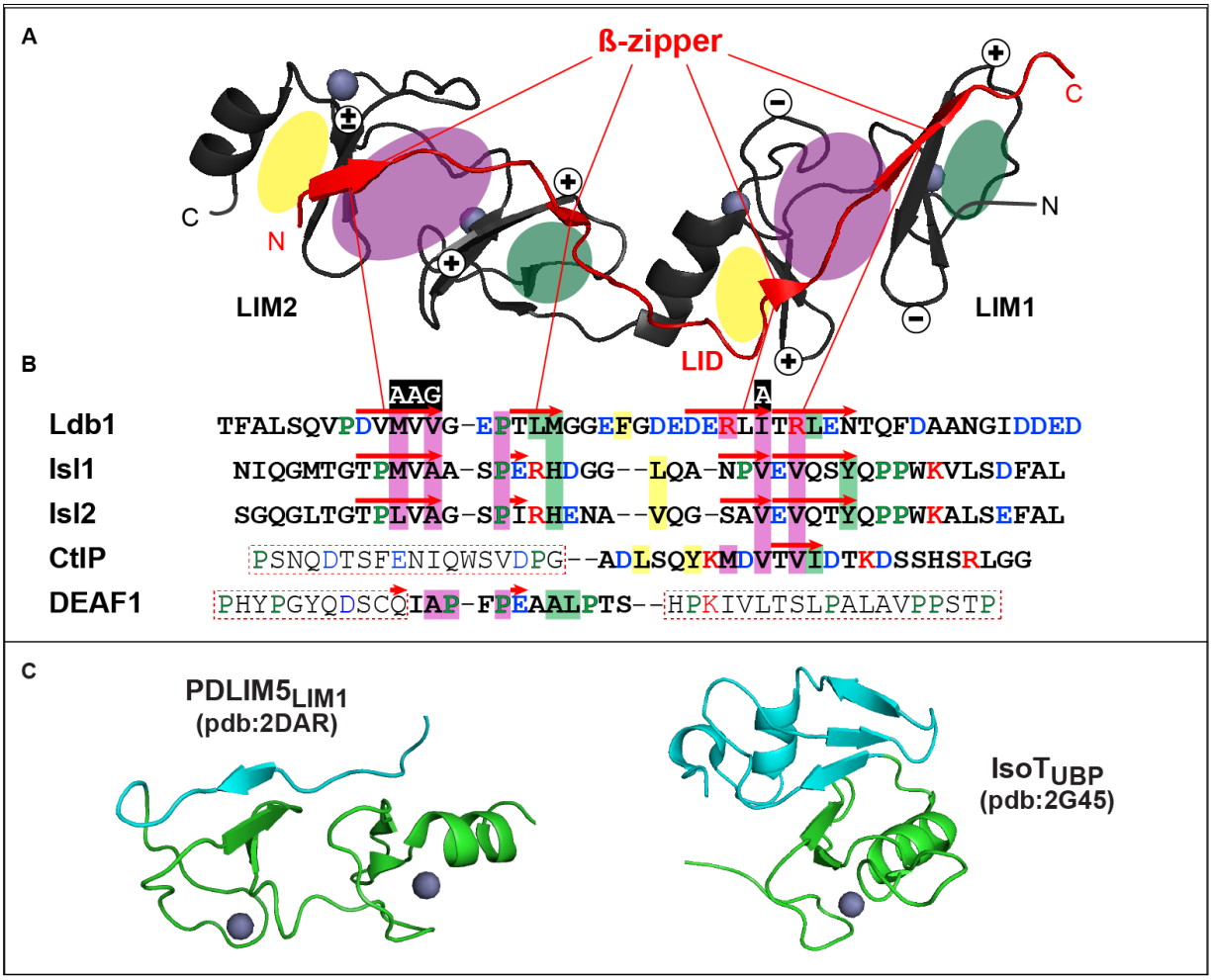


Figure 1.9 Linear motifs form β-zipper motifs at 'front' interfaces of LIM domains. A) Ribbon representation of the β-zipper structure of the LMO2 (black):$Ldb1_{LID}$ (red) complex (pdb code:2YPA). In each LIM domain, a hydrophobic pocket created by the loop of the first treble clef motif is indicated in green, the LIM core in purple, and core of the second treble clef in yellow. Coordinated zinc atoms are grey spheres. Positions of conserved charged residues are indicated. B) Alignment of structurally characterised LIM-HD/LMO-interacting peptide domains. β-sheets are shown as red arrows, and hotspot residues that interact with specific LIM surfaces or pockets are highlighted in the same colour as the corresponding hydrophobic region. Acidic residues are shown in light blue, basic residues are in red, and prolines are green. Parts of LIM-binding motifs for which the structure has not been determined are in rectangles with red dashed borders. Alanine or glycine mutations that significantly decrease binding are shown above the alignment (white font with black highlights). C) Examples of ZnF domains that exhibit intramolecular β-augmentation are shown in cartoon representation. Complementing β-strands segments are shown in cyan.

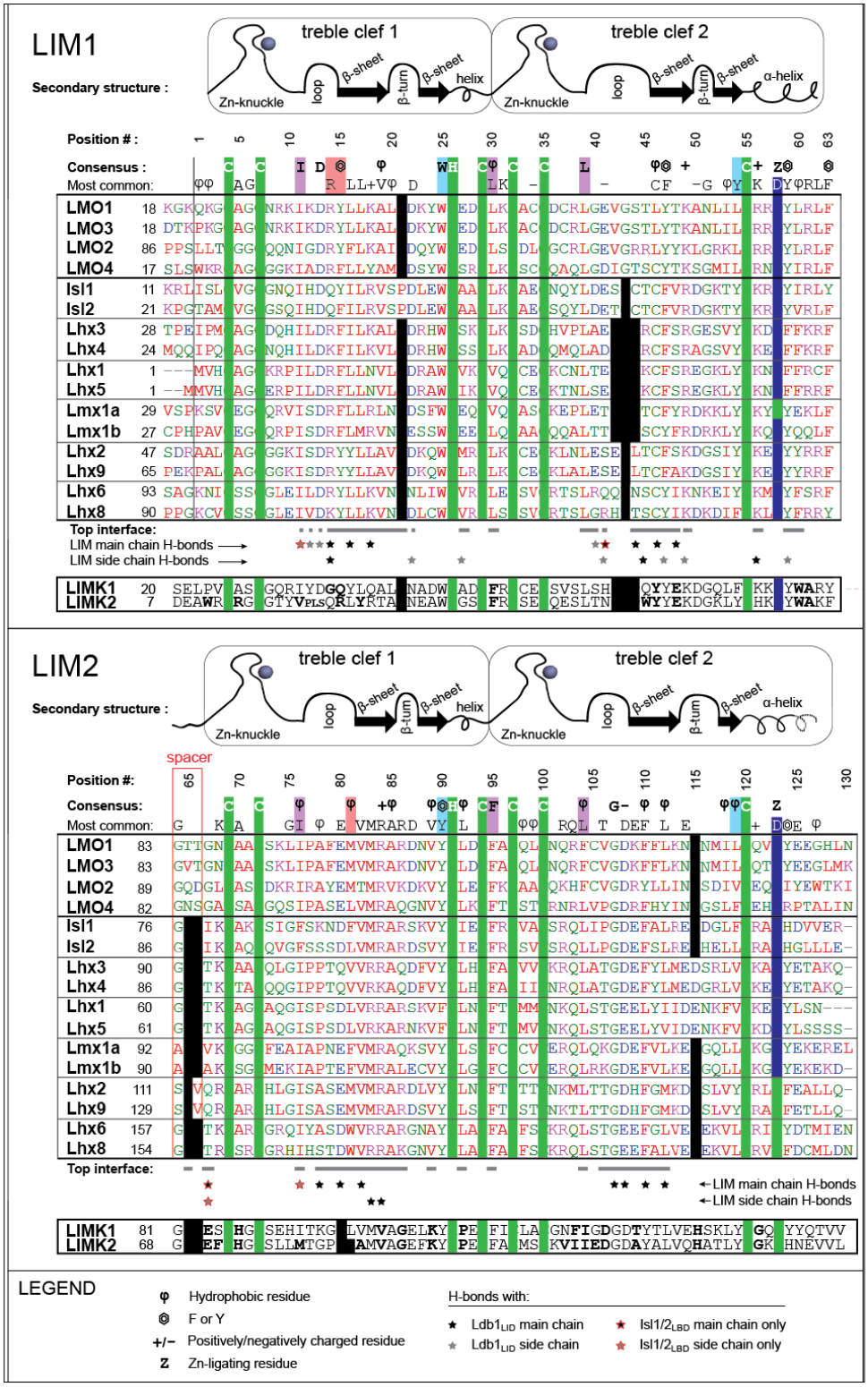


Figure 1.10 Comparison of LIM domain amino acid sequences from LIM-HD/LMO and LIMK1/2. LIM1 alignment. Residue position 1 was set 3 residues before the first Zn-ligating cysteine. LIMK1/2 residues in bold font indicate potentially significant differences compared to LIM-HD/LMO residues at the same positions.

Tandemly arrayed LIM domains (up to five) exist in many LIM-domain proteins but the spacers between them vary from ~50 in CRP1/2 and 8–10 residues in some other LIM proteins to 1–3 residues in LIMK1/2, LIM-HDs and LMO proteins (Figure 1.10) [108]. Other multi-LIM proteins also contain closely spaced LIM domains (e.g., TES LIM2 and LIM3). As there are generally few contacts between the LIM domains, it is thought that in the unbound state the LIM domains flex relative to one another, but close spacing creates an essentially continuous interaction interface, and the binding peptides provide interdomain constraints and influence the relative angle between the LIM domains [19, 165, 168].

1.10 Potential interactions between other LIM-HD proteins

The coexpression of Lhx3 and Isl1 proteins in the developing spinal cord is only one of many examples of overlapping expression patterns of LIM-HD proteins. Therefore, it seems feasible that other LIM-HD proteins could interact in a similar fashion. Indeed, as mentioned above, Lmx1a showed some evidence of binding to Isl1 in co-immunoprecipitation experiments [132], and Y2H experiments performed in our laboratory indicated that an Lmx1a:Isl1 interaction is mediated by $Lmx1a_{LIMs}$ and $Isl1_{LBD}$ [169]. Other LIM-HD proteins contain sequences with unidentified functions, especially in the C-terminal regions, so it is possible that LIM-domain binding motifs exist in LIM-HD proteins other than Isl1/2.

As noted at the beginning of this chapter many HD-proteins can cooperatively bind DNA through direct HD-HD interactions [12, 170, 171]. This could point to a mechanism for the interaction of LIM-HD proteins, independent of, or in addition to, interactions via the LIM domains.

It is also possible that intramolecular contacts between the LIM domains and LBDs from the same protein can take place. Such interactions may reversibly affect localization and/or protein function of some LIM domain proteins [172, 173]. The autoinhibitory effects of some LIM domains on HD-mediated DNA-binding have been reported [174] and recent work from our laboratory showed that the LIM and LDB domains in Isl1/2 do interact in an intramolecular fashion, which may protect against non-specific binding and non-productive complex formation in the absence of other binding partners [162].

1.11 *In vitro* competition experiments can investigate binding affinities of competing LIMs:LID/LBD complexes

Given the poor solution properties of isolated LIM domains from LIM-HD and LMO proteins, it is not possible to measure binding affinities between their LIM domains and peptide binding partners using standard methods. However, it is feasible to use homologous competition experiments to estimate binding affinities for LIM domain interactions with peptide partners.

In the past, homologous competition ELISAs were used to estimate binding affinities for this family of proteins. For those ELISAs, GST-LIMs-LID constructs (containing LIMs from LMO2, LMO4, Lhx3 and Isl1) were generated that contained a protease site in the tether between the LIMs and LID. The purified proteins were treated with the protease, and the cut complexes bound to ELISA plates that had been pre-coated with anti-GST antibodies. Increasing concentrations of $Ldb1_{LID}$ peptide containing a C-terminal FLAG-tag were added. After washing, bound FLAG-tagged peptides were detected using an anti-FLAG antibody conjugated to a peroxidase. The initial study resulted in binding curves suggesting that this format of ELISAs can be used to compare in vitro binding of $LMO4_{LIMs}$ to wild type and mutant $Ldb1_{LID}$ peptides, and gave an estimate of IC_{50} = 1 nM for the wild type LMO4:$Ldb1_{LID}$ interaction [156]. After more experiments using lower starting concentrations of $Ldb1_{LID}$-FLAG, this estimate was revised to ~10 nM and the IC_{50} for the interactions of $Ldb1_{LID}$ with three other LIM domain proteins were reported: 20 nM for LMO2, 35 nM for Lhx3 and 90 nM for Isl1 [18, 167].

In these experiments, it was assumed that non-specific binding is not significant, binding affinities are not affected by protein tags and that surface-adsorbed complexes can approximate low concentrations of complexes in solution. It was further assumed that incubation allows for the system to come to equilibrium and that subsequent washing steps do not significantly affect this equilibrium [175, 176]. Based on these assumptions, the IC_{50} is a reasonable approximation of K_d. However, it is likely that these assumptions do not hold, especially through numerous washing steps. Further, the method cannot measure weaker interactions, and data can be inconsistent. Thus, part of the project described in this thesis focussed on exploring approaches to generate a more robust assay for this system.

1.12 Aims of the thesis

LIM-HD proteins are involved in many developmental and disease processes but questions remain about their mechanisms of action, especially with respect to overlapping expression patterns and their potential to form multi-protein transcriptional complexes. Given that LIM-domains of Lhx3/4 interact with LBD-domain of Isl1/2, do other LIM-HD proteins interact with Isl1/2, or with each other? How do the different LIM domain modules contribute to binding $Ldb1_{LID}$ and similar polypeptide sequences? What are the binding affinities of the different tandem LIM-domains to $Ldb1_{LID}$? If there is a competition between LIM-HD proteins for their common partner Ldb1, could differences in binding affinities explain the formation of specific complexes? Answering these questions should lead to a better understanding of cell specification associated with this family of LIM domain proteins. This information could be important for the development of treatments to inhibit neoplasms associated with the overexpression of various LIM-HD and LMO proteins, to regenerate damaged tissues with appropriately induced stem cells, or to induce proper axon pathfinding in recovering nerves. The key aims of this thesis were to develop a number of experimental procedures that could provide some of these answers.

The specific aims, as broken down into results chapters were:

1. To apply Y2H screening approaches based on yeast mating to detect interactions between LIM-HD proteins (Chapter 3).
2. To establish the contributions from individual and tandem LIM-domains of LIM-HD proteins to interactions with $Ldb1_{LID}$ (Chapter 4).
3. To develop fluorescence-based complementation assays in yeast to address some of the limitations of Y2H experiments (Chapter 5).
4. To develop strategies for the specific N-terminal fluorescent labelling of tethered LIM-LID complexes for use in solution-based homologous competition assays as an alternative to ELISAs (Chapters 6 and 7).

Chapter 2 Materials and methods

2.1 Materials

2.1.1 Chemicals, enzymes and kits

Table 2.1 Chemicals.

Chemical	Manufacturer/Source
2,2-Dimethyl-2-silapentane-5-sulfonic acid (DSS)	Fluka
$^{2}H_2O$ (Deuterium oxide)	Aldrich
2-Mercaptoethanol	Sigma-Aldrich
2-Log DNA standard	New England Biolabs
3-Amino-1,2,4-triazole (3-AT)	Sigma
3,5-Dimethoxy-4-hydroxycinnamic acid (sinapinic acid)	
4-Mercaptophenylacetic acid (MPAA)	
5-Bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α- gal)	GOLD Biotechnology
α-Cyano-4-hydroxycinnamic acid (CHCA)	Sigma
Agarose	Sigma-Aldrich
Ampicilin (Amp), sodium salt	
Arabinose,	
Bis-Tris	
Bovine serum albumin (BSA, monomeric)	
Chloramphenicol	GOLD Biotechnology
Complete EDTA-free protease inhibitor cocktail tablets	Roche
Coomassie Brilliant Blue G-250	Bio-Rad
Deoxyribonucleotide triphosphates (dNTP for PCRs)	Bioline
Dimethylformamide (DMF)	Sigma-Aldrich
Dithiothreitol (DTT)	GOLD Biotechnology

Deoxyribonuclease I (DNaseI)	Roche
EDTA	Sigma-Aldrich
Ethidium bromide	
Ficoll	GE Life Sciences
Glutathione (GSH)	Sigma-Aldrich
HEPES	Sigma-Aldrich
Imidazole	Sigma
Immersion oil	Cargille
Isopropyl-β-D-thiogalactopyranoside (IPTG)	GOLD Biotechnology
Lithium Dodecyl Sulphate (LDS)	Sigma-Aldrich
Maltose	
Mark 12 protein standard	Life Technologies
$MgCl_2$	Sigma-Aldrich
MOPS	
Paraformaldehyde (PFA)	
Polyethylene glycol (PEG) 4000	
Rainbow™ molecular weight marker	GE Life Sciences
Sorbitol	Sigma-Aldrich
Salmon sperm (10 mg/mL)	Thermo Fisher Scientific
Trifluoroacetic acid (TFA)	Sigma-Aldrich
Tricine	
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl)	Sigma

All other standard chemicals and reagents were obtained from Boehringer Mannheim (Castle Hill, NSW), ICN Biochemicals (Seven Hills, NSW), Promega (Madison, WI), Sigma (Castle Hill, NSW) or Univar (Auburn, NSW).

Stock solutions of simple sugars (D-glucose, maltose and arabinose), sorbitol, Ficoll and glycerol were made as syringe-filtered 40% or 50% stocks. $MgCl_2$ was made as a 2 M stock and sterile filtered.

Table 2.2 Enzymes and antibodies.

Enzyme	Manufacturer
Restriction endonucleases	New England Biolabs
Factor Xa	
HRV3C protease	Produced in house by Angela Nikolic, SMB. Specific activity not specified.
Lysozyme	Sigma
Pfu Turbo DNA polymerase	Stratagene
T4 DNA ligase	New England Biolabs
Taq DNA polymerase	Boehringer Mannheim
TEV protease	Produced as a part of this thesis, section 2.5.3.2
Thermosensitive Shrimp Alkaline Phosphatase (TSAP)	Promega
Sequencing Grade Trypsin	
Rabbit anti-GFP polyclonal IgG	Amersham biosciences
HRP-conjugated anti-rabbit IgG	

Table 2.3. Kits used in this study.

Kit	Manufacturer
QIAprep Miniprep Kit	QIAGEN GmbH
QIAquick PCR Purification Kit	
QIAquick Gel Extraction Kit	
Expand Long Template PCR System	Roche
QuikChange II Site-Directed Mutagenesis Kit	Agilent Technologies
SNAP-Cell Starter Kit	New England Biolabs
Pierce™ ECL Western Blotting Substrate kit	Life Technologies

2.1.2 Buffers and solutions

All reactions performed with commercial DNA-modifying enzymes used the appropriate commercial buffers recommended by the manufacturer. Buffers used in cell lysis, protein purification, proteolysis and labelling are shown in Tables 2.4–2.7. All aqueous buffers and solutions were made with MilliQ water (Millipore). Protein purification buffers were filtered with the 0.45 µm filter, whereas NCL labelling reaction buffer, Factor Xa buffer, competition assay buffer and solutions used for bacterial or yeast cultures were sterile filtered using a 0.22 µm syringe filter. All buffers were made fresh, and cooled to 4 °C before use where necessary.

Table 2.4. Lysis and affinity purification buffers.

Yeast breaking buffer		20 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 1 mM EDTA, 5% glycerol, 1 mM DTT, 0.3 mM $(NH4)_2SO_4$, 1 mM PMSF and 1 tablet/50 mL buffer of Complete EDTA-free protease inhibitor cocktail
GSH-affinity purification buffers	**GST-Lysis buffer**	50 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM PMSF, 0.1 mg/mL lysozyme, 10 µg/mL DNase I, 10 mM $MgCl_2$
	GST-Wash buffer	50 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM DTT, 1 mM PMSF
	GST-Proteolysis buffers (HRV3C or TEV)	20 mM Tris-HCl pH 7.5–8.0, 150 mM NaCl, 1 mM DTT,
	GST-Elution buffer	20 mM Tris-HCl pH 7.5–8.0, 150 mM NaCl, 1 mM DTT and 10 mM freshly prepared reduced GSH
Amylose-affinity purification buffers	**MBP-Lysis buffer**	20 mM Tris-HCl pH 7.4, 0.2 M NaCl, 1 mM DTT, 1 mM PMSF, 1 mM EDTA, 0.1 mg/mL lysozyme, 10 µg/mL DNase I, 10 mM $MgCl_2$
	MBP-Wash buffer	20 mM Tris-HCl pH 7.4, 0.2 M NaCl, 1 mM DTT, 1 mM PMSF, 1 mM EDTA
	MBP-Elution buffer	MBP-Column buffer with 10 mM maltose
Ni-NTA affinity purification buffers	**Ni-NTA-Lysis buffer**	20 mM Tris-HCl pH 8.0, 0.4 M NaCl, 20 mM imidazole, 0.1% Triton X-100, 10 mM 2-mercaptoethanol, 1 mM PMSF
	Ni-NTA-Wash buffer	20 mM Tris-HCl pH 8.0, 1 M NaCl, 50 mM imidazole, 10% glycerol, 10 mM 2-mercaptoethanol, 1 mM PMSF
	Ni-NTA-Elution buffer	Ni-NTA-Wash buffer with 250 mM imidazole
	Ni-NTA-Proteolysis buffer	20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM 2-mercaptoethanol

Table 2.5. HPLC buffers.

Size exclusion chromatography buffer	20 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM DTT	
Anion-exchange chromatography buffers	10 mM Tris-HCl pH 7.4–8 or 10 mM sodium-phosphate pH 7.4, with 1 mM DTT	
	Buffer A: 0 M NaCl	Buffer B: with 1 M NaCl
Reverse phase chromatography buffers	Buffer A: 0.1% TFA	Buffer B: 0.1% TFA in acetonitrile

Table 2.6. Electrophoresis buffers.

50×TAE buffer	40 mM Tris-base, 5.7% glacial acetic acid, 10 mM EDTA (diluted 1-in-50 for use)
SDS-PAGE running buffer	1-in-20 dilution of Life Technologies NuPAGE MES SDS Running Buffer (20×). At 1× = 50 mM MES, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.3
10×Bis-Tris-Tricine buffer	200 mM Bis-Tris, 200 mM Tricine, pH 6.8 (diluted 1-in-10 for use)

Table 2.7. Specific buffers and solutions.

NCL labelling reaction buffer	0.1 sodium-phosphate pH 7.5, 0–150 mM NaCl, 0.5–10 mM TCEP
SNAP-tag labelling buffer	20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT
Western blot transfer buffer	SDS-PAGE Running Buffer, 20% methanol
TBST buffer	50 mM Tris-base (pH 7.4), 150 mM NaCl, 0.05% (v/v) Tween-20
4×LDS (SDS-PAGE loading buffer)	0.3125 M Tris base pH 6.8, 10% LDS, 50% glycerol, 25% DTT, 0.5% bromophenol blue in aqueous solution
Native gel loading buffer	3% (w/v) Ficoll in Competition assay buffer (from 15% stock) or 5–10% glycerol in Competition assay buffer (from 50% stock)
Coomassie blue stain solution	0.125% (w/v) Coomassie Brilliant Blue G-250, 40% (v/v) methanol, 7% (v/v) acetic acid
Gel fixing and destaining solution	30% (v/v) methanol, 10% (v/v) acetic acid
20×Na-borate buffer	47 g/L boric acid, 200mM NaOH, pH 8.7 (diluted 1-in-20 for use)
Factor Xa buffer	Competition assay buffer with 2 mM $CaCl_2$
10×TE buffer	100 mM Tris-HCl, 10 mM EDTA, pH 8.0
10×LiAc solution	1 M LiAc, pH 7.5
1×TE/LiAc buffer	1×LiAc solution, 1×TE buffer

PEG/LiAc buffer	40% PEG, 1×LiAc solution, 1×TE buffer
K-phosphate/sorbitol buffer	1.2 M sorbitol (diluted from a 2 M stock) 0.1 M potassium phosphate pH 7.5 (diluted from a 1 M stock)
Yeast fixation solution	4% paraformaldechyde (made in warm water, NaOH 1M added dropwise with stirring until dissolved), 3.4% sucrose
Competition assay buffer	20 mM sodium phosphate buffer pH 7.5, 150 mM NaCl, 1 mM DTT

Table 2.8. Buffers for preparing competent *E. coli* cells.

TFB1	100 mM RbCl, 50 mM $MnCl_2$, 10 mM potassium-acetate, 10 mM $CaCl_2$, 15% glycerol, pH 5.8, sterile-filtered (0.22 μm)
TFB2	10 mM MOPS, 10mM RbCl, 75 mM CaCl2, 15% glycerol, adjusted to pH 6.8 with KOH, sterile-filtered (0.22 μm)
Resuspension buffer (100 mL)	10% PEG, 5% DMSO, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 10% glycerol, LB broth, pH 6.1

Table 2.9. Solutions for preparation of MALDI-TOF samples.

Wetting Solution	0.1% TFA, 50% acetonitrile
Wash Solution	0.1% TFA
Matrix Solution	0.1% TFA, 70% acetonitrile, 10 mg/mL matrix (CHCA or sinapinic acid)

2.1.3 General media components

Table 2.10. General media components.

Component	Manufacturer
Bacto Peptone	Difco
Bacto Tryptone	Difco
Yeast extract	Difco, Affymetrix
Yeast nitrogen base	Difco
Amino acids and nucleobases	Sigma-Aldrich
Agar	Amyl Media

2.1.4 Resin and other material

Table 2.11. Resin

Resin	Manufacturer
Glutathione (GSH)-sepharose resin	GE Life Sciences
Ni-NTA resin	Invitrogen, Life Technologies, Qiagen
Amylose resin	New England Biolabs

Table 2.12. Other disposable and reusable materials.

Material	Manufacturer
Glass beads, 425–600 µm	Sigma-Aldrich
SnakeSkin dialysis tubing (various molecular weight cut-offs)	Life Technologies
PD-10 desalting columns	GE Life Sciences
Millex-GP Syringe Filter (0.22 µm)	Sigma-Aldrich
Nanosep MF centrifugal filters (0.45 µm)	Pall
ZipTips C18	Merck Millipore
Vivaspin sample concentrators (3 , 5 or 10 kDa cut-offs)	Sartorius Stedim Biotech
Concentrators	Thermo Scientific Pierce
Microscope slides and cover glass	Sail Brand
NativePAGE Novex Bis-Tris precast gels (3–12%)	Life Technologies
4–12% SDS-PAGE precast gels	Invitrogen
Biotrace^{TM} NT pure nitrocellulose transfer membrane	Pall Gelman Laboratory
Velveteen	Purchased at Lincraft
Petri dishes (90 and 140 mm)	Thermo Scientific
96-well plates	Life Technologies
Replica plating apparatus	Made by the School of Molecular Biology Workshop

2.1.5 Bacterial and yeast strains

Table 2.13. *E. coli* strains, genotypes and their purpose. The pLysS plasmid is in bold.

Bacterial strain	Genotype	Purpose	Source
DH5α	F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, *hsdR17*(r_K^- m_K^+), λ–	Plasmid cloning	Bethesda Research Laboratories
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*q*ZΔM15* Tn*10* (Tetr)]		Stratagene
BL21(DE3)	F⁻ *ompT gal dcm lon hsd*S_B(r_B^- m_B^-) λ(DE3 [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*])	Protein overexpression	Novagen
***BL21-AI™**	F⁻ *ompT gal dcm lon hsd*S_B(r_B^- m_B^-) *araB*::T7RNAP-*tetA*		Invitrogen
Rosetta™ 2 (DE3)pLysS	F⁻ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) **pLysSRARE2 (Cam^R)**		Novagen

*Tight regulation of the T7 RNA polymerase by the arabinose-inducible *araBAD* promoter and high overexpresed protein yields make BL21-AI cells optimal for expression of toxic proteins (BL21-AI™ One Shot® Chemically Competent *E. coli* manual, Invitrogen) [177]. These cells can be used in combination with any T7 promoter-based vector.

Table 2.14. *S. cerevisiae* yeast two-hybrid (Y2H) strains, genotypes and reporters. Reporters in bold were used in this thesis.

Y2H strain	Genotype	Reporters
AH109	*MAT**a**, trp1-901, leu2-3, leu2-112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3, GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2, URA3::MEL1$_{UAS}$-MEL1$_{TATA}$-lacZ, MEL1*	***HIS3*** ***ADE2*** *lacZ* ***MEL1***
Y187	*MAT**α**, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, leu2-112, met15Δ, gal4Δ, gal80Δ, URA3::GAL1$_{UAS}$-GAL1$_{TATA}$-lacZ, MEL1*	*lacZ* ***MEL1***

2.1.6 Bacterial and yeast culture media

Table 2.15. Media for bacterial cultures. LB broth and all solid media were autoclaved.

LB media (1L)	Bacto-Peptone/Tryptone 10 g, Yeast Extract 5 g, NaCl 5 g, agar (for plates only) 15 g, deionized water added to 1L, autoclaved
SOC media	Bacto-Tryptone 8 g, Yeast Extract 2 g, NaCl 0.2 g, KCl 4 mL of 250 mM stock, deionized water to 390 mL, adjusted to pH 7 and autoclaved. 2 mL of 2 M $MgCl_2$ and 8 mL of 1 M D-glucose was added from filtered stocks

Antibiotics as appropriate were added to LB broth after cooling to ~37 °C, and to agar-containing media after cooling to 40–50 °C, after which the agar-containing media was poured into plastic petri dishes and further cooled to solidify. LB_{Amp} broth and solid media contained 100 µg/mL ampicillin, while $LB_{Amp+Cam}$ broth and solid media contained 50 µg/mL ampicillin and 100 µg/mL chloramphenicol.

Table 2.16. Rich and minimal media for yeast growth. YPD broth and all solid media were autoclaved.

YPD Media	20 g/L bacto peptone, 10 g/L yeast extract, 2% (w/v) agar (for solid media)
Yeast Synthetic Drop-out (SD) Media	6.7 g/L yeast nitrogen base solution (autoclaved), 1×DO solution (see below for 10×DO solution), 2% D-glucose, 1× appropriate supplemental nutrients (Table 2.18), 1.5% (w/v) agar (for solid media)

Table 2.17. Components of 10×Dropout (DO) Solution.

Component	Concentration (mg/mL)
L-Isoleucine	300
L-Valine	1500
L-Adenine hemisulfate salt	200
L-Arginine HCl	200
L-Histidine HCl.monohydrate	200
L-Leucine	1000
L-Lysine HCl	300
L-Methionine	200
L-Phenylalanine	500
L-Threonine	2000
L-Tryptophan	200
L-Tyrosine	300
L-Uracil	200

Table 2.18. Supplement nutrient solutions added to SD media. All solutions were filter-sterilized.

Component	Concentration (mg/mL)
100× Uracil	2000
100× L-Histidine	2000
100× Adenine hemisulfate	2000

2.2 Constructs

2.2.1 LIM-HD full-length constructs

Unless otherwise indicated, all proteins are from *M. musculus*. Mouse proteins were chosen because of their high similarity to human proteins and because they allow testing in mouse models. The sequences used are those identified as isoforms current at the time of starting the project. In the interim several other isoforms have been added to the NCBI Consensus CDS (CCDS) protein set of LIM-HD proteins, of which all except Lhx3b, Lhx9.1 and Lhx9.2 were analysed in this thesis (see Appendix A). All four isoforms of Lhx6 were used in Y2H mating arrays. $Lhx4_{24–390}$ (hereafter referred to as Lhx4) was used to represent full-length protein in all assays. The Isl2 construct used in this thesis was based on the NCBI Reference Sequence NP_081673.1 and contains the mutations S157T, H185D, L267F and R282G. This version was succeeded by the NP_081673.2 NCBI Reference Sequence.

Full-length Lhx1, Lhx2, Lhx9.3 and Lmx1a were cloned by Vanessa Craig (Matthews laboratory, SMB, University of Sydney) from mouse brain mRNA. Isl1α and Lhx3a full-length constructs were obtained from Prof Ingold Bach (University of Massachusetts Medical School) as cloned cDNA and Lhx4, Lhx5, Lmx1b, Isl2, Lhx6.1 and Lhx6.2 were obtained from Codon Devices (Cambridge, MA 02139, USA) and Lhx8 was obtained from GenScript (Piscataway, NJ 08854,USA) as synthetic genes. Y2H vectors containing Lhx1, Lhx2, Lhx9.3 and Lmx1a constructs were obtained from Vanessa Craig and Lhx5 and Lhx6.2 were obtained from Dr Sally Eaton (Matthews laboratory, SMB, University of Sydney). Lhx6.3 and Lhx6.4 variants were generated by PCR from Lhx6.1 and Lhx6.2 templates, and were cloned along with Lhx4, Lhx6.1, Lhx8, Lmx1a and Isl2 into Y2H vectors as a part of this thesis. Note that in this thesis, the protein variants (Isl1α, Lhx3a, Lhx6.1, etc.) are specified only in case of full-length proteins.

2.2.2 Tandem and separate LIM domain constructs

Tandem and separate LIM constructs of Lhx3, Lhx4 and Isl1 in Y2H vectors and NpGBT9-$Lmx1a_{LIMs}$ plasmid were obtained from Amy Nancarrow, Dr Mugdha Bhati, Vanessa Craig, Dr Cy Jeffries, Dr Morgan Gadd and Dr Sally Eaton (all from within the Matthews

laboratory, SMB, University of Sydney). Tandem and separate constructs of Lhx1, Lhx2, Lhx5, Lhx6, Lhx8, Lhx9, Lmx1b, Isl2 and separate LIM constructs of Lmx1a were created and cloned into Y2H vectors as a part of this thesis.

Both tandem LIMs and separate LIM-domains contained 6 residues before the first Zn-ligating cysteine (N-terminal extension) and 6 residues after the last Zn-ligating residue (C-terminal extension), except for the N-terminal extensions of the following LIMs and LIM1 constructs: Lhx5, 4 residues; Lhx1, 3 residues (these represent the full extent of the native protein sequence); and Lmx1a, 2 residues. Two versions of $Lhx1_{LIMS}$ were used. An initial $Lhx1_{LIMs}$* construct did not contain any C-terminal extension and was replaced by a 3-residue extended construct part way through experiments as detailed in section 3.6.1.

2.2.3 LIDs and other constructs

$Ldb1_{LID}$ and $Isl1_{LBD}$ constructs in Y2H vectors were obtained from M. Bhati in the laboratory, and an extended LID domain (residues 331–375) construct was generated by PCR and cloned it into the pIH1121 plasmid for expression with an N-terminal Maltose binding protein (MBP)-tag (see section 6.10.1). Plasmids containing constructs encoding other ZnF proteins (LIMs of proteins LMO1–4, ZnF 1 and 9 of U-shaped, N-terminal ZnF of Pannier) were sourced from the laboratory. Human SSB1 in Y2H vectors was obtained from Dr Liza Cubbedu (Matthews laboratory, SMB, University of Sydney) for use as a non-related protein negative control in mating arrays (section 3.4). Full-length coding sequence of rat Pit1 protein (98% identical to the mouse protein) was cloned into Y2H vectors. $DEAF1_{404–438(T435D)}$, cloned into the pRSET vector was obtained from Dr Soumya Joseph (Matthews laboratory, SMB, University of Sydney).

2.2.4 Tethered LIMs-LID constructs with the Factor Xa site

Tethered $LMO4_{LIMs}$-$Ldb1_{LID}$ DNA construct [156] and $Lhx3_{LIMs}$-$Ldb1_{LID}$ DNA construct [18] were designed to code for the Factor Xa protease cleavage site in the linker between the LIMs and LID domains. These LIMs-LID constructs were originally cloned into pGEX-2T vectors (GE Healthcare) for expression with an N-terminal glutathione-S-transferase (GST)-tag

that could be removed by treatment with thrombin. The LIMs-LID DNA constructs containing LIMs-coding sequences of the remaining 11 LIM-HD proteins were cloned in pGEX-2T by Dr Mitchell O'Connell (Mackay laboratory, SMB, University of Sydney).

2.3 Cloning methods

2.3.1 PCR-based cloning and mutagenesis

PCR-based cloning was used to generate new constructs (wild type and mutant, fragments or fusion/tethered constructs), and introduce/remove restriction sites in plasmids.

2.3.1.1 Primer design

Primers were designed such that regions complementary to the template, or over overlapping regions of primers used for overlap extension PCR, had optimal melting temperatures (Tm, estimated as Tm = [2 × #(A+T) + 4 × #(G+C)]°C). Primers that formed the termini of PCR products intended for ligation were engineered with restriction sites (e.g., BamHI for the forward and EcoRI for the reverse primer). The reverse primer included 3 consecutive stop codons (transcribed as UGAUAAUAG) in front of the restriction site, except when the resulting PCR fragment was intended as an N-terminal part of a fusion or tethered construct. To increase the efficiency with which restriction enzymes digested PCR products, primers were designed with extra bases on the terminal sides of restriction sites as recommended by the manufacturer. See Appendix A for details of primers used in this thesis.

2.3.1.2 Standard polymerase chain reaction (PCR)

All PCR reactions were performed in an Eppendorf Mastercycler (Eppendorf AG, Hamburg, GER), Hybaid PCRsprint thermal cycler (Thermo Electron Corp., Waltham, MA, USA) or the T3000 Thermocycler (Biometra, GER). Reaction mixtures (50 or 100 µL) contained two primers (0.02 nM each), 2.5 mM each of dATP, dCTP, dGTP and dTTP, 8% (v/v) DMSO (Dimethyl sulfoxide), 1× volume of manufacturer-supplied or home-made *Pfu Turbo* PCR buffer, 10–500 ng template DNA and 1 U *Pfu Turbo* DNA polymerase. The reaction mixtures

were subjected to 25 cycles of the temperature sequence: 94 °C (1 min), [Tm – 2] °C (1 min), 72 °C (1 min), followed by a final elongation step at 72 °C (10 min). A 'hot-start' version of the method was used for most reactions, where a reaction without the enzyme was initially subjected to a long denaturation step (5 min at 95 °C), after which the enzyme was added and the standard PCR programme initiated. Completed reactions were stored at 4 °C or -20 °C until required.

2.3.1.3 Inverse PCR

The Expand Long Template PCR System Kit was used to perform inverse PCR. For this method, two primers with matching restriction sites at their 5' ends were designed to extend away from each other using a circular plasmid template. When the primers were designed to be complementary to sites on each side of the sequence that is intended to be removed, inverse PCR results in amplification of an entire plasmid except for the sequence between the primer-complementary sites. The linear product was digested with the primer-specific restriction enzyme(s) and the plasmid re-circularized by ligation. Reaction mixtures (50 µL) contained ~10 ng of plasmid DNA as template, primers at 300 nM, 5 µL of provided 10×Buffer 3 and 3.75 U (0.75 µL) of the supplied enzyme mix (Taq and Tgo DNA polymerases) [178]. The PCR program included a hot start step, followed by addition of the enzyme mix and 10 cycles of: 94 °C (30 s), 52 °C (30 s) and 68 °C (4 min). The programme continued with 20 cycles of: 94 °C (30 s), 52 °C (30 s) and an elongation step at 68 °C that increased in length from 4 min by 20 s with each successive cycle, with an additional final elongation step of 7 min.

The QuikChange Site-Directed Mutagenesis kit was used to perform a version of the inverse PCR where the forward and reverse primers had complementary (overlapping) 5' ends that contained the designed mutation, whereas the 3' primer ends were complementary to the surrounding sequence in the template on each side. Reactions contained 125 ng (125 nM) of each primer, 1 µL of dNTP mix, 5 µL of the 10× Pfu buffer, 1 µL of Pfu polymerase (2.5 U/ µL) and 10 ng of the plasmid template in 50 µL reaction volume. The product of PCR reaction had complementary sticky 5' ends introduced by primers and circularized but did not ligate until it was transformed into the XL1-Blue supercompetent *E. coli* strain and the nick in the plasmid has been repaired. DpnI digestion after PCR was used to degrade the original methylated template plasmid leaving the newly generated non-methylated template intact.

2.3.1.4 Overlap extension PCR

Overlap extension PCR was used to introduce mutations or remove restriction sites from DNA sequences. Pairs of precisely overlapping oligonucleotides were designed that contained mutations flanked by 12 bp of wild type sequence on each side. Each oligonucleotide from a pair was used in a separate reaction with a primer corresponding to the 5' or 3' ends of the full-length construct (end primers). The overlapping fragments were gel purified and aliquots (5–10 µL) were combined and used as templates in a second round PCR reaction which contained end primers. In cases where multiple distant mutations were needed, the product of this reaction was gel-purified and used as a template in following rounds of PCR reactions with different pairs of overlapping primers.

2.3.1.5 Purification of PCR products

DNA from inverse PCR reactions was purified by ethanol precipitation. Sodium-acetate (5 µL of a 3 M stock solution) was added to a 50 µL PCR reaction, followed by 100 µL chilled ethanol. The tube was incubated on ice for ~30 min followed by centrifugation at 12000 × g for 20–30 min at 4 °C. The supernatant was carefully removed and the pelleted DNA dried in a SpeedVac concentrator. DNA from other PCR reactions was either purified by the QIAquick PCR Purification Kit or by agarose gel electrophoresis followed by the use of the QIAquick Gel Extraction Kit. In all cases the purified DNA was resuspended or eluted in ~30 µL of MilliQ water.

2.3.1.6 Restriction endonuclease digestion of PCR products

PCR products were digested by restriction endonucleases to prepare them for directional cloning into plasmids. Purified PCR product (section 2.3.1.6), 4 µL of the recommended 10× restriction enzyme buffer, 4 µL of 10×BSA, and 1 µL of each enzyme were made up to a total volume of 40 µL with MilliQ water, and the reaction was incubated 37 °C for 2–4 h.

2.3.1.7 Agarose gel electrophoresis of digested DNA

DNA samples from restriction digestion or PCR reactions were mixed with 6× loading dye (Fermentas Life Sciences, Ontario, Canada) and loaded into 1–2% (w/v) agarose gels

containing ethidium bromide (1 μg/mL) that were prepared in 1×TAE buffer, and subjected to electrophoresis at 80–100 V for 45–60 min. Gels were visualised under long-wave UV light and target bands were excised from the gel, stored at -20 °C until required, and purified using the QIAquick gel extraction kit according to the manufacturer's instructions (Qiagen, Valencia, CA, USA).

2.3.2 Plasmid-based cloning

2.3.2.1 Restriction endonuclease digestion of plasmid DNA

Plasmid DNA was treated with the same restriction enzymes as used for the PCR products to prepare a target (empty) plasmid for ligation with a new DNA fragment, or to excise a previously cloned DNA fragment from a plasmid, as described in section 2.3.1.6 but the final volumes were 30 μL (with 3-μL aliquots of 10× solutions) with 5–10 μL of miniprep purified plasmid (see section 2.3.2.6).

2.3.2.2 Ligation reactions

All ligation reactions were performed in a total volume of 10 μL. Reactions contained 1 μL of commercially supplied 10× T4 DNA ligase buffer (400 mM Tris, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP) and 1 Weiss unit of T4 DNA ligase (1 μL). Restriction enzyme digested plasmid DNA was combined with an excess of similarly digested insert DNA and incubated at room temperature for 4 h or in icy water overnight. The reactions were usually heated at 65 ºC for 20 min to deactivate the ligase. The total reaction mixture was transformed into DH5α *E. coli* cells.

2.3.2.3 Plasmid amplification

Most plasmids were amplified by propagation in *E. coli* strain *DH5α* (Table 2.13). Plasmid products of QuickChange PCR reactions were transformed into the XL1-Blue supercompetent cells (QuikChange® II Site-Directed Mutagenesis Kit). Bacteria were plated on a solid LB_{Amp} media and cultured in Luria-Bertani (LB) broth.

2.3.2.4 Preparation of competent E. coli

$MgCl_2$ method. Antibiotic-free SOC media (400 mL) was inoculated with a 10 mL overnight SOC culture of a strain, then grown at 37 °C until OD_{600} = 0.4. The culture was cooled for 20 min by placing the tube in an ice/water slurry and the cells were subsequently centrifuged at 1000 × g for 15 min in cold 50 mL tubes. The pellet was resuspended in 20 mL autoclaved Resuspension buffer. The resuspended cells were divided into small (50–100 µL) aliquots, snap frozen in liquid N_2 and stored at -80 °C. DH5α, BL21(DE3) and Rosetta 2(DE3)pLysS strains were made competent for transformation by this method.

RbCl method (BL21-AI cells) [179]**.** LB broth (50 mL) was inoculated with a single colony of BL21-AI cells and was grown, with shaking at 37 °C to an OD_{600} of 0.3–0.5. The culture was cooled on ice for 5 min, the cells centrifuged (4000 × g) at 4 °C, and the supernatant was discarded. Cells were gently resuspended in ice-cold TFB1 buffer (15 mL) and incubated on ice for 90 min. The cells were collected by centrifugation (5 min, 4000 × g, 4 °C) and the cell pellet carefully resuspended in 2 mL ice-cold TFB2 buffer, aliquoted (50 µL) into sterile microcentrifuge tubes, frozen on dry ice, and stored at -80 °C.

2.3.2.5 Transformation of E. coli

Plasmid DNA (1 µL of prepared plasmid DNA) or 10 µL ligation mixture was added to 30 and 100 µL of competent *E. coli* cells, respectively, and incubated on ice for ~15 min. Cells were heat-shocked at 42 ºC for 5 min and 200 µL of sterile LB broth added. Cells were allowed to propagate for a further 1.5 h at 37 ºC before being spread onto LB_{Amp} plates. Transformed Rosetta 2 cells were spread onto $LB_{Amp+Cam}$ platesand incubated overnight at 37 ºC.

2.3.2.6 Amplification and preparation of plasmids

A single colony of DH5α or XL1-Blue *E. coli* cells transformed with plasmid DNA (section 2.3.2.5) was used to inoculate 5 mL of LB_{Amp} broth and incubated overnight at 37 ºC. The cells were then collected by centrifugation (5000 × g, 4 ºC, 10 min) and plasmid DNA prepared using a QIAprep Miniprep Kit according to the manufacturer's instructions. Final solutions of plasmid DNA were eluted in 50 µL elution buffer and stored at –20 ºC.

2.3.3 Verification by DNA sequencing

All plasmid inserts generated by PCR were sequenced to confirm their identities at SUPAMAC, University of Sydney, NSW or Australian Genome Research Facility (AGRF). Samples were prepared according to SUPAMAC or AGRF guidelines.

2.4 Plasmids

2.4.1 Plasmids for bacterial protein over-expression

See Appendix B for details of design strategies for bacterial overexpression plasmids in this thesis.

2.4.1.1 Plasmids for generating MBP-tagged constructs

Plasmids that code for N-terminally-tagged MBP constructs were derived from the pIH1119 plasmid, which was in turn derived from the pMAL-c2x plasmid (New England Biolabs) such that the EcoRI and BamHI restriction sites were swapped (MBP-linker-BamHI-EcoRI in pIH1119). The pMAL-stop plasmid contains a stop codon after the MBP sequence and it was used to produce MBP-only protein controls for competition assays (Chapters 6 and 7).

The pIH1121 plasmid was created by replacing the Factor Xa protease site in pIH1119 with an HRV3C protease site. This was achieved by digesting the pIH1119 plasmid with AvaI and BamHI restriction enzymes to excise the Factor Xa protease site, and ligating annealed and phosphorylated complementary oligonucleotides that encoded the HRV3C protease site into the AvaI and BamHI sites in. An extended variant of $Ldb1_{LID}$ spanning residues 331–375 was cloned into the pIH1121 plasmid and was used for expression of MBP-$Ldb1_{LID}$.

2.4.1.2 Plasmids for GFP-tagging

The GFPsol full-length sequence was created by joining the NGFPsol and CGFP split-GFP fragments by overlap extension PCR. Note that the C-terminal fragment is the same in GFPuv and GFPsol and is therefore referred to simply as CGFP. NGFPsol, the N-terminal

fragment of the GFPsol was created by introducing F64L and S65T mutations in the NGFPuv fragment from a BiFC plasmid template (see section 2.4.3) by QuickChange PCR. Plasmid pET15b, which allows proteins to be expressed with an N-terminal His tag, was modified to remove an EcoRI restriction site positioned between HindIII and AatII sites, and an out-of frame BamHI site that lay close to an NdeI site. Standard PCR was performed in which the new in-frame BamHI and EcoRI sites were incorporated into the sequence of the forward primer, which was complementary to the template BamHI site with additional downstream sequence. The reverse primer was complementary to sequence just upstream of the template EcoRI site and contained an AatII site on the 5' end. The resulting plasmid was named pET15bBE. A BglII-GFPsol-BamHI-TAATGA-EcoRI construct was created using PCR that contained two stop codons in frame before the EcoRI site to allow expression of GFPsol-only constructs. This construct was digested with BglII and EcoR1 and inserted between the BamHI and EcoRI sites of a pHisGB1 plasmid that contained an HRV3C site in place of the original thrombin site (created by Dr. Soumya Joseph, Matthews laboratory, SMB, University of Sydney). The construct was similarly inserted into pET15bBE and pGEX-6P. The resulting plasmids have a hybrid BamHI/BglII that is not cuttable by either enzyme before the GFPsol sequence and only the downstream BamHI-EcoRI site is available for further cloning. Tethered constructs of $LMO4_{LIMs}$, $Lhx3_{LIMs}$ and $Isl1_{LIMs}$ with $Ldb1_{LID}$ (with Factor Xa protease sites in the linkers) were subcloned into these vectors.

2.4.1.3 SNAP-tag plasmid with a cleavable His-tag

The pSNAP-tag(T7) Vector (New England Biolabs) was modified by standard PCR to include an N-terminal His-tag for Ni-NTA purification and a TEV protease site in the linker for removal of the tag. The $LMO4_{LIMs}$-$Ldb1_{LID}$ construct, which was initially cloned into the pSNAP-tag(T7) vector between the BamHI and EcoRI sites, was amplified by PCR for which the forward primer contained a coding sequence for the TEV site and the NdeI restriction enzyme site on its 5' end. The PCR product and the pSNAP-tag(T7) were digested with NdeI and EcoRI, and the two ligated. The resulting plasmid was named pT7-His-SNAP-$LMO4_{LIMs}$-$Ldb1_{LID}$.

2.4.1.4 Plasmids that allow production of proteins with an N-terminal cysteine

The pGEX 2TKE plasmid (a modified version of pGEX 2T from GE Healthcare in which a pseudo thrombin site in the GST moiety was removed by mutation) was edited by overlap extension PCR to generate pGEX-TEVC in which the thrombin protease site was replaced by a modified TEV protease site (ENLYFQ|**C**) where "|" indicates the proteolysed bond. Removal of the GST tag with TEV protease leaves an N-terminal cysteine residue. Tethered constructs of $LMO4_{LIMs}$, $Lhx1_{LIMs}$, $Lhx2_{LIMs}$, $Lhx3_{LIMs}$, $Lhx4_{LIMs}$ or $Isl1_{LIMs}$ with $Ldb1_{LID}$ (with Factor Xa protease sites in the linker) were subcloned into this plasmid. See Appendix C for details of protein sequences of tethered LIMs-LID constructs with Factor Xa-cleavable linkers used in this thesis.

2.4.2 Plasmids used in yeast two-hybrid assays

Constructs for yeast two-hybrid experiments were expressed as fusion constructs with the GAL4 DNA-binding domain (DBD) and GAL4 activation domain (AD). These fusion proteins were generated using NpGBT9 and pGAD10 vectors, respectively. NpGBT9 is a modified version of pGBT9 (Clontech, CA, USA) in which the BamHI and EcoRI sites within the MCS were reversed.

2.4.3 Plasmids used in BiFC assays in yeast

See Appendix D for DNA and protein sequences of BiFC plasmids used in this thesis and Appendix E for details of ADH1 promoters used in these plasmids.

pW397 and pL397 plasmids were created by inverse PCR from the NpGBT9 and pGAD10 Y2H plasmids. The Gal4 domains and related NLS sequences were replaced with a NdeI restriction site. NGFPuv fragments with $(GGGGS)_2$ linkers on either the N- or C- termini, generated by PCR, were cloned into pW397 (see Appendices B and C), allowing N- or C-terminal tagging of inserted proteins (pW397-$NGFP_{\mathbf{N}}$ and pW397-$NGFP_{\mathbf{C}}$ plasmids). The $NGFP_N$ fragment was cloned between the NdeI and BamHI sites, whereas the $NGFP_C$ contained the stop codons and was cloned between the EcoRI and PstI sites. CGFP fragments with

$(GGGGS)_2$ linkers on N- or C- termini, generated by PCR, were cloned into the pL397, allowing N-or C-terminal tagging of inserted proteins (pL397-$CGFP_{\mathbf{N}}$ and pL397-$CGFP_{\mathbf{C}}$ plasmids). The $CGFP_N$ fragment was cloned between the NdeI and BamHI sites, whereas the $CGFP_C$ contained the stop codons and was cloned between the EcoRI and BglII sites. $LMO4_{LIMs}$ and $Ldb1_{LID}$ constructs with or without the stop codons were subcloned into the appropriate plasmids, producing 8 possible combinations of fusion construct pairs. Other LIM-domains and ZnFs were subcloned only into the pW397-$NGFP_{\mathbf{N}}$ plasmid, and their binding partners were subcloned into pL397-$CGFP_{\mathbf{C}}$. Full-length LIM-HD proteins Lhx3a and Isl1α were subcloned into pW397-$NGFP_{\mathbf{N}}$ and pL397-$CGFP_{\mathbf{C}}$, respectively.

pW397-$NGFP_{\mathbf{N/C}}$ and pL397-$CGFP_{\mathbf{N/C}}$ plasmids were modified by replacing the short ADH1 (yeast alcohol dehydrogenase 1 gene) promoter (397 bp) with the full-length ADH1 promoter (1472 bp), creating pWAFL-$NGFP_{\mathbf{N/C}}$ plasmids and pLAFL-$CGFP_{\mathbf{N/C}}$ plasmids. Full-length Lhx3a and Isl1α constructs were subcloned into these plasmids. pW397-$NGFP_{\mathbf{N}}$ and pL397-$CGFP_{\mathbf{N}}$ plasmids were modified to make pW397-NLS-$NGFP_{\mathbf{N}}$ and pL397-NLS-$CGFP_{\mathbf{N}}$ plasmids in which an SV40 NLS sequence, created by PCR, was inserted into the NdeI site that codes for the N-terminus of each related GFP-fragment. $LMO4_{LIMs}$ and $Ldb1_{LID}$ constructs were cloned in these plasmids, respectively.

PCR mutagenesis (QuickChange PCR) was used to modify the NGFPuv fragment in the pW397-$NGFPsol_{\mathbf{N}}$ and pW397-$NGFPsol_{\mathbf{C}}$ plasmids through introduction of F64L and S65T mutations in the NGFPuv fragments. Full-length Lhx3a and Isl1α constructs were subcloned into both of these plasmids.

Finally, pW397-His-GFPsol and pL397-His-GFPsol plasmids were created by inserting the full-length GFPsol construct, created by overlap extension PCR (see section 2.3.1.4), between the NdeI and BamHI sites in the plasmids, with the 5' primer containing the His-tag.

2.5 Yeast two-hybrid (Y2H) assays

2.5.1 Preparation of yeast competent cells

Yeast was cultured overnight at 30 °C with shaking (200 rpm) in 50 mL YPD media and then added to 300 mL YPD and grown for 4–5 hours under the same conditions. Cells were

collected in 7×50 mL falcon tubes by centrifugation at 1000 × g for 5 min. The cells were washed in 25 mL sterile water, centrifuged, collected and resuspended in 3–4 mL of fresh sterile 1×TE/LiAc buffer. Competent yeast cells were stored at 4 °C and used within 4 days.

2.5.2 Transformation of yeast competent cells

Plasmid DNA (~0.5 μg) and salmon sperm DNA (~100 μg) were combined with 0.1 mL of competent yeast, 0.6 mL of PEG/LiAc (40% PEG-4000, 100 mM LiAc, 10 mM Tris (pH 8.0), 1 mM EDTA) and then mixed by vortexing. The cells were incubated at 30 ºC with shaking (200 rpm) for 30 min. DMSO (70 μL) was added and the mixture inverted 3–4 times. The cells were heat shocked at 42 ºC for 15 min with gentle mixing every 5 min. The cells were chilled on ice for 2 min and pelleted by centrifugation (15 000 × g) for 5 s. The supernatant was removed and the cells were resuspended in 300 μL of sterile 1×TE buffer (10 mM Tris (pH 8.0), 1 mM EDTA), spread onto appropriate selective SD-agar plates and incubated at 30 ºC for 3–4 days. In all further protocols yeast were incubated at 30 °C, and liquid cultures were grown with shaking at 200 rpm, except where otherwise specified.

2.5.3 Spot-tests with haploid yeast

Experiments were carried out using inserts cloned into pGAD10 and NpGBT9 plasmids, which were co-transformed into AH109 cells (Clontech). Successfully co-transformed cells were selected by growth on media lacking leucine and tryptophan (SD-L-W). Yeast cultures were diluted to an absorbance of 0.2 at 600 nm in a 1 cm pathlength (OD_{600}), and two serial 1-in-10 dilutions were made for each sample. Yeast suspensions (2 μL) were spotted in arrays on different solid selection media (Table 3.1) and reporter activation was followed and analysed over the next several days.

2.5.4 Y2H mating arrays

2.5.4.1 Replica plating

Transfer of cells from one plate to another in mating arrays was done by replica plating (Figure 2.1). For this purpose, a metal stamp was constructed by the SMB workshop that fitted

140 mm petri dishes (plates). This stamp was covered with pieces of sterilized velveteen. The velveteen was carefully tightened over the stamp and fixed with rubber bands to prevent wrinkles in the fabric. Plates were dried prior to transfer and were gently pressed onto a stamp.

2.5.4.2 Cross-mating and spot-mating Y2H arrays on solid media

In yeast mating experiments AH109 cells were transformed with NpGBT9 constructs and Y187 cells were transformed with pGAD10 constructs. After a 72 h incubation colonies were picked from transformant plates and grown overnight in 3 mL of appropriate liquid SD media. Cultures of AH109 and Y187 cells at $OD_{600} \approx 1.0$ were either:

a) centrifuged for 5 min at 1500 × g, the supernatant removed and pellets streaked in parallel on a SD-W or a SD-L plate, with each streak representing a single transformant (Figure 2.1, left). These plates were incubated for ~24 h, and then replica plated at right angles to each other on a mating (YPD) plate. Mating occurred at each crossover point after overnight incubation.

b) spotted in small aliquots (0.5 µL) on top of each other in an array on a mating plate and incubated for ~16 h. Spots containing mated diploid cells were replica plated onto different interaction selection media (Table 3.1) and incubated for ~48 h (Figure 2.1, middle). In a short protocol, yeast was replica plated directly from the mating plate onto interaction selection plates.

2.5.4.3 The 96-well mating arrays

The 96-well Y2H mating array was carried out manually (Figure 2.1, right). Solid media was pipetted into each well (200−300 µL/well using 1 mL tips for easy transfer) while the media was still warm. SD-selective media broth was used to grow transformants overnight in a separate plate. A set of 8×12 pipette tips in a tip storage holder was dipped in wells containing one strain, and then dipped in wells containing YPD broth. The same procedure was repeated with fresh tips to inoculate transformants of the opposite mating type into the same plate for mating. After overnight cultivation, fresh tips were used to dip and replica plate the mated yeast on to the surface of the solid selection media in the 96-well plate. Colony growth and colour was followed for the next several days.

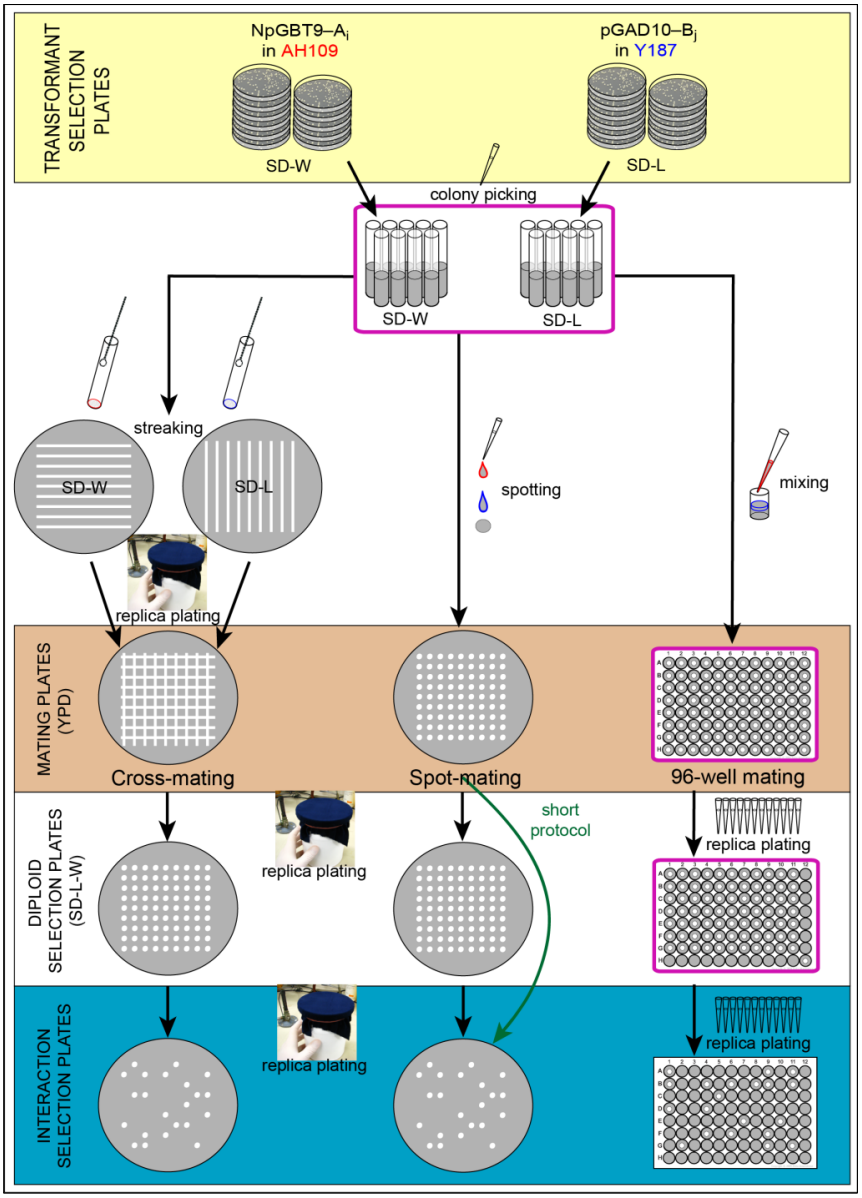


Figure 2.1. Flowchart of Y2H mating array protocols. Yeast is depicted as white spots or streaks on grey coloured media whereas arrows show transfers of yeast between media. Tubes and plates with liquid media are marked by purple frames. Colonies of AH109 strain (red) and Y187 strain (blue) transformants carrying a NpGBT9-A_i plasmid (i=1→n) or a pGAD10-B_j plasmid (j=1→m), respectively, were picked with sterile pipette tips and appropriate liquid media were inoculated. In the cross-mating protocol (left), cell pellets of overnight cultures were streaked with the inoculation loop as parallel streaks on appropriate selection plates. Haploid yeast streaks were replica plated orthogonally to each other onto a mating plate, via a velveteen stamp. In the spot-mating protocol (middle), haploid yeast cultures were spotted on top of each other in an array of spots on a mating plate. Both cross-mating and spot-mating protocols employed velveteen stamp replica plating in all subsequent steps. An alternative short protocol applied replica plating from the spot-mating plate directly onto interaction selection plates. In the 96-well mating protocol (right), mating and diploid selection steps were performed in liquid media, whereas arrays of sterile pipette tips were used for replica plating. Note that interaction selection plates included SD-L-W-H plates with X-α-gal and different amounts of 3-AT or SD-L-W-H-A plates.

2.6 Bimolecular fluorescence complementation (BiFC) in yeast

2.6.1 Yeast strain and culture media

Yeast cells were co-transformed with pairs of complementary BiFC plasmids as described above (section 2.5.2). Transformed yeast was cultured on solid media as described above (section 2.5.3). Adenine hemi-sulfate salt (60 µg/mL) was added to the plates and liquid media to prevent auto-fluorescence of AH109 cells. Transformed cells were grown at 30 °C for 3 days on SD-L-W plates. Yeast transformant colonies were used to inoculate liquid SD-L-W media, grown overnight at 30 °C with shaking, and used to inoculate fresh media. Protein expression was performed in secondary 10 mL overnight cultures until the OD_{600} = 1.0 (~ 20 h) [180].

2.6.2 Yeast fixation

Yeast cultures were centrifuged (1000 × g, 5 min at room temperature), supernatant was removed and cell pellets were resuspended in 1 mL pH 7.5 K-phosphate/sorbitol buffer (Table 2.7). The resuspended cells were transferred to microtubes, centrifuged and pellets resuspended in 100 µL of yeast fixation solution. The cells were vortexed and incubated at room temperature for 15 min. This was followed by a wash with 1 mL K-phosphate/sorbitol buffer.

2.6.3 Epifluorescence microscopy

Detection and localization of BiFC fluorescence was carried out using epifluorescent microscopy on an Olympus BX51 System fluorescent microscope with a mercury UV-lamp, FITC and DAPI filter sets and a 100× objective was used for visualization. Yeast cells were either fixed (in trial experiments only) or were only washed with 1 mL K-phosphate/sorbitol buffer, without fixing. Cells were mixed in a 1:1 ratio with Mobial mountant on the microscopic slide. The best results were obtained when a dense 0.5–1 µL cell suspension was gently smeared by the cover glass and flattened into an approximately single layer of cells. Areas with tightly packed yeast cells in a single layer were analysed to standardize the cell density per slide. All

micrographs were created using the same magnification and scale bars were not added to micrographs.

2.6.3.1 Image acquisition

AnalySIS LS Starter software was used for image acquisition. Camera settings were controlled in the Camera Control window and the starting exposure was 100–200 ms. For some samples exposure times were increased to 500 ms or 1 s. The Intelligent Control feature was used to set the optimal signal intensity range. Micrographs were created on two (FITC and bright field) or three channels (FITC, bright field and DAPI) and composite images were generated by the software.

2.6.4 Yeast lysis

Yeast BiFC co-transformants were grown on selective media overnight. After confirming the BiFC fluorescence signal in positive controls using the epifluorescent microscope (section 2.6.3), yeast were pelleted by centrifugation at 1000 × g for 5 min. The resulting pellets (1 volume) were washed with 1 volume of water and dissolved in 3 volumes of Yeast breaking buffer on ice. Glass beads (4 volumes) were added and the samples were vortexed vigorously (5 × 1 min). The resulting extracts were either analysed directly using SDS-PAGE (section 2.8.1) or the cells were centrifuged in a microcentrifuge at 12000 × g for 2 minutes to separate the soluble and insoluble fractions.

2.7 Overexpression and purification of recombinant proteins

2.7.1 Protein overexpression in *E. coli*

All proteins, except SNAP-tagged constructs, were overexpressed in BL21(DE3) cells; but some initial trials with GFP-labelled proteins were carried out in Rosetta 2 cells. Glucose (1%) was added all cultures for MBP-tagged constructs to repress the expression of bacterial amylases that tend to degrade amylose-based purification resins. The overexpression of His-SNAP-LMO4$_{LIMs}$-Ldb1$_{LID}$ in BL21-AI cells was induced with 0.001% arabinose. Small scale overexpression experiments (expression trials) were performed in order to find the optimal

conditions for overexpression, stability and solubility of each protein or class of proteins. These trials usually involved 10–15 mL bacterial cultures, in 50 mL conical centrifuge tubes. An overnight culture was started from a single colony in 10 mL LB broth with the appropriate antibiotic(s) and grown at 37 °C with shaking (180 rpm) for 16–18 h. The culture was diluted in fresh media to OD_{600} = 0.1 and grown to OD_{600} = 0.4, when the temperature was lowered to 18–25 °C (usually 20 °C). Overexpression of BL21(DE3) cells was usually induced at OD_{600} of 0.5–0.6 with 0.1 mM for GST-GFP-tagged proteins, 0.3 mM for MBP-tagged proteins, or 0.5 mM IPTG for other proteins. Overexpression continued at lower temperatures for 16–18 h (overnight) in order to minimize proteolytic degradation that is common in constructs expressed from pMal-type vectors, and decrease aggregation/insolubility in other cases. Large scale overexpression experiments were usually carried out in 1L cultures, following scaled-up trial expression procedures.

2.7.2 Bacterial cell lysis

Bacterial cells were harvested by centrifugation at 5500 × g (12–15 minutes at 4 ºC) and thoroughly resuspended in appropriate lysis buffers. In large scale experiments lysis buffers were used at 4–40 mL/g cell pellet (usually 5 mL/g for GST- and His-tagged proteins, and 20–40 mL/g for MBP-tagged proteins), and in small scale experiments at ~1 mL/10–15 mL culture. The suspension was then placed at -20 °C (to promote partial cell lysis through slow freezing) and lysed, either immediately or after storage at -20 °C, by 2–3 freeze-thaw cycles using liquid N_2 for freezing and thawing in a water bath. Lysozyme (0.1 mg/mL) and Dnase I (10 µg/mL) were either present in the lysis buffer from the start or were added after thawing. Samples were sonicated on ice (3 series of 10 × 1 s bursts, with 1 min breaks between series) using a macrotip probe for large volumes or step tip probe for small volumes, at 70% the maximum sonicator output (Sonifier® S-250A, Branson ultrasonics, CT, USA). The lysate was then incubated for >30 min at 4 °C with mixing, and the insoluble fraction removed by centrifugation at 20000 × g for 30 min at 4 °C.

2.7.3 Protein purification by affinity chromatography

Small scale purification was performed using microtubes and large scale purification was performed in glass or plastic columns fitted with glass wool filters at the bottom of the column. All resins were first washed with 5 column volumes (CV, which corresponds to the volume of the resin in the column) of MilliQ water. After equilibration with the appropriate lysis buffer, the resin was mixed with the soluble fraction of the lysate and incubated in a sealed column/flask/tube for 1 h on a rotating wheel at 4 °C. The resin was then run through the column, at ~1 mL/min and the non-bound components (flow-through) collected. The same flow rate was used in subsequent washing and elution steps. Fractions containing purified protein were combined, concentrated, snap-frozen in liquid N_2 and stored at -80 °C.

2.7.3.1 GSH-affinity chromatography

GSH resin (3–10 mL/L culture or 2 mL/g cell pellet, 50 µL for trials) was equilibrated by running 10 CV of GST-Lysis buffer through the column. After binding, the GSH resin was washed with 20 CV of GST-Wash buffer, followed by either elution from the column by 3–5 × CV of GST-Elution buffer (some GST-tagged proteins in trial experiments), or washing with 10 CV of an appropriate GST-Proteolysis buffer prior to proteolysis step (most preparations). GSH in the GST-Elution buffer competes with the resin for binding the GST-tag, thereby specifically releasing a GST-tagged construct from the resin.

2.7.3.2 Native Ni-affinity chromatography

All His-tagged constructs of TEV protease and SNAP-LMO4$_{LIMs}$-Ldb1$_{LID}$ were initially purified by Ni-NTA affinity chromatography under native conditions. Resin was used at the ratio of 2 mL resin per 1 L of culture. Ni-NTA-Lysis buffer included 20 mM imidazole to minimise non-specific binding of different bacterial proteins to the Ni-NTA resin, as the imidazole competes with His for Ni-binding. Relatively higher concentrations of NaCl (0.4–1 M in the Ni-NTA buffers compared with 300 mM in GST-purification buffers) served a similar purpose. The purification was carried out as described for GSH-affinity purification, except that 4 × 2 CV Ni-NTA-Wash buffer was used for washing and 4 × 2 CV of Ni-NTA-Elution buffer was used for

elution steps. Size exclusion chromatography (SEC) was used to remove imidazole and finely purify the protein.

2.7.3.3 Amylose-affinity chromatography

MBP-only and MBP-Ldb1$_{LID}$ constructs were initially purified by amylose-affinity chromatography. Amylose resin (~15 mL per 1 L culture) was equilibrated with 8 CV of MBP-Lysis buffer. After binding, the resin was washed with 12 CV of the MBP-Wash buffer and the protein was eluted with 3 × CV of the MBP-Elution buffer. Maltose from the MBP-elution buffer competes with the amylose resin for MBP-binding, thereby specifically releasing a MBP-tagged construct from the resin. Elution fractions containing the protein were concentrated, exchanged into a 20 mM Tris-Cl pH 8.0 buffer (no NaCl) with PD-10 columns, and further purified using anion-exchange chromatography. No additional purification steps were required for MBP-only.

2.7.4 Protein cleavage by proteases

Affinity tags were removed by proteolysis. In most cases, this step was carried out when the protein was still immobilised on the affinity resin. The resin was mixed with ~2 CV of a proteolysis buffer and the appropriate amount of enzyme in the purification column, which was sealed and incubated as indicated on a rotating wheel. After treatment with the protease as described below, the freshly cleaved samples were collected and more cleaved protein was further washed off from the resin with proteolysis buffer (~2 CV). For some experiments proteolysis was carried out in solution, after the protein was eluted from the affinity resin. Fractions containing cleaved or eluted proteins were combined, concentrated and purified further or snap-frozen in liquid N_2 and stored at -80 °C.

Proteolysis of GST-GFP-tagged constructs was conducted with the HRV3C protease. HRV3C protease (50 µL of in-house product) was used per 1 mL resin for 4–5 h for trials or overnight for large scale experiments at 4 °C.

Constructs expressed from pGEX-TEVC plasmids were treated with ~50 µg of TEV protease per 1 mL of resin, and incubated overnight, at room temperature.

2.7.5 High Performance Liquid Chromatography (HPLC)

Most proteins were further purified using size-exclusion chromatography (SEC) and/or ion exchange chromatography. All samples were passed through 0.22 μm filters prior to loading. The major fractions containing purified proteins were combined and stored at 4 ºC for immediate or short term use, or snap-frozen and kept at - 80 ºC for long-term storage.

2.7.5.1 Size-exclusion chromatography (SEC)

HiLoad Superdex 75 16/60 or HiLoad Superdex 200 16/60 size-exclusion columns (GE Healthcare) were used for preparative SEC of concentrated, affinity-purified samples (1–5 mL/injection), at a flow rate of 1 mL/min on a BioLogic DuoFlow (Bio-Rad Laboratories Inc., CA, USA). Elution of target proteins was monitored by absorbance at 280 nm.

2.7.5.2 Anion-exchange chromatography

A 6 ml, prepacked monolith ion exchange column UnoQ6 (BioRad) was used for purification of MBP-Ldb1$_{LID}$, fluorescein-labelled and GFP-tagged tethered constructs. Protein samples were desalted into Anion-exchange chromatography buffer A and loaded onto the column that was equilibrated in the same buffer. The proteins were eluted using different buffer B gradients of 0–100% over 20 min, with a typical flow rate of 1 mL/min in the gradient and 2 mL/min in other steps of the program. Elution of target proteins was monitored by absorbance at 280 nm or 215 nm.

2.7.5.3 Reverse-phase (RP) chromatography

Reverse phase high performance liquid chromatography (RP-HPLC) was used by Dr Brendan Wilkinson (from the research group of Prof. Richard Payne, School of Chemistry, University of Sydney) for purification of the fluorescein-2MES. The reagent was purified by applying a 0–100% methanol gradient with 0.1% (v/v) trifluoroacetic acid (TFA) on a Waters Symmetry C4 column with 0.2 mL/min flow rate. Fluorescein 2-MES eluted at ~40% methanol.

2.7.6 Buffer exchange, desalting and removal of low MW components

2.7.6.1 Protein dialysis

Samples were loaded into dialysis buttons (100 µL to 5 mL), microtube lids (≤100 µL) or dialysis bags (larger volumes). Dialysis membranes (SnakeSkin® dialysis tubing 3.5 or 10 kDa molecular weight cut off) pre-wetted in new buffer were used to cover the samples in buttons or lids and make dialysis bags. Samples were sealed inside the buttons with O-rings, in lids with the tubes with the bottom cut-out, and the bags were sealed with specialized dialysis clips (Spectra/Por Closures). These apparatuses were placed in buffer using buffer-to-sample volume ratios of ≥250:1 for >3 h with gentle stirring at the desired temperature. The buffer was changed 2–3 times with the final dialysis step generally being carried out overnight at 4 °C.

2.7.6.2 Gravity flow PD-10 columns

Gravity flow PD-10 columns (GE-Healthcare) were used to rapidly exchange buffers and remove unbound reagents in NCL labelling reactions (section 2.10.2.4) according to the manufacturer's instructions.

2.8 Analytical methods for protein detection and quantification

2.8.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were mixed with 4×LDS loading dye (Table 2.7), heated at ~85 °C for 5 min (except for fluorescein-labelled and GFP-tagged proteins), loaded into precast SDS-PAGE gels (Table 2.12) and run at 170–200 V in 1×MES SDS-PAGE running buffer for 45–60 min, or until the dye front had reached the bottom of the gel. Gels were fixed with Gel fixing and destaining solution for 20 min, stained with Coomassie blue stain solution for ~1 h and destained in Gel fixing and destaining solution (Table 2.7).

2.8.2 Western blot analysis

Western blot analysis was performed with yeast lysate samples from BiFC assay cultures. Following SDS-PAGE, the yeast lysates or soluble fractions were transferred onto Biotrace™ NT nitrocellulose membranes using a Hoefer TE22 Mini Tank Blotting Unit (Fisher Scientific) in Western blot transfer buffer. Transfer took place at 200 mA for 2–3 h at 4 °C. After transfer the membrane was blocked in 5% skim milk powder in TBST overnight with rocking at room temperature. The membrane was washed with TBST (15 min), and incubated with polyclonal anti-GFP antibody diluted 1:1000 in TBST with rocking for 1 h at room temperature. The membrane was then washed in 10 mL TBST (4 × 15 min), and subsequently incubated with the secondary anti-rabbit HRP-conjugate antibody diluted 1:4000 in TBST for 1 h with rocking at room temperature. Final washing was carried out in TBST (4 × 15 min). Antibody detection was carried out using the ECL Western Blotting Substrate kit, and chemiluminescence was detected by exposure to X-ray film (10 s to 2 min). The film was then developed using Kodak reagents.

2.8.3 Spectrophotometric determination of concentration

Protein and DNA concentrations were determined via UV-spectrophotometry using a ND-1000 spectrophotometer (NanoDrop Technologies Inc., 3411 Silverside Road, Wilmington, DE 19810, USA). Buffers were used as blanks. The absorbances of protein samples were measured at 280 nm and concentrations were calculated using the Beer-Lambert equation:

$$A_\lambda = \varepsilon_\lambda \cdot c \cdot l \qquad \text{(Equation 2.1)}$$

where A_λ is the sample absorbance at wavelength λ, ε_λ is the molar extinction coefficient (M^{-1} cm^{-1}) at wavelength λ, c is the concentration of the sample (M) and l is the path length (cm). Molar extinction coefficients were estimated using the program ProtParam [181] and are shown in Appendix C (Table C.1).

For protein constructs that did not contain tryptophan residues the concentration was calculated using the Equation 2.2 [182]:

$$144 \cdot (A_{215\ nm} - A_{225\ nm}) = \text{protein } \mu g \cdot mL^{-1} \qquad \text{(Equation 2.2)}$$

Concentrations of plasmid solutions and other dsDNA samples were determined by absorbance readings at 260 nm using the same spectrophotometer, assuming ε_{260nm} = 50 ng/µL/cm.

2.8.4 Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF)

Mass spectrometry analysis was done using the Matrix-assisted laser desorption/ionization - time of flight mass spectrometry (MALDI-TOF) method on the Voyager DE STR MALDI-TOF instrument (Applied Biosystems Inc., Foster City, CA USA) at the Sydney University Proteome Research Unit (SUPRU)/ Mass Spectrometry Core Facility (MSCF).

To generate peptides for protein identification by MALDI-TOF, bands were excised from Coomassie-stained SDS-PAGE gels, destained in gel fixing and destaining solution (Table 2.7), washed with 100% acetonitrile, dried, and then incubated with 15 µL of sequencing grade trypsin (12 ng/µL in 50 mM NH_4HCO_3) for 1 h at 4 °C. Excess trypsin was removed and the sample digested at 37 °C overnight in 15 µL 20 mM ammonium bicarbonate. All samples were desalted with C18 ZipTips. Subsequently, targets on the MALDI plate were first spotted with 1 µL of the appropriate matrix solution, CHCA for trypsin digest samples (peptides) and with sinapinic acid for proteins. Several concentrations of protein and peptide samples were prepared (1/10 dilutions in Matrix Solution, starting from ~ 1 µM), spotted on top of the previously deposited matrix on the MALDI plate, and mixed with matrix by pipetting. After drying (~30 min), the plate was loaded into the instrument for data acquisition. The results were analysed in Data Explorer (Microsoft), and Mascot Server (Matrix Science) was used for the identification and characterisation of proteins based on trypsin fingerprinting.

2.8.4.1 C18 ZipTips desalting

MALDI-TOF samples were desalted with C18 ZipTips (Millipore). The C18 tips were preincubated with 10 µL of Wetting Solution, washed with 10 µL Wash solution, and the 10 µL of protein/peptide solutions were bound to columns by pipetting them through the tip 5–10 times.

After washing the column 1–3 times with Wash Solution, the samples were eluted with 1.3 or 10 µL Matrix Solution by pipetting it through the tip 1–5 times.

2.9 Methods for protein characterisation

2.9.1 Far UV-Circular Dichroism (CD) spectropolarimetry

Samples for Far UV-CD experiments (13 µM protein in 20 mM Na-phosphate buffer pH 7.5, 20 mM TCEP) were placed in a 1-mm path length quartz cell seated in a peltier temperature controlled cell holder. CD spectra were recorded at 25 ºC on a Jasco J-815 spectropolarimeter (Jasco Inc., Easton, MD, USA). CD data were collected over the wavelength range 260–190 nm in a continuous scanning mode, with a speed of 20 nm/min, step resolution of 1 nm, bandwidth of 1 nm and a response time of 1 s. Data were omitted if the total signal at any given wavelength exceeded 500 V. Final spectra were the average of three scans, which was then baseline corrected. The corrected data were converted to mean-residue ellipticity ([θ], deg/cm^2/dmol) using the following equation:

$$[\theta]_{M,\lambda} = M \cdot \theta_\lambda / (10 \cdot d \cdot c) \qquad \text{(Equation 2.3)}$$

where M is the mean residue weight, θ_λ is the observed ellipticity (mdeg) at the wavelength λ, *d* is the path length (cm), and *c* is the concentration (g/mL).

2.9.2 Multi-angle laser light scattering (MALLS)

Analytical SEC for SEC-MALLS experiments was performed using a Superose 12 column on an Äkta Basic liquid chromatatography system (GE Healthcare) at a flow rate of 0.5 mL/min in Competition Assay Buffer. Protein samples of 250 µL at ~2 mg/mL were used. The SEC setup included an in-line miniDAWN light scattering detector and an interferometric refractometer (Wyatt Technologies, CA, USA). Light scattering analysis was performed using a 690 nm wavelength laser. Voltage and light scattering intensity were calibrated with toluene yielding a constant of 8.534×10^{-6} for this experiment. Molecular weights were estimated using ASTRA software (Wyatt Technologies). MALLS analysis of monomeric BSA was also conducted to check the calibration of the instrument

2.9.3 Nuclear magnetic resonance (NMR) spectroscopy of fluorescein-2-MES

The sample contained 2.7 mM fluorescein-2-MES in 20 mM Na-phosphate buffer pH 7.5 to which 5% (v/v) D_2O and 2 μM DSS (Table 2.1) were added. 1D ^{1}H-NMR spectra (64 scans) were recorded with a spectral width of 12 ppm at 25 ºC on an Avance III 800 MHz Bruker spectrometer (Bruker AXS GmbH, Karlsruhe, Germany) equipped with a 5 mm triple-resonance TCI cryoprobe and processed using Topspin 2.1 (Bruker). Water suppression was achieved using pulsed-field gradients. Spectra were processed using TOPSPIN (Bruker) running on Linux workstations. Spectra were referenced to DSS at 0.00 ppm.

2.10 Site-specific protein labelling with synthetic fluorophores

2.10.1 Labelling of SNAP-tagged proteins

pT7-His-SNAP-LMO4$_{LIMs}$-Ldb1$_{LID}$ was overexpressed in BL21-AI cells, and purified by Ni-affinity chromatography (see section 2.7.3.2), followed by anion-exchange chromatography (see section 2.7.5.2). The protein was labelled with BG-488 reagent in SNAP-tag labelling buffer, according to the manufacturer's instructions (protocol for labelling in solution, SNAP-Cell Starter Kit, New England Biolabs). Dialysis (see section 2.7.6.1) was performed to remove free reagent from solution and the degree of labelling was estimated (see section 2.10.3).

2.10.2 Labelling of proteins by Expressed Protein Ligation (EPL)

2.10.2.1 Production of the fluorescein-2MES reagent

5-(and 6-)Carboxyfluorescein succinimidyl ester (96 mg) and sodium 2-mercaptoethanesulfonate (375 mg) were combined and dissolved in 10 mL of 100 mM Na-borate buffer, pH 8.7. The reaction was stirred at room temperature for 2 h and the product purified by RP-HPLC (see section 2.7.5.3).

2.10.2.2 Production and purification of TEV protease

The S219V mutant of TEV protease was overexpressed as a His-tagged protein in BL21(DE3) cells. In the initial stages of the project the protein was purified only by Ni-affinity chromatography as described previously [183, 184]. In later stages of the project the enzyme was additionally purified by SEC. After purification, the enzyme was concentrated to ~0.5 mg/mL, diluted with glycerol to ~0.25 mg/mL (50% glycerol stocks), snap-frozen as 100 µL aliquots, and stored at -80 °C.

2.10.2.3 Production and purification of proteins with an N-terminal cysteine residue

Tethered LIMs-LID constructs and the DEAF1 construct in pGEX-TEVC vectors were overexpressed in BL21(DE3) cells, purified by GSH-affinity chromatography, treated with TEV to remove the tag and expose the N-terminal cysteine residue, and subsequently purified by anion-exchange chromatography.

2.10.2.4 Native chemical ligation (NCL) reaction

Labelling of N-terminal cysteines with fluorescein was performed by native chemical ligation (NCL). Tethered constructs with an exposed N-terminal cysteine residue were concentrated to 1–4 mg/mL in NCL buffer and the solutions used to resuspend dry fluorescein 2-MES reagent to a final concentration of 0.2–0.5 mg/mL. MESNA (or MPAA) thiol in buffer (0.4 M) was added to this tube to a final concentration of 20–100 mM thiol. The reaction was incubated in the dark at room temperature as indicated (5–92 h). Long reactions (>48 h) with MESNA were used to obtain products without significant levels of nonspecific binding (Figure 6.13). The reaction was stopped by adding 1 volume of buffer containing 100 mM DTT and incubated for 30 min at room temperature. Unreacted reagents were removed, and the buffer changed to anion-exchange buffer A, using PD-10 desalting columns (section 2.7.6.2). Samples were analysed by SDS-PAGE and fluorescence scanning on Typhoon FLA9000 scanner (GE Healthcare), and absorbances at 280 nm and 494 nm measured on the ND-1000 after each step.

2.10.3 Calculating the degree of labelling

Purified labelled protein was diluted in buffer, and absorbance was measured at 280 nm (A_{280}) and 506 nm (BG-488-labelled proteins, section 2.10.1) or 494 nm (fluorescein-labelled proteins, section 2.10.2.4) and the concentration of labelled protein calculated using Equation 2.4:

$$c_p = 1/\varepsilon_{280} \cdot (A_{280} - CF \cdot A_{exc}) \cdot \text{dilution factor} \qquad \text{(Equation 2.4)}$$

where ε_{280} is the molar extinction coefficient of the unlabelled protein at 280 nm (for specific values see Appendix C), A_{exc} is the absorbance measured at the appropriate excitation wavelength (506 nm for BG-488-labelled and 494 nm for fluorescein-labelled proteins), and CF is a correction factor (~0.3 for most proteins), which compensates for absorption of the bound fluorophore (fluorescein) at 280 nm. The degree of labelling was calculated using the following equations:

$$\text{dye per protein} = A_{exc}/(\varepsilon_{exc} \cdot c_p) \qquad \text{(Equation 2.5)}$$

or

$$\%\ \text{labelled protein} = 100 \cdot A_{exc}/(\varepsilon_{exc} \cdot c_p) \qquad \text{(Equation 2.6)}$$

where ε_{exc} is the extinction coefficient of BG-488 at 506 nm or fluorescein at 494 nm (~70 000 $M^{-1}cm^{-1}$ for both).

2.11 Quantitative competition binding assays

2.11.1 The proteolysis of tethered complexes

Factor Xa was used to cut the specific IEGR|G site in tethers that connect the LIMs and LID domains in tethered complexes. The reaction was usually performed with 1–5 µM protein in the appropriate proteolysis buffer (Table 2.7), containing 20 µg/mL Factor Xa, for 6 h at room temperature. The reaction was stopped with 1 mM PMSF.

2.11.2 Fluorescence anisotropy (FA)

All solutions used in fluorescence spectroscopy were filtered through 0.22-µm filters to remove light-scattering particles. Increasing concentrations of MBP-$Ldb1_{LID}$ were mixed with solutions containing <10 nM of Factor Xa-cleaved, fluorescent LIMs:LID complexes in 300 µL

reaction volumes, and incubated at room temperature for at least 30 min up to several hours. Each sample was transferred to a quartz cuvette and the fluorescence intensity monitored: excitation at 475 nm and emission at 510 nm for GFP-tagged constructs; excitation at 492 nm and emission at 520 nm for fluorescein-labelled constructs. Data were recorded in order from low to high MBP-Ldb1$_{LID}$ concentration samples. The Fluorescence spectrophotometer (Cary Eclipse, Varian Instruments, Mulgrave VIC) was fitted with a manual polariser (Varian Instruments), and Long Pass filters (Coherent Scientific Pty. Ltd.) for excitation and emission. The temperature was maintained at 25 ºC using a block temperature controller (Varian Instruments, Mulgrave, VIC), slit widths were set to 10 nm and each data point was averaged over 15 s. In control assays, MBP-only was used instead of MBP-Ldb1$_{LID}$ to evaluate non-specific binding.

Anisotropy values (r) were calculated from the formula

$$r = (I_{vv} - G \cdot I_{vh})/(I_{vv} + 2 \cdot G \cdot I_{vh}) \qquad \text{(Equation 2.7)}$$

where:

I_{vv} = emission light intensity when both the excitation and emission filters transmit vertically polarized light

I_{vh} = emission light intensity when the excitation filter transmits vertically polarised and the emission filter transmits horizontally polarised light

Any bias towards vertically or horizontally polarised light caused by the fluorescence spectrophotometer was corrected by including a G-value:

$$G = I_{hv}/I_{hh}, \qquad \text{(Equation 2.8)}$$

where:

I_{hv} = emission light intensity when the excitation filter transmits horizontally polarised and emission filter transmits vertically polarised light

I_{hh} = emission light intensity when both the excitation and emission filters transmit horizontally polarized light.

The G-value was 1.7–1.9 for different FA competition assays performed in this study.

2.11.3 Fluorescent gel shift assays

Fluorescently labelled LIMs-LID tethered constructs were treated with Factor Xa to cut the linker between the LIMs and LID domains, generating un-tethered (or cut) complexes for use in competition binding assays. Increasing concentrations of MBP-$Ldb1_{LID}$ were added to compete with $Ldb1_{LID}$ from the original construct for binding to the LIM domains. The ~42-kDa MBP tag means that newly formed complexes are larger, which can be detected in different binding assays.

2.11.3.1 Native gel electrophoresis (protein gel shift assays)

Samples of cut GFP-tagged LIMs:LID complexes were mixed with increasing concentrations of MBP-$Ldb1_{LID}$ in a total reaction volume of 0.5 mL. Following incubation for 1 h at room temperature, these samples were mixed with the Native gel loading buffer and loaded into wells of a precast gradient native gel (NativePAGE Bis-Tris gel system, Life Technologies). The gels were subjected to electrophoresis at 4 °C in the dark, in Blue Native PAGE or Clear Native PAGE formats, according to the manufacturer's instructions. In both formats Bis-Tris-Tricine buffer was used, but in the Blue Native format the cathode buffer contained 0.4% Coomassie Brilliant Blue G-250. After 1 h of electrophoresis at 150 V, the voltage was increased to 250 V and the gel was run for a further 45–50 min.

2.11.3.2 Detection and quantification of fluorescence

Blue and clear native gels were scanned on a Typhoon FLA9000 imager (GE Healthcare Life Sciences) using EGFP excitation/emission settings. The Clear Native gels were subsequently stained with Coomassie blue stain solution. The obtained images were saved in TIFF format and ImageJ [185] was used to analyse these images and quantify the intensity of signal in each shifted fluorescent band. These values were plotted to obtain binding curves in Origin9.1 (OriginLab, Northampton, MA). Calculation and curve fitting is explained in more detail in Appendix F.

2.12 Bioinformatics, structure models and image processing

ProtParam, PeptideCutter and some other Expasy online tools [181] were used to predict characteristics, structure, and identify proteins analysed in this study. NCBI BLAST search tools Clustal W2 and Clustal Omega alignment were used to compare protein sequences of different proteins and find conserved residues and motifs [186, 187]. The STRING protein-protein interaction database was used to investigate functional connections and possible interactions between LIM-HD proteins and their partners [116]. The Weblogo online tool was used to generate a diagram of conserved residues in HDs of LIM-HD proteins (http://weblogo.berkeley.edu; [188]). Data on gene positions, clusters, paralogues and ortologues, phylogenetics, protein isoforms, mutations etc. was obtained from the UCSC Genome Browser (http://genome.ucsc.edu) [189] and ENSEMBL database [190]. Protein expression and localization data was obtained from the mouse Gene Expression Database (GXD) [86], Gene Expression Evolution database (Bgee) [191], or the Human Protein Atlas [192].

Figures in this thesis were made with Adobe Illustrator CC (Adobe Systems) and CorelDraw Graphics Suite X7 (Corel Corporation) software packages. Image models of various structures were analysed and prepared in the PyMOL Molecular Graphics System, version 1.7.6.3 (Schrödinger, LLC.) Where available, RCSB protein data bank (PDB) accession codes were indicated [193].

Chapter 3 Optimising Y2H mating arrays for interactions between LIM-HD proteins

3.1 Introduction

This chapter describes experiments aimed to detect protein-protein interactions between all LIM-HD proteins and specifically if LIM-HD proteins other than Lhx3/4 and Lmx1a interact with Isl1/2.

Protein-protein interactions in this chapter were exclusively investigated by yeast two-hybrid analysis. For this reason, the approach and the two ways the approach was used in this chapter are explained in detail. More information on yeast strain genotypes, yeast plasmid maps, media composition and experimental procedures can be found in the Materials and Methods section of this thesis.

3.1.1 General approach

Yeast-two hybrid (abbreviated hereafter as Y2H) has been a standard *in vivo* method for detecting protein-protein interactions for more than 20 years [194]. The approach is based on the idea that if you express two interacting proteins fused to specific complementary active domains, upon binding, this hybrid complex will generate some detectable output *in vivo*. For standard Y2H, this function is the transcriptional activation of reporter genes. Many related two- (and three-) hybrid methods in yeast and other organisms have been developed that are based on enzyme activity, fluorescence and various other functions [195-198].

3.1.2 Plasmids, yeast strains and reporter genes

The experiments presented in this chapter used the well-defined Gal4 system (Figure 3.1). In this system, the two domains to be tested are cloned as fusion constructs with each of the DNA-binding domain (DBD) and activating domain (AD) of yeast transcription factor Gal4. Each construct is present in a separate expression vector. DBD-fusion constructs were expressed from the NpGBT9 plasmid, and AD-fusion constructs were expressed from the pGAD10

plasmid, both from a short ADH1 promoter (396 bp) that gives very low levels of constitutive expression during the exponential growth phase [199].

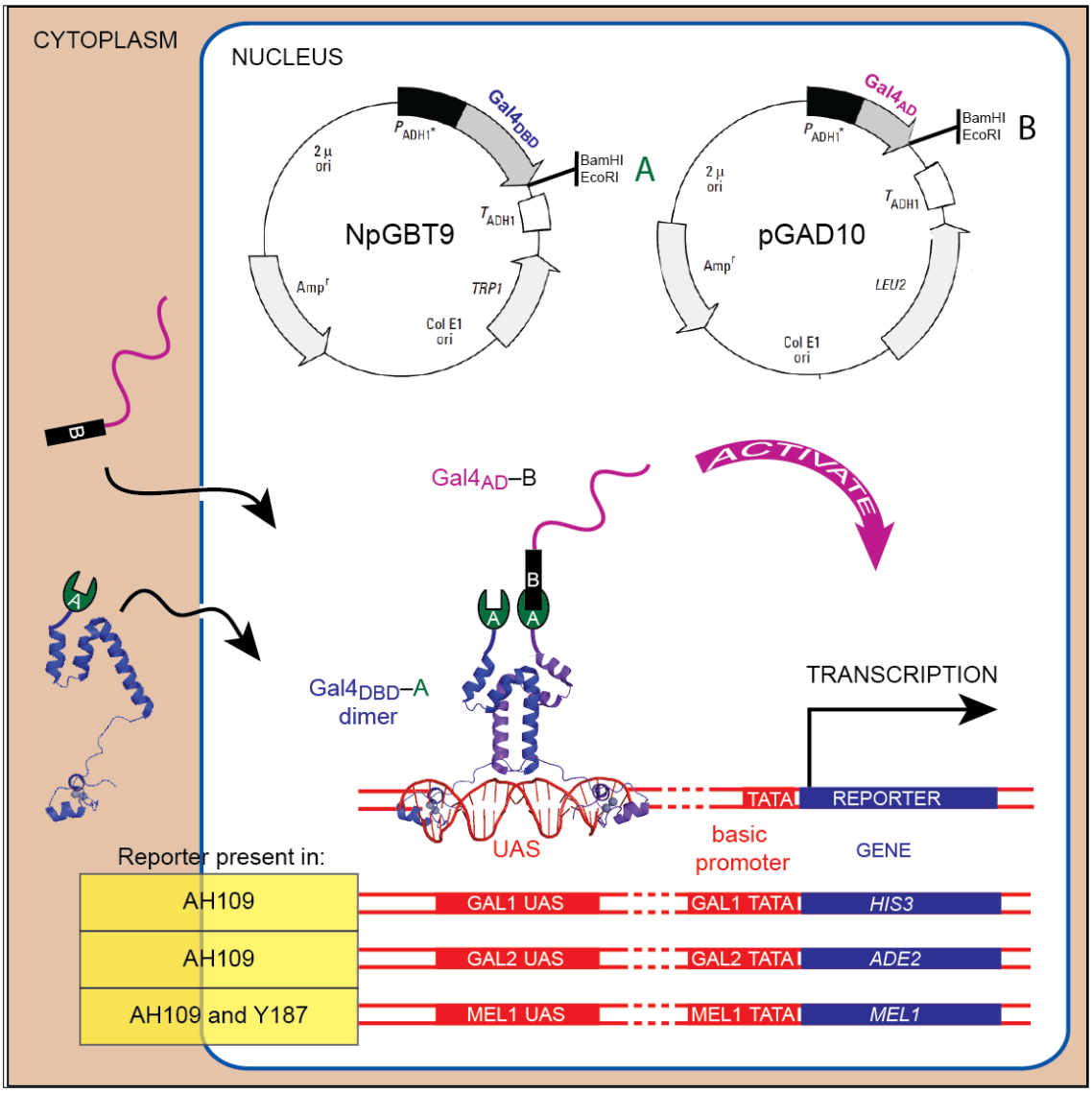


Figure 3.1. The Gal4 yeast two-hybrid system. The DNA-binding domain (DBD) of the Gal4 yeast transcription factor is fused to protein A and the activating domain (AD) of Gal4 is fused to protein B. The two fusion constructs are co-expressed in a yeast cell and localized in the nucleus. If proteins A and B interact, reconstituted Gal4 dimers bind UAS sites and activate the *HIS3*, *MEL1* and *ADE2* reporter genes, used in this thesis. Inserted yellow table shows which of the reporters are present in AH109 and Y187 strains.

The assays used commercially available (Clontech) strains of yeast *Saccharomyces cerevisiae* AH109 (in all experiments) and Y187 (in mating arrays only). These yeast strains

have both been made auxotrophic (cannot grow in the absence of a given substance) for the amino acids leucine (L), tryptophan (W) and histidine (H), and the nucleobase adenine (A).

NpGBT9 and pGAD10 contain selective markers (genes that express the missing enzymes) that allow growth on tryptophan-deficient (-W) and leucine-deficient (-L) media, respectively. When transformed with both types of plasmids, AH109 and Y187 strains should grow on SD-L-W media. The AH109 genome also carries reporter genes that compensate for auxotrophy in histidine (*HIS3*) and adenine (*ADE2*) biosynthesis. Both AH109 and Y187 carry the colourimetric reporter gene *MEL1*, which encodes an α-galactosidase that creates a blue product from the substrate X-α-Gal.

Upstream activating sequences (UAS) in promoter regions of these reporter genes contain binding sites for the Gal4 protein. When the test proteins interact, the hybrid complex will be recruited to Gal4-binding sites via the $Gal4_{DBD}$ and interact with the basal transcriptional machinery via the $Gal4_{AD}$. The hybrid complex activates transcription of reporter genes, allowing growth on selective media without histidine and adenine (Table 3.1). Up to 100 mM of 3-aminotriazole (3-AT), a competitive inhibitor of the *HIS3* reporter product, is often added to media to increase selection pressure. In this thesis, 0.5–10 mM 3-AT was added to interaction selection plates (Table 3.1). However, 3-AT was not used to further increase selection pressure under double selection conditions.

Table 3.1. Yeast selection conditions used to detect an interaction. Different media were classified according to their assumed selection pressure into weak, moderate or double selection media. All media require activation of the *HIS3* auxotrophic reporter while double selection media requires the additional activation of the *ADE2* auxotrophic reporter. Weak selection media does not contain 3-AT. Moderate selection media contains 3-AT (0.5–10 mM). X-α-Gal, a substrate for the *MEL1* colourimetric reporter was added in weak and moderate selection media. Any single assay used a subset of these conditions. SD media content is listed in Table 2.16. SD-L-W-H(-A) is SD media lacking amino acids L, W, H (and adenine nucleobase).

Weak selection	SD-L-W-H, X-α-Gal
Moderate selection	SD-L-W-H, X-α-Gal, 0.5–10 mM 3-AT
Double selection	SD-L-W-H-A

Yeast growth and *MEL1* reporter activation were followed in parallel in all experiments except the double selection conditions where X-α-Gal was not included in the media. However, colour is used to report on a given interaction only when differences in colour development could not be explained by differences in growth, or when the growth was not consistent.

3.1.3 Background growth: causes, inhibition and controls

In all Y2H experiments positive and negative controls were included on every plate. The main negative controls comprised each of the DBD- and AD-fusion proteins tested against the AD-only or DBD-only domains, respectively, expressed from the appropriate 'empty' plasmids (i.e., the AD or DBD are produced in the absence of a fusion partner). Where a negative control gave rise to yeast growth it was considered as assay background for the related interaction. Yeast spots that did not grow significantly above background were classified as false positives. Weak background growth can be caused by leaky expression of the *HIS3* reporter in the absence of Gal4 activity. Indeed, under weak selection conditions this was often observed, so most experiments reported use 3-AT (i.e., moderate selection conditions) to eliminate this phenomenon.

Some proteins, when expressed as DBD-fusion constructs, can directly activate (‘autoactivate’) reporter genes and lead to false positive results. Autoactivation can also be caused by AD-fusion constructs of DNA-binding proteins that bind non-specifically to promoters of reporter genes [198, 200, 201]. Autoactivation is the most common cause of false positives and can be confirmed using additional negative control experiments in which a single plasmid containing either an AD- or DBD-fusion construct of a protein in question is transformed into yeast. Several moderate selection conditions were used in this work to compensate for different levels of autoactivation caused by some of the tested constructs.

When testing a pair of constructs A and B for an interaction, the Y2H assay commonly uses two combinations of fusion constructs: DBD-A vs. AD-B; and, DBD-B vs. AD-A. Although carrying out experiments that test both combinations is preferred, autoactivation can often prevent detection in one of the combinations.

3.1.4 Yeast mating

There are two strategies for inserting both DBD- and AD-containing plasmids into yeast cells: co-transformation (adding both plasmids during transformation) and mating [202, 203]. The latter method takes advantage of the ability of haploid yeast *S. cerevisiae* to exist in either of two complementary mating types, MATa or MATα. The two mating types can mate with each other and thereby fuse into a diploid form (Figure 3.2). In this study AH109 (MATa) cells were transformed with DBD-containing plasmids and Y187 (MATα) with AD-containing plasmids. The haploid transformants were mated to create diploid yeast cells that co-express both fusion constructs. A big advantage of mating is the simplification of transformation when large numbers of interactions are tested [204, 205].

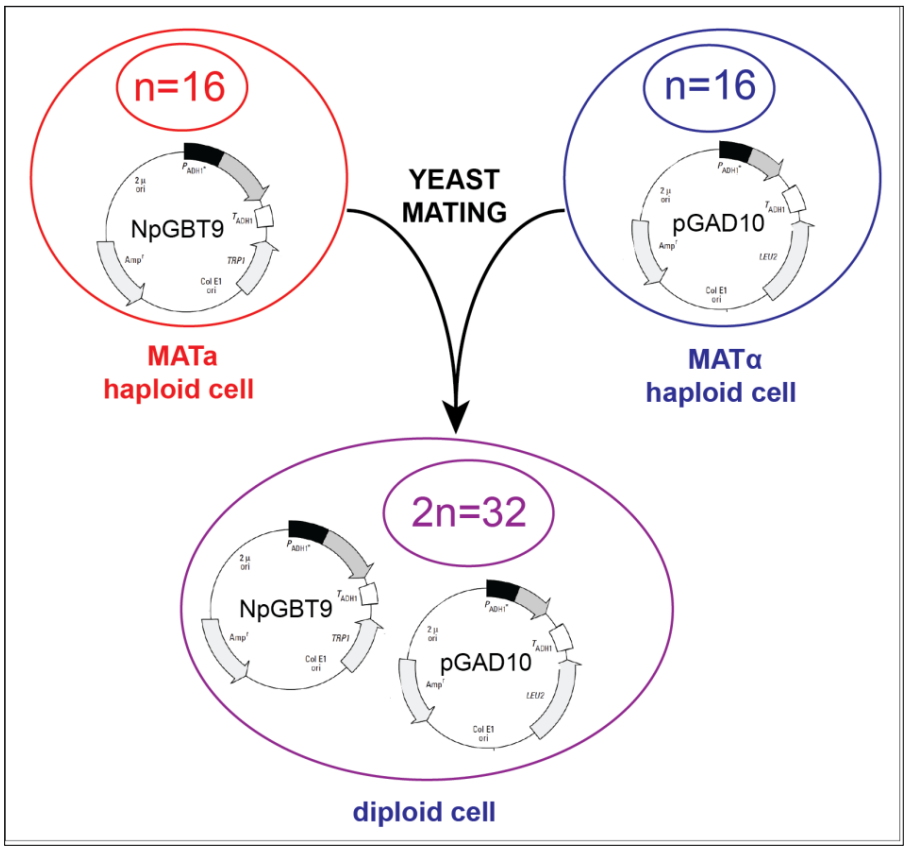


Figure 3.2. Yeast mating. Haploid strains of opposite mating types are each transformed with one of the complementary Y2H plasmids. MATa and MATα cells fuse together to make a diploid yeast cell (i.e., with two sets of 16 yeast chromosomes) that carries both Y2H plasmids.

3.1.5 Arrays and replica plating

Y2H mating arrays often test defined numbers of known protein constructs against each other in an ordered matrix of spots on the same plate [206-208]. Mating arrays can also be used for library screens [209]. In mating arrays, replica plating is used to transfer yeast from haploid plates to a mating plate, and then transfer diploid yeast to different selection plates. Replica plating maintains the relative positions of yeast colonies, spots or stripes from starting plates ('master' plates) on the subsequent plates (replica plates) [209, 210]. Where appropriate, in these experiments manual velveteen stamp replica plating was used in which a sterile piece of velveteen cloth is applied to the first plate and then touched onto the replica plates.

3.2 Identification of autoactivating domains and experimental design

Full-length (or near full-length) constructs of several mouse LIM-HD proteins in Y2H assays were first tested for autoactivation as previous studies in the laboratory have indicated that autoactivation is a potential problem for this family of proteins. All constructs showed some level of autoactivation, although the extent of autoactivation varied. For example, when expressed as a DBD-fusion construct, Lhx3a showed strong background growth in the negative control under all selection conditions (Figure 3.3A) and background growth when expressed as an AD-fusion construct only under weak selection conditions, as did most other members of the family (e.g., Figure 3.3B). This autoactivation likely arises from the C-terminal regions of these proteins as several LIM-HD proteins are reported to contain transcriptional activation domains in this region [136, 211-213], although the HDs also appear to contribute to this affect (data not shown). In contrast, although DBD-fusion constructs of tandem LIM domains (LIMs) can show background growth under weak selection, this is generally eliminated by the inclusion of up to 10 mM 3-AT in the media, as shown for DBD-Isl2$_{LIMs}$, and DBD-Lhx2$_{LIMs}$ (Figure 3.3C). Other LIMs constructs have been used previously in Y2H experiments as DBD-fusion constructs, with little or no background growth [18, 156, 167].

The tests confirmed that DBD-full-length LIM-HD constructs can cause background growth that could make weak interactions difficult to detect. Therefore it was decided to use only LIMs constructs in DBD-fusion constructs and test them for binding against full-length

constructs expressed as AD-fusion constructs (DBD-LIMs vs. AD-full-length combination). There is good precedent for using these constructs, as $Lhx3_{LIMs}$ mediates an interaction with the $Isl1_{LDB}$ domain. It should be noted that with this assay design non-LIM mediated interactions within the LIM-HD family cannot be assayed. The mating arrays were carried out with up to 10 mM 3-AT on moderate selection plates, and on double selection plates.

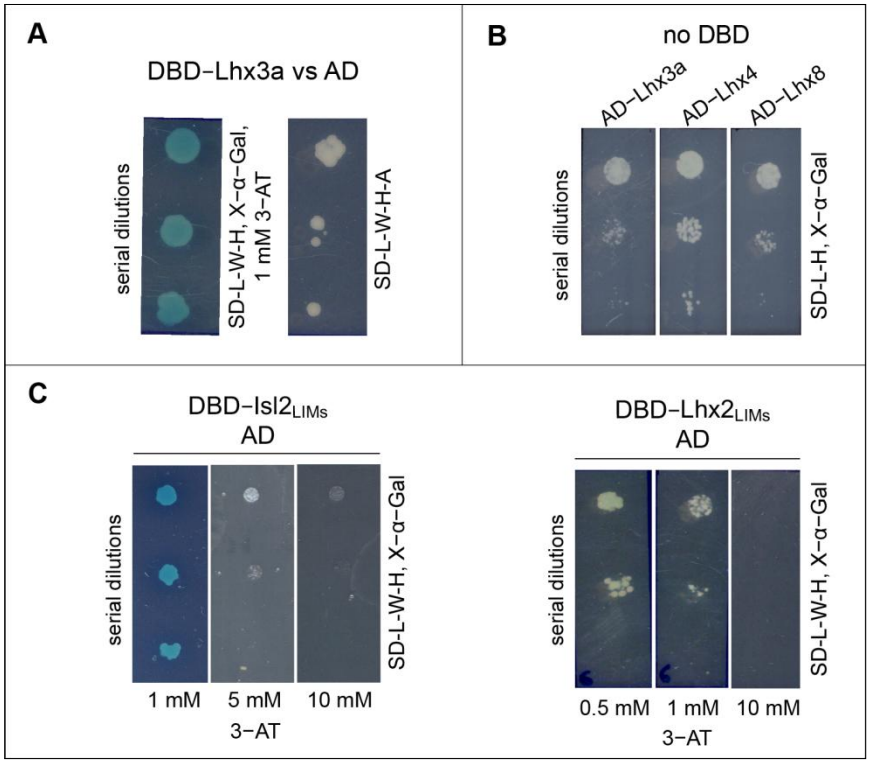


Figure 3.3. False positives from autoactivation. (A) Autoactivation produced by DBD-Lhx3a in the presence of the Gal4 AD on moderate (SD-L-W-H, X-α-Gal, 1 mM 3-AT) and double (SD-L-W-H-A) selection plates. (B) Autoactivation produced from AD-LIM-HD fusion constructs. Yeast transformants expressing AD-Lhx3a, AD-Lhx4 or AD-Lhx8 (but not the DBD-domain), spotted on weak (SD-L-H, X-α-Gal) selection plates. (C) Yeast co-expressing AD-only with either DBD-$Isl2_{LIMs}$ on the left or DBD-$Lhx2_{LIMs}$ on the right on SD-L-W-H + X-α-Gal plates with a range of 3-AT concentrations. Y2H autoactivation and negative control assays (A-C) were performed using cell-normalized dilution spot-tests.

3.3 Preliminary experiments

Prior to the start of this thesis project, a series of Y2H spot-test assays were carried out to test possible interactions between some LIM-HD proteins [169]. These experiments covered only some of the possible interactions within the family of LIM-HD proteins as not all protein

constructs were available at the time. The results of these assays are presented in Table 3.2, and showed previously detected interactions between Isl1α, and $Lhx3_{LIMs}$ and $Lhx4_{LIMs}$. In addition this work reported some novel, apparently weaker interactions between Lhx3a and the tandem LIM domains of several LIM-HD proteins, and between $Lmx1a_{LIMs}$ and $Isl1_{LBD}$. Some of these apparently weaker interactions could not be detected by an Honours Student in the laboratory, Kim Tieu, at a later date [214]. This disparity possibly arose because weaker interactions often give varying levels of growth from experiment to experiment as they are close to the limit of detection. Small variations in media, growth conditions and handling can push them either side of the detection limit. Y2H assays on weak selection plates are particularly sensitive to these effects. It is also possible that some of these interactions were in fact false positives due to additive autoactivation effects of DBD-LIMs and AD-Lhx3a fusion constructs.

Table 3.2. The results of preliminary Y2H spot-test assays from [169]**.** The table shows a summary of relative yeast growth "on SD-L-W-H plates containing X-α-Gal and 3-AT (0.5 mM or 1 mM). (+++) indicates strong growth and blue colour, (++) signifies growth and faint/absent blue colour; (+) indicates weak growth and no blue colour; (-) indicates absence of yeast growth on SD-L-W-H plates containing X-α-Gal and 3-AT (1 mM)" [169]. The scaling system used for these results was different from the system applied in experiments performed for this thesis. Light grey fields mark the constructs and interactions that were not tested.

+++ strong ++ moderate + weak - no interaction		AD–fusions												
		Lhx1	Lhx2	Lhx3	Lhx4	Lhx5	Lhx6	Lhx8	Lhx9	Lmx1a	Lmx1b	Isl1	$Isl1_{LBD}$	Isl2
DBD–fusions	$Lhx1_{LIMs}$	-	++	++					+	-		-	-	
	$Lhx2_{LIMs}$	-	-	+					-	-		-	-	
	$Lhx3_{LIMs}$	-	-	+					-	-		+++	+++	
	$Lhx4_{LIMs}$	+	-	+					-	-		+++	+++	
	$Lhx5_{LIMs}$													
	$Lhx6_{LIMs}$													
	$Lhx8_{LIMs}$													
	$Lhx9_{LIMs}$	-	-	+					-	-		-	-	
	$Lmx1a_{LIMs}$	-	-	-					-	-		+	+	
	$Lmx1b_{LIMs}$													
	$Isl1_{LIMs}$	-	-	-					-	-		-	-	
	$Isl1_{LBD}$	+	-	+++					-	++		-	-	
	$Isl2_{LIMs}$													

3.4 Advantages of a larger scale approach with appropriate controls

To minimise experimental variation, an aim of this project was to simultaneously test all mouse LIM-HD proteins (including all Lhx6 isoforms), for pair-wise interactions. Given the

large number of possible interactions and corresponding controls, a Y2H mating array was designed that would take advantage of replica plating and enable all the interactions to be tested at the same time, on a single plate. Hits from this screen would be confirmed using Y2H spot-tests.

Since the $Ldb1_{LID}$ interacts with LIM domains of all LMO and LIM-HD proteins, it was used as a positive control (both DBD- and AD-$Ldb1_{LID}$) in all arrays. $Isl1_{LBD}$ was chosen as a second positive control because it is known to interact with $Lhx3_{LIMs}$, and to determine if other LIM-HD proteins could interact with this domain. DBD- and AD-fusion constructs of rat Pit1 protein, a putative partner of Lhx3 [215], were included in some of the assays but this interaction is not a focus of this thesis. Human SSB1 protein, which was shown previously not to interact with LMO proteins in Y2H assays (Dr Liza Cubeddu, personal communication) was used as a non-related protein control. Protein isoforms and constructs used in the assays are described in Chapter 2.

3.5 Y2H mating arrays

3.5.1 A cross-mating array

NpGBT9 constructs were individually transformed into AH109 yeast maintained on SD-W media, whereas pGAD10 constructs were individually transformed into Y187 cells maintained on SD-L media. The first rounds of mating arrays were performed using a cross-mating procedure (section 2.5.4.1). On mating plates, horizontal stripes from the AH109 transformants were pink compared with mostly cream coloured vertical stripes of the Y187 transformants (Figure 3.4A). The red pigment is an oxidized intermediate of adenine synthesis (AIR) and is known to appear in yeast strains that carry the *ade2* and/or *ade1* mutant genes [216-219] or the *ade2* gene replaced by the ADE2 reporter (as in AH109 strain, [220]), when growing on media with low but permissive adenine content. The Y187 strain does accumulate the same red pigment, but probably at a different rate than AH109.

Only spots with mated diploid yeast that carried auxotrophic selection markers for both -L and -W grew on the diploid selection plate (Figure 3.4B). Yeast cells that were transferred but did not grow because they lacked appropriate plasmids were visible as opaque spots on the

diploid selection plate after incubation. The same effect can be seen on all interaction selection plates. The negative mating controls showed no yeast growth, as expected, because the two mating partners carried the same plasmid type, and the resulting diploid cells had only one of the two auxotrophic markers necessary for survival on the SD-L-W plate. Diploid yeast were further replica plated onto 1 mM and 10 mM 3-AT moderate selection plates (Figure 3.6). Double-selection conditions were also used in cross-mating arrays (not shown) but since the resolution (separation between neighbouring spots) became unacceptable after two replica platings from the same diploid selection plate, moderate conditions were given priority.

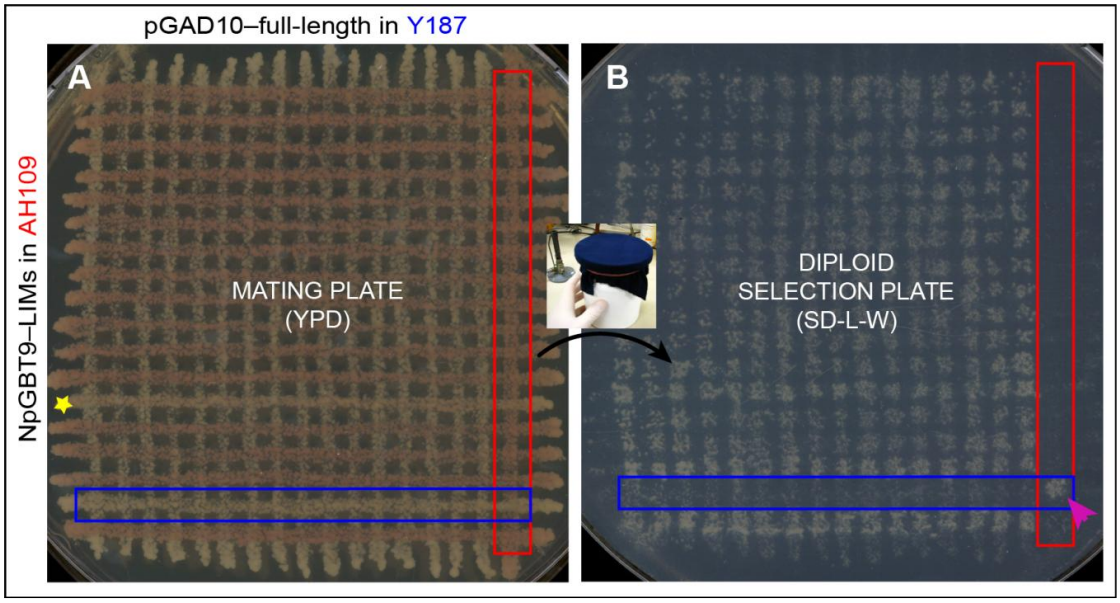


Figure 3.4. Cross-mating and diploid yeast selection. Black arrow shows the replica plating direction and the small photograph behind the arrow shows the replica plating apparatus (velveteen stamp). Note that although the figure emphasizes strains and plasmids with LIMs and full-length LIM-HD constructs, the empty plasmids and other controls were included in this array as explained in the text. (A) Cross-mating plate showing both haploid strains (and mated diploids at crossing points) growing on a YPD plate after being replica plated from each of the single-plasmid selection plates with streaks arrayed orthogonally to each other. The yeast was left to grow and mate overnight at 30 °C. (B) Cross-mated yeast was replica plated on to SD-L-W medium to select for diploids containing both plasmids. Mating controls are indicated by rectangles. Blue rectangles show the position of Y187 yeast transformed with an empty NpGBT9 plasmid and transferred onto the YPD plate from the horizontally streaked SD-W plate. Red rectangles show the position of AH109 yeast transformed with an empty pGAD10 plasmid and transferred onto the YPD plate from the vertically streaked SD-L plate. A purple arrowhead points to a positive mating control spot (B). A yellow star indicates the horizontal streak of AH109 transformed with NpGBT9-Isl2$_{LIMs}$ (A).

In the cross-mating arrays, the final selection plates (1 and 10 mM 3-AT moderate selection conditions) were scanned every day for 4 days. Each of the two plates were analysed

(scored for interactions) at incubation times that were the most informative (i.e., gave the best signal-to-noise ratio and resolution). Background growth made 1 mM 3-AT moderate selection plates difficult to analyse after 2–3 days. Fortunately, most positive controls could be detected on day 1 of the incubation.

Although replica plating is a well-established method for this type of interaction array, several issues were noted. The amount of yeast cells transferred to each spot by velveteen stamp replica plating cannot be normalized prior to transfer and is probably much higher than the highest amounts commonly used in Y2H spot-tests (~10^4 cells). Therefore, it is likely that in the absence of autoactivation all weak interactions detectable in a particular Y2H system would appear on weak selection plates using this method.

Yeast spot size on an interaction selection plate is determined by spot size on the previous plate, efficiency of replica plating transfer and media consistency, in addition to reporter activation. Consecutive replica plating steps led to decreased resolution of yeast spots. Based on initial observations of how closely spaced streaks could be prior to fusion of spots and/or diffusion of blue pigment to neighbouring spots, this cross-mating protocol was estimated to have an upper limit of 16 streaked lines per column/row in a 140 mm plate. However, the complete LIM-HD vs. LIMs array needed at least 18 columns for all AD-fusion constructs and controls.

3.5.2 A spot-mating Y2H array

To improve the resolution of the assay, instead of streaking and crossing the yeast streaks, small aliquots of overnight cultures of AH109 and Y187 transformants in the appropriate selective media were spotted manually on top of each other on a mating plate (section 2.5.4.2.). Mating control was not deemed necessary in this version of the assay. Mated yeast spots were then replica plated on to a diploid selection plate (Figure 3.5), which was used as a template for replica plating onto interaction selection plates. The interaction spots were scanned every 24 h (Figures 3.7 and 3.8).

The spot-mating arrays showed improved resolution, which allowed analysis of up to 20 × 20 spots in 140 mm plates. Using relatively drier YPD and selection plates allowed cleaner transfers and did not negatively affect the yeast growth but improved resolution of spots. At this point

attempts were made to identify additional weaker interactions under weak selection conditions. A variation of the spot mating protocol was attempted to further reduce the loss in resolution induced by consecutive replica plating. In the short protocol (Figure 3.5, green arrow) replica plating was done only once to transfer yeast spots directly from the YPD plate on to the interaction plates. The resolution was slightly improved but the overall growth rate of yeast on the weak selection plate was still very high, which made the background growth in all spots appear early in the incubation period. For this reason, this plate was monitored for only 48 h (Figure 3.7C and D).

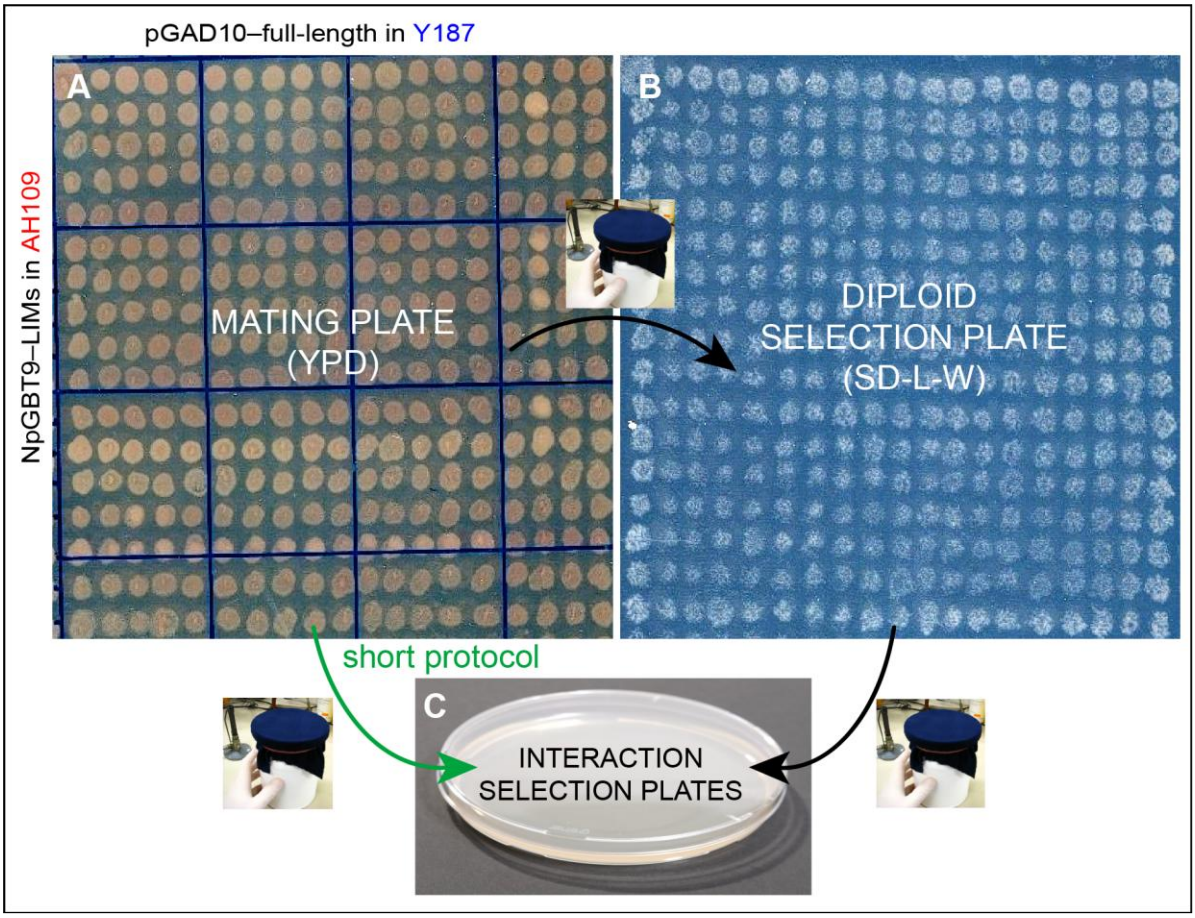


Figure 3.5. Y2H spot-mating flowchart. Haploid strains were manually spotted on top of each other on a YPD mating plate and incubated for 24 h at 30 °C. AH109 transformants expressing DBD-fusion constructs (rows) were mated with Y187 transformants expressing AD-fusion constructs (columns) (A). The successfully mated diploid yeast spots were selected on a SD-L-W plate (B). Standard (black arrows) and short (green arrow) protocols were performed for transferring yeast to the various interaction selection plates (C). The small photographs represent replica plating. Note that although the figure emphasizes strains and plasmids with LIMs and full-length LIM-HD constructs, the empty plasmids and other controls were included in this array as explained in the text.

3.6 Results from the cross-mating and spot-mating Y2H arrays

The following section summarises all of the interaction data taken from Figure 3.6 (cross-mating method; 1 and 10 mM 3-AT moderate selection), Figure 3.7A–B (spot-mating, standard protocol; 1 mM 3-AT moderate selection) and Figure 3.8 (spot-mating, short protocol; double selection). Note that references to colour only refer to weak and moderate selection plates, where X-α-Gal was included in the media.

3.6.1 Controls

In terms of interpreting the interaction plates, it was first necessary to evaluate the negative and positive controls to establish if constructs show different levels of background growth in diploid cells compared to the haploid cells that were tested for autoactivation in section 3.2, and to establish if known interactions were detected by this method.

Examination of DBD negative controls (expressing DBD-fusion constructs and AD-only) confirmed the results from the preliminary autoactivation assays. The pattern of reproducible background growth on 1 mM 3-AT moderate selection plates was consistent with that expected from preliminary autoactivation assays.

DBD-Isl2$_{LIMs}$ autoactivation was particularly strong in these arrays and the corresponding yeast grew on the 10 mM 3-AT moderate selection plate but did not develop colour (Figure 3.6B). This construct, as well as DBD-rPit1, allowed some weak growth due to autoactivation under double selection conditions (Figure 3.8). Of the full-length fusion constructs, AD-Lhx1, Lhx2, Lhx3a and Lhx4 showed low levels of growth when tested against DBD-only (AD negative controls) under 1 mM 3-AT moderate selection conditions in the streak-mating method, which was eliminated under 10 mM 3-AT moderate selection conditions. Other negative controls in this combination were essentially clear.

For the positive controls: the interaction plates showed evidence of an interaction (yeast growth and evolution of blue colour in the presence of X-α-Gal) between all AD-full-length LIM-HD constructs and DBD-Ldb1$_{LID}$. Thus, all of the longer AD-full-length constructs appear to have reasonable levels of expression and stability, with each containing functional LIM-domains. Similarly, AD-Isl1α and AD-Isl2 showed evidence of an interaction with DBD-

$Lhx3_{LIMs}$ and DBD-$Lhx4_{LIMs}$, indicating that AD-Isl1α and AD-Isl2 are stably expressed, with each containing a functional LBD. Apart from the Lhx1 and Lhx5 interactions (which appeared to be weaker under some conditions, with variability from experiment to experiment), all DBD-LIMs constructs showed evidence of an interaction with AD-$Ldb1_{LID}$. Note that some early experiments (e.g., Figure 3.6) used $Lhx1^*_{LIMs}$, which was found to extend only as far as the last zinc-coordinating aspartate residue, D117. LIM domains tend to have ~4–6 structured residues beyond this point, so an extended construct (to S123) was generated and used in later experiments and is referred to as $Lhx1_{LIMs}$.

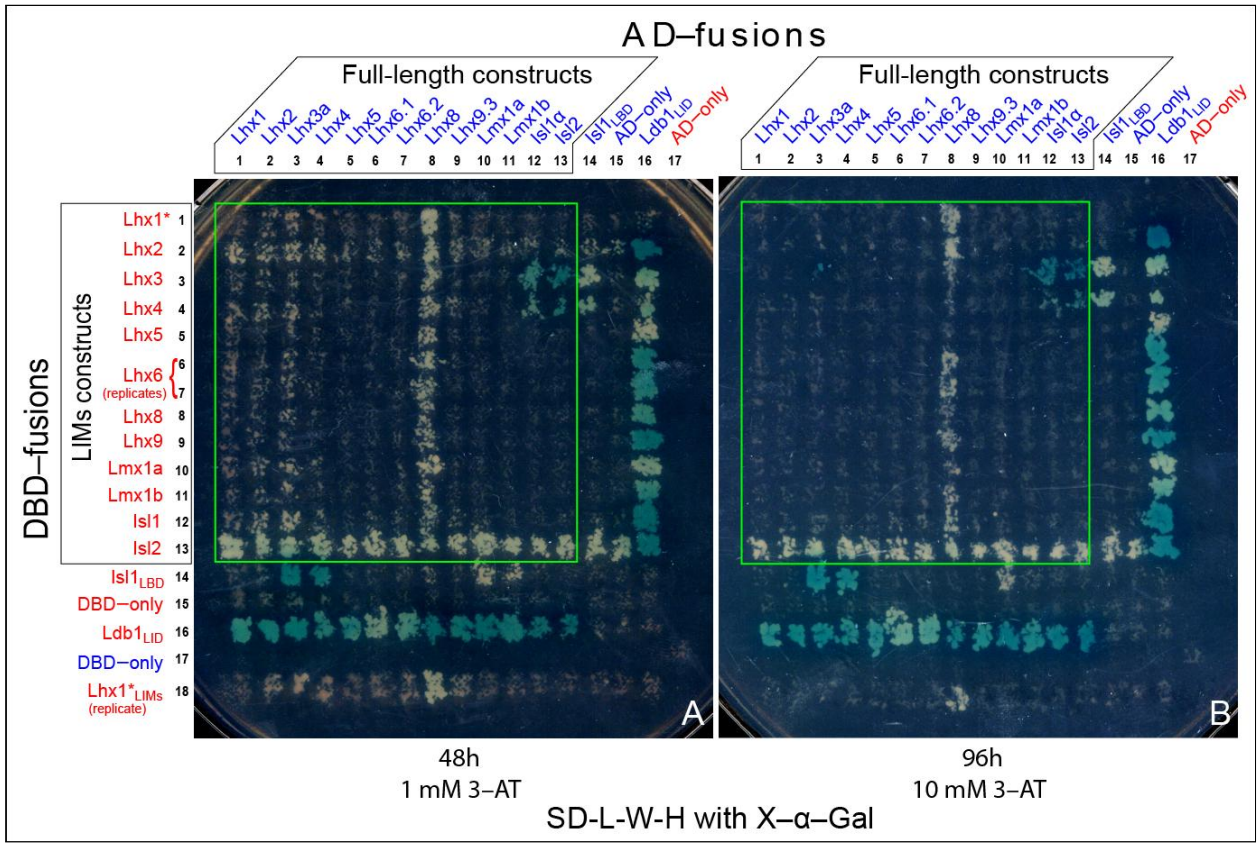


Figure 3.6 Results of a cross-mating Y2H array. A SD-L-W diploid selection plate was replica plated onto SD-L-W-H with X-α-Gal and with either 1 mM (A) or 10 mM 3-AT (B). DBD-fusion constructs (horizontal) were tested against AD-fusion constructs (vertical). AH109 transformants are shown in red and Y187 transformants are shown in blue. $Ldb1_{LID}$ and $Isl1_{LBD}$ were used as positive controls in both combinations. DBD-fusion constructs of $Lhx1_{LIMs}$ and $Lhx6_{LIMs}$ were tested in duplicate. The 1 mM 3-AT moderate selection plate (A) is shown at day 2 and the 10 mM 3-AT moderate selection plate (B) is shown at day 4. Red and blue font colours differentiate the two mating strains used. AD-full-length and DBD-LIMs fusion constructs of LIM-HD proteins are shown in black boxes; Yeast spots that correspond to interactions between these constructs are shown in green boxes. The $Lhx1^*_{LIMs}$ construct extends only as far as the last Zn-coordinating residue and was replaced in later assays by a longer construct.

To assess reproducibility within an experiment for the streak mating experiments (Figure 3.6) duplicates of DBD-Lhx1$_{LIMs}$ and DBD-Lhx6$_{LIMs}$ expressing yeast were generated from separate colonies from the same transformation plates. Both pairs of replicates displayed identical results.

3.6.2 LIM-HD vs. Ldb1$_{LID}$ interactions

As noted above, evidence for interactions were observed between all full-length AD-LIM-HD proteins and DBD-Ldb1$_{LID}$, and all DBD-LIMs constructs and AD-Ldb1$_{LID}$. It was noted that despite extending the DBD-Lhx1$_{LIMs}$ construct, both Lhx1 and Lhx5 appeared to interact less strongly that the other LIM-HD proteins, as evidenced by weaker growth and/or slower colour development under some conditions, and variability between experiments, which are characteristics of weak interactions analysed by Y2H as noted above (section 3.3). The pattern of apparent interaction intensities (based on colour development) in spots with DBD-LIMs fusion constructs did not exactly reflect the apparent interaction intensities in spots with respective AD-full-length fusion constructs (Figures 3.6–3.8). Lhx1 and Lhx5, whose DBD-LIMs fusion constructs gave the weakest activation, resulted in strong activation when AD-full-length fusion constructs were tested. A similar effect was observed for Lhx3a/4, Lmx1a/b and Lhx8. This resembles results from Y2H spot-test experiments with LIMs vs. LID/LBD interactions from our laboratory, where interactions with DBD-LIMs vs. AD-LID/LBD gave weaker reporter activation than with the AD-LIMs vs. DBD-LID/LBD combination [18, 19]. The opposite effect was observed for Lhx2, the DBD-LIMs fusion of which interacted very strongly but the AD-Lhx2 fusion of which did not produce such high levels of reporter activation. Also, AD-fusion constructs of Lhx6 isoforms produced noticeably lower activation of *MEL1* compared to DBD-Lhx6$_{LIMs}$.

3.6.3 LIM-HD vs. LIM-HD interactions

All previously identified interactions between AD-Isl1α/Isl2 full-length and DBD-Lhx3/4$_{LIMs}$ fusion constructs and between the DBD-Isl1$_{LBD}$ and AD-Lhx3/4$_{LIMs}$ fusion constructs, were detected in these mating assays. However, AD-Isl2 gave rise to a more intense

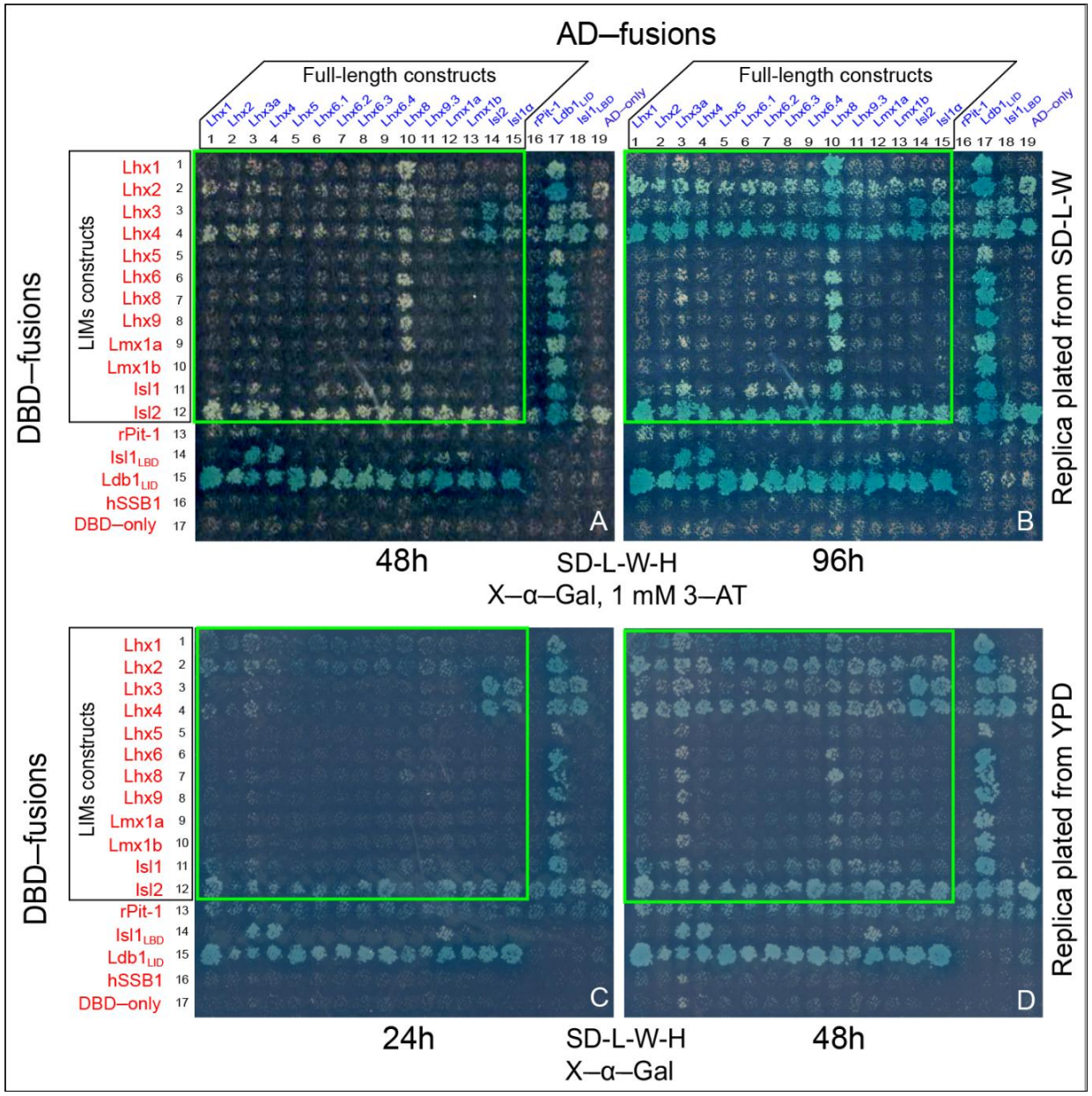


Figure 3.7 Results of a spot-mating Y2H array. Standard protocol (A and B): The YPD plate with spot-mated yeast spots was first replica plated onto the SD-L-W plate. After 24 h at 30 °C, the spots were replica plated on to the SD-L-W-H plate with X-α-Gal and 1 mM 3-AT. The figure shows scans after 48 h (A) and 96 h (B). Short protocol (C and D): The YPD plate with spot-mated yeast was replica plated directly on to the weak selection SD-L-W-H plate with X-α-Gal, The figure shows scans after 24 h (C) and 48 h (D). Red and blue font colours differentiate the two mating strains used. AD-LIM-HD and DBD-LIM-HD$_{LIMs}$ constructs are shown in black boxes; Yeast spots that correspond to interactions between these constructs are shown in green boxes.

activation in the interaction assays than AD-Isl1α, with the former showing increased levels of blue colour under weaker selection conditions plus growth under double selection conditions for both DBD-Lhx3$_{LIMs}$ and DBD-Lhx4$_{LIMs}$, compared to less colour under weaker selection conditions and no growth under double selection conditions for AD-Isl1α. In contrast, the interactions involving the isolated Isl1$_{LBD}$ (AD-Isl1$_{LBD}$) showed good levels of colour and robust growth under the same conditions (Figures 3.6–3.8). It is possible that the full-length Isl1α protein interacts more weakly with Lhx3/4$_{LIMs}$ compared to the isolated LBD domain due to intramolecular contacts between the LIMs and LDB in Isl1α that could compete for intermolecular binding events [162]. On the other hand, Isl2 has a slightly different LBD that might not form weak intramolecular interactions. As was the case with Ldb1$_{LID}$ interactions, AD-Lhx3a/4 vs. DBD-Isl1$_{LBD}$ were apparently stronger than interactions of DBD-Lhx3/4$_{LIMs}$ with AD-Isl1$_{LBD}$.

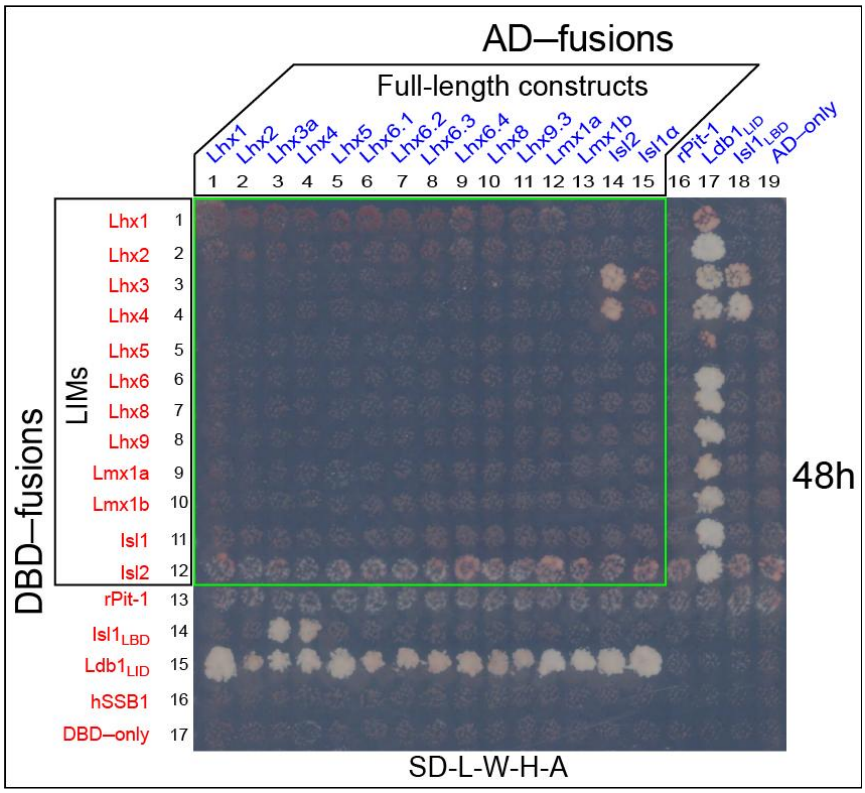


Figure 3.8. Results of a spot-mating Y2H array on a double selection plate (short protocol). The YPD plate with spot-mated yeast spots was replica plated directly on to the double selection SD-L-W-H-A plate. The plate was scanned after 48 h. Red and blue font colours differentiate the two mating strains used. AD-full-length and DBD-LIMs constructs of LIM-HD proteins are shown in boxes; Yeast spots that correspond to interactions between these constructs are shown in the green box.

DBD-Lmx1a$_{LIMs}$ and DBD-Lmx1b$_{LIMs}$ showed some evidence of an interaction with AD-Isl1$_{LBD}$, with limited growth, but no blue colour being evident on the 1 mM 3-AT moderate selection plate. The DBD-Lmx1a$_{LIMs}$/AD-Isl1α spot was also evident on the 10 mM 3-AT moderate selection plate in the streak-mating array. Neither set of diploid transformants grew under double selection conditions.

AD-Lhx8 appeared to interact with all DBD-LIMs constructs with weak to moderate intensity (but not with other constructs), as evidenced by some growth but no colour development on both plates. In contrast, AD-Lhx8$_{LIMs}$ did not show evidence of interaction with other full-length LIM-HD proteins.

Pit1 did not show evidence of an interaction with any construct in the arrays. This was later followed up by additional spot tests with Lhx3a and Lhx3$_{LIMs}$ (not shown) and although weak selection conditions allowed some growth in the interaction between Lhx3$_{LIMs}$ and POU-domains of Pit1, this interaction could not be confirmed.

3.7 A 96-well Y2H mating array

Towards the end of this project, another version of the Y2H mating array was assessed that used a 96-well plate format with liquid and agar-based media [210]. This method avoids both smearing of the template and target plates, and loss of resolution from successive replica plating steps.

In this 96-well mating protocol, singly-transformed strains were grown overnight at 30 °C/ 220 rpm in the appropriate media, as in the previous mating assays. Then 96-well plates containing rich liquid media (YPD) were co-inoculated with 2-µL aliquots from appropriate pairs of overnight cultures, and left to mate for up to 16 h at 30 °C/ 220 rpm. Replica plating was performed by dipping a set of EasyLoad 200-µL tips into the plates containing the liquid rich media culture, and inoculating a new 96-well plate containing liquid SD-L-W media for selection of diploid yeast. After overnight incubation, the yeast was replica plated using the same method onto 96-well plates containing 0.5 mM 3-AT moderate selection media (agar). The plate was scanned every 24 h for 4 days to monitor yeast growth and blue colour development.

In this mating array, in addition to the LIMs and full-length LIM-HD constructs, the four mouse LMO proteins were also included as DBD-fusion constructs and tested against the AD-full-length LIM-HD constructs. As each yeast transformant is grown in a separate well of the 96-well plate, there are no opportunities for cross-contamination during transfer, fusion of neighboring spots, or diffusion of blue pigment over time. These properties can allow the successful scoring of different intensity interactions on the same plate over a longer time period (e.g., up to 5–7 instead of 1–2 days). The use of multiple sets of 96-well plates also presents an advantage over the spatial limitation of 20 × 20 arrays in a single Petri dish.

Figure 3.9 shows the growth and colour of yeast in the wells after 4 days, estimated to best represent the relative differences in signal intensities from different wells. Overall these interaction data are very similar to the plate method, with the following exceptions:

1) Previously detected false positives (such as the DBD-Lhx2$_{LIMs}$ with all LIM-HD proteins in Figure 3.9) were difficult to confirm here due to lack of growth in the controls, likely because of inefficient yeast transfer.

2) The AD-only negative controls for DBD-Lhx4$_{LIMs}$ that grew on the plates did not appear here. It is possible that there were inefficiencies or inconsistencies associated with simple manual transfer method to the selection plates. These possible issues might be overcome in the future by using a robotics approach with measured drop sizes.

3) DBD-Lmx1b$_{LIMs}$ showed evidence of an interaction with both full-length AD-Isl1α/Is12 protein constructs while the DBD-Lmx1a$_{LIMs}$ did not, whereas the Lmx1a interactions were apparently stronger in the cross-mating and spot-mating methods. This may be a problem with drop size variability, or may simply be indicative of weak interactions for both Lmx1a and Lmx1b with Isl1α and Isl2.

4) AD-Lhx6 showed a weak propensity to interact when tested against some LIMs constructs, as was seen previously in the plate assays only for AD-Lhx8. AD-Lhx8 showed evidence of an interaction with many LIMs constructs, with apparently stronger interactions with the LMO proteins that were additionally tested in this approach. Note that these preliminary data need to be retested using precise drop sizes before they could be considered significant.

5) Manual replica plating likely resulted in the lack of growth in some controls, which made some interactions/false positives difficult to confirm on the basis of this assay alone.

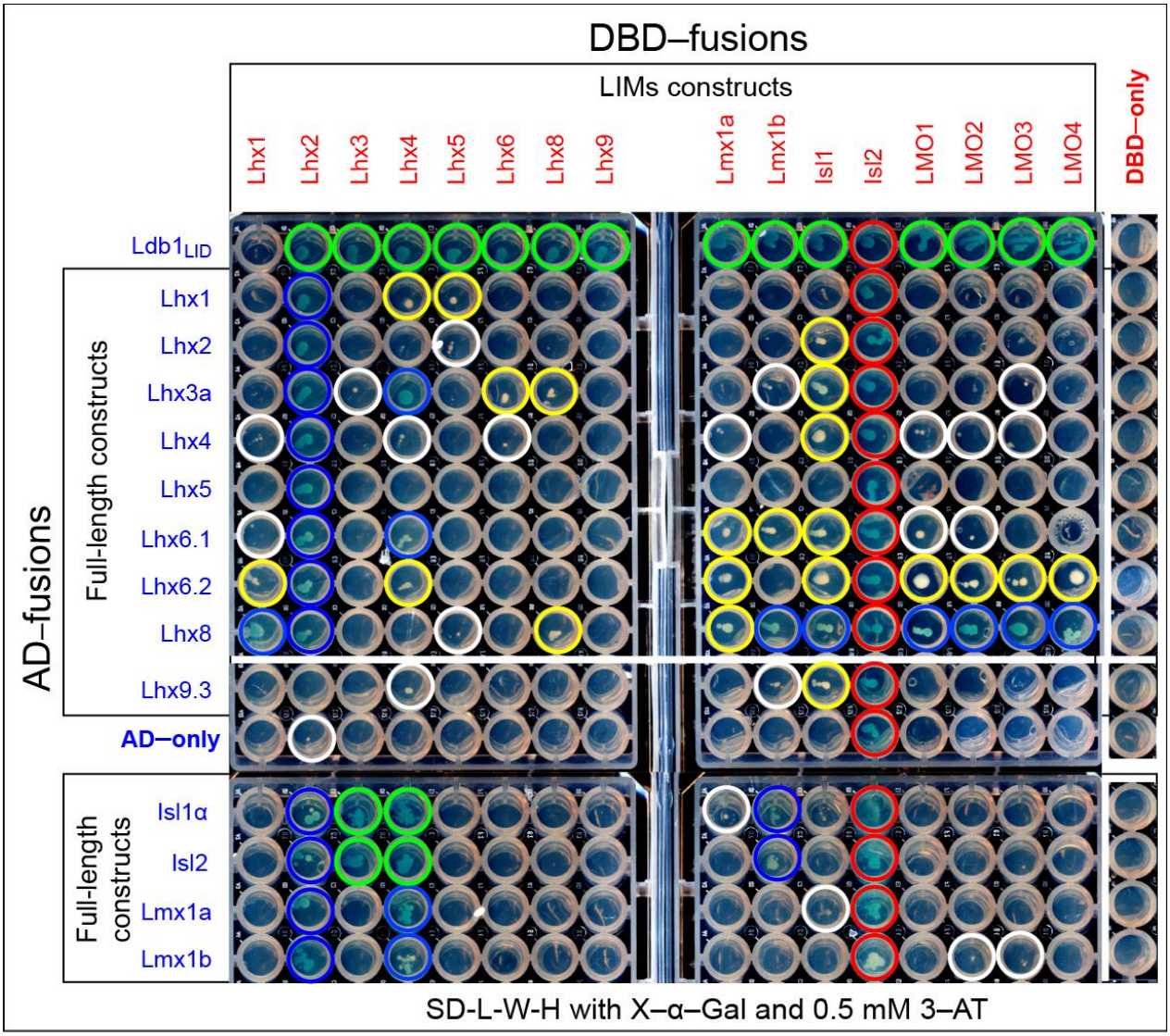


Figure 3.9. Results of the 96-well plate-based Y2H mating array. The figure shows AD-fusion constructs in rows interacting with DBD-fusion constructs in columns of the plate. The 96-pipette tip set was dipped in the 96-well plate with overnight mated yeast cultures growing in SD-L-W media. The tips were gently brought in contact with the agar in the 96-well 0.5 mM 3-AT moderate selection plate. The agar plate was incubated at 30 ºC for 4 days and then scanned. The figure was pieced together from two plates made with the same media and incubated together and white lines show where those pieces were connected. Green circles indicate known interactions detected in previous studies that were confirmed in this assay. Red circles indicate false positives detected in this plate (significant growth in AD-only control). White, yellow and blue circles indicate very weak, colourless and coloured yeast spots respectively.

3.8 Discussion

The experiments presented in this chapter show that among the mammalian LIM-HD proteins, high affinity interactions appear to be mediated only between the LBD domains from Isl1 and Isl2 with the LIM domains from Lhx3 and Lhx4. In addition to Lmx1a, Lmx1b protein was also found to exhibit some Isl_{LBD}-binding activity, but these interactions were apparently of much weaker affinity. These interactions are summarised and the likely biological relevance of interactions between LIM domain proteins is discussed in sections 8.2 and 8.3. The discussion in this section is restricted to technical details of Y2H assays.

Overall Y2H mating arrays proved to be a useful technique for performing larger scale experiments and this method could be further improved by the use of robotics and 96–768-well plates to standardize experiments, prevent manual errors, reduce assay-to-assay variability, and allow medium-to-high throughput experiments [210, 221].

Slight differences between results from the yeast mating arrays and Y2H spot-tests could result from differences in expression profiles or different properties between cells transferred from a colony compared with cells spotted from an overnight culture, or differences in diploid versus haploid states of Y2H transformants. For example, diploid yeast are said to be more vigorous than haploid strains and can better tolerate expression of toxic proteins [222]. A previous report suggested that yeast mating reduces background caused by autoactivation [203]. Only AH109 carries the auxotrophic HIS3 and ADE2 reporters, so diploid nuclei contain twice as much DNA per copy of these reporters, reducing the likelihood of their activation. In contrast the MEL1 reporter is present in both strains and should have resulted in stronger activation in diploid cells.

3.8.1 Key limitations of Y2H assays for detecting LIM-HD interactions

The restriction of possible interactions to the LIM domains of the LIM-HD proteins was necessary to minimize background growth in the assay.; This prevented detection of interactions that are not mediated by LIM domains. It should be noted that the Y2H system holds several limitations that could have prevented detection of LIM-HD vs. LIM-HD interactions including:

the sensitivity of the assay, stability of heterogenous proteins in yeast and non-native posttranslational modifications.

It has been reported that Y2H is not sensitive enough to detect interactions weaker than $K_d \approx 10^{-6}$ M [222, 223]. However, the ability of a reconstituted Y2H construct to activate *HIS3*, *MEL1* or the *ADE2* reporter might not be directly related to K_d values of tested protein pairs [223], and there are some reports of interactions with K_d values close to 10^{-4} M being detected by the Gal4 Y2H system [224, 225]. The ADH1-derived promoter used in the vectors used to date drives very low constitutive expression, so it might be possible to enhance the detection of weak interactions by increasing the copy number of the plasmids used and/or use promoters that give higher levels of expression of fusion proteins. Such an approach would probably simultaneously increase autoactivation, which was already common in this system, but if higher levels of 3-AT were used to compensate for autoactivation, the remaining *HIS3* activity could conceivably still be higher and more easily detected.

The ability to detect an interaction will likely also depend on the stability in yeast of the proteins being tested and steric hindrance resulting from fusion constructs to the Gal4 domains. For example, intrinsically disordered regions of proteins can give rise to false negatives, possibly caused by proteolytic degradation of these constructs. Stabilization of these constructs by additional structured elements can sometimes lead to higher stability and allow detection [163, 164, 226]. This type of approach might be useful for testing interactions involving the C-terminal domains of LIM-HD proteins.

Steric hindrance could prevent the formation of functional Y2H complexes. In some of these cases, interactions might be detected by finding the optimal combination(s) of fusion constructs. There are alternative Y2H plasmids that allow DBD-/AD-fragments to be fused to C-termini of tested proteins so that 6 additional combinations can be tested [227, 228]. An example of this type of optimization is described in the context of a modified assay in Chapter 5.

Budding yeasts, including the strains used for Y2H, lack complete families of important tyrosine kinase genes present in vertebrates and therefore have a limited ability to phosphorylate tyrosines [229, 230]. SH2 domains that recognize phospho-tyrosines (present in some Ub-protein ligases) are also absent, possibly preventing further modifications [231, 232]. If tyrosine phosphorylation is required for an interaction, that interaction will not be detected by Y2H. In

other cases, the lack of appropriate posttranslational modifications could affect protein lifetimes and effective concentrations in yeast cells. At present, there is no data to suggest that posttranslational modifications affect interactions of LIM-HD proteins in vivo.

In an ideal Y2H system, an increase in binding affinity of tested proteins would result in a corresponding increase in reporter activation (Figure 3.10). Unfortunately, this is often not the case [222, 223], possibly in part due to autoactivation and the attempts to reduce it. However, some classes of interactions could be quantified through the activation of colourimetric LacZ/MEL1 reporters or fluorescent yEGFP reporters in well-plates combined with colourimetric scanning or flow cytometry [233-235]. This type of approach might provide a better way of quantifying LIM-HD interactions.

3.8.2 Autoactivation and false positive interactions in LIM-HD proteins

The autoactivation seen in Y2H assays with full-length LIM-HD proteins could arise from two main sources. The first source is the DNA-binding activity of the HDs. These domains target AT-rich sequences, more than 10 of which exist in the GAL1 promoter, suggesting that the HDs could directly target these sites to trigger autoactivation. The second source is the transactivation-like domains in the C-terminal regions (tails) of the LIM-HD proteins. Protein sequences likely to cause transactivation as DBD-fusion constructs in yeast are often acidic, acidic amphipathic helices, and/or are rich in proline or glutamine residues [236]. Tails of LIM-HD proteins are rich in prolines and four proteins contain glutamine clusters (Isl1α, Isl2, Lmx1a and Lmx1b) which could explain some of the observed effects. In most cases it has not been established if these regions have transcriptional transactivation activity in nature.

It should also be noted that proteins that have evolved as part of obligate complexes like LIM-HD proteins are often naturally 'sticky' and can cause false positives by non-specific interactions with many different targets, likely through exposed hydrophobic surfaces [237].

3.8.3 Possible explanations for differences in reporter activation between two combinations of fusion constructs (vectors)

As reported in this chapter, interactions between LIMs and LID/LBDs often appear stronger in one combination of yeast vectors. The same phenomenon has been reported for other systems (e.g., Myc and Max bHLH-LZs; [223]). A possible explanation extends a mechanism that was proposed to explain the difficulties in detection of parallel coiled coil homodimers by Y2H [238, 239]. $Gal4_{DBD}$ forms homodimers [240, 241], and homodimerization of that domain was thought to promote the specific homodimerization of the coiled-coil domains and thereby prevent interactions with AD-fusion partners, thereby lowering reporter activation (Figure 3.10) [238, 239]. In a similar manner, $Gal4_{DBD}$ homodimerisation could promote non-specific binding of LIMs and thereby prevent interaction with the AD-LIDs. As the 2µ plasmids that carry the TRP1 selection marker were reported to give higher expression of Y2H fusion constructs than those that carry the LEU2 marker [242], the concentration of DBD-LIMs could be higher than that of AD-LID, and that could increase the negative effects of $Gal4_{DBD}$ homodimerisation.

The structured DBD domain could also provide some protection of intrinsically disordered LIM domain partners from proteolysis, and increase the lifetimes of reconstituted Y2H complexes on DNA. In contrast monomeric AD-fusion constructs of LIM-partners which are likely to be unstructured are likely to be more accessible to proteolysis [164]. Thus, DBD-LID/LBDs vs. AD-LIMs interactions would appear stronger than in DBD-LIMs vs. AD-LID/LBDs interactions.

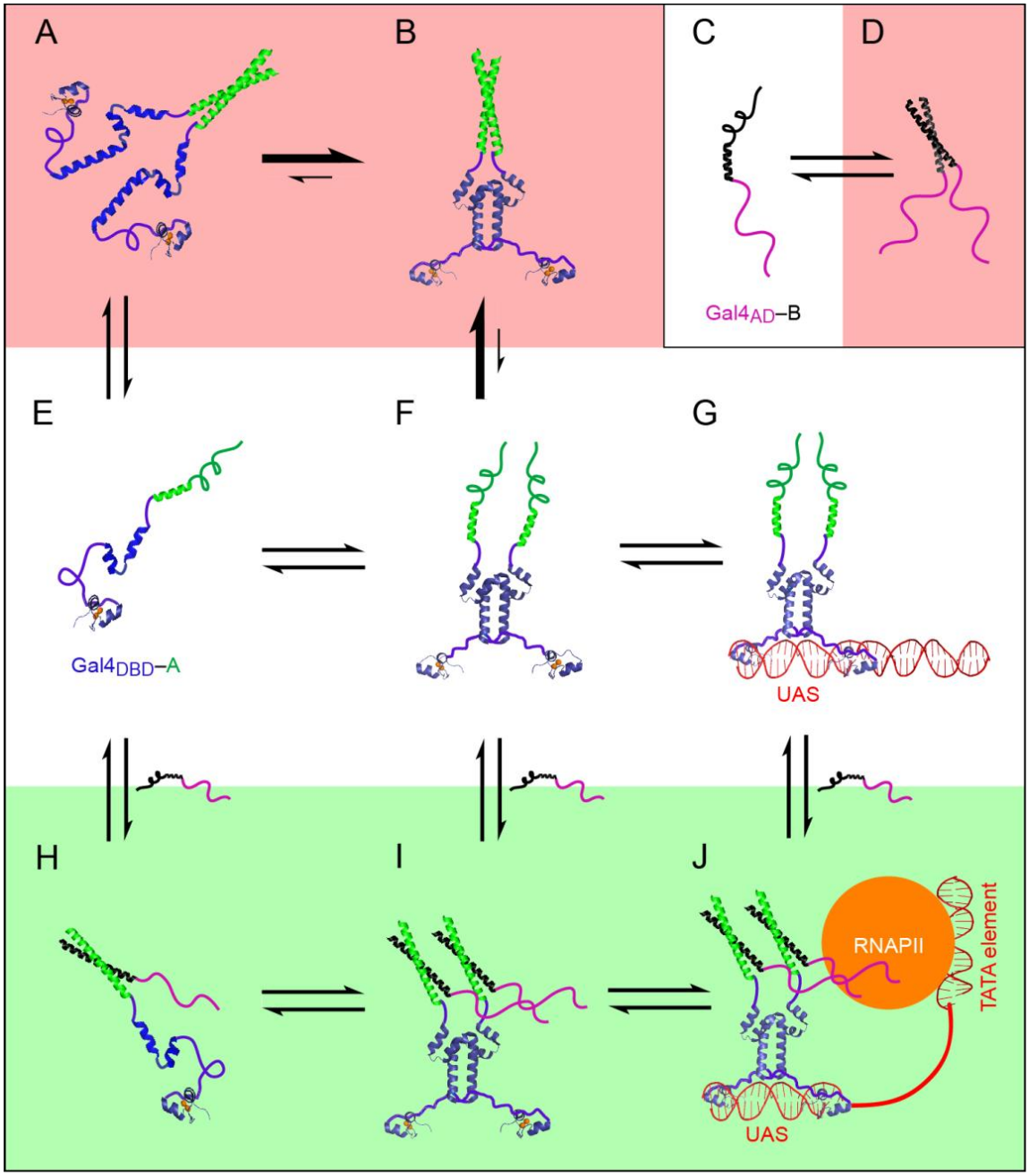


Figure 3.10. Homodimerization of hypothetical constructs in the Gal4 Y2H system. The example shows Y2H tests with hypothetical green and black proteins that can form homodimers (specific or unspecific) as well as heterodimers (green-black, green or black coiled coils). DBD-fusion constructs (E) can dimerize via DBD (F), dimerize via tested protein (A), or bind AD-fusion constructs to reconstitute the full Gal4 protein (H). These reconstituted Gal4 monomers can dimerize via DBD (I) and all complexes with dimerized DBD domains (I, F and B) can bind UAS elements but activation of RNAPII requires bound AD-fusion constructs (J). Reconstituted Gal4 complexes (free or bound to a UAS element) are shown in green fields (H, I and J), while fusion protein homodimers are shown in red (A, B and D). Chelate effect (cooperative binding) between DBD-fusion constructs would tip the equilibria towards a species that cannot cause activation (B). Figure is adapted from Hu *et al*., 2000 and extended to include all species and complexes.

Chapter 4 The contribution of individual LIM domains in binding to Ldb1

4.1 Introduction

It is generally considered that both of the tandem LIM domains of LIM-HD and LMO proteins are required for high affinity binding to Ldb1 (Figure 4.1), but in some cases either one LIM domain or the other has been shown to be more important for the interaction [131, 136]. It was decided to use the spot-mating Y2H array (as described in sections 2.5.4.2 and 3.4) to investigate modular binding of $Ldb1_{LID}$ for the whole panel of individual LIM domains from LIM-HD proteins to determine if there are patterns to the LIM-specific binding across this family of proteins.

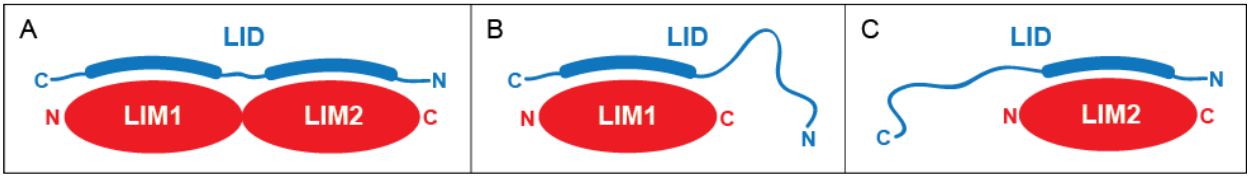


Figure 4.1. A scheme showing tandem (A) and separate LIM domains (B and C) binding the $Ldb1_{LID}$. LIM domains are represented as ovals and $Ldb1_{LID}$ (LID) as a ribbon. Binding segments of the LID peptide are represented by thicker segments of the ribbon. N- and C-terminal ends of each partner are labelled.

4.2 Assay design

Individual LIM domain (LIM1 or LIM2) constructs were designed to contain six flanking residues at either end of the first and last zinc-coordinating residues, meaning that the constructs overlap by 3 or 4 residues between LIM1 and LIM2 depending on the length of the intervening sequence (Figure 4.2). Previous experiments with separate LIM domains of LMO2, LMO4 and Lhx3 indicated that these intervening sequences help maintain structural integrity of separate domains but do not directly contribute to binding of Ldb1 [18, 156]. These constructs were generated by PCR and cloned as both DBD- and AD-fusion constructs. Interactions between tandem and the individual LIM domain constructs were tested for binding to $Ldb1_{LID}$ in both combinations (i.e., as DBD-fusion constructs as shown in Figure 4.3, and AD-fusion constructs as shown in Figure 4.4). In addition, AD-$Isl1_{LBD}$ and AD-Isl2 were tested for interaction with the

LIM domain constructs from Lhx3 and Lhx4 as shown in Figure 4.4. Three levels of selection conditions were used to assess the interactions: weak, moderate (with 1 mM 3-AT) and double selection (as in Table 3.1). However, only moderate and double selection conditions were used in the DBD-LIM/AD-(LID/LBD) arrays. Weak selection conditions were not used for this combination because autoactivation tests (section 3.2) showed that DBD-LIMs fusion constructs gave rise to background growth under weak selection conditions.

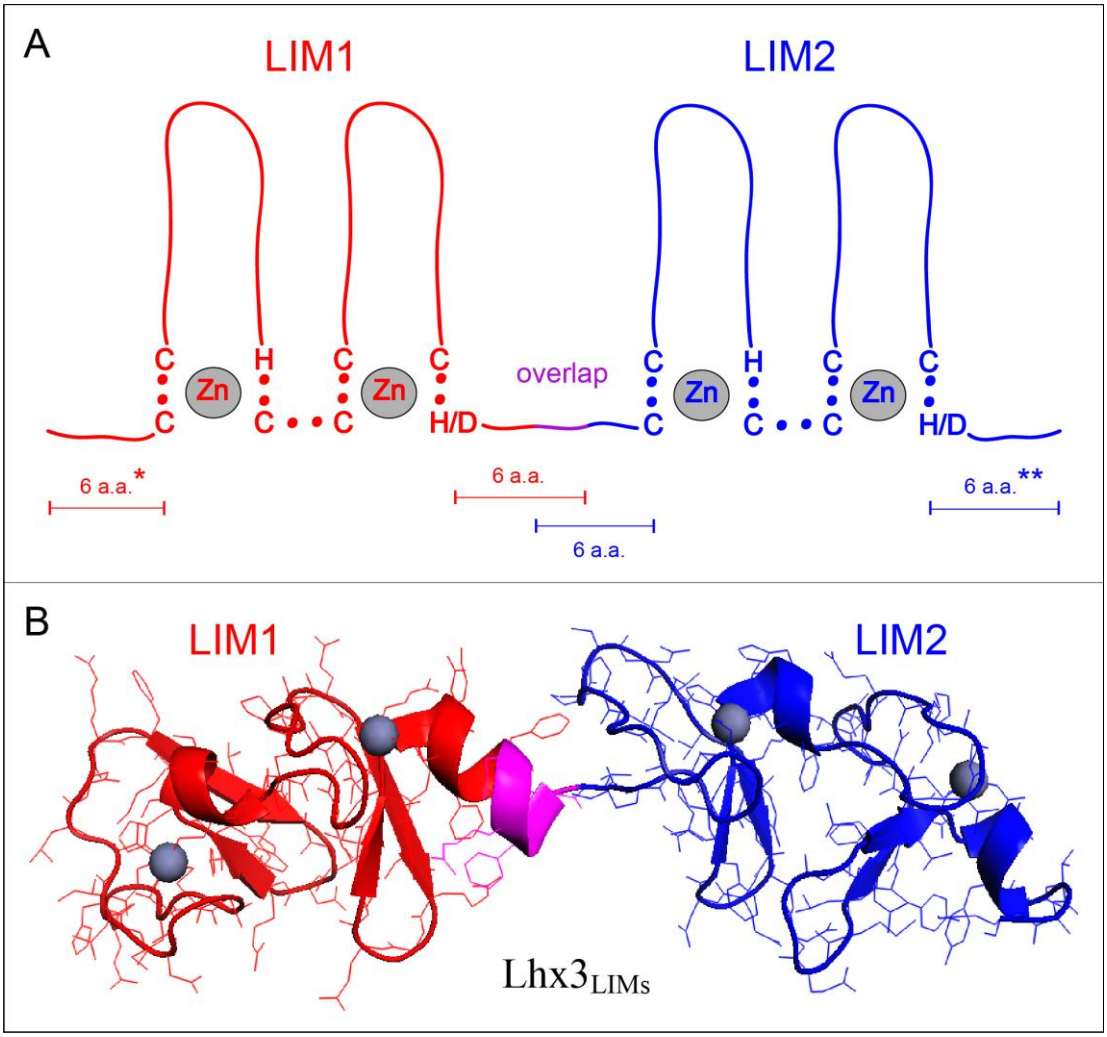


Figure 4.2. Design of individual LIM domain constructs. (A) The LIM1 and LIM2 domains are shown in red and blue, respectively. These constructs were designed to contain six residues before the first and after the last Zn-ligating residues, with the exceptions that the N-terminal 'extension' is (*) 4 residues in Lhx5, 3 residues in Lhx1 and 2 residues in Lmx1a LIM1 and LIMs constructs, and there were no residues (**) beyond the last Zn-ligating residue in Lhx1 LIM2 and LIMs constructs. The overlap between separate LIM constructs was 3–4 residues, depending on the spacing between the LIM domains in each protein. (B) The LIM domains from the structure of $Lhx3_{LIMs}$-$Isl1_{LBD}$ (2RGT) coloured as in panel A.

4.3 Comparison of relative strengths of LIM1 and LIM2 interactions with LID/LBD domains

DBD-LIM vs. AD-LID combination. The spot-mated yeast on selection plates were initially scanned after three days of incubation. At this time the moderate selection plate showed evidence of interactions of $Ldb1_{LID}$ with all tandem LIM constructs, but not with any individual LIM domain (Figure 4.3A). The yeast growth seen for the $Lhx2_{LIM1}$ construct in the interaction spot control matched that of the empty vector and was disregarded. Some background growth was also detected for the $Isl2_{LIMs}$ construct but reporter activation in the interaction spot was much stronger. A longer incubation time resulted in the loss of resolution on this plate and no additional interactions could be detected (not shown). On the double selection plate, only a subset of spots were visible after 3 days, corresponding to interactions of $Lhx2_{LIMS}$, $Lhx6_{LIMs}$, $Lhx8_{LIMs}$, $Lhx9_{LIMs}$, $Lmx1b_{LIMs}$, $Isl1_{LIMs}$ and $Isl2_{LIMs}$ with $Ldb1_{LID}$ (Figure 4.3B). These double selection plates were scanned again after 6 days of incubation, when spots corresponding to interactions of $Lhx3_{LIMs}$, $Lhx4_{LIMs}$, $Lmx1a_{LIMs}$ with $Ldb1_{LID}$, appeared. After 14 days of incubation, the spots for $Lhx1_{LIMs}$ and $Lhx5_{LIMs}$ also appeared, as well as that for $Isl2_{LIMs}$ background in the negative control. The strongest growing and the earliest appearing spots were considered the strongest interactors.

Interactions of $Lhx3_{LIMs}$ and $Lhx4_{LIMs}$ with both $Isl1_{LBD}$ and Isl2 were detected on the moderate selection plate after 3 days of incubation (Figure 4.4). On the double selection plate, both $Lhx3_{LIMs}$ and $Lhx4_{LIMs}$ interacted with Isl2 while only $Lhx3_{LIMs}$ interacted with $Isl1_{LBD}$ and resulted in yeast growth after 6 days. A few colonies appeared for the $Isl1_{LBD}$ interaction with $Lhx3_{LIMs}$, but only after 14 days of incubation. The individual LIM domain interactions were not detected on any plate.

AD-LIM vs. DBD-LID combination. Under double selection conditions, yeast growth for all AD-LIMs fusion constructs vs. DBD-$Ldb1_{LID}$ appeared by day 2 of incubation and developed into strong spots by day 5 (Figure 4.5, bottom panel). Of the tests with separate LIM domains, only the LIM1 domains of Isl1 and Isl2 showed a robust interaction with $Ldb1_{LID}$ on the double selection plate, while the interaction with $Lhx2_{LIM2}$ was lower and with $Lhx9_{LIM2}$ produced only a few colonies. The AD-$Lhx3_{LIMs}$/$Lhx4_{LIMs}$ vs. DBD-$Isl1_{LBD}$ interactions showed several

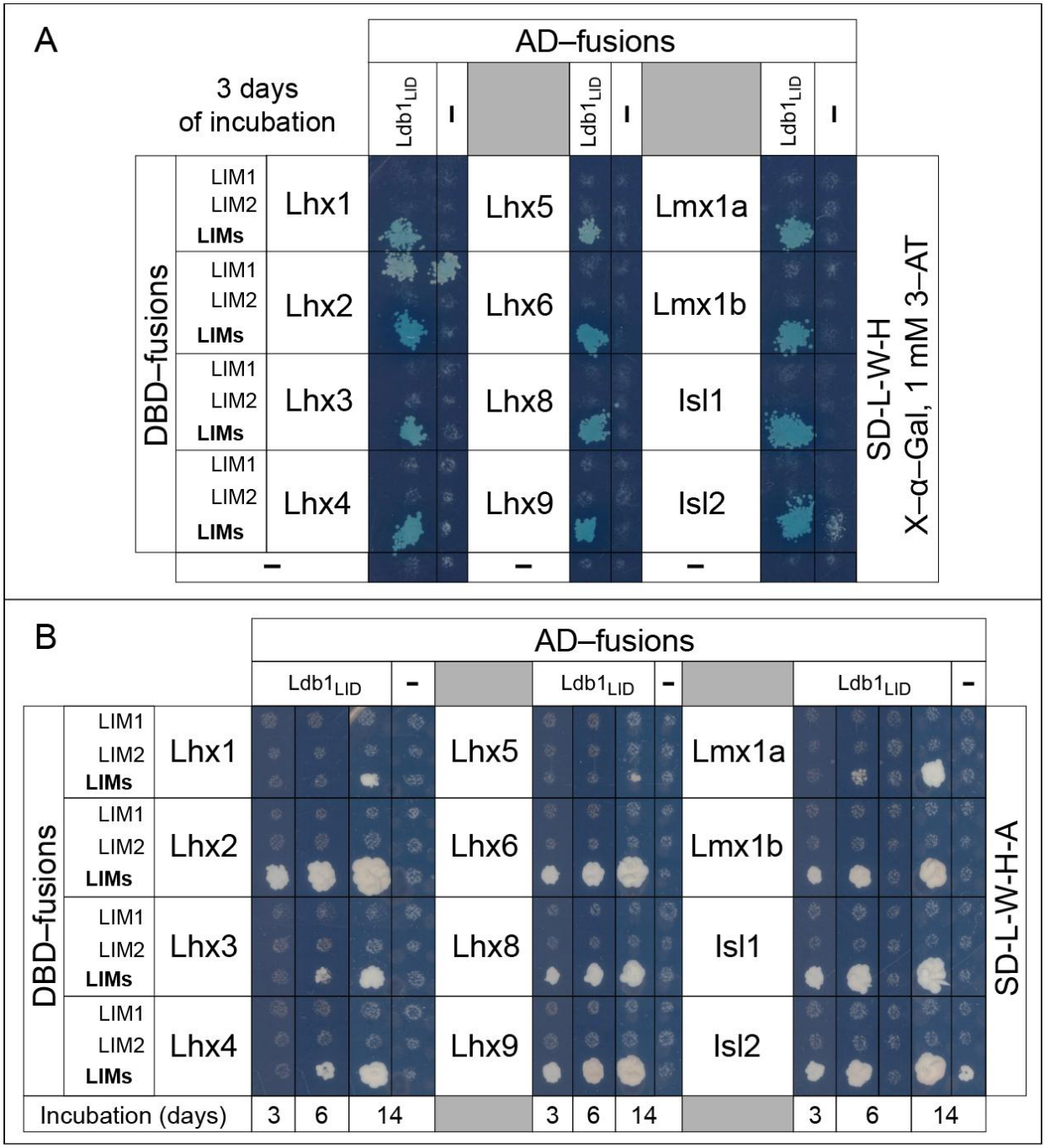


Figure 4.3. Y2H spot-mating array for DBD-LIM vs. AD-LID. All LIM-domains were expressed as DBD-fusion constructs and tested for interaction against AD-$Ldb1_{LID}$. The figure compares the Y2H mating arrays on moderate (A) and double (B) selection media and displays the chronological appearance of signals on the double selection plate. (C) DBD-fusion constructs of the Lhx3 and Lhx4 domains were also tested for interaction against AD-$Isl1_{LBD}$ and AD-Isl2 on both selection plates.

colonies by day 2, which developed into spots by day 5. Under moderate selection conditions (Figure 4.5, top panel), all interactions that later appeared on double selection plates were visible after overnight incubation. Later, interactions between $Ldb1_{LID}$ and $Lhx2/9_{LIM2}$ also appeared, followed by interactions of $Ldb1_{LID}$ with $Lmx1a/b_{LIM2}$ and $Lhx6/8_{LIM2}$. Some growth was also detected for $Lhx2_{LIM1}$ and only a few colonies for interactions between $Lhx4_{LIMs}$ and $Ldb1_{LID}$ or $Isl1_{LBD}$.

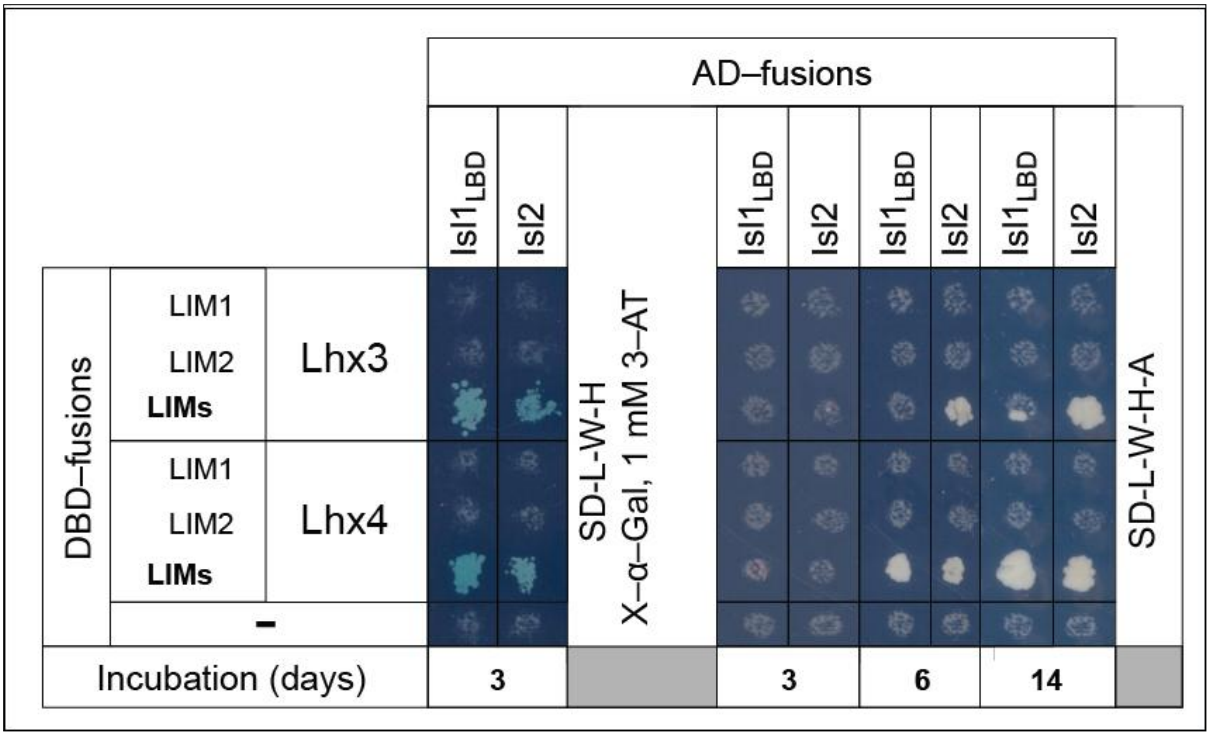


Figure 4.4. Y2H spot-mating array for DBD-LIM vs. AD-LID. DBD-fusion constructs of the Lhx3 and Lhx4 domains were additionally tested for interaction against AD-$Isl1_{LBD}$ and AD-Isl2 on moderate and double selection plates.

Under weak selection conditions (Figure 4.6), after 2 days of incubation, the same pattern of growth was observed as on the moderate selection plate after 5 days, but with an additional yeast spot detected for the $Lhx3_{LIM2}$ interaction with Ldb1. Under this selection condition reporter activation for the $Lhx3_{LIMs}$ interactions were similar to that of $Lhx4_{LIMs}$. Again, none of the separate LIM domains from Lhx1 and Lhx5 showed strong evidence of an interaction under any conditions at day 2. Longer incubation (5 days) resulted in some background growth in all spots expressing the DBD-$Ldb1_{LID}$ fusion construct and in the appropriate negative control as well.

Minor evidence of an interaction was detected for $Ldb1_{LID}$ with $Lhx1_{LIM2}$ due to some colour development. An interaction between $Lhx3_{LIM2}$ and $Isl1_{LBD}$ was indicated by weak growth compared to no growth in the DBD-$Isl1_{LBD}$ negative control.

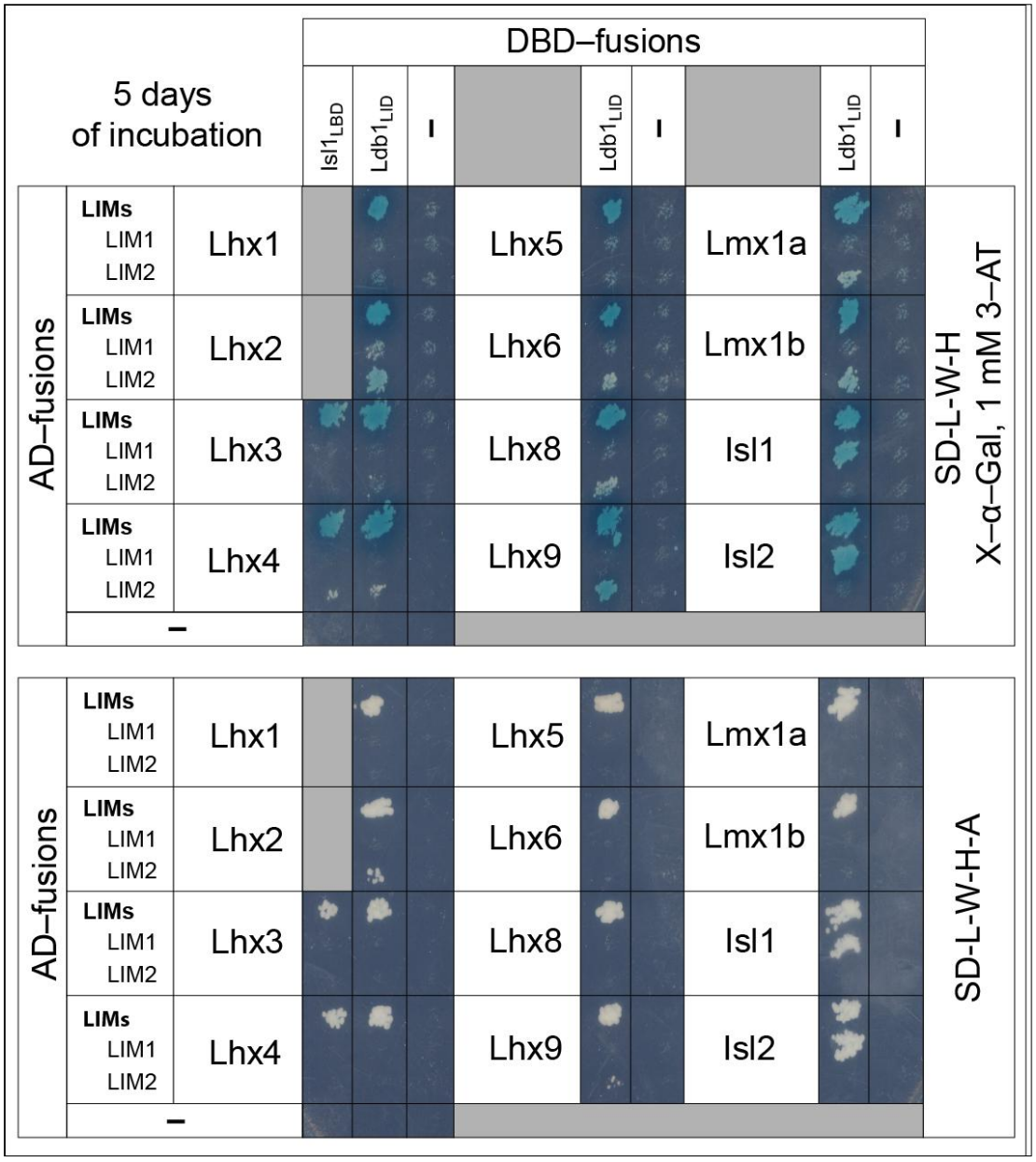


Figure 4.5. Y2H spot-mating array for AD-LIM vs. DBD-LID. All LIM-domains were expressed as AD-fusion constructs and tested for interaction against DBD-$Ldb1_{LID}$. The LIM domains from Lhx3 and Lhx4 were also tested for interaction against DBD-$Isl1_{LBD}$. DBD-only and AD-only controls were included ('-'). The top panel shows the moderate selection plate and the bottom panel shows the double selection plate after 5 days of incubation at 30 °C.

Arrays with AD-LIM vs. DBD-(LID/LBD) combination of fusion constructs showed greater sensitivity than the opposite (DBD-LIM vs. AD-(LID/LBD)) combination of fusion constructs. This observation is in line with previous experiments in our laboratory [18, 19], and some results from the previous chapter (section 3.7.2). This increased sensitivity revealed a binding preference of Ldb1 and Isl1/2 for individual LIM domains. The negative controls were free of significant background growth, except on weak selection plates after 5 days of incubation.

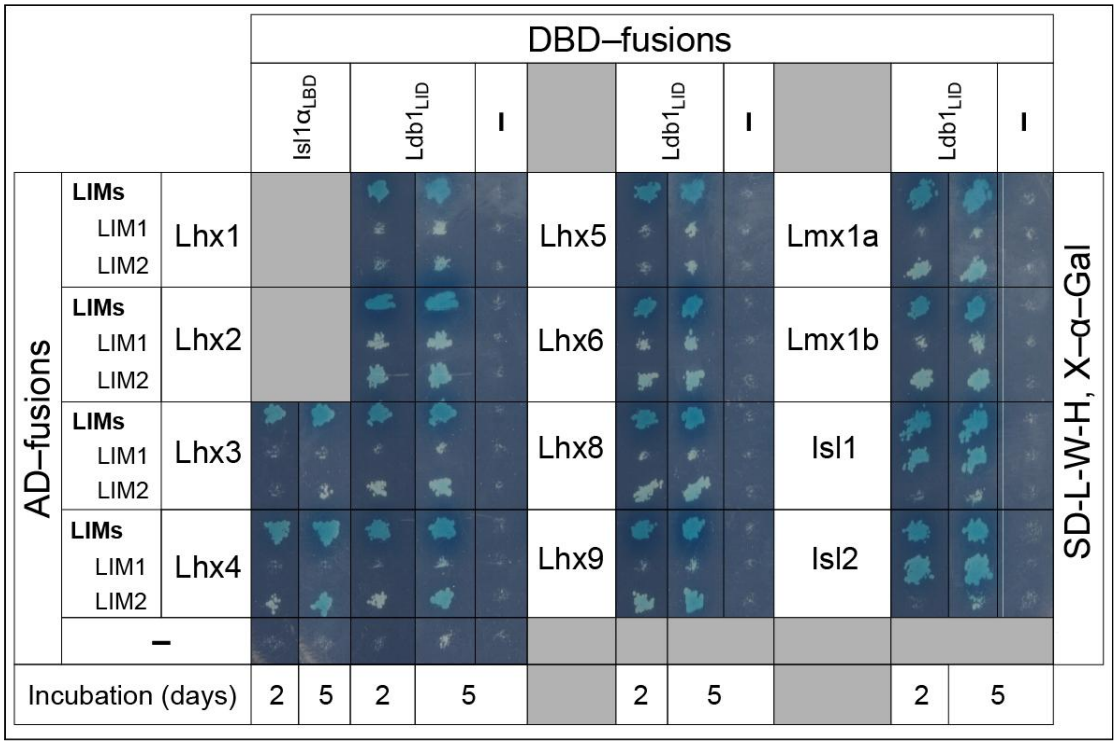


Figure 4.6. Signal development under weak selection conditions for AD-LIM vs. DBD-LID in Y2H spot-mating arrays. The figure shows the spot-mating array of tandem and separate LIM domain AD-fusion constructs vs. the DBD-Ldb1$_{LID}$ fusion construct. The LIM domains from Lhx3 and Lhx4 were also tested for interaction against DBD-Isl1$_{LBD}$ and included in the figure. Results from day 3 and day 5 of incubation at 30 °C are shown. DBD-only and AD-only controls were included ('-').

Y2H spot-tests using the mated yeast diploids were carried out (without serial dilutions) to ensure equal loading of yeast cells and better gauge the relative apparent binding affinities. The selection data are largely identical, but in this spot-test form are easier to visualise (Figure 4.7). Apparent interaction intensities were classified according to growth rates and colour development in spots on different selection plates. All tandem LIM vs. Ldb1$_{LID}$ interactions and

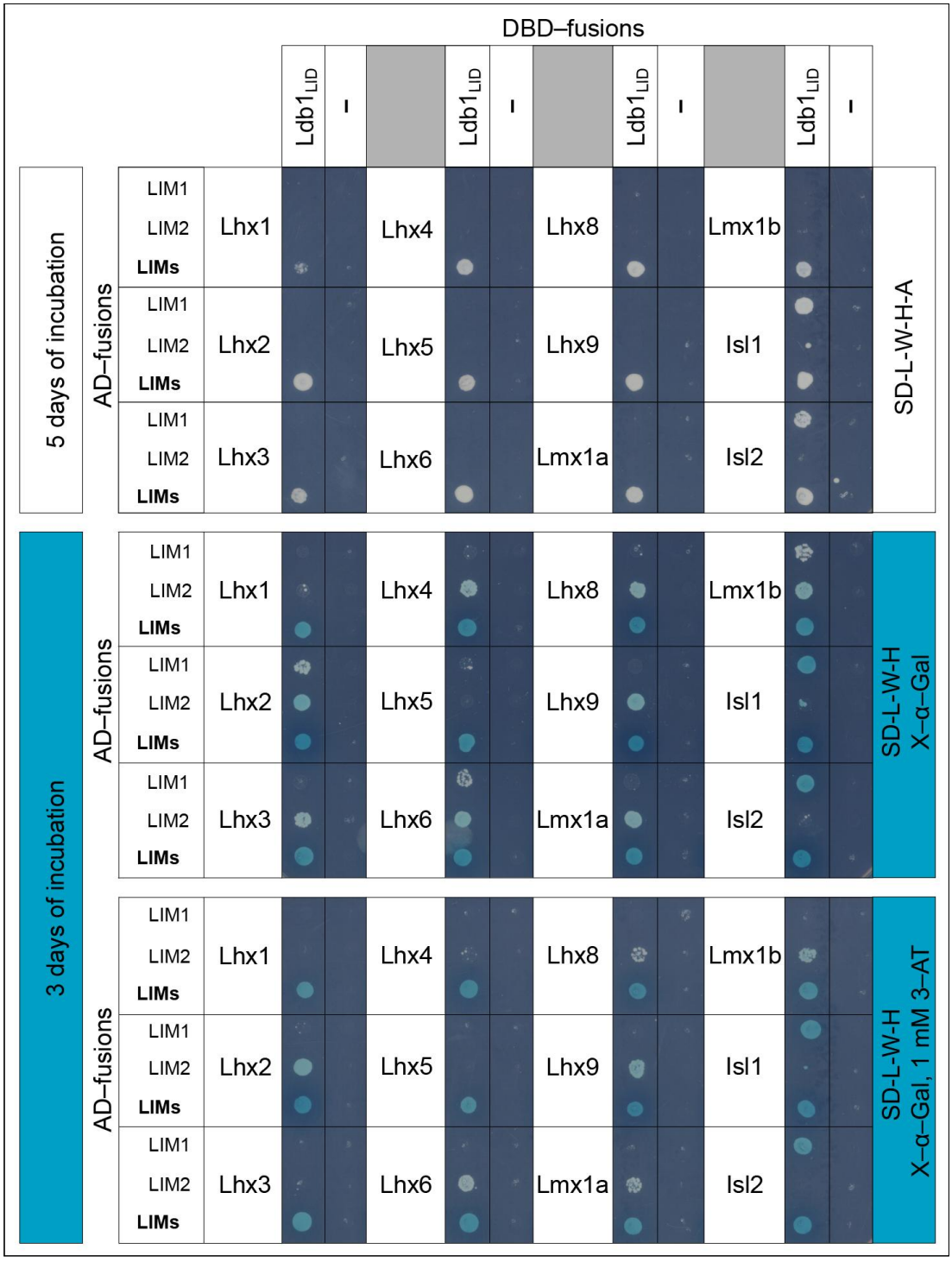


Figure 4.7. Y2H spot-tests of tandem and separate AD-LIM fusion constructs vs. the DBD-Ldb1$_{LID}$ in diploid cells. Yeast growth on double selection plates on day 5 of incubation (top panel) and yeast spots on low and moderate selection plates, on the third day of incubation (middle and bottom panel). '-' indicates DBD domains expressed from empty vectors. The yeast cultures were normalized to $OD_{600} \approx 0.2$ before 2 µL aliquots of yeast cell suspensions were spotted onto the plates (a single spot per interaction). All plates were incubated for 3 days at 30 °C.

Isl1/2$_{LIM1}$ vs. Ldb1$_{LID}$ interactions were strong as shown by robust growth under double selection conditions for all proteins except Lhx1, which showed more modest yeast growth. Moderate interactions were seen between Ldb1$_{LID}$ and LIM2 domains of Lhx2, Lhx6, Lhx8, Lhx9, Lmx1a and Lmx1b (1 mM 3-AT moderate selection conditions) and weaker interactions were seen for Lhx3/4$_{LIM2}$ vs. Ldb1$_{LID}$. Weak interactions were detected in tests between Ldb1$_{LID}$ and LIM1 domains of Lhx2, Lhx6 and Lmx1b (under weak selection conditions).

4.4 The red-white selection based on the ADE2 reporter activation

Activation of the *ADE2* reporter was additionally assessed by the colour of spots growing on rich YPD media that was not supplemented with adenine (Figure 4.8). Under these conditions

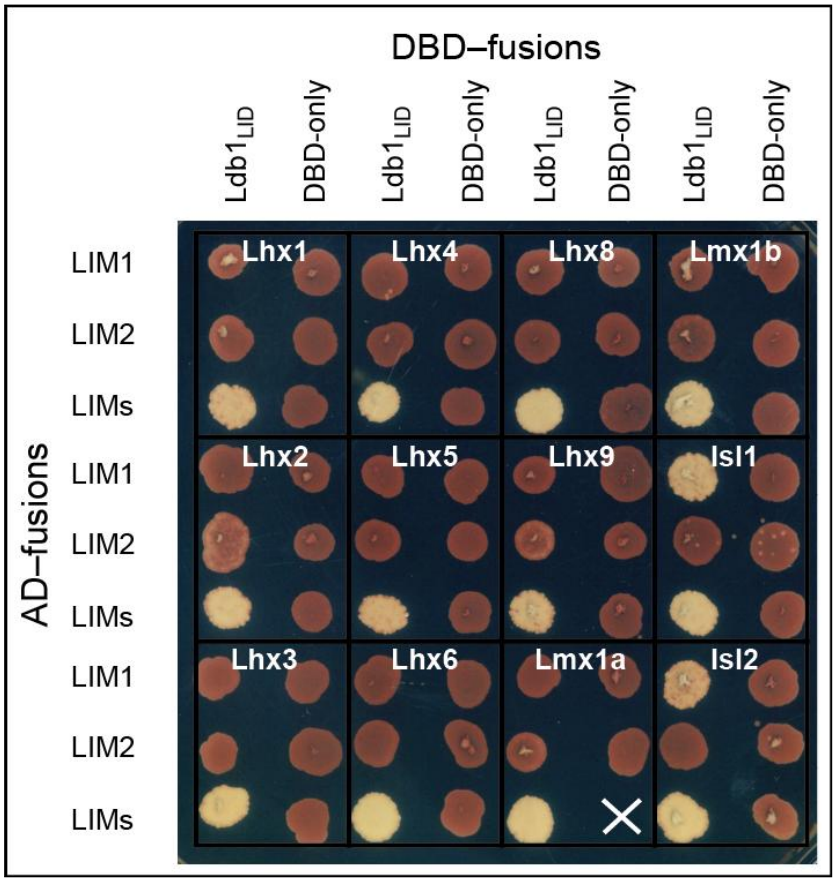


Figure 4.8. A YPD plate spotted with indicated yeast transformants. The plate was spotted with transformants and incubated for 3 days at 30 °C. The white spots are diploid yeast (derived from the mating of AH109 and Y187 cells) expressing the interacting constructs that activate the *ADE2* reporter. DBD-Lmx1a$_{LIMs}$ vs. DBD-only control did not grow on this plate (white cross).

there is no selection pressure so all the transformants grew, but only yeast that activated *ADE2* turned white (section 3.5.1). The assay reports the same interactions as seen under double selection conditions in Figure 4.7, with the additional observation of a slight pigment reduction for $Lhx2_{LIM2}$ vs. $Ldb1_{LID}$. The colour differences appeared early, after overnight incubation. The colour was followed for two additional days as the differences became more pronounced, but because the pigment accumulated quickly in the absence of stronger interactions, no additional interactions could be detected with prolonged incubation. As this method didn't provide additional information, it wasn't further used.

4.5 Summary of results

The results of Y2H arrays presented in this chapter show that most LIM-HD proteins appear to have a tighter (or dominant) binding and a weaker binding LIM-domain in interactions with $Ldb1_{LID}$ (Table 4.1). In most LIM-HD proteins the dominant binder is LIM2, except for Isl1/2 where LIM1 is the dominant binder, and Lhx1/5 where no significant binding was observed for either individual domain, indicating that if there is a dominant binder the affinity of binding lies below the detection limit of the assay. The LIM domains from paralogous pairs of LIM-HD proteins show similar modes and affinities of binding to $Ldb1_{LID}$ in Y2H assays. Based on the assays described in this chapter the tandem LIMs of Isl1/2 and Lhx2/9 have the strongest binding overall, followed by Lhx6/8, Lmx1a/b and Lhx3/4, with $Lhx1/5_{LIMs}$ constructs being the weakest binders. $Isl1/2_{LIM1}$ appeared to be the strongest interacting single domains, which were consistently detected under double selection conditions. Somewhat weaker interactions of $Lhx2/9_{LIM2}$ were detected under double selection conditions only in one combination of fusion constructs. $Lhx6/8_{LIM2}$ and $Lmx1b_{LIM2}$ apparently bound more strongly to $Ldb1_{LID}$ than $Lmx1b_{LIM2}$ and $Lhx4_{LIM2,}$ followed by $Lhx3_{LIM2}$.

Some apparent binding of LIM1 domains was detected for proteins where LIM2 was the dominant domain ($Lhx2_{LIM1}$, $Lhx6_{LIM1}$ and $Lmx1b_{LIM1}$) as evidenced by growth and colour development under 1 mM 3-AT moderate selection ($Lhx2_{LIM1}$ only) and weak selection conditions. $Isl1_{LBD}$ interactions with $Lhx3/4_{LIMs}$ resulted in reporter activation at levels lower than that for $Ldb1_{LID}$ with $Lhx3/4_{LIMs}$ and only slightly lower than that for $Ldb1_{LID}$ interaction

with LIM1 domains of Isl1/2. Finally, the range of interactions detectable by red-white selection was shown to be similar to interactions detected on double selection media.

The potential biological relevance of the LIM domain contributions are discussed in section 8.4.

Table 4.1. Summary of Y2H array results testing the abilities of separate LIM-domains to bind $Ldb1_{LID}$. For this comparison, reporter activation was quantified by a scale from 1 (lowest under weak selection conditions) to 12 bars (strongest under moderate and double selection conditions) based on growth and colour development in all 3 selection conditions, using both combinations of Y2H fusion constructs. Dashed bars indicate potential interactions that caused yeast growth at or slightly above background level on weak selection plates. Interactions consistently detected under double selection conditions in at least one combination are shown in yellow fields and interactions detected only under weak selection conditions are in light grey fields. Red bars indicate interactions which were detected in both combinations of fusion constructs, and black bars indicate interactions which were detected only in the AD-LIMs vs. DBD-LID combination. The column on the right shows which LIM domain is dominant (if any) in interactions with $Ldb1_{LID}$.

Protein	LIMs	LIM1	LIM2	Dominant LIM domain
Lhx1				neither one
Lhx5				neither one
Lhx3				LIM2
Lhx4				LIM2
Lmx1a				LIM2
Lmx1b				LIM2
Lhx8				LIM2
Lhx6				LIM2
Lhx9				LIM2
Lhx2				LIM2
Isl1				LIM1
Isl2				LIM1

Chapter 5 Development of a bimolecular fluorescence complementation (BIFC) assay for detecting interactions of LIM-domain proteins

5.1 Protein-fragment complementation assays (PCAs)

It has been known for decades that some proteins can fold into native or native-like tertiary structures when the protein has been split into two separate polypeptides fragments [243-245]. This idea is referred to as protein complementation. Johnsson and Varshavsky developed an in vivo assay based on this approach for detecting protein-protein interactions through complementation of split-ubiquitin fragments [246]. Since then, a number of similar methods have been developed (reviewed in [198, 247-250]). They are commonly known as protein-fragment complementation assays (PCAs) or split-protein assays (Figure 5.1A). PCAs rely on direct interactions between the split-protein fragments. Note that the interactions involve protein binding and folding rather than docking of pre-folded fragments [248].

The types of signal generated in PCAs can include transcriptional activation, survival selection, enzyme activity, bioluminescence or fluorescence. Common PCA methods employ split ubiquitin, DHFR, intein, luciferase and GFP (Figure 5.1B). The latter is referred to as Bimolecular Fluorescence Complementation (BiFC).

5.2 BiFC

BiFC is based on the Aquarea victoria green fluorescent protein (GFP) and its numerous variants. These fluorescent proteins (FPs) have a β-barrel structure comprised of 11 anti-parallel β-sheets, surrounding a deformed helical sequence that stretches through the interior to form an autocatalytic fluorophore that is created only when the protein is folded (Figure 5.2A and B). Fluorophore maturation involves residues S(T/G)65, Y66 and G67 undergoing a series of cyclization, dehydration and oxidation reactions to create a functional fluorophore [251, 252]. After maturation, the FP fold is practically irreversible under physiological conditions [253]. In BiFC two test proteins are expressed as fusion constructs of complementary fragments from an FP. Binding of the test proteins induces folding of the two complementary fragments and

creation of the fluorophore (Figure 5.2C). Split GFPs are usually split in the loops between the sheets 7 and 8 (this study), 8 and 9, or, 10 and 11 (Figure 5.2B). In some cases, the complementary fragments have an overlap as in the case for $NYFP_{1-172}$ and $CYFP_{155-238}$ [254]. Studies investigating GFP folding indicated that N-terminal fragments could be partially folded when expressed independently, but the C-terminal fragment likely does not form secondary structure prior to complementation [255, 256]. BiFC assays have now been carried out in a wide variety of cells from bacteria, mammalian cells, Fungi, Nematoda, fruitfly and plants (reviewed in [247, 257]).

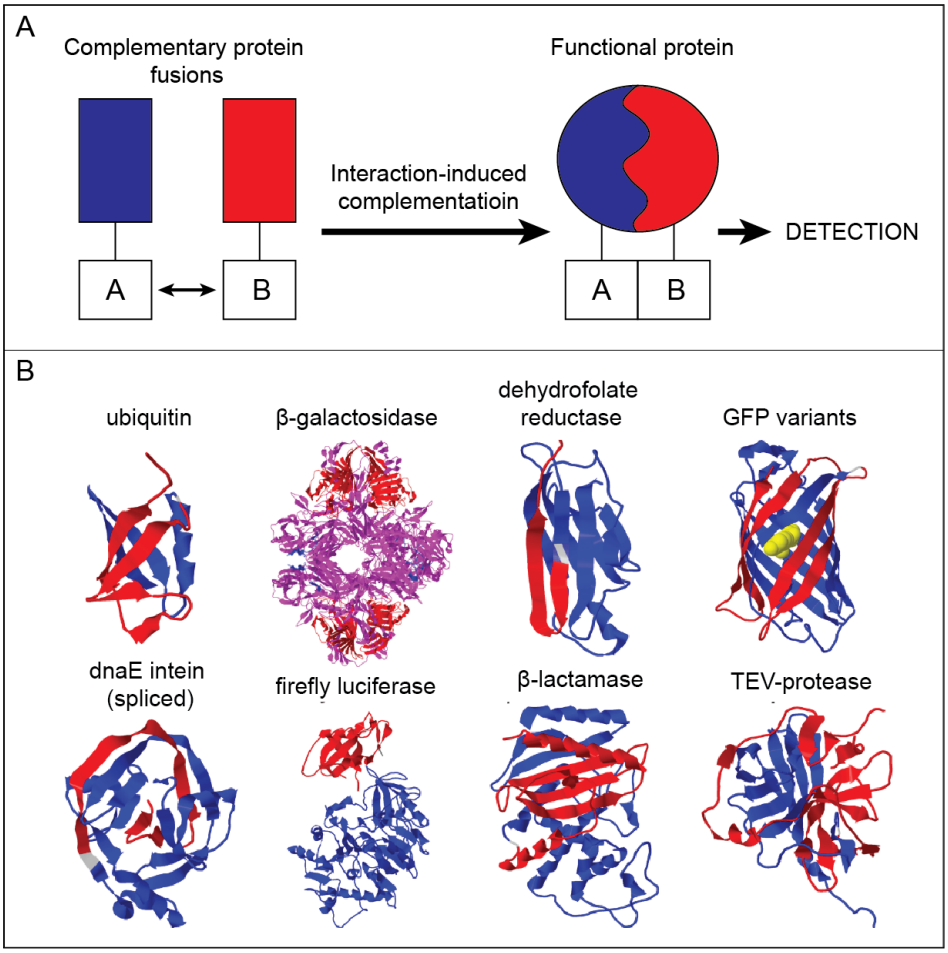


Figure 5.1. Protein complementation assay (PCA). (A) The principle of PCA. Two proteins (A and B) fused to non-functional fragments of a reporter protein interact with each other to facilitate the association and folding of non-functional fragments into a functional reporter protein. (B) Split-proteins commonly used in PCA experiments (taken from Kerppola, 2006). Two colours in backbone cartoon representations show two complementary fragments in their folded/bound state. Magenta indicates overlap between fragments. The fluorophore from GFP variants is shown in yellow.

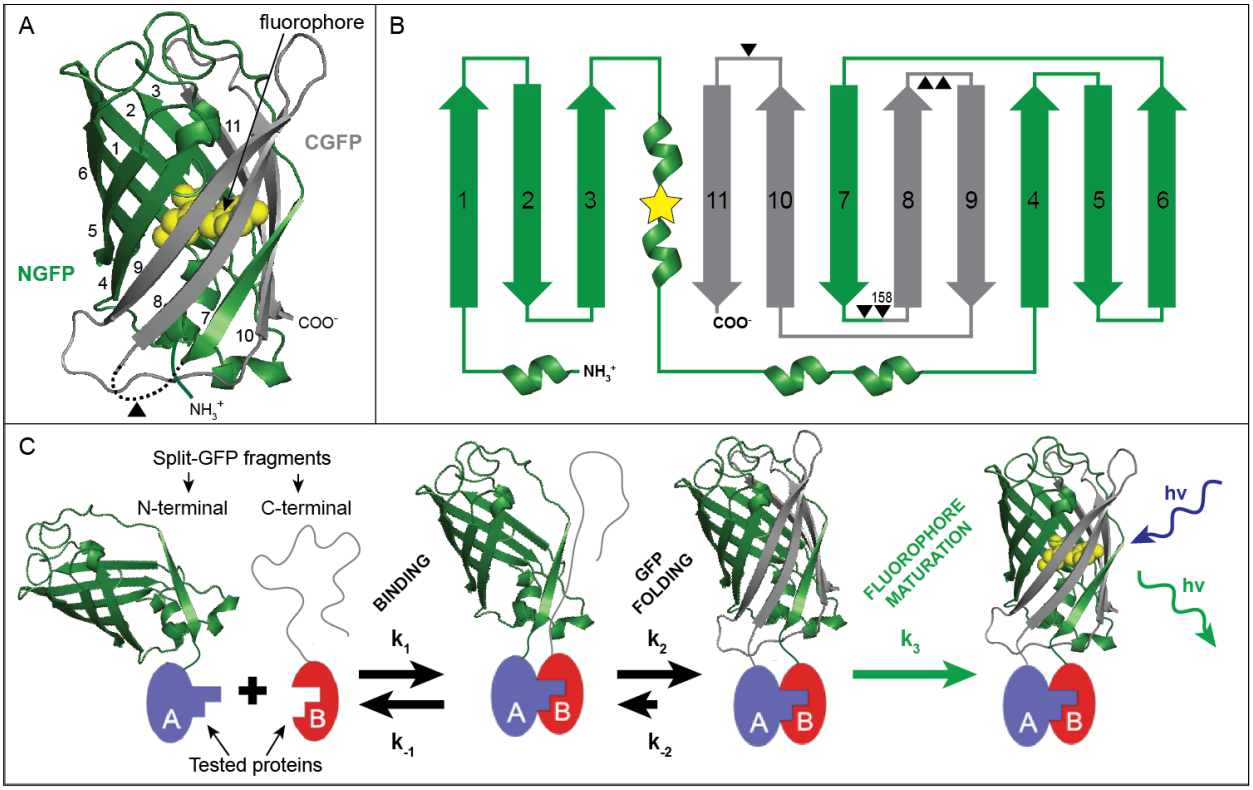


Figure 5.2. Split-GFP and the mechanism of Bimolecular Fluorescence Complementation (BiFC). (A) Cartoon representation of GFP (pdb code: 1B9C). The fluorophore atoms are indicated by yellow spheres inside the β-barrel. (B) Topology map of GFP showing secondary structure where arrows with numbers show β-sheets, helices show α-helical regions and line segments show loops. In this study the split site immediately followed residue 158. (C) BiFC mechanism as described in the above text based on cartoon representations of FP fragments and the complete FP protein. The yellow star shows the position of the fluorophore. In (A) and (B) the split-GFP structure is shown with its N-terminal (NGFP, dark green) and C-terminal (CGFP, grey) fragments. β-sheets are numbered according to their position in the primary sequence. Black triangles mark the commonly used split sites.

Our laboratory wanted to explore using a yeast-based BiFC assay to try and overcome two problems associated with the standard Y2H assay for LIM-HD proteins. The first problem is that the Y2H assay is a transcription based assay that can show high levels of background activation for transcription factors, which was particularly problematic for many LIM-HD proteins (see section 3.2), and prevented comprehensive screening of interactions that could be mediated by the C-terminal domains of the LIM-HDs. Whereas Y2H assays take place in the yeast nucleus, BiFC assays can be cytoplasmic and potentially avoid the problem of background transcriptional activation. The second problem is that it can be difficult to detect weak interactions. Because the final BiFC product is a stable molecule, it may be better suited to the detection of weaker interactions. Of the possible eukaryotic systems that could be used, *S.*

cerevisiae was selected for use because it is relatively simple and has proven to be a good platform for testing LIM-HD interactions in the past (apart from the previously mentioned problems, see section 3.8.1).

5.3 Plasmid editing and protein fusion constructs

A split-GFP sequence that coded for an optimized form of GFP (GFPuv, alphaGFP) was obtained. This version belongs to the same spectral group as the wild type protein, but contains the mutations F99S, M153T and V163A [258]. It has similar folding kinetics to wild type GFP but is less prone to aggregation resulting in higher yields and improved relative fluorescence at 37 °C [252, 259, 260]. The protein was split between residues 158 and 159, in the loop between β-sheets 7 and 8, into non-overlapping NGFP (1–158) and CGFP (159–238) fragments. The assay design involved the use of modified Y2H vectors in which the Gal4 domains were replaced with NGFP and CGFP. Two NpGBT9-based BiFC vectors were generated: pW397-$NGFP_N$, which contained NGFP and a linker ($NGFP_N$) upstream of the cloning site to generate N-terminally tagged constructs; and, pW397-$NGFP_C$, which contained a linker followed by NGFP ($NGFP_C$) downstream of the cloning site to generate C-terminally tagged constructs. Similarly two pGAD10-derived BiFC vectors were generated: pL397-$CGFP_N$ and pL397-$CGFP_C$, to generate N- and C-terminally tagged constructs, respectively (Figure 5.3). The 10-residue linker $(GGGGS)_2$ is designed to allow flexibility, enhance solubility, and was used successfully in earlier BiFC studies [261].

As part of this process two additional empty vectors that lacked any GFP-tags (pW397 and pL397) were created. An NGFP-CGFP fusion construct in which the GFP fragments were connected by a 24-residue long linker (contains 4 repeats of the GGGGS sequence, see Appendix B) was cloned into each of these plasmids for use in control experiments. As the complementary GFP fragments are joined by the linker, this construct was expected to display maximal fluorescence intensity. Note that it was not expected that the same levels of fluorescence intensity would be reached by actual BiFC interactions as such signals are typically < 10% of the full-length GFP [247].

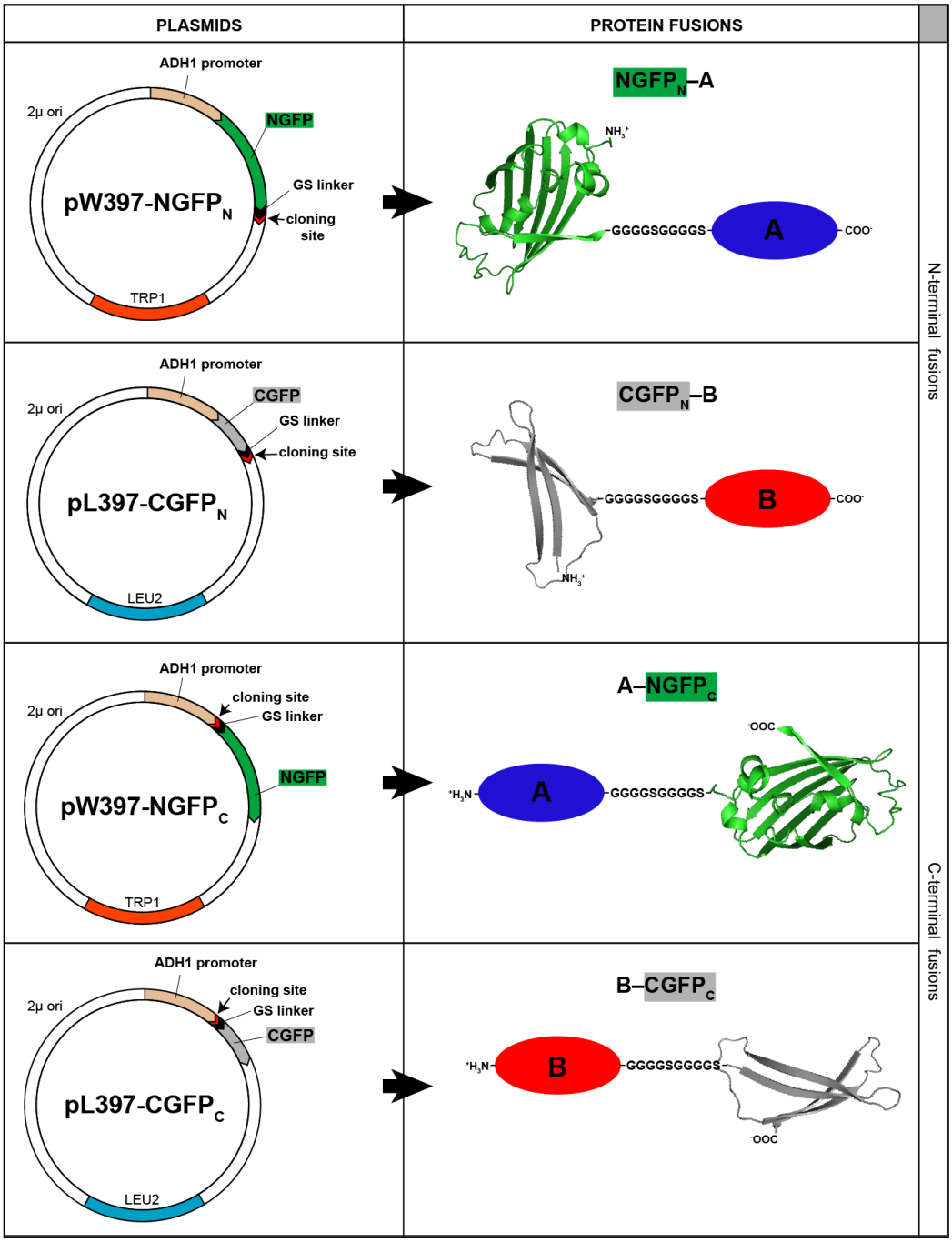


Figure 5.3. BiFC plasmids and fusion constructs. Plasmids are shown on the left and the corresponding constructs are shown on the right. Each yeast BiFC plasmid has the 2µ origin, a truncated ADH1 promoter (397 bp), and was modified to contain a GFP fragment cDNA plus a linker before, or after, the BamHI-EcoRI cloning site. pW397 and pL397 plasmids contain the *trp1* and *leu2* auxotropic section markers, respectively. The N or C in the subscript of the plasmid name indicates that the GFP fragment is expressed N-terminal or C-terminal of the tested protein. In this figure only a subset of the test proteins constructs are depicted as each of A and B are cloned into each vector.

5.4 Autofluorescence

The Y2H AH109 yeast cells, which contain the optimal auxotrophic markers required for the designed BiFC plasmid system, showed high levels of autofluorescence in the initial tests compared to some other strains such as EGY48 (Figure 5.4). This autofluorescence was localized mainly in the vacuole and comes from the red AIR pigment that accumulates due to a mutation in *ade2* under adenine-depleted conditions (section 3.5.1). By supplementing the media with adenine-hemisulfate (60 µg/mL) and by growing the transformants in two consecutive overnight cultures, it was possible to significantly reduce autofluorescence, although some older and dying cells still displayed autofluorescence [262]. Several other studies have reported the successful use of *ade2* mutant yeast strains in BiFC experiments, supporting this as a reasonable approach [263-265].

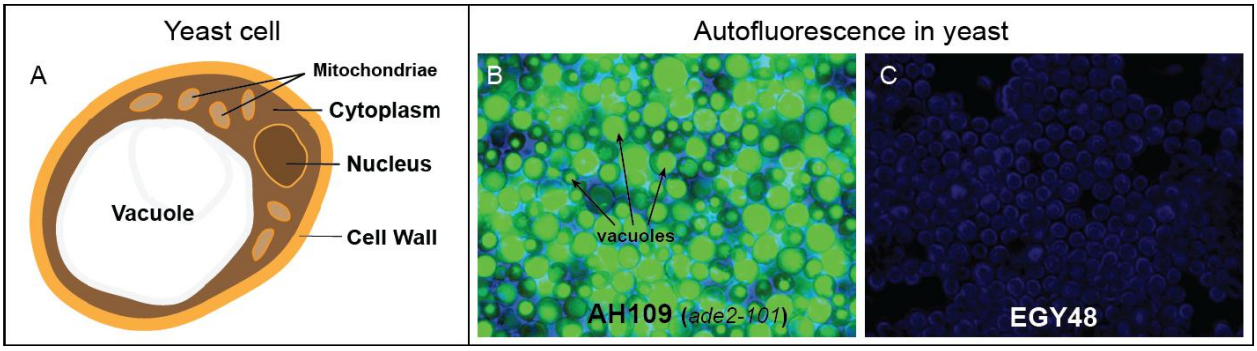


Figure 5.4. Autofluorescence in yeast. (A) Cartoon of a yeast cell. The vacuole, nucleus, mitochondria, cytoplasm and cell wall are labelled. (B) Micrograph of the untransformed AH109 strain (*ade2* mutant) showing strong vacuolar autofluorescence. (C) Untransformed EGY48 (*ade2 wt*) strain showing no autofluorescence. Both strains were grown on YPD plates without added adenine. The images are composites of FITC (green) and bright field (blue) micrographs.

5.5 Optimization of the $LMO4_{LIMs}$:$Ldb1_{LID}$ interaction in the BiFC system

The well characterised interaction of $LMO4_{LIMs}$ vs. $Ldb1_{LID}$ [152, 156, 167] was selected as a model system for BiFC assay development. Because BiFC requires that complementing fragments are brought together in an orientation that favours folding of the GFP, the first step in optimizing BiFC for previously untested groups of proteins is to find the combination of fusion constructs that gives the highest increase of BiFC signal over controls (i.e., an optimal signal to

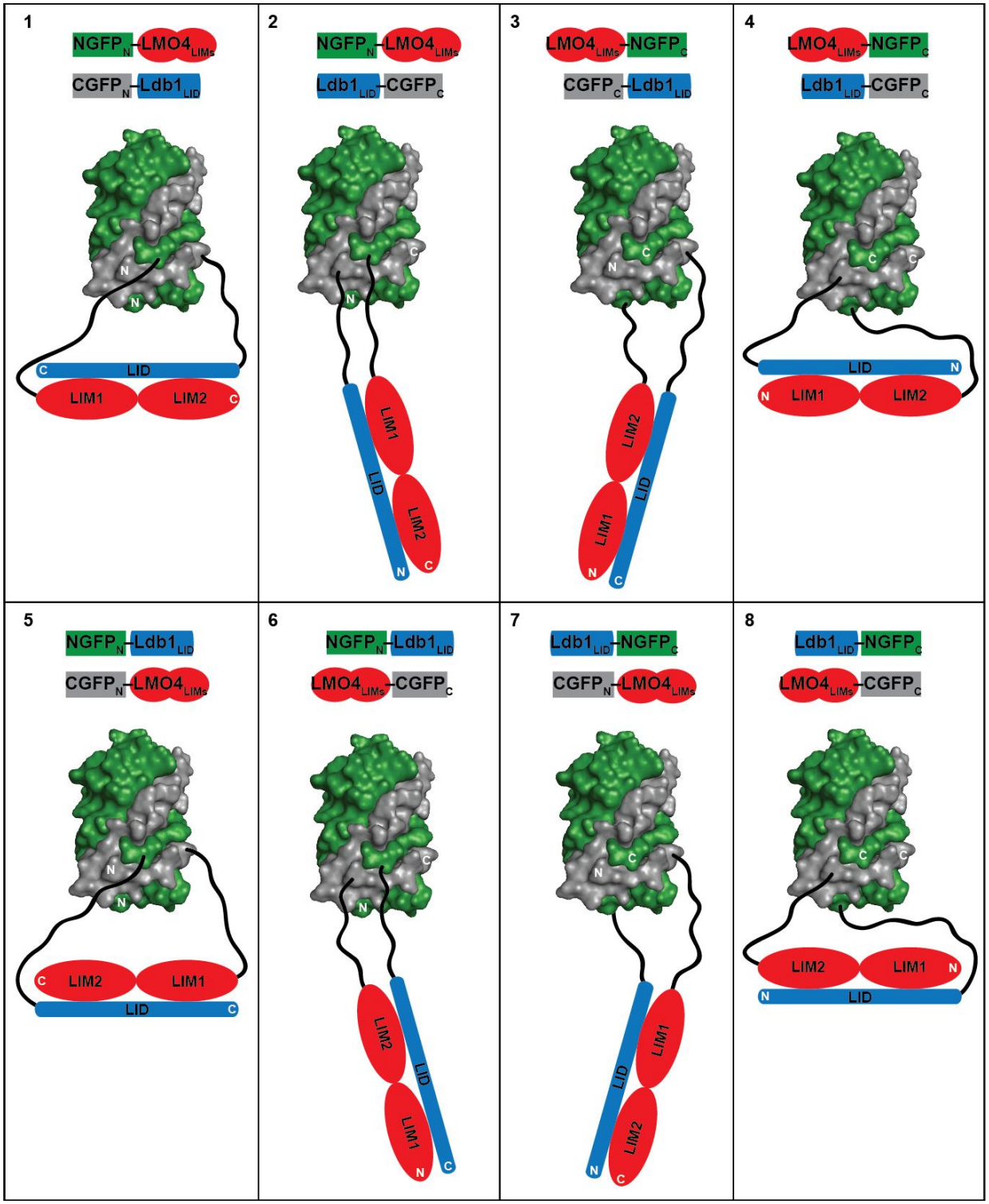


Figure 5.5. Models of possible assemblies formed from protein fusion constructs shown in Figure 5.3. The combinations or assembly modes are numbered. GFP fragments are shown as parts of the folded GFP in dark green (NGFP) and grey (CGFP). The flexible linkers are shown as black lines, $LMO4_{LIMs}$ as tandem red ovals, and $Ldb1_{LID}$ as a blue rod.

noise ratio, S/N) [254]. Eight combinations of fusion constructs that could allow for eight different assembly modes were tested (Figure 5.5).

Every interaction pair was compared to its two relevant negative controls (a NGFP- or CGFP-tagged test protein against $CGFP_{N/C}$ or $NCFP_{N/C}$, respectively). All NGFP- and CGFP-only pairs were tested for self assembly in the absence of interacting test proteins. An autofluorescence control was also included in which cells were co-transformed with pW397 and pL397 (i.e., did not contain any GFP fragments). The results of these assays are presented in Figures 5.6 and 5.7.

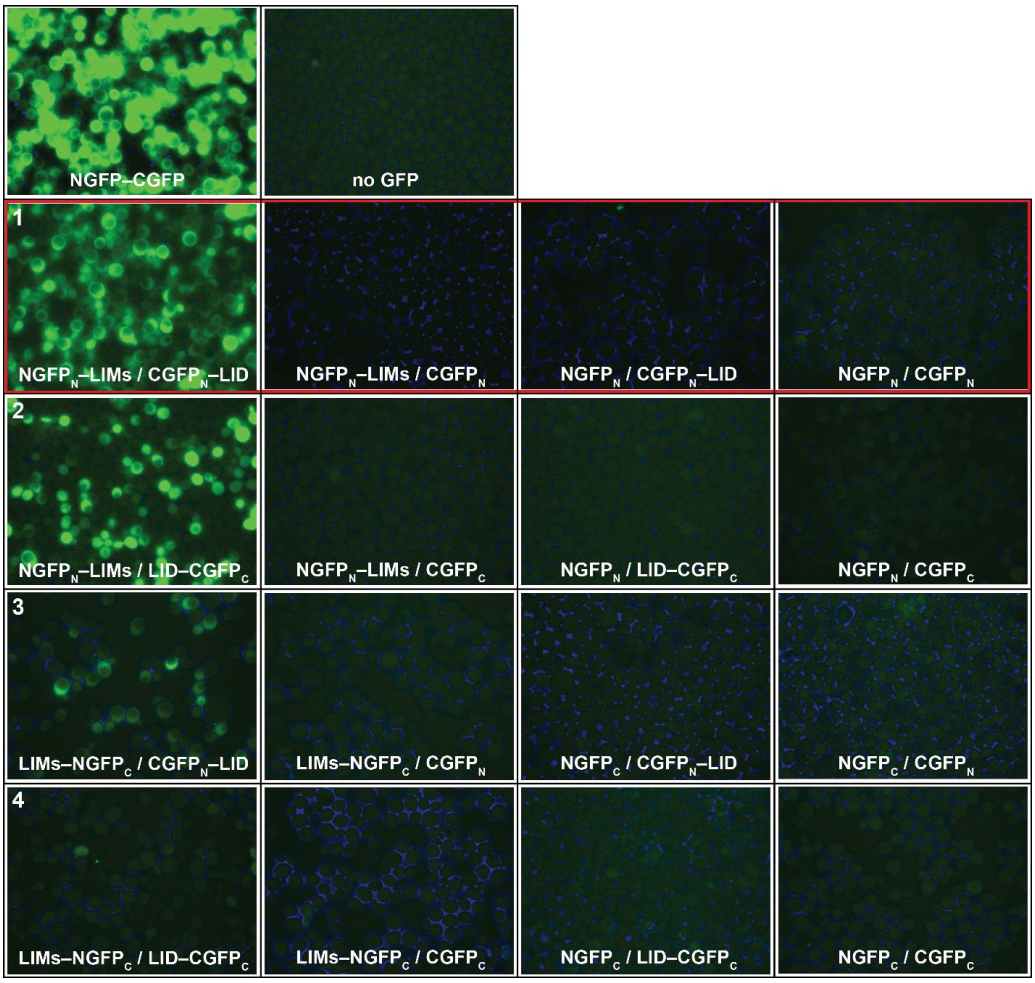


Figure 5.6. Optimizing the orientation of GFP-fragment tags for the $LMO4_{LIMs}$ vs. $Ldb1_{LID}$ BiFC assay. Combinations 1–4. Yeast cells were visualized with a 100× objective and FITC filter at 200 ms exposure, unless otherwise indicated. Representative images are composites of FITC (green) and bright field (blue) micrographs. The NGFP-CGFP fluorescence control is shown at 100 ms exposure. No GFP refers to pW397/ pL397 (i.e., no GFP fragments). The 16 images below show combinations 1–4 (on the left, numbered) and the rest are appropriate controls. Combination 1 and its controls are shown in the red rectangle.

Importantly the self assembly and autofluoresence controls showed no significant fluorescence. Except for combination 4, BiFC fluorescence was detected in all sets of $LMO4_{LIMs}$/$Ldb1_{LID}$ interaction pairs with the average fluorescence intensity of BiFC-positive cells being at least an order of magnitude weaker than that of the NGFP-CGFP positive control (Figure 5.6, top left). The combinations in which $LMO4_{LIMs}$ was fused to $NGFP_C$ or $CGFP_C$ gave relatively weaker BiFC signals compared to $LMO4_{LIMs}$ fused to $NGFP_N$ or $CGFP_N$. This is particularly evident for combinations involving the $LMO4_{LIMs}$-$NGFP_C$ fusion construct (combinations 3 and 4). The strongest signals were detected for combination 1 involving $NGFP_N$-$LMO4_{LIMs}$ and $CGFP_N$-$Ldb1_{LID}$ (magenta rectangle in Figure 5.6) and combinations 2, 5 and 7. Combination 1 was chosen as a benchmark for other LIM-domain interactions and was further characterized.

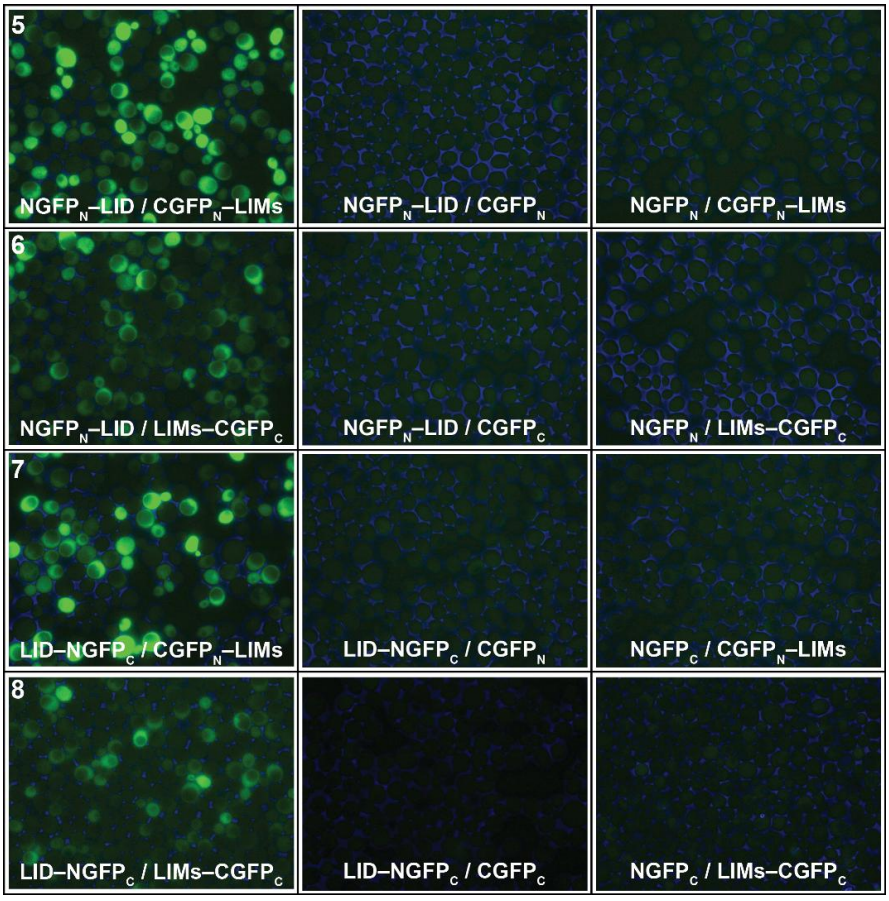


Figure 5.7. Optimizing the orientation of GFP-fragment tags for the $LMO4_{LIMs}$ vs. $Ldb1_{LID}$ BiFC assay. Combinations 5–8. Presentation is the same as Fig. 5.6, but for combinations 5–8. The images are composites of FITC (green) and bright field (blue) micrographs.

All tested combinations and the NGFP-CGFP control showed BiFC signals with the same cytoplasmic-nuclear localization. This was confirmed with Hoescht staining of the cells on the slide and visualization under a DAPI filter to detect nuclear DNA (Figure 5.8B). Interestingly, the cells that exhibited stronger BiFC were less likely to take in the Hoescht stain.

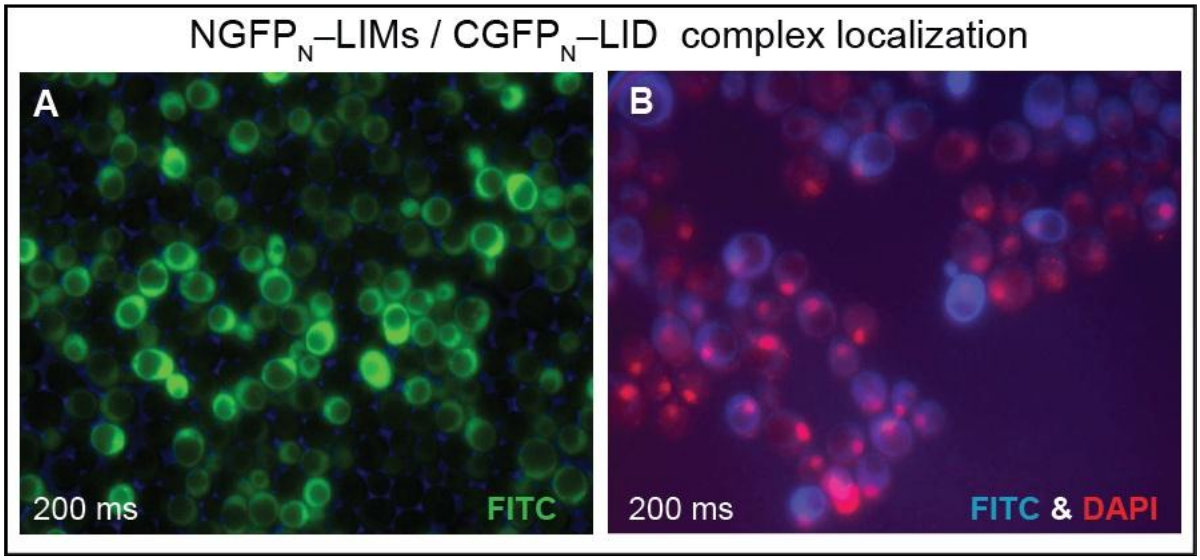


Figure 5.8. Localization of BiFC signal in the $LMO4_{LIMs}$ vs. $Ldb1_{LID}$ interaction. AH109 yeast cells co-transformed by $NGFP_N$-$LMO4_{LIMs}$ vs.$CGFP_N$-$Ldb1_{LID}$.(A) Composite image made of bright field (blue) and FITC (green) micrographs (B) Composite image made of DAPI (red) and FITC (blue) micrographs of the Hoechst-stained yeast slide.

5.6 Selectivity and sensitivity of the assay

In order to test the selectivity of BiFC assays for detecting strong LIM-LID interactions an $Ldb1_{LID}$ mutant that had previously shown to completely disrupted the growth of yeast in the Y2H assay [167] was tested for interaction against LMO4. The m30 variant of $Ldb1_{LID}$ is a double mutant (V339A/I358A) in which the mutations affect binding to $LMO4_{LIM2}$ and $LMO4_{LIM1}$, respectively. The orientation of GFP fragments was $NGFP_N$ vs. $CGFP_N$ as selected above. Cells expressing m30 had very little fluorescent complementation signal compared to the cells expressing wild type $Ldb1_{LID}$ (Figure 5.9A and B). Since the fixation procedure in trial experiments produced lower BiFC fluorescence, higher autofluorescence and often disrupted vacuolae and nuclei (not shown), the cells were not fixed before microscopy. Therefore, the cells

were used fresh to minimize degradation and were not additionally incubated to allow recently formed BiFC complexes to mature. As this procedure was sufficient for the full-length GFP protein to show fluorescence in appropriate positive controls, it was estimated that the relatively slower growth of yeast in SD media at 30°C (~140 min doubling time), allowed enough time for maturation. However, when the m30-expressing cells were incubated for 5 h at 4 °C, the fluorescence increased to reach a similar level to that of the wild type $Ldb1_{LID}$ (Figure 5.9C), which itself did not significantly change (not shown). The data indicates that this BiFC system can discriminate between LIM:LID interactions with wild type sequences and interactions involving specific mutants designed to affect binding, but that temperature and time of incubation are important factors in distinguishing interactions of differing affinity. These properties are consistent with the temperature-dependence of the GFP fluorophore maturation and the apparent irreversibility of complementation [252, 254]. Previous studies have similarly used lower temperatures and longer maturation times prior to detection to increase fluorescence signal [264, 266].

The binding of m30 for LMO4 (binding affinity is not known) can be detected by low moderate selection conditions by Y2H but not double selection conditions [156]. Previous reports of high sensitivity in BiFC assays (down to 1 mM) were attributed to the accumulation of reconstituted BiFC complexes [253, 267].

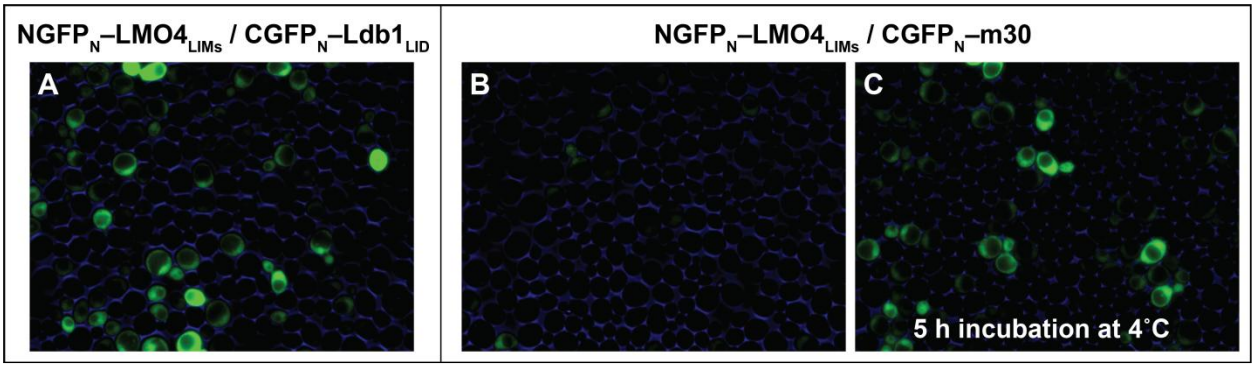


Figure 5.9. Selectivity of the BiFC assay for a LIM:LID interaction and the effect of cold incubation. $LMO4_{LIMs}$ was tested against (A) wild-type $Ldb1_{LID}$ (no incubation) and (B) the m30 mutant of $Ldb1_{LID}$ containing V339A/I358A double mutation (no incubation). (C) The $NGFP_N$-$LMO4_{LIMs}$/$CGFP_N$-m30 combination was incubated at 4 °C for 5 h and re-visualized. The images are composites of FITC (green) and bright field (blue) micrographs.

To further explore the sensitivity of this system in the context of LIM family and related proteins, a study was carried out in which the ability of several ZnF domains (including LIM domains) to interact with their binding partners in the yeast BiFC assay was tested (Figure 5.9). These interactions cover a range of previously reported binding affinities stretching from $K_d \approx 10^{-8}$ M to $\approx 10^{-4}$ M (Table 5.1). The experiments were repeated with some variation in fluorescent signal intensities, especially for the lower affinity interaction with the U-shaped ZnFs 1 and 9 (USHF1 and USHF9) with the N-terminal ZnF from Pannier (PNR-NF). Overall, the LMO4-Ldb1 interaction appeared to give the strongest signals by a small margin and other LIM-LID interactions were of similar average intensity (Figure 5.10). The fluorescence intensity was lowest for the weakest interaction involving USH_{F9}. Overall, these data show that BiFC in yeast can detect interactions in a wide range of binding affinities, including weak interactions of close to K_d = 100 μM.

Table 5.1. Binding affinities of construct pairs tested in the BiFC yeast system. Binding affinities are expressed in the form of dissociation constant K_d. All interactions were previously detected with Y2H and the K_d values were determined by biophysical methods in vitro.

Proteins (domains)		***K_d*(M)**	**Reference**	**Detected by Y2H**
$LMO4_{LIMs}$	**$Ldb1_{LID}$**	10^{-8}	[167]	yes
$LMO2_{LIMs}$	**$Ldb1_{LID}$**	2×10^{-8}	[167]	yes
$Lhx3_{LIMs}$	**$Ldb1_{LID}$**	3.4×10^{-8}	[18]	yes
$Isl1_{LIMs}$	**$Ldb1_{LID}$**	9×10^{-8}	[18]	yes
USHF1	**PNR-NF**	10^{-4} to 6.9×10^{-6}	[268]	yes
USHF9	**PNR-NF**	10^{-4} to 5.2×10^{-5}	[268]	yes

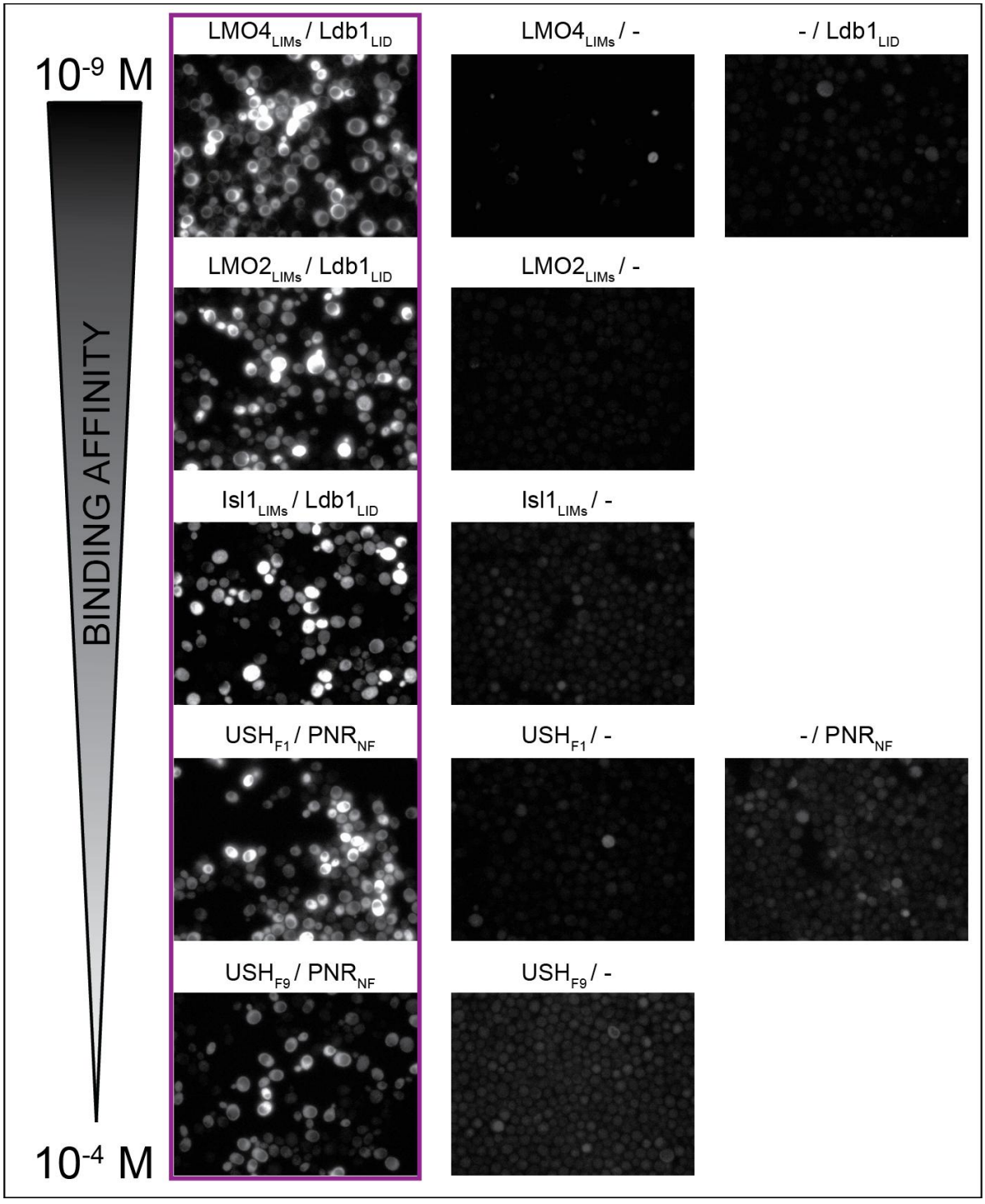


Figure 5.10. Sensitivity of the BiFC assay in protein-protein interactions mediated by ZnFs from LIM, GATA and FOG proteins. Mouse $LMO4_{LIMs}$, $LMO2_{LIMs}$ and $Isl1_{LIMs}$ constructs were tested for interaction against $Ldb1_{LID}$, whereas USF1 and USF9 were tested against PNR-NF. Only the FITC signal is shown. All micrographs on the left (magenta box) show tests between $NGFP_N$-tagged and $CGFP_N$-tagged proteins. Other micrographs show the appropriate controls with the relevant GFP fragment. The reported binding affinities of interactions decrease from top (low K_d) to bottom (high K_d) as shown by the triangle on the left.

5.7 Full-length LIM-HD proteins in the BiFC assay and the impact of a stronger promoter

The well characterised Lhx3 vs. Isl1α interaction (section 1.8) was chosen as a model to test full-length LIM-HD proteins in BiFC. Lhx3a was cloned into the pW397-$NGFP_N$ vector, whereas Isl1α was cloned into the pL397-$CGFP_N$ vector. $Lhx3_{LIMs}$ and $Isl1_{LBD}$ were also cloned into appropriate plasmids to provide a domain vs. domain interaction control for comparison with the full-length protein interaction. Full-length protein vs. $Ldb1_{LID}$ interactions were also tested as a comparison. The domain-domain interaction $NGFP_N$-$Lhx3_{LIMs}$ vs. $CGFP_N$-$Isl1_{LBD}$ was detected (Fig 5.11C), but no fluorescence signal was detected for the full-length proteins ($NGFP_N$-Lhx3a vs. $CGFP_N$-Isl1α) under the same experimental conditions (Figure 5.11F). It was noted, however, that all tests involving the $CGFP_N$-Isl1α fusion construct failed to develop fluorescence (Figure 5.11F, D and H), which could indicate a problem with the expression and/or stability of this construct. In contrast, tests of $NGFP_N$-Lhx3a with $CGFP_N$-$Ldb1_{LID}$ and (to a lesser extent) $CGFP_N$-$Isl1_{LBD}$ showed nuclear localization of signal in some cells (Figure 5.11G and E).

A western blot using a polyclonal anti-GFP antibody was carried out to test for proper expression of full-length LIM-HD and LMO4/Ldb1 constructs (Figure 5.11I). A strong band corresponding to the NGFP-CGFP construct was detected, but no GFP-fragments were detected in other samples, despite detection of GFP fluorescence in the cells containing $NGFP_N$-$LMO4_{LIMs}$/$CGFP_N$-$Ldb1_{LID}$ and $NGFP_N$-$Lhx3_{LIMs}$/$CGFP_N$-$Isl1_{LBD}$ plasmids. The non-detection of protein suggests that GFP-fragments and fusion constructs in yeast cells have much lower levels of stable expression than that of the full-length GFP (NGFP-CGFP).

The 397 bp ADH1 (truncated) promoter that was part of the parent vectors was designed to give low levels of constitutive expression. The 397 bp promoter was replaced with a 1473 bp full-length ADH1 promoter from the pGADT7 plasmid in the relevant BiFC plasmids to generate pWAFL-$NGFP_N$ and pLAFL-$CGFP_N$ (where AFL indicates the full-length ADH1 promoter) to try and increase expression levels of GFP-fusion proteins and the experiment was repeated. This promoter should give high expression in mid-log phase under aerobic conditions (Clontech, Yeast Protocols Handbook). The modified plasmids did result in significantly

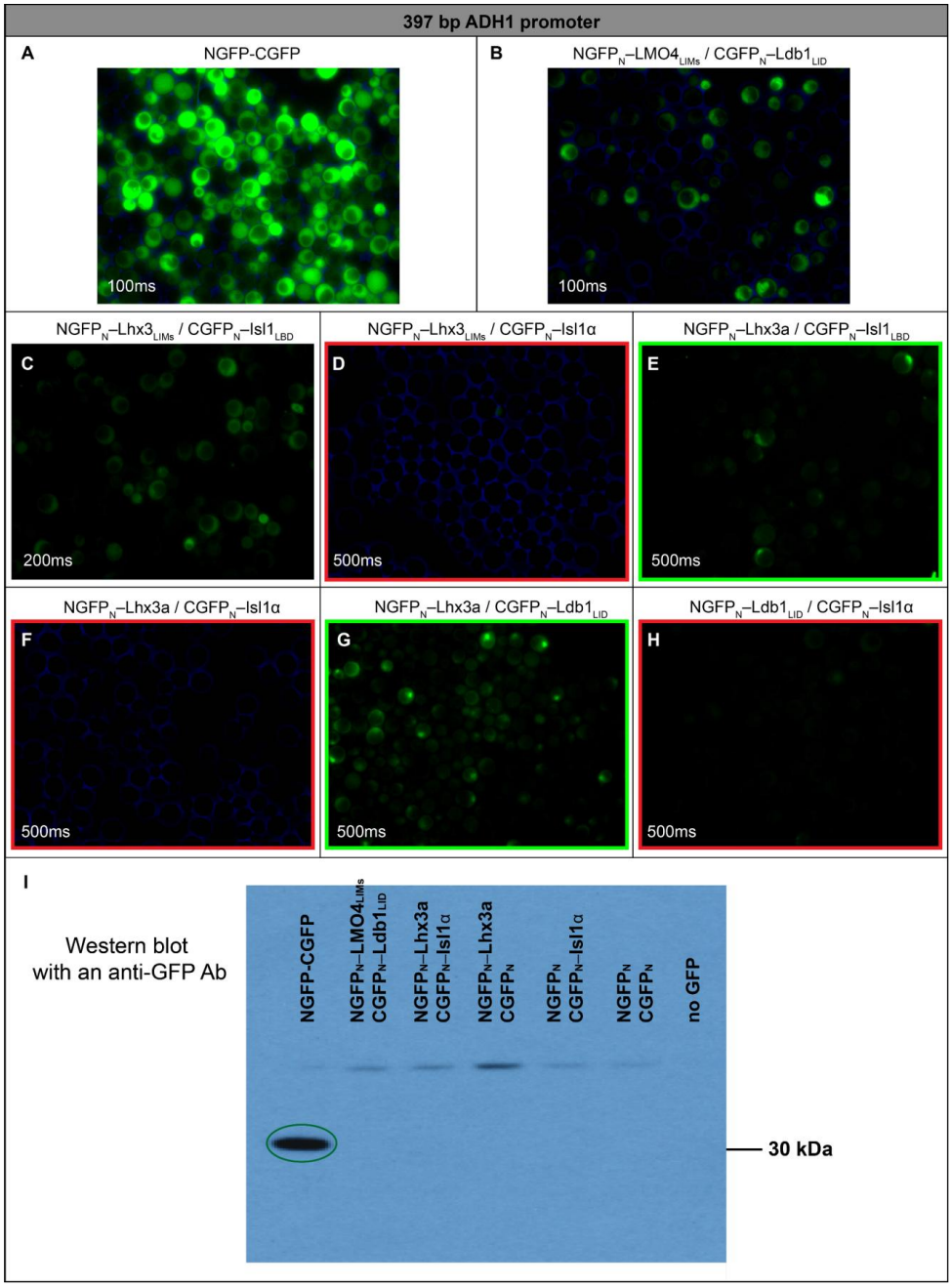


Figure 5.11. Investigating interactions between Lhx3a and Isl1α proteins. (A) Fluorescence control. (B) Positive interaction control. (C) interaction between isolated $Lhx3_{LIMs}$ and $Isl1_{LBD}$ domains, (D) and (E) show tests between isolated domains and full-length proteins, (F) test between full-length proteins, (G) and (H) tests between $Ldb1_{LID}$ and full-length proteins. All proteins were expressed using the 397 bp ADH1 promoter. BiFC images (A), (B), (D) and (G) are composites of FITC (green) and bright field (blue) micrographs while other images show only FITC channels. Exposure times are shown in the bottom left corner of every image. Images framed in red show no or very weak fluorescence and images framed in green detect nuclear localization of fluorescence (I) Western blot of yeast lysates from indicated transformants. NGFP-CGFP construct (28.6 kDa; circled). Rabbit Anti-GFP polyclonal primary Ab and goat anti-rabbit secondary Ab was used. An unspecific band is seen in all lanes of the gel at higher MW that does not correspond to any GFP-fragments or complemented proteins.

higher levels of fluorescence for yeast cells expressing the NGFP-CGFP control, the $NGFP_N$-$LMO4_{LIMs}$ vs. $CGFP_N$-$Ldb1_{LID}$ BiFC control and $NGFP_N$-$Lhx3_{LIMs}$ vs. $CGFP_N$-$Isl1_{LBD}$ (Figure 5.12A–C). In the case of full-length Lhx3a and Isl1α BiFC, a weaker signal that was predominantly localised to the nucleus was observed (Figure 5.12D–E). Full-length Lhx3a tested against $CGFP_N$-$Ldb1_{LID}$ gave rise to nuclear localization (Figure 5.12F) as was observed previously (Figure 5.11G). These observations suggested that some full-length Isl1α and Lhx3a proteins are being directed to the nucleus where the interaction may take place. However, the levels of fluorescence were very low compared to interactions between separate domains. In some replicate experiments cells expressing both fusion constructs failed to grow, which could imply the toxicity due to higher expression levels.

A western blot of samples from this experiment showed higher levels of background, but more importantly, showed clear bands that correspond to the expected sizes of several NGFP-fragments (Figure 5.12G). For example, NGFP-CGFP (28.6 kDa) in lanes 1 and 6, $NGFP_N$ (20.4 kDa) in lanes 2 and 4, and $NGFP_N$-$LMO4_{LIMs}$ (33.2 kDa) in lane 5 were seen. However, no bands corresponding to CGFP-containing fragments were evident. Possible degradation products of the NGFP-CGFP were also detected in lanes 1 and 6. These data indicate that the stronger promoter did result in higher levels of protein expression, but suggest that at least some of the protein constructs are susceptible to degradation.

A new set of plasmids for expression of His-tagged GFP-fusion constructs in yeast were designed (pWAFLHisGFP-Lhx3a and pLAFLHisGFP-Isl1α) to permit visualisation of localization of individual proteins while facilitating their detection with anti-His-tag antibodies in a western blot. Unfortunately, yeast transformed with these plasmids failed to grow, or grew very slowly, which suggested that the His-GFP-Lhx3a and His-GFP-Isl1α constructs were probably toxic. Yeast transformed with pLAFLHisGFP-Isl1α displayed very weak fluorescence in a majority of cells, indicating low expression or misfolding of constructs. The signal in these cells was stronger in the cytoplasm and slightly depleted in the nuclei (Figure 5.13). Yeast transformed with pWAFLHisGFP-Lhx3a showed nuclear fluorescence in a small percentage of cells, but most cells showed no signal. When two proteins were co-expressed, the result was similar except that some cells showed weak cytoplasmic fluorescence. The apparent toxicity of

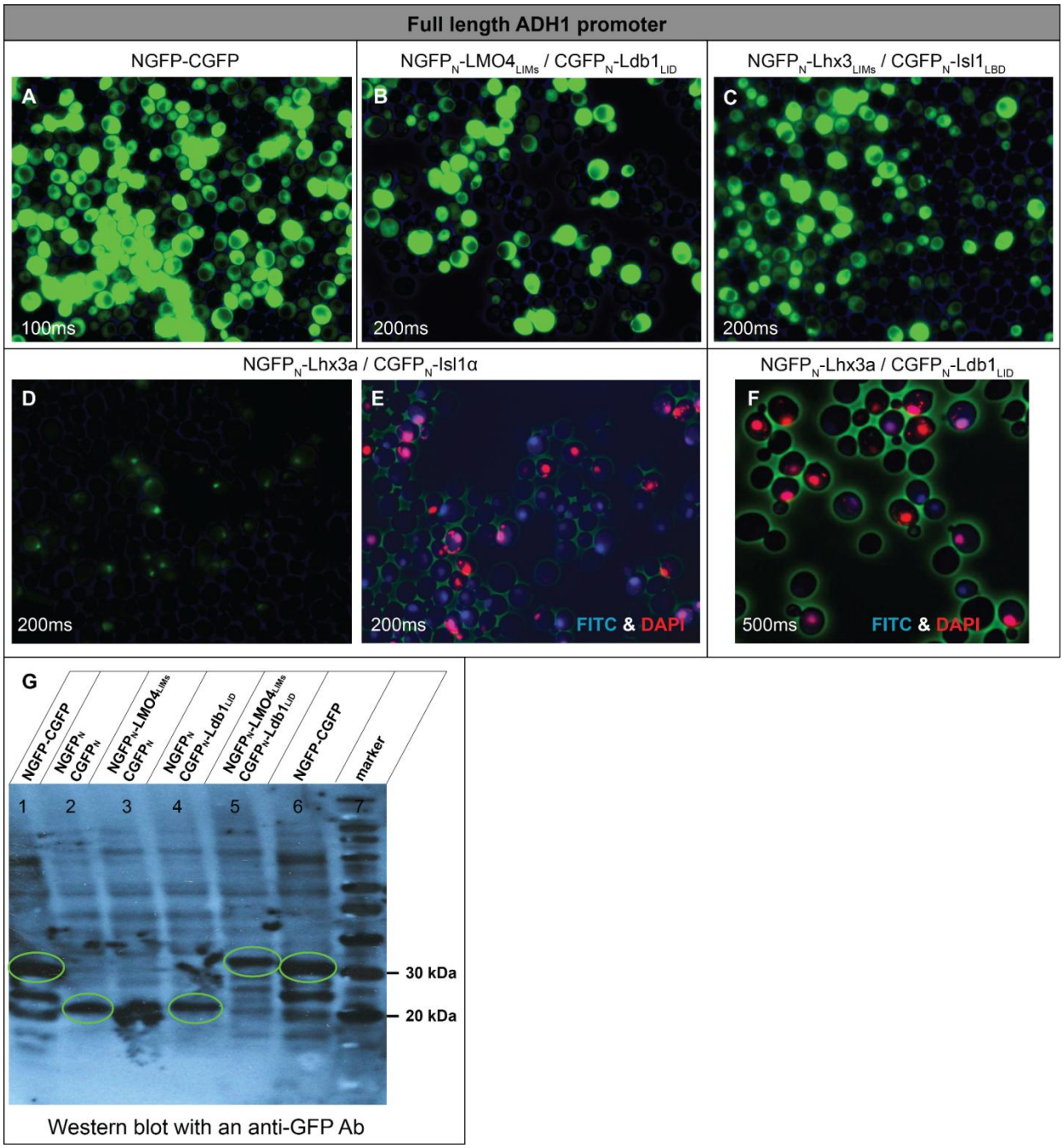


Figure 5.1. The effects of the full-length ADH1 promoter on fluorescence intensity. Constructs under the control of the full-length ADH1 promoter were expressed from pWAFL-NGFP$_N$ and pLAFL-CGFP$_N$ vectors. (A-D) BiFC images are composites of FITC (green) and bright field (blue) micrographs. (E-F) BiFC signal localization. FITC (blue), DAPI (red) and bright field (green) micrographs were overlayed. Purple indicates nuclear localization of complexes. Exposure times are shown in the bottom left corner of every image. (G) Western blot of yeast cell lysates from indicated transformants using an anti-GFP polyclonal antibody as the primary antibody. Relevant bands (see text) are indicated with green ovals.

His-tagged constructs prevented any meaningful comparison in expression levels, so no western blot experiments were carried out for these constructs.

In previous studies, a BiFC signal in yeast was detected by fluorescence methods using flow cytometry [264, 269-273]. This approach filters out and/or rapidly quantifies background cells that are not fluorescent because of a lack of interaction, in contrast to Y2H assays where those cells do not grow. As an alternative approach, solution state fluorimetry was used to try and distinguish between samples that contain interacting proteins and negative controls. Intact yeast cells that co-expressed $NGFP_N$-$LMO4_{LIMs}$ and $CGFP_N$-$Ldb1_{LID}$ under the control of full-length ADH1 promoter were resuspended in PBS buffer, normalised for cell density and scanned for bulk fluorescence in solution, and compared with a NCFP-CGFP positive control and a $NGFP_N$ vs. $CGFP_N$ self-assembly control (Figure 5.14). There was no evidence of GFP-fluorescence in the self-assembly control and GFP-fluorescence at 510 nm was more than 3-fold weaker in the $NGFP_N$-$LMO4_{LIMs}$ vs. $CGFP_N$-$Ldb1_{LID}$ interaction sample than in the NGFP-CGFP positive control. These results showed that the interaction was detectable by standard fluorimetry, and confirmed that the average BiFC fluorescence in case of a strong interaction was only ~30% of the wtGFP (NCFP-CGFP).

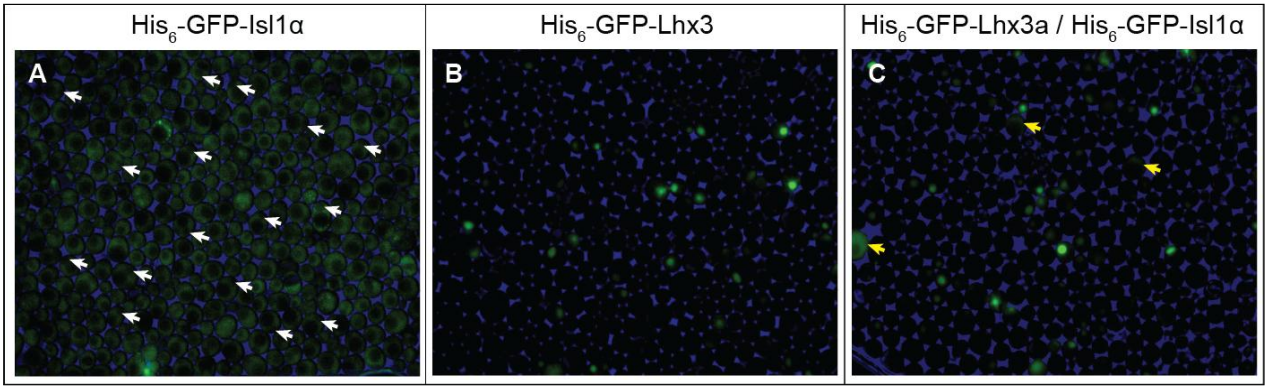


Figure 5.2. Fluorescence in cells expressing His_6-GFP-tagged full-length proteins from the full-length ADH1 promoter. The proteins were either expressed separately (A) and (B) or co-expressed (C). White arrows indicate some nuclei with depleted fluorescence and yellow arrows indicate cells with mostly cytoplasmic fluorescence.

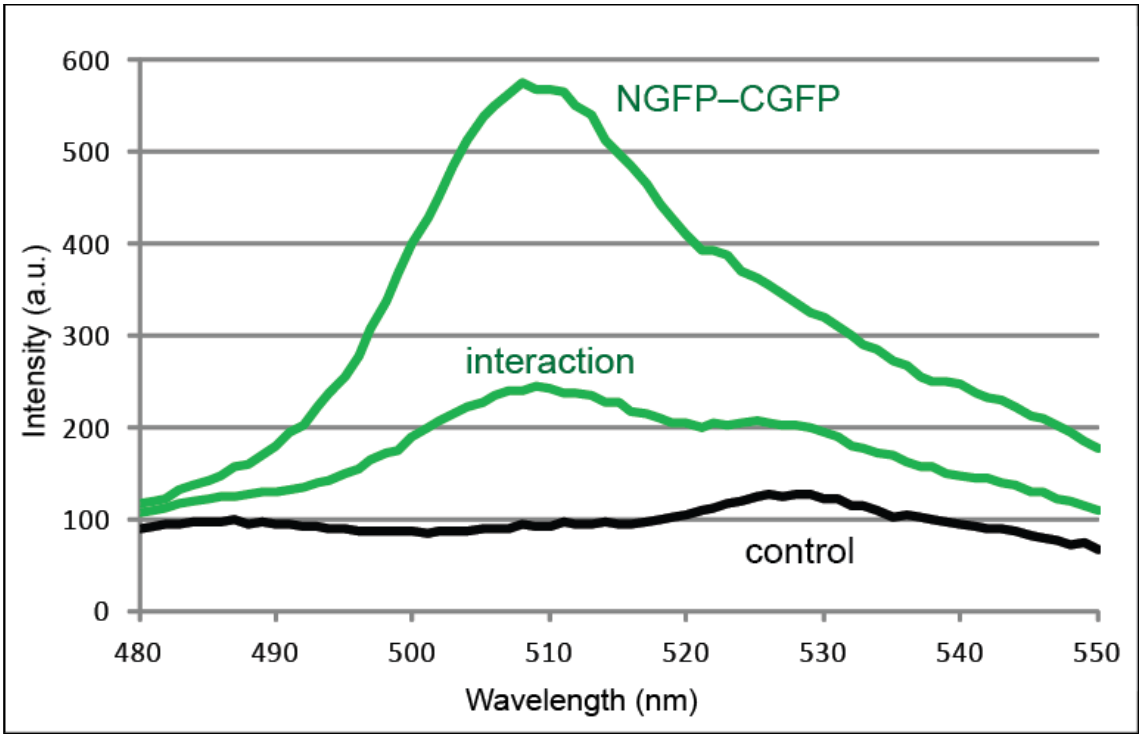


Figure 5.3. *In vivo* detection of a BiFC interaction. All constructs were expressed under the control of full-length ADH1 promoter. The yeast cell suspensions were excited at 475 nm and the emission spectra was recorded from 480–550 nm. The emission spectra for $NGFP_N$-$LMO4_{LIMs}$ vs. $CGFP_N$-$Ldb1_{LID}$ interaction and NGFP-CGFP positive control are shown in green and the $NGFP_N$ vs. $CGFP_N$ self-assembly control is in black. The cell suspension buffer was used as the blank.

5.8 Nuclear BiFC

The interactions between the full-length Lhx3a and Isl1α were associated with nuclear localisation (Fig 5.12 and 5.13). Whereas Lhx3a did seem to localize to the nucleus (Fig 5.13B), Isl1α did not (Fig 5.13B). Differences in sub-cellular localization are obvious features that could give rise to false negative results. For example, if one partner is directed to the nucleus and the other to the cytoplasm, there is a possibility that GFP complementation would be below the detection level because the fragments have little or no opportunity to encounter one another. Thus, a nuclear-localized version of the BiFC assay was designed by introducing the SV40 NLS sequence at the N-termini of the pW397-$NGFP_N$ and pL397-$CGFP_N$ plasmids (Figure 5.15). The resulting plasmids were named pW397-NLS-$NGFP_N$ and pL397-NLS-$CGFP_N$.

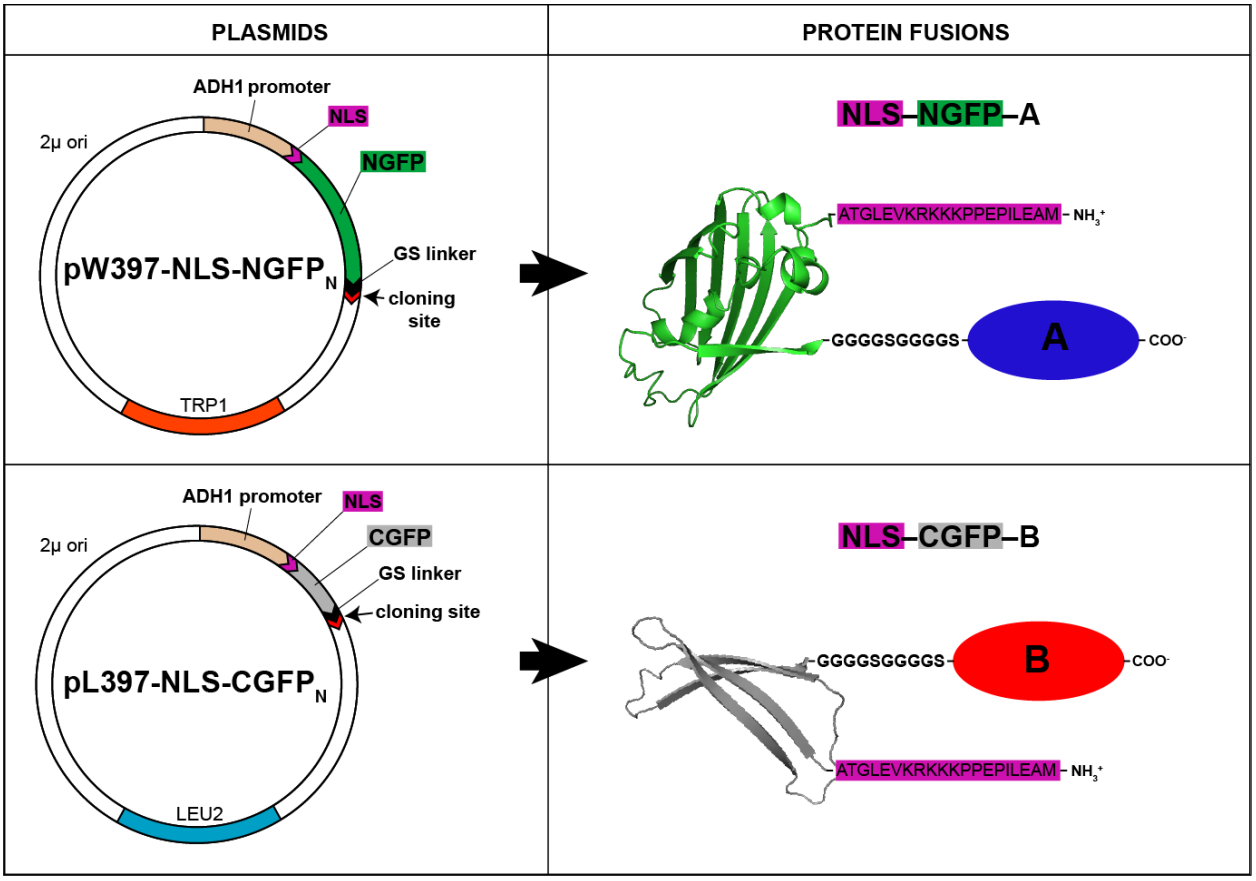


Figure 5.4. NLS-containing BiFC plasmids (left column) and models of the resulting protein fusion constructs (right column). The SV40 NLS sequence is coloured purple.

Yeast co-transformants containing NLS-NGFP$_N$-LMO4$_{LIMs}$ and NLS-CGFP$_N$-Ldb1$_{LID}$ and a self-assembly control, NLS-NGFP$_N$ vs. NLS-CGFP$_N$, were tested for fluorescence (Figure 5.16A and B). The resulting complexes showed strong and localized signals in the nuclei for both co-transformants. Unfortunately, however, the self-assembly control showed the same strong level of fluorescence in a similar proportion of the cell population. This phenomenon is presumably caused by the increased local concentrations of GFP-fragments leading to self-assembly, and indicates that NLS-tagging was unsuitable for further experiments.

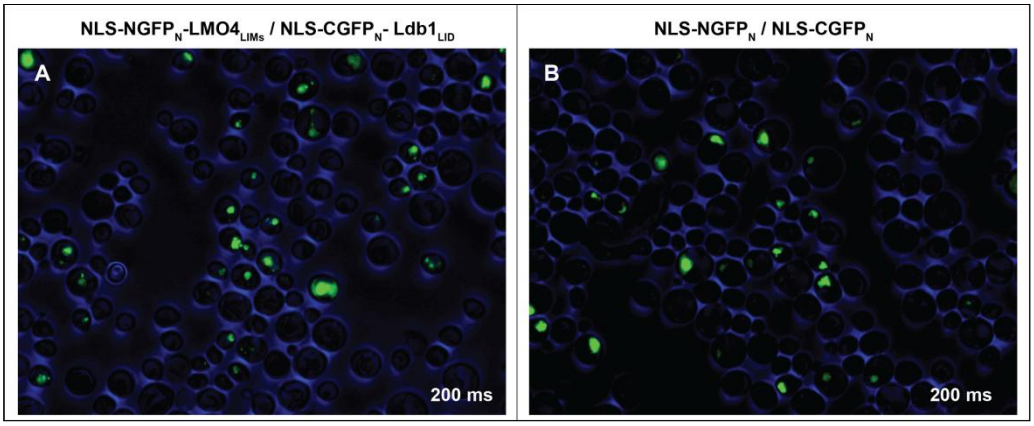


Figure 5.5. Co-localization-induced self assembly of the GFP-fragment fusion constructs. (A) NLS-NGFP$_N$-LMO4$_{LIMs}$ vs. NLS-CGFP$_N$-Ldb1$_{LID}$ (B) NLS-NGFP$_N$ vs NLS-CGFP$_N$ self-assembly control. The images are composites of FITC (green) and bright field (blue) micrographs. Exposure in both was 0.2 s.

5.9 Split-GFPsol variant in the BiFC system

In an effort to improve the BiFC signal in these assays, a switch was made from using the split-GFPuv protein to a variant that contains the mutation S65T to improve the fluorescent properties of GFP, and F64L to increase the folding efficiency [252]. When these mutants are present on the background of the wild-type protein the resultant protein is called enhanced GFP (EGFP), and when on the background of GFPuv the resultant protein is referred to as GFPsol (or GFPuv3/GFP+). GFPsol is reported to have faster maturation and eight times more brightness than GFPuv [274, 275] and is more compatible with sensitive detection using standard FITC filters (Olympus BX51 System microscope manual). Thus, the mutations S65T and F64L were introduced into the split-GFPuv constructs in appropriate NGFP-containing vectors. The resulting vectors were named pW397-NGFPsol$_N$ and pW397-NGFPsol$_C$.

The GFPsol-BiFC system was used to retest the interaction between full-length Lhx3a and Isl1α. The NGFPsol$_N$ vs. CGFP$_N$ orientation that was optimal for the head-to-tail interaction between LIMs and LBD domains might not be compatible with the full-length constructs as the GFPsol fragments might be too far apart to be able to complement. Thus, all eight different combinations of the fusion constructs were tested to find the optimal orientation for this interaction. Unfortunately, all self-assembly controls showed detectable fluorescence in cytoplasm (Figure 5.17D, H, L and P), and some of the other controls had signals comparable or

even stronger than those in the interaction tests (Figure 5.17G, O and Figure 5.18C, H, I and K). This background signal probably represents nonspecific interactions between a split-FP fragment and sticky full-length LIM-HD proteins (Figures 5.17 and 5.18).

False positives were more common in cells with CGFP-containing fusion constructs of either protein compared to cells containing NGFPsol-fusion constructs (Figure 5.17G, O and Figure 5.18C, I), and they include all cases where Isl1α has a C-terminal tag (whether it be NGFPsol or CGFP).

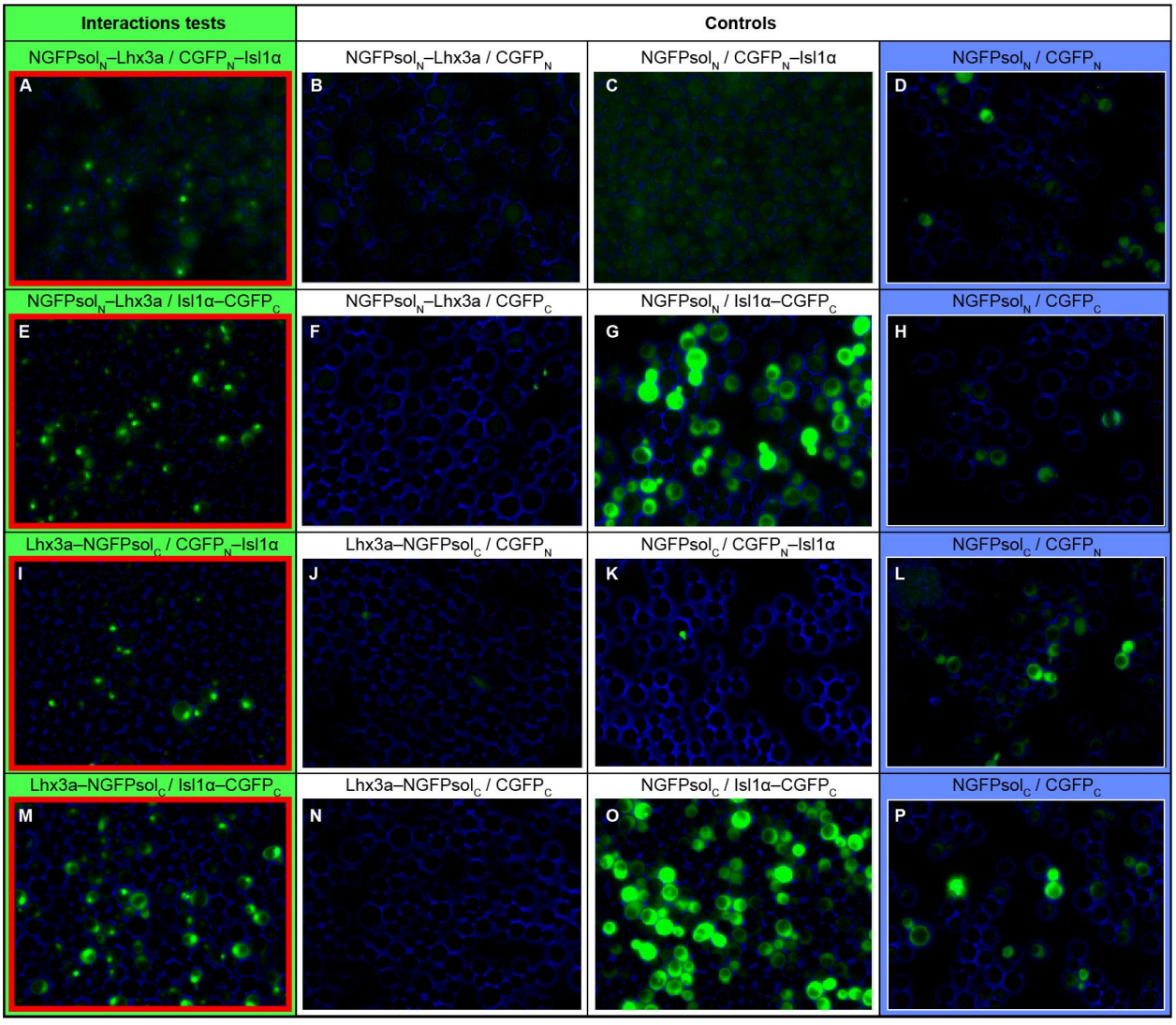


Figure 5.6. Optimization of orientations in the split-GFPsol BiFC assay - First four combinations. Interaction tests are shown in the first column on the left (light green background) and the relevant controls are in rows next to each interaction. The rightmost column (light blue background) contains self-assembly controls. Images showing nuclear-localised signals are in red rectangles. The images are composites of FITC (green) and bright field (blue) micrographs. Exposure was 1 s.

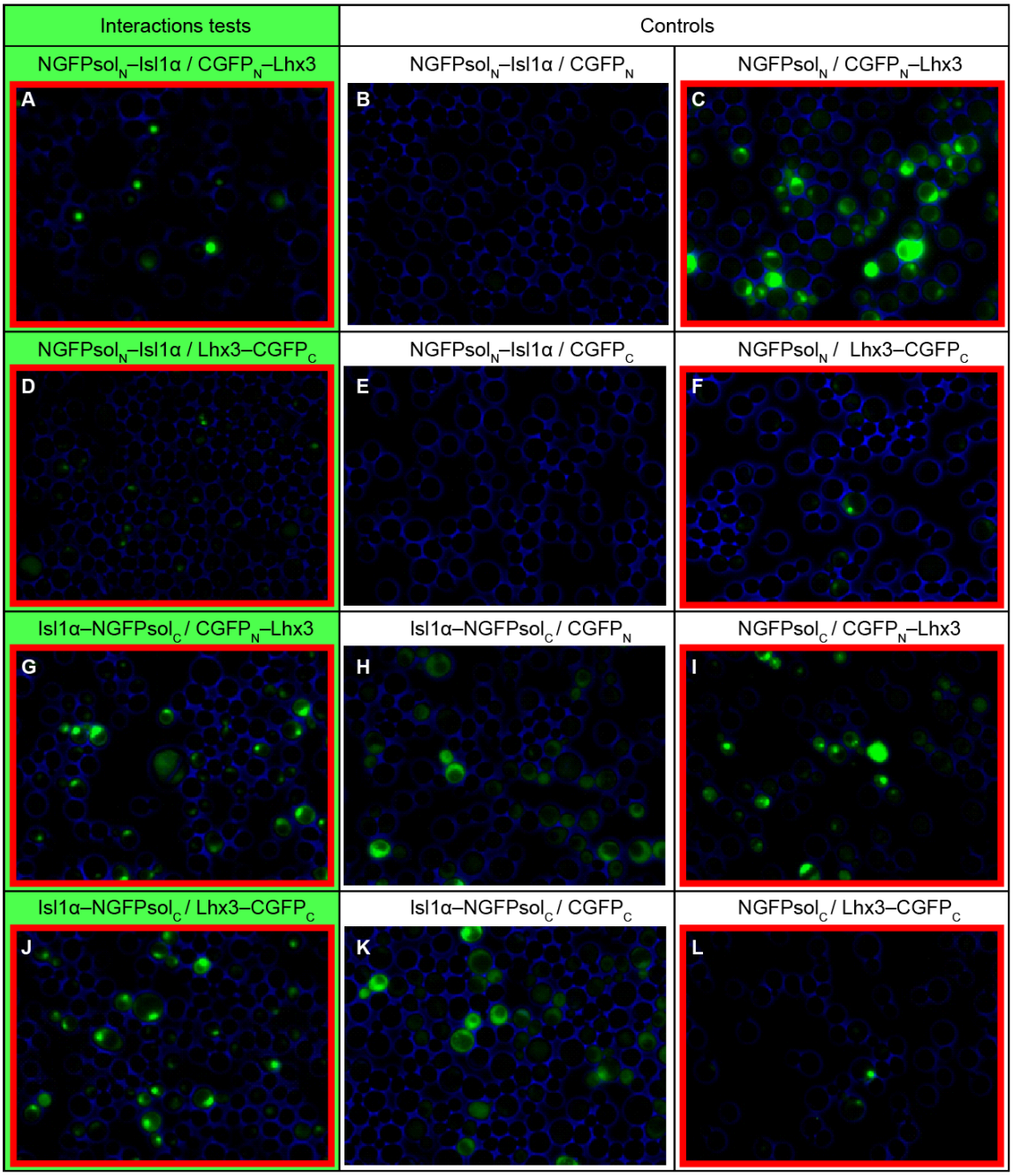


Figure 5.7. Optimization of orientations in the split-GFPsol BiFC assay - Last four combinations. Interaction tests are shown in the first column (light green background) on the left and the relevant controls are in rows next to each interaction. The self-assembly controls are shown in the previous figure. Images showing nuclear-localised signals are in red rectangles. The images are composites of FITC (green) and bright field (blue) micrographs. Exposure was 1 s.

Interestingly, nuclear localization was observed in interaction tests in all orientations, but only some controls showed a stronger ratio of nuclear to cytoplasmic signal. These controls all involved Lhx3a (Figure 5.18C, F, I and L). Fluorescing controls that involved Isl1α-containing constructs showed signals that were not specifically localized to the nucleus (i.e., were the same as the NGFP-CGFP construct alone).

5.10 Discussion

5.10.1 Comparison of BiFC and Y2H assays for detecting interactions between LIM-HD proteins

In this chapter, a BiFC assay was developed for detection of interactions between heterologous proteins in yeast. This assay was successful in detecting known interactions across a broad range of binding affinities for mouse and insect ZnF proteins including LIM domains. Although BiFC was able to detect an interaction between full-length Lhx3a and Isl1α proteins, the signals were not as robust as those observed for interactions between isolated domains. In BiFC assays that attempted to detect interactions between full-length Lhx3a and $Ldb1_{LID}$, full-length Isl1α or $Isl1_{LBD}$, fluorescent signals were predominantly localized in nuclei.

Yeast BiFC assays appear to be a viable approach for detecting LIM-peptide interactions, but ultimately did not reach the selectivity, signal to noise ratio and reproducibility achieved by standard Y2H assays. The assays described in this chapter apply epifluorescent microscopy for detection of protein-protein interactions, which limits the throughput of this assay format, in contrast to high throughput potential of Y2H assays. Although BiFC assays can detect an interaction, low confidence should be given to any analysis of relative interaction intensities or analysis of small mutagenesis effects based on this assay. The reason for this lies in the multitude of factors influencing the assay output. The lack of growth selection, irreversibility of BiFC complexes, variable levels of self-complementation, variable fluorophore maturation of different BiFC complexes (Figure 5.9) [276], greater sensitivity to steric effects, lack of proper negative controls, and different intracellular localization effects make BiFC assays in yeast inferior to Y2H assays for detection of heterologous protein interactions.

BiFC assays in yeast can be used as an alternative when autoactivation prevents detection in Y2H assays. Results from this chapter show that well optimized BiFC assay in yeast can be applied to test if separate constructs of LIM-HD C-terminal regions interact with LIM domains. In contrast to HD-containing constructs, subcellular localization is unlikely to cause false negative or false positive results when BiFC fusion constructs of C-terminal and LIM domain constructs are used.

5.10.2 Possible improvements of BiFC assays

Since this project began, numerous BiFC assays have been developed, which may provide possible improvements for the BiFC assays associated with LIM-HD proteins. The main areas for consideration are cell-to-cell variation in levels of fluorescence, detection, stability of constructs, and further consideration of localisation effects.

Cell to cell variation of fluorescence can be attributed in part to stochastic processes and in part to copy number variation from the multi-copy 2µ plasmids used in these assays [277, 278]. Copy number variation could be limited through the use of CEN/ARS plasmids or chromosome-integrated genes, although this approach tends to also lower overall fluorescence [279]. The temperature and time of yeast incubation are important factors that affect fluorophore maturation in BiFC assays. Automated fluorescence microscopy could be applied in time course experiments to compare the development of fluorescence between samples. However, variations in fluorescence intensity observed between replicates could lead to ambiguity. Some studies suggest that the population of highly fluorescent cells is higher when tested proteins interact and that flow cytometry can be applied to detect such differences between yeast cultures, leading to detection of protein-protein interactions [264, 270-272].

Levels of protein expression could be increased by optimizing the codon usage of the split GFP fragments for expression in yeast. The constructs used in this chapter had been optimised for expression in *E. coli.* Yeast-optimized split-yEGFP can give significantly higher expression levels [280] and was used in a previous BiFC study [281]. An additional approach would be to optimize the translation initiation site in BiFC plasmids [282]. Note, however, that overly high concentrations of fusion proteins can lead to self-assembly, which would lead to

more ambiguity in the assay outputs. Furthermore, high levels of recombinant proteins could be toxic to cells as might have occurred with His-tagged constructs in section 5.7.

Improved detection and stability of constructs could also help in the identification of interactions between unstable or transiently interacting constructs. Some attempts were made to improve brightness of FPs in this study and lower the interaction detection limit. Faster folding FPs could provide more protection of unstable proteolysis-prone fragments but could also lead to higher self-assembly rates. In these cases proper controls for self-assembly would be particularly important. Split-EYFP (enhanced YFP variant) could be applied instead of split-GFP, as it is reported to give about two times higher fluorescence and have relatively low self-assembly rates [254, 283].

Sub-cellular localisation. The main advantage of BiFC compared to Y2H experiments presented in chapters 3 and 4 is the potential for in situ, non-invasive, visual detection of protein interactions. Whereas Y2H vectors are directed into the nucleus via NLS tags from the plasmids, it is assumed (but usually not demonstrated) that both fusion constructs locate to the nucleus. In BiFC assays, the intrinsic properties of the proteins themselves play an important role for subcellular localisation. Fluorescence arising from interactions between domains that do not have known NLS sequences (LIMs vs. LID/LBD) were found in both the cytoplasm and the nucleus. Other constructs preferentially localised to the nucleus, most likely because of intrinsic NLS and HD sequences direct the protein into the nucleus where retention is enhanced by binding to DNA [284, 285]. In the case of full-length Isl1α:Lhx3a interactions, Isl1α could be protected from nuclear localisation because of intramolecular interactions (section 1.10) whereas Lhx3a is directed to the nucleus; different subcellular localisation prevents a robust interaction signal. Rather than promoting nuclear localisation, it might be possible to promote cytoplasmic localisation, through addition of nuclear export signals [286-288]. However, if some sequestered fusion constructs are reflecting in vivo protein localization, forced co-localization of such fusion constructs could only bring about biological artefacts (false positives that do not interact in vivo).

Note that even if improvements could be made for the BiFC assay, it is important to confirm and quantify interactions identified by this approach by orthogonal methods. The following chapters focus on attempts to develop assays to quantify LIM-partner interactions in vitro.

Chapter 6 Site-specific labelling of a LIMs-LID construct with fluorescein and competition fluorescent anisotropy binding assays

A part of this research focussed on determining the binding affinities of LIM-HD/LMO proteins for Ldb1. As the LIM domains of these proteins have limited solubility and tend to aggregate in solution, it was decided to take the approach of using fluorescence-based homologous competition assays based on tethered LIMs-LID complexes that contained protease sites in the linker (section 1.9.1). Thus, it was necessary to label the tethered LIMs-LID complexes. This chapter details the different strategies used to label the N-terminus of the LIM domains in the LIMs-LID constructs and attempts to use these proteins in fluorescence anisotropy experiments.

6.1 Methods for labelling proteins with synthetic fluorophores

The design of fluorescence-based assays requires selection of a fluorophore that is suited to monitor a particular property of a protein. For example, tryptophan fluorescence can be very useful to study the folding, conformational dynamics, and interactions of a protein, but it is limited to proteins that contain tryptophan residues and is complicated by the photophysical properties of tryptophan [289, 290]. The most frequent approach used to make fluorescent proteins is the attachment to the proteins of interest of fluorophores with emission spectra in the visible range of the electromagnetic spectrum. Another method is to use protein engineering to create a fusion construct in which an autonomously (intrinsically) fluorescent protein, such as the green fluorescent protein (GFP) or its variants, is tethered to the protein of interest (as described in Chapter 7). An alternative approach is to chemically attach a synthetic fluorescent compound to the protein. The advantage of this approach is that synthetic probes can be small, chemically stable and photo-stable. Compared to GFP proteins, they can have higher quantum yield (brightness), little-to-no nonspecific binding, are resistant to irreversible photobleaching, less sensitive to buffer conditions, and have simpler and better defined photophysical properties.

Table 6.1. Common methods for chemical labelling of proteins. Labelling of native functional groups. Reagents for labelling primary amines, thiols or carboxyl groups are shown.

	Target groups	Reagents
Labeling of native functional groups	Amine	Isothiocyanates NHS esters Sulfonated esters Tetrafluorophenyl (TFP) esters SDP esters Carboxyl azides Sulfonyl chlorides
	Thiol	Maleimides Iodoacetamides Benzylic halides Bromo methylketones
	Carboxyl	Hydrazines, hydroxylamines or amines mediated by carbodiimide (e.g. EDC)

A large number of small fluorophores are commercially available, of which various fluorescein derivatives are most commonly used. Different derivatives have specific photophysical properties and contain specific reactive groups that target appropriate functional groups in proteins (Table 6.1) [291-293]. Fluorescein isothiocyanate (FITC) and NHS-fluorescein are often used for labelling primary amines (N-terminal amine groups and lysine side chains) and fluorescein-maleimide is used for labelling cysteine side-chains. These, and several other compounds, label their target functional groups in native proteins but generally lack site-specificity and may therefore not be optimal for investigating protein-protein interactions.

There are several approaches that do allow for site-specific (usually N- or C-terminal) chemical labelling of proteins [290-292, 294]. This is achieved with the introduction of unique target groups, usually via expressed tags modified by an enzymatic or autocatalytic reaction (Table 6.2).

The most common covalent labelling strategies involve native chemical ligation, different self-labelling enzymatic tags, enzyme-mediated labelling (especially with CP-tags and PPTase) or CLICK chemistries. Proteins can also be labelled site-specifically through non-covalent labelling by chelation of metal-/metaloid-dye conjugates (e.g., through His_6 or tetracysteine tags) or other types of protein-ligand binding (e.g., enzyme-inhibitor interactions of DHFR).

Table 6.2. Site-specific chemical labelling of proteins. Target groups (tags) and reagents are shown. In green and red font are the two methods applied in this study for labelling LIMs-LID constructs *in vitro*. EPL and IPL stand for Expressed and Intein-mediated protein ligation, respectively. H and S in superscript mark azide-alkyne Huisgen cycloaddition and Staudinger ligation, respectively.

			Target groups	Reagents
Site-specific labeling	Intein-mediated	Expressed Protein Ligation	N-terminal cysteine	Thioester-derivatives
			C-terminal intein junction or C-terminal thioester	Cysteinyl/ Cys-peptidyl derivatives
		Protein trans-splicing	N-terminal intein C	labeled peptide (amino acid) - intN
			C-terminal intein N	intC - labeled peptide (amino acid)
	Self-labeling enzyme tags		SNAP or CLIP	O^6-benzylguanine or O^2-benzylcytosine derivatives
			HaloTag	Primary alkylhalide-derivatives
	Enzymatic labeling		Carrier protein (CP) tags	CoA derivatives & PPTases
			Q-tag	Cadaverin derivatives & Transglutaminase
			Farnesylation motif	Geranyl alkynediphosphate derivatives & farnesyltransferase
			Sulfatase aldehyde tag	Formyl glycine generating enzyme & aminooxy or hydrazide derivatives
			N-terminal $G_{1\text{-}5}$ / $A_{2\text{-}5}$ tag	Labeled LPXTG/A peptide & sortase
	Non-covalent labeling		Biotinelated tag (BirA ligase)	Labeled streptavidin
			Tetracysteine tag	FlAsH-EDT2 or ReAsH-EDT2
			Fluorette tag	Texas red
			FKBP12 (F36V) tag	SLF-derivatives
			eDHFR tag	Methotrexate-derivatives
	CLICK chemistry		Azide or alkyne^H	Complementary alkyne or azide-derivatives
			Azide^S	Phosphine / phosphite-derivatives
			Oxyamine	Ketone-derivatives

6.2 Specific labelling of a SNAP-tagged LIMs-LID construct

The SNAP-tag was designed as a tool for the specific labelling of fusion proteins in living cells [295-297]. This labelling strategy depends on the ability of human DNA repair protein O^6-alkylguanine-DNA alkyltransferase (AGT), or its engineered version (SNAP-tag), to covalently label itself with a single molecule of an O^6-benzylguanine (BG) derivative (Figure 6.1A). For the BG-88 reagent, excitation and emission maximum are specified to be 506 nm and 526 nm, respectively [298].

For this project a commercially available pSNAP-tag(T7) vector was modified to introduce a His-tag (to aid in purification) and a TEV-protease site (for removal of the His_6-tag; Figure 6.1B) before the SNAP-tag. Both the original vector and the modified vector (named pT7HT-SNAP) appeared to be toxic to the standard cloning strain (DH5α), which made cloning very difficult (not shown). Successful transformants were extremely rare and any cell growth was slow, resulting in small colonies that often died.

The $LMO4_{LIMs}$-$Ldb1_{LID}$ construct was chosen for the initial labelling trials as it is stable and well characterized. The corresponding DNA construct was sub-cloned into the pT7HT-SNAP vector (Figure 6.1B). Toxicity continued to be a problem in our standard bacterial expression strains, BL21(DE3) and Rosetta 2 cells, probably due to leaky expression. Fortunately, BL21-AI cells (Table 2.13 and section 2.3.2.4) were able to tolerate the vector and were successfully used in expression trials of the SNAP-tagged construct. The resultant protein was largely insoluble (estimated >80%; Figure 6.1C); however, some protein could be purified on Ni-NTA agarose beads. The His_6-tag could not be cleaved on beads by TEV protease (not shown) so the tag remained on the protein in subsequent steps. Further purification using anion-exchange chromatography resulted in >90% pure protein as estimated by Coomassie-stained SDS-PAGE (Figure 6.1C). Purified SNAP-tagged protein was labelled with the BG-488 reagent and the free reagent was removed by dialysis (section 2.10.1). Absorbances at 280 and 506 nm were measured and the labelling efficiency was calculated to be ~27% (section 2.10.3).

A fluorescence emission spectrum was recorded for the 50 nM sample of the partially labelled protein (excitation wavelength 506 nm; Figure 6.1D). The first peak in each emission

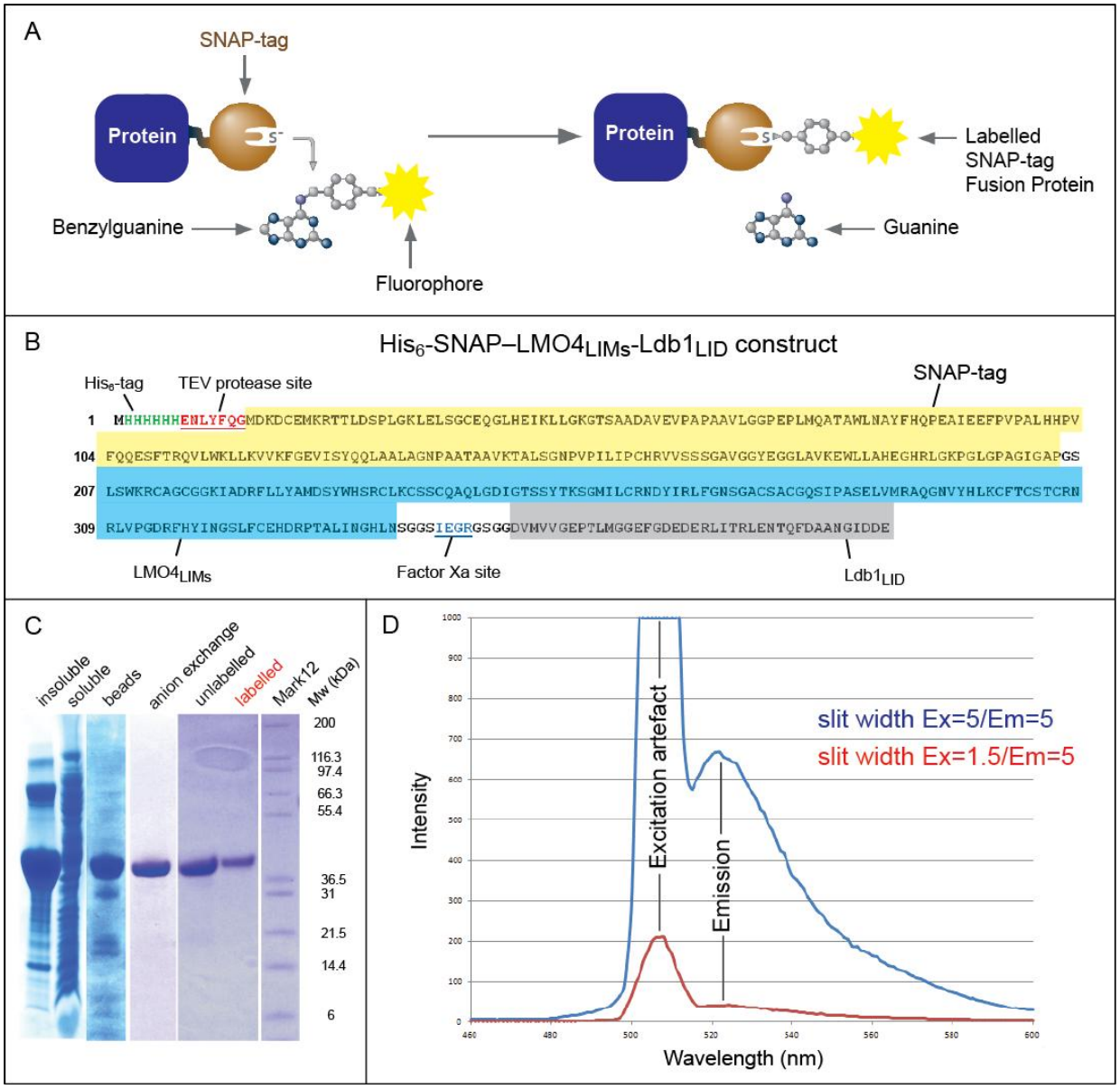


Figure 6.1. SNAP-tag enzymatic labelling of the $LMO4_{LIMs}$-$Ldb1_{LID}$ **construct with the BG-488 reagent.** (A) SNAP-tag labelling mechanism. SNAP-tag reacts with a benzylguanine derivative of a chosen fluorophore. Guanine is released and the SNAP-protein fusion construct becomes fluorescently labelled. (B) Amino acid sequence of the expressed protein construct His_6-SNAP-$LMO4_{LIMs}$-$Ldb1_{LID}$. Different domains are colour coded: the His_6-tag is in green font, the TEV protease site is in red font and the Factor Xa site is in blue font, SNAP-tag is highlighted in yellow, $LMO4_{LIMs}$ is highlighted in blue and $Ldb1_{LID}$ is highlighted in grey. (C) SDS-PAGE gels showing the solubility and purification of the construct with Ni-NTA agarose resin and anion-exchange chromatography. (D) Emission scans of the labelled and dialysed 50 nM sample, excited at 506 nm with different excitation slit widths. In both spectra, the first peak is an artefact of excitation and the second peak is the emission.

spectrum at 506 nm was likely due to reflected excitation light whereas the second peak at 526 nm is characteristic of the BG-488 label. The reflected emission peak, the intensity of which does not depend on the protein concentration, was significantly stronger than the emission peak at low labelled protein concentrations tested.

Unfortunately, the production of His_6-SNAP-$LMO4_{LIMs}$-$Ldb1_{LID}$ could not be repeated in subsequent attempts (not shown). It was considered that the low solubility of the expressed protein was an indicator of poor stability, including a tendency to aggregate, prompting the use of alternative labelling strategies.

6.3 Expressed protein ligation as a method for fluorescent labelling of proteins

Native chemical ligation (NCL) is a method commonly used in solid-phase peptide synthesis to join two synthetic unprotected peptides through formation of a native peptide bond, thereby producing an extended polypeptide. The C-terminal thioester of one molecule reacts with an N-terminal cysteine of the other. The first step of this process involves a reversible transthioesterification that is followed by a second, irreversible, S→N acyl shift step (Figure 6.2). Recombinant proteins can also be subjected to this general process in which case the method is called Expressed Protein Ligation (EPL). Both NCL and ECL are non-enzymatic, chemo-selective reactions that occur at physiological pH in aqueous solutions. These methods can also be used to attach fluorophores to protein termini in a site-specific manner [299].

Whereas C-terminal labelling involves attachment of a protein-thioester via a Cys-fluorophore reagent, proteins with N-terminal cysteines can be N-terminally labelled with fluorophore-thioesters. The former method usually requires an intein-mediated step to generate the thioester. The rate-limiting step in native chemical ligation is transthioesterification with the thiol group of the N-terminal cysteine [300, 301]. The most common thioesters used in NCL reactions are alkyl-thioesters as they are easy to produce and less susceptible to hydrolysis but are less reactive in transthioesterification.

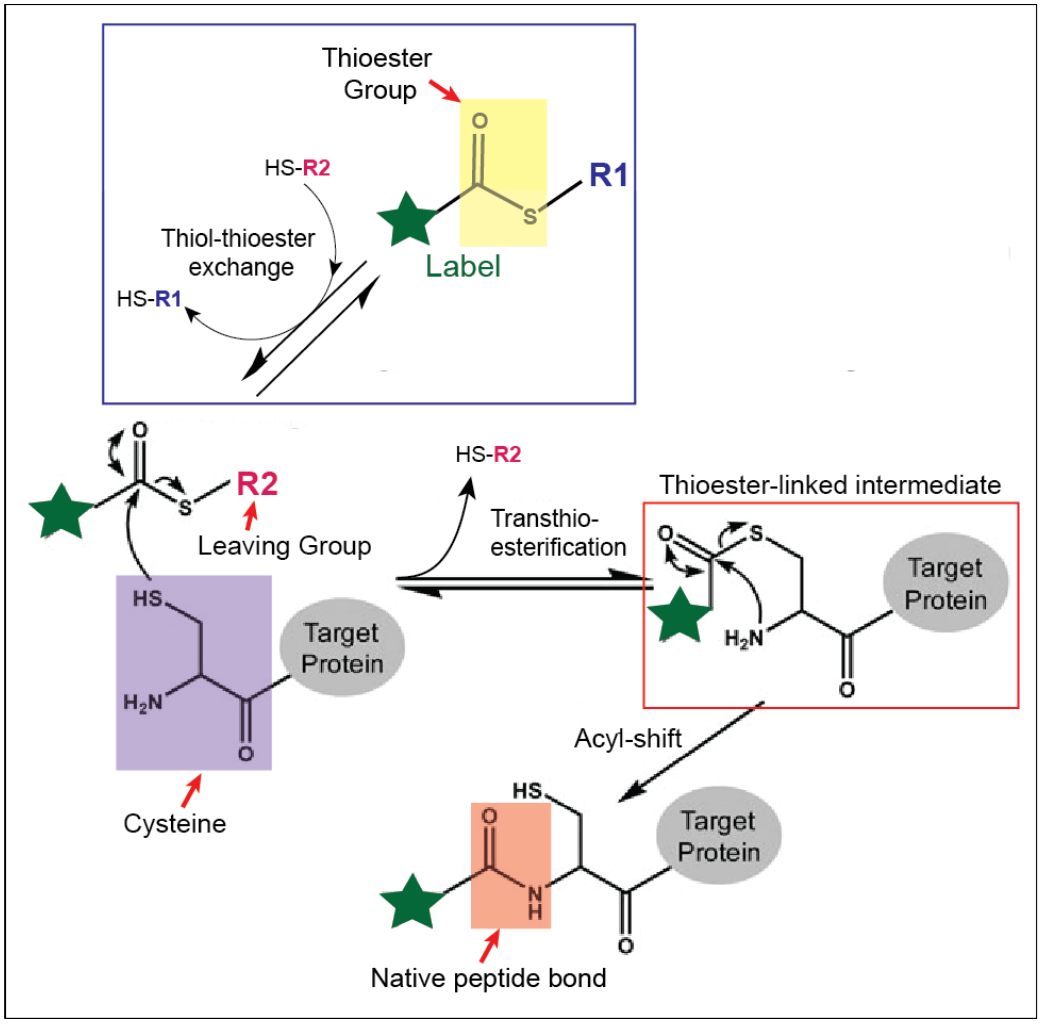


Figure 6.2. The mechanism of Native Chemical Ligation used for N-terminal protein labelling. The reaction consists of two consecutive steps: a reversible transthioesterification and an irreversible acyl shift. The thioester-linked intermediate is in the red box. The blue box shows an optional step of thiol-thioester exchange when a thiol with a better leaving group is added to the reaction.

Sodium 2-sulfanylethanesulfonate (MESNA) is a thiol reagent commonly used for making alkyl-thioseters. In NCL reactions it acts as a reducing agent and reverses any non-productive transthioesterification with the thiol moieties of side chains from internal cysteine residues [300]. Added thiols, if more active than the one that generated the thioester, can also promote the in situ formation of thioesters with better leaving groups by thiol-thioester exchange (blue box in Figure 6.2) and thus increase the kinetics of ligation [301].

The requirement for an N-terminal cysteine is a potential limitation of N-terminal labelling by EPL. Fortunately, such cysteine residues can be introduced in several ways. If the protein is expressed with a methionine at the N-terminus and a cysteine as the second residue endogenous methionylaminopeptidases (MAP) can remove the methionine leaving cysteine as the N-terminal residue. This approach has been achieved with variable efficiency depending on the downstream amino acid residues [302, 303]. For example, intein-mediated protein splicing can produce Cys-proteins (Figure 6.3A) if the inteins have been mutated to induce cleavage at the C-terminal splice junction [289]. However, the most common method for generating Cys-proteins is the generation of fusion proteins in which modified protease sites for Factor Xa, tobacco etch virus (TEV) protease and some other proteases, are introduced into the linker at the N-terminus of the target protein (Figure 6.3B). These proteases will tolerate a cysteine residue in the P1' site, so their proteolytic cleavage of constructs results in an N-terminal cysteine residue [289].

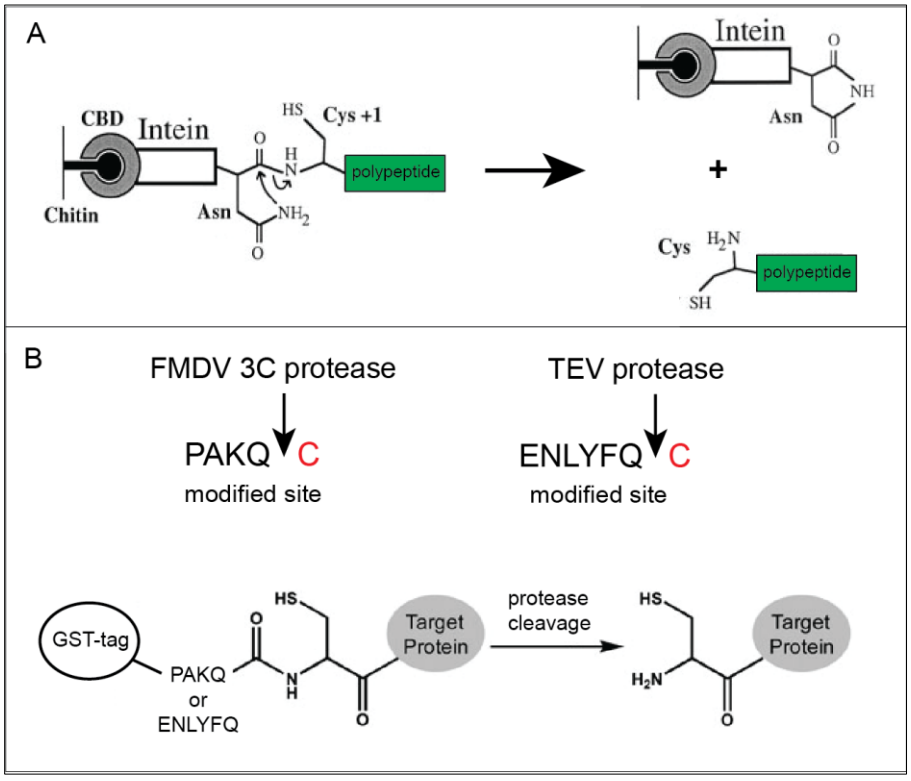


Figure 6.3. Methods for creating proteins with N-terminal cysteines. (A) Intein-mediated cleavage of expressed constructs on beads produces free Cys-polypeptides (green). (B) Modifications of FMDV 3C or TEV protease sites that generate an N-terminal cysteine upon cleavage, and the proteolytic cleavage of the GST-fusion constructs on beads.

6.4 Making the proteases and vectors

Initially it was intended to use the Foot and Mouth Disease Virus (FMDV) 3C protease, which had been used previously to generate Cys-proteins (Figure 6.3B)[304]. However, attempts to express FMDV3C protease were not successful. Thus, TEV protease was used [305]. His_6-tagged TEV protease was expressed in BL21(DE3) cells and purified on Ni-NTA agarose beads followed by size-exclusion chromatography (Figure 6.4). Target proteins were generated by editing the pGEX-6P vector to replace the HRV3C site with modified cleavage sites for TEV protease (ENLYFQC site; pGEX-TVC vector; Figure 6.3A). For details see Appendix B.

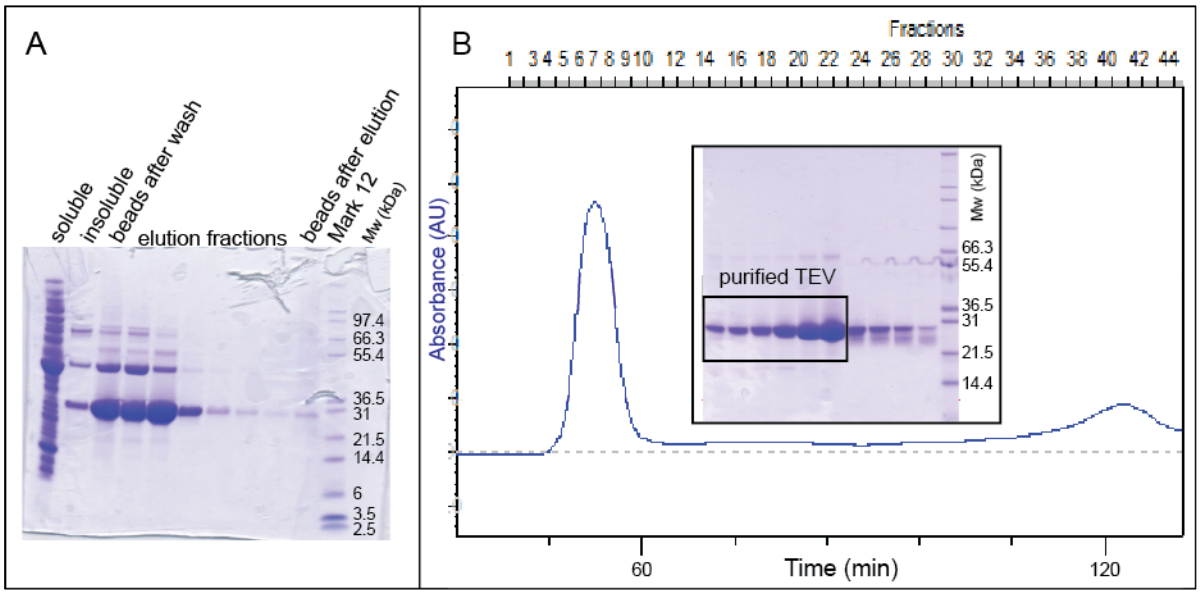


Figure 6.4. Purification of the His-tagged TEV protease. (A) Affinity purification on Ni-NTA agarose beads and elution with 250 mM imidazole. (B) Size-exclusion chromatography (SEC) of the His_6-TEV construct on a Superdex 200 column. The blue trace shows absorbance at 280 nm. The inserted photo shows an SDS-PAGE of SEC fractions.

6.5 Production of tethered constructs with N-terminal cysteines

Cleavable LIMs-LID constructs that contained either LMO4, or Lhx1–4 (for later experiments) were sub-cloned into the pGEX-TVC vector. GST-tagged constructs were expressed in BL21(DE3) cells and purified on GSH-agarose beads. TEV was added to remove the GST-tags and the free Cys-LIMs-LID constructs were collected. Proteolytic cleavage by

TEV was confirmed by SDS-PAGE (Figure 6.5A). Subsequent anion-exchange chromatography resulted in >90% pure protein as judged by Coomassie-stained SDS-PAGE (Figure 6.5B).

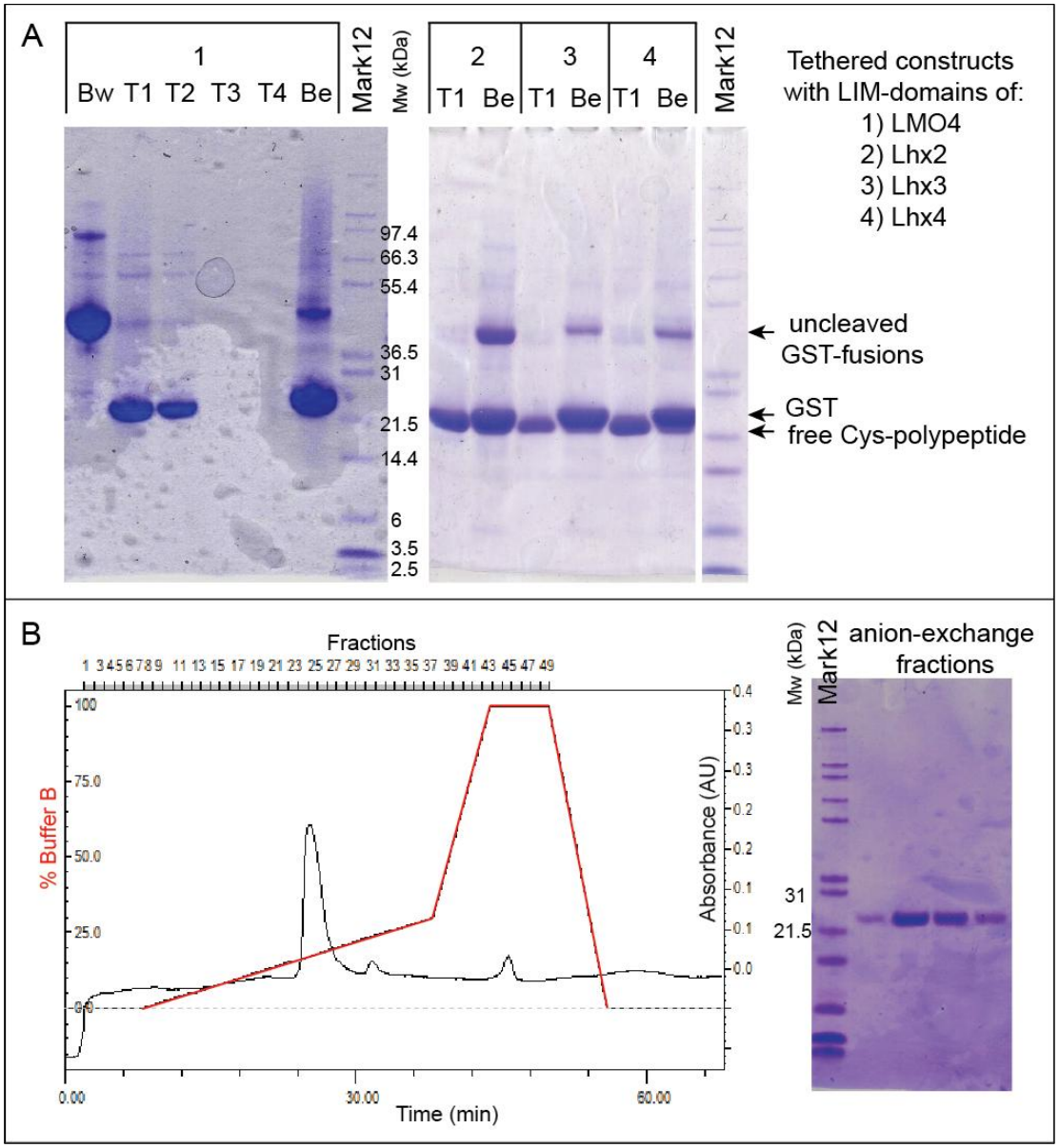
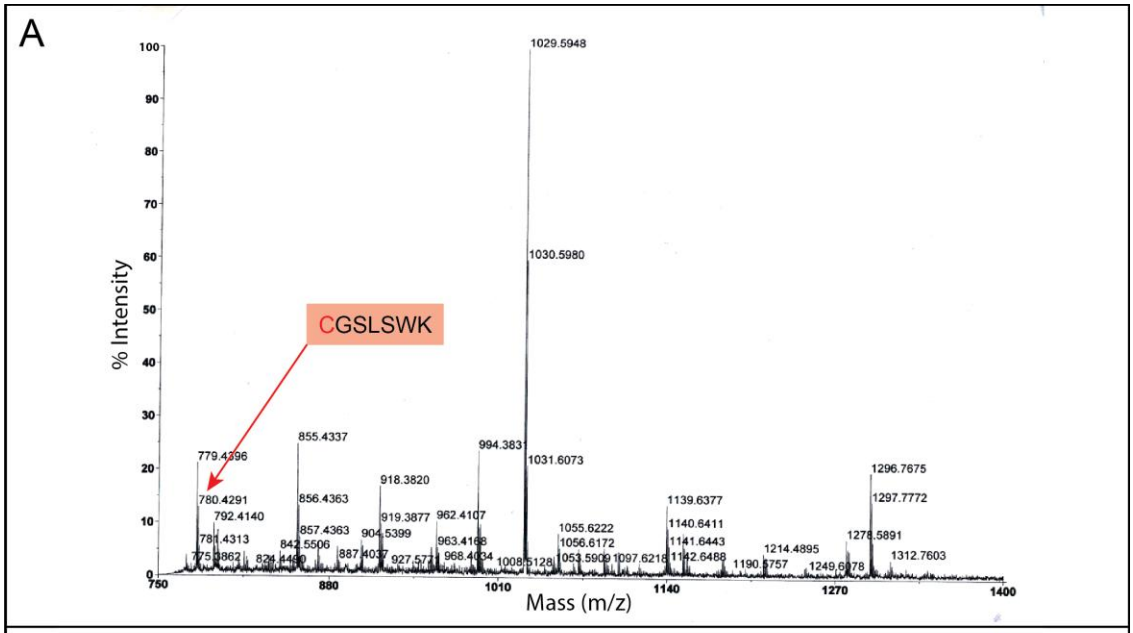


Figure 6.5. Purification of Cys-LIMs-LID constructs. (A) SDS-PAGE of samples from the affinity purification of GST-Cys-LIMs-LID constructs (labelled as GST-fusion constructs in this figure) immobilised on GSH-beads, subsequent overnight cleavage with TEV protease, and collection of untagged Cys-LIMs-LID constructs (labelled as Cys-polypeptides in this figure). Bw - fusion construct on beads after washing, Be - beads after cleavage, M - protein standard Mark12. (B) Anion-exchange chromatography of Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$. Black trace is the absorbance at 280 nm and red trace shows % buffer B (with 1 M NaCl). Fractions 23–27 contained the pure protein, as indicated by SDS-PAGE (right).



B

Position of cleavage site	Name of cleaving enzyme(s)	Resulting peptide sequence	Peptide length [aa]	Peptide mass [Da]	Cleavage probability
7	Trypsin	CGSLSWK	7	779.909	100 %
8	Trypsin	R	1	174.203	100 %
15	Trypsin	CAGCGGK	7	594.701	100 %
19	Trypsin	IADR	4	473.530	100 %
32	Trypsin	FLLYAMDSYWHSR	13	1688.921	100 %
35	Trypsin	CLK	3	362.488	100 %
53	Trypsin	CSSCQAQLGDIGTSSYTK	18	1849.017	100 %
60	Trypsin	SGMILCR	7	778.983	100 %
65	Trypsin	NDYIR	5	679.731	100 %
88	Trypsin	LFGNSGACSACGQSIPASELVMR	23	2298.632	100 %
97	Trypsin	AQGNVYHLK	9	1029.164	100 %
105	Trypsin	CFTCSTCR	8	920.084	100 %
107	Trypsin	NR	2	288.307	100 %
113	Trypsin	LVPGDR	6	655.752	100 %
127	Trypsin	FHYINGSLFCEHDR	14	1737.910	66.9 %
145	Trypsin	PTALINGHLNSGGSIEGR	18	1792.969	100 %
170	Trypsin	GSGGDVMVVGEPTLMGGEFGDEDER	25	2540.713	100 %
174	Trypsin	LITR	4	501.627	100 %
189	**end of sequence**	LENTQFDAANGIDDE	15	1651.660	-

Figure 6.6. Detection of the N-terminal peptide of Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ by MALDI-TOF. (A) MALDI-TOF mass to charge (m/z) vs. % intensity diagram. (B) PeptideCutter prediction of peptides generated by the trypsin digest of Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$. The CGSLSWK N-terminal peptide peak is indicated in each panel.

To confirm that the proteolysis did generate an N-terminal cysteine residue, a band expected to contain the Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ construct was excised from an SDS-PAGE gel and subjected to in-gel tryptic digestion and MALDI-TOF mass spectrometry. The MALDI-TOF spectrum showed a 780.4291 m/z peak that most likely corresponds to the 7-residue N-terminal

CGSLSWK peptide (highlighted in red; Figure 6.6A). This value agrees with the mass predicted by the PeptideCutter tool (779.91, highlighted in red; Figure 6.6B) [306].

Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ was characterized by far-UV circular dichroism (Figure 6.7A). The spectrum was essentially identical to spectra of similar tethered LIMs-LID complexes previously obtained in our laboratory, such as the spectrum of LMO2$_{LIMs}$-Ldb1$_{LID}$ (Figure 6.7B), indicating that the N-terminal cysteine did not significantly alter the folding of Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ [19, 150, 307].

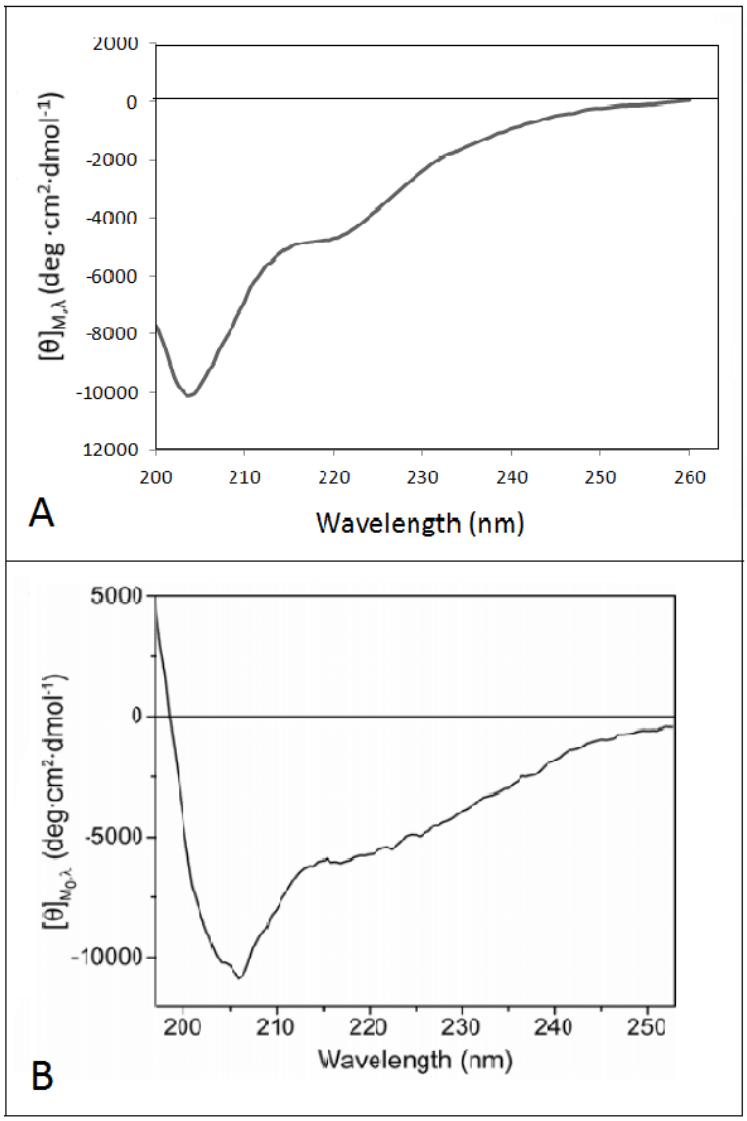


Figure 6.7. Characterisation of the Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ construct by far-UV circular dichroism. (A) spectrum of Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$. (B) The spectrum of LMO2$_{LIMs}$-Ldb1$_{LID}$ recorded by Dr Daniel Ryan (D. Ryan PhD thesis, Sydney University, 2005)

6.6 Production and purification of Fluorescein-2MES

The labelling reagent fluorescein-2MES (fluorescein-MESNA; Figure 6.8) was synthesised by combining 5-(6)-carboxyfluorescein succinimidyl ester (CFSE) with MESNA.

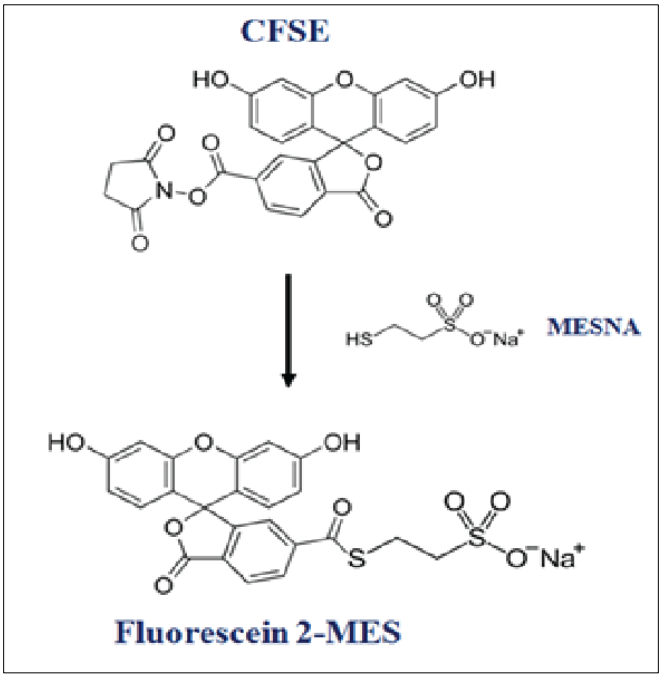


Figure 6.8. Synthesis of the fluorescein-2MES reagent. CFSE reacts with the thiol MESNA to create the fluorescein-2MES product.

Initial attempts at synthesis attempts and purification (Figure 6.8, more details in section 2.10.2.1), included reverse-phase HPLC using a C18 column with different phase gradients of water-acetonitrile (0.1% TFA) or water-methanol (0.1% TFA). Under all conditions the separation of products was poor (Figure 6.9A), and MALDI-TOF mass spectrometry of collected samples showed the presence of multiple products, but not fluorescein-2MES. Given limited facilities for carrying out chemical synthesis in the laboratory, a collaboration was established with Dr Brendan Wilkinson (Payne laboratory, School of Chemistry, University of Sydney) to produce Fluorescein-2MES in DMF under an inert argon atmosphere, and subsequently purify the reagent using a C4 column with a water-methanol gradient (Figure 6.9B). Subsequent confirmation of the identity of the product by 1D 1H NMR was carried out as part of this project (Figure 6.9C).

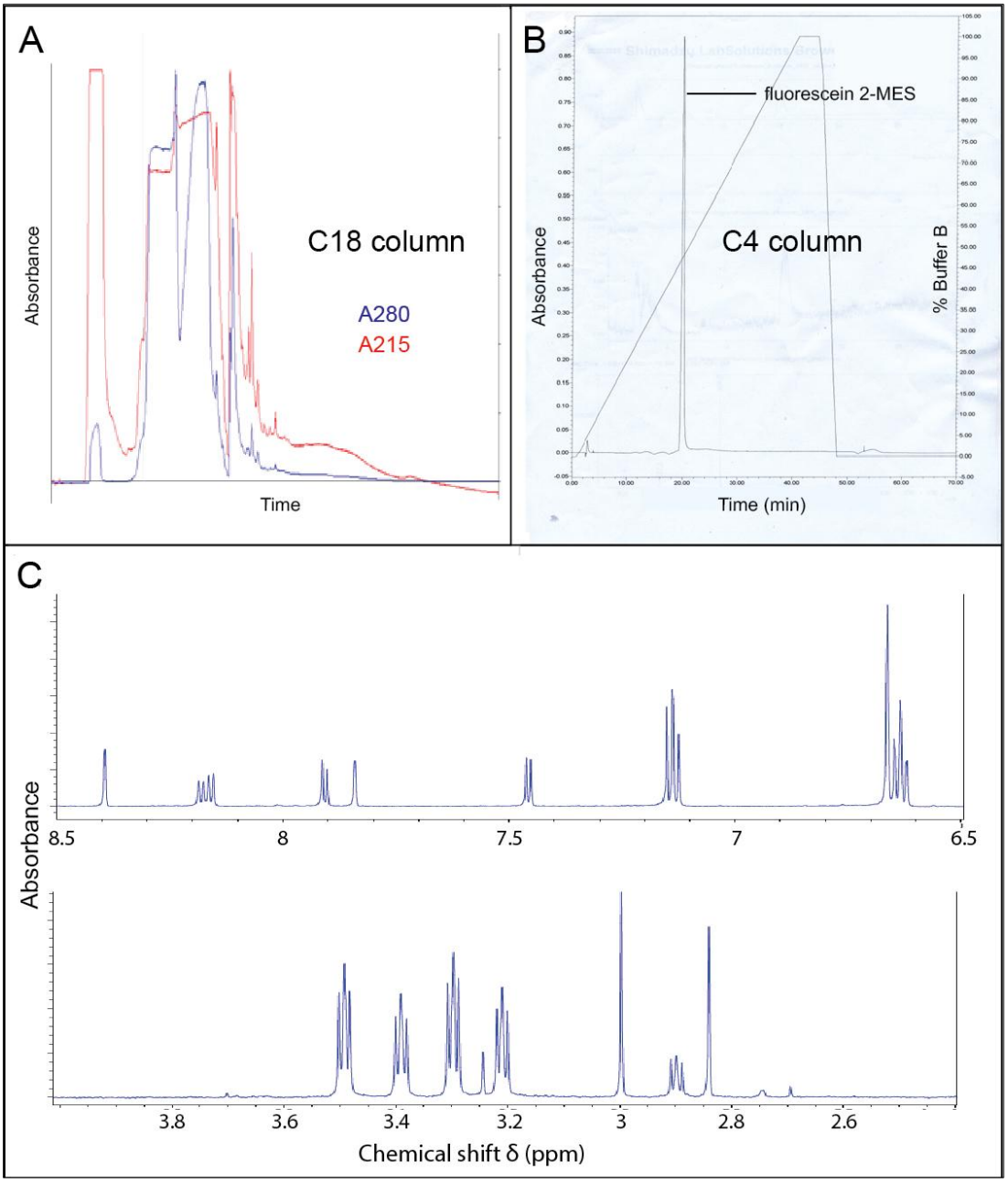


Figure 6.9. Purification and detection of the fluorescent reagent fluorescein-2MES. (A) The reaction was performed as described in section 2.10.2.4 and the purification was attempted by reverse phase HPLC with a C18 column in an 0–1 M acetonitrile (0.1% TFA) gradient. (B) Fluorescein-2MES was successfully purified using a C4 column. (C) ^{1}H NMR spectrum of the purified product. The chemical shifts in two ranges 2.4–4 ppm (bottom) and 6.5–8.5 ppm (top) are shown.

6.7 Labelling and purification of fluorescein-Cys-LIMs-LID constructs

Labelling of Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ was initially performed as described in Gentle et al., (2004) [303] (see section 2.10.2), except that the protein was not lyophilized and subsequently re-solubilized to avoid possible folding problems associated with these steps. Fluorescence scans (using FITC filter sets) of the SDS-PAGE gels showed fluorescein-labelled Cys-constructs of approximately correct sizes (Figure 6.10, green stars). However, two strong fluorescent bands of low MW were also observed (red stars, Figure 6.10A).

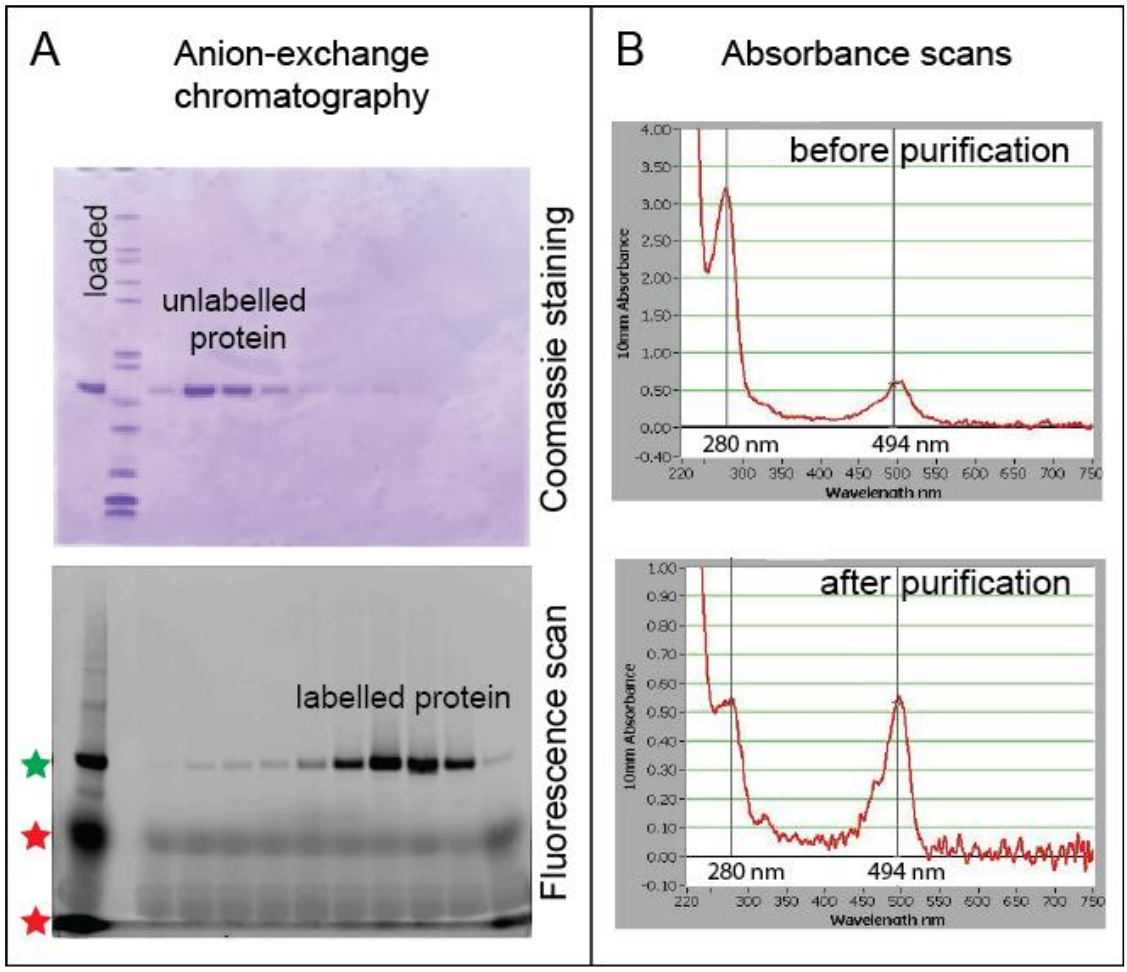


Figure 6.10. Removal of non-covalently bound fluorescent contaminants by anion-exchange chromatography. (A) Anion exchange chromatography of column-desalted, labelled Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$. The upper panel shows a Coomassie-stained SDS-PAGE gel of anion exchange fractions; the lower panel shows the fluorescence scan of the same gel. Labelled Cys-LIMs-LID is marked with a green star. Red stars indicate fluorescent bands of lower MW. (B) Absorbance scans of protein samples before and after the desalting and ion exchange. Vertical lines mark absorbance at 280 nm and 494 nm.

Attempts were made to remove unligated dye and purify the protein from the reaction mixture by various approaches including RP-HPLC, dialysis, use of a PD-10 desalting column (GE Healthcare), and buffer modifications. The best results were obtained when the stopped reaction was desalted with a PD-10 column and then purified by anion-exchange chromatography. The procedure removed a high percentage of the contaminating species and showed that the peak containing the labelled protein eluted with higher salt than the unlabelled protein. However, SDS-PAGE in combination with Coomassie staining and fluorescence scanning following anion-exchange chromatography showed that only a small percentage of protein was labelled, as fractions that showed fluorescence could not be detected by Coomassie staining (Figure 6.10A).

The labelled protein fractions were concentrated and absorbance scans at 280 nm and 494 nm of the samples before and after anion-exchange chromatography were compared (Figure 6.10B). The labelled protein now corresponded to ~40% of the total. Purified labelled proteins were desalted using Zip-tips (Millipore) and subjected to MALDI-TOF. Surprisingly, spectra showed only the unlabelled protein peaks (Figure 6.11). MALDI-TOF has been successfully

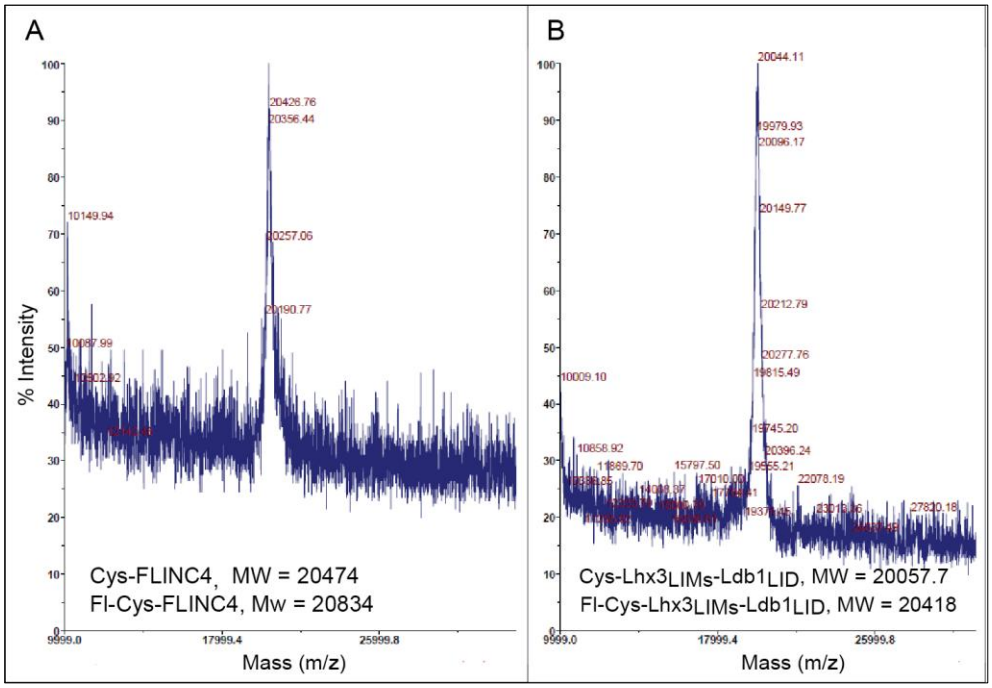


Figure 6.11. MALDI-TOF spectra of labelled protein samples. (A) Labelled Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ and (B) Labelled Cys-Lhx3$_{LIMs}$-Ldb1$_{LID}$. Spectra show only the unlabelled protein peaks.

used to detect other fluorescently labelled proteins [303, 304], so it is possible that fluoresein-labelled LMO4-LID may not be readily detected by MALDI-TOF. ESI-MS may be a better approach for detection of labelled products [308, 309], but access to this approach was limited during this study.

Labelling of other purified Cys-LIMs-LID constructs (Appendix C) has been attempted using the same approach, but all showed similar inefficiency of labelling and difficulties in unligated fluorescein removal.

6.8 Labelling trial with a small unstructured protein

To determine if the unexpectedly inefficient labelling and difficulties in removal of unligated dye was a result of problems with the protocol or the reagents used, or if it related to the proteins (Cys-LIMs-LID tethered constructs) being labelled, a smaller unstructured peptide was labelled under the same conditions. $DEAF1_{404–438(T435D)}$ [164] was cloned into the pGEX-TVC vector, expressed and a Cys-tagged construct was purified as described above (Figure 6.12A). In this case the labelling of the Cys-DEAF1 peptide was very efficient (Figure 6.12B).

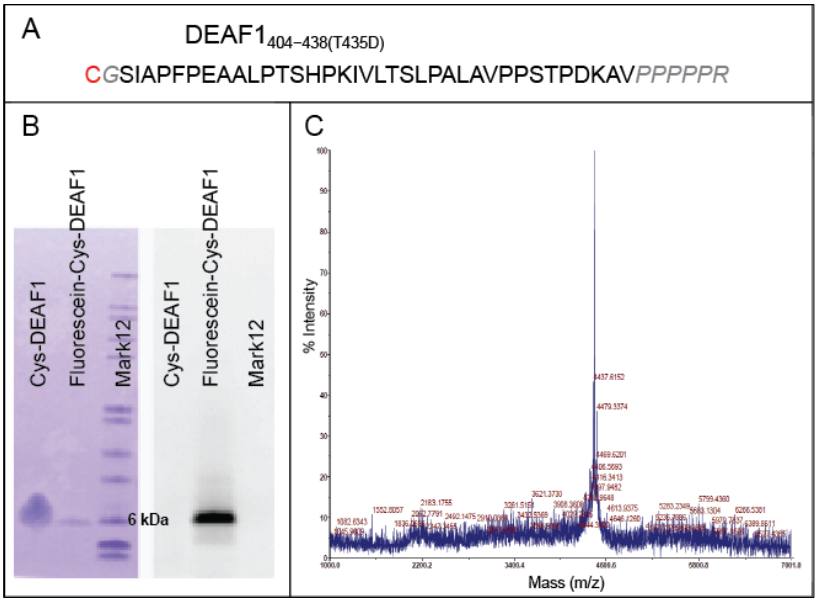


Figure 6.12. Fluorescent labelling of DEAF1 peptide by EPL. (A) Cys-$DEAF1_{404–438(T435D)}$ peptide sequence. The construct contains additional residues (in grey). (B) Scanned and Coomassie-stained gel with unlabelled and labelled peptide samples. Mark12 molecular weight markers were used. (C) MALDI-TOF of the labelled sample shows a single 4448.5 m/z peak corresponding to unlabelled Cys-DEAF1 peptide.

The desalted sample gave rise to a single strong band by SDS-PAGE corresponding to the expected size of the labelled protein, with a labelling efficiency estimated to be >90% (section 2.10.3).

The differences in labelling between the DEAF1 construct and tethered constructs could arise from the intrinsically disordered nature of the former, and the folded nature of the latter influencing the availability of the N-terminal cysteine. However, in the tethered constructs at least 4-5 unstructured residues separated the N-terminal cysteine and structured residues, making such an effect less likely.

6.9 The effects of different thiol reagents and protein concentration

The choice of a thioester type is an important factor for an NCL reaction rate (Figure 6.13A) [301]. Although MESNA-thioesters are thought to represent a good compromise between stability and reactivity, NCL reaction rates can be increased by the addition of more potent thiols. The effects of different Cys-protein concentrations and thiol reagents on NCL reactions were tested. Different molar ratios of Cys-protein:fluorescein-2MES (1:2–1:5), reaction times (2–16 h) and thiol regents (MESNA and MPAA) were trialled.

Labelling trials of Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ were carried out in the presence of excess MESNA and MPAA in parallel for 5 h (Figure 6.13B). The more reactive MPAA gave higher levels of labelling, but also stronger signals attributed to non-covalently bound dye, or products of cross-reactions with buffer components [310, 311]. Some differences in the relative amounts of non-covalently bound dye and its separation from the protein in PD-10 columns are evident when different concentrations of Cys-proteins were used.

Finally, a Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ labelling reaction with fluorescein-2MES in the presence of MESNA was incubated for > 48 h at room temperature. SDS-PAGE analysis of the PD-10 column-desalted sample showed strong labelling of the construct without the low MW fluorescent components that were characteristic of previous reactions (Figure 6.13C). The labelling efficiency was estimated to be >90% (section 2.10.3).

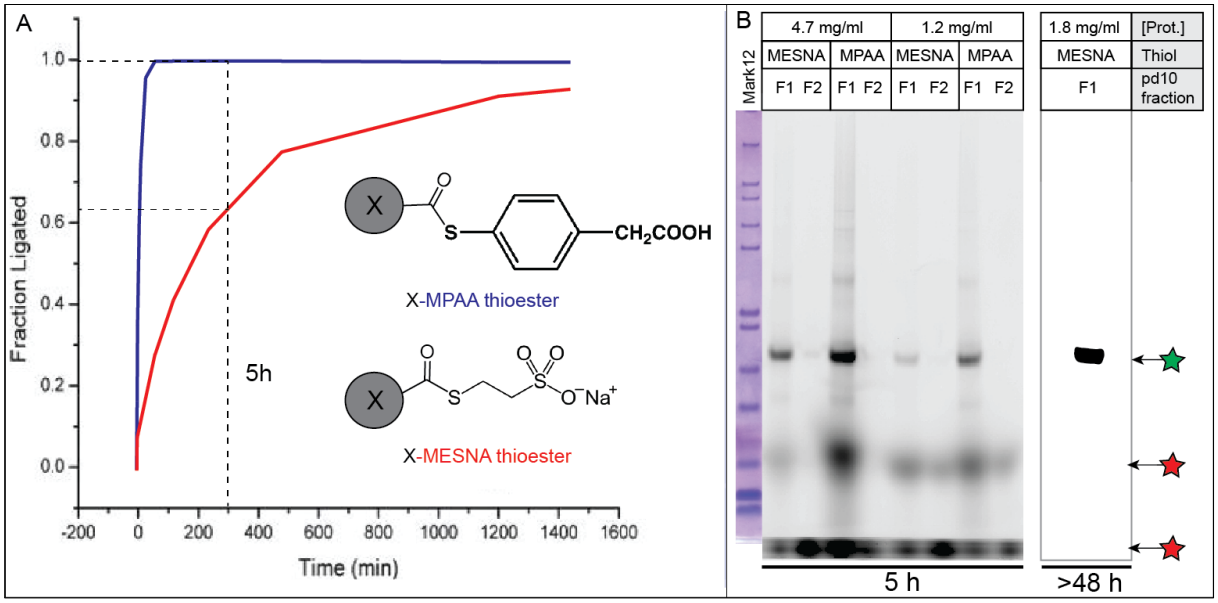


Figure 6.13. SDS-PAGE of labelling trials for Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ with MESNA or MPAA as thiol reagents. (A) A comparison of reaction rates of two different thioesters with a Cys-peptide (adopted from Johnson et al., 2006 [301]). The formulas of peptide X-MESNA and -MPAA thioesters are shown. Dashed lines mark yields observed after 5 h at room temperature. (B) SDS-PAGE gel of desalting column fractions F1 (and F2) obtained after labelling reactions with either MESNA or MPAA thiol present. The effects of incubation time as well as of a lower and a higher concentration of Cys-protein in the reaction are shown. The gel was scanned for fluorescence. The green star indicates labelled Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ and red stars indicate fluorescent species from unreacted dye and side-reactions with buffer components. Note that the gel on the right was scanned at lower intensity setting.

Although the yields of purified labelled proteins were low, a sufficient amount of Fl-Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ was obtained, which allowed the homologous competition binding experiments to be performed using fluorescence anisotropy.

6.10 Fluorescence anisotropy

Fluorescence-based methods are routinely used in protein binding affinity studies due to their ease of use, safety, sensitivity, time scale (in µs), reproducibility and adaptability to different assay formats [312]. Commonly used steady-state and time-resolved approaches include: fluorescence anisotropy (or polarization), fluorescence resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS) and related fluorescence cross-correlation spectroscopy (FCCS).

Both fluorescence anisotropy assays (FA assays), which measure anisotropy values (r), and polarization assays (measure P) report the same phenomenon, and are interconvertable, but

are calculated in slightly different ways. The phenomenon involves excitation by polarized light, which results in a partially polarized emission of the sample because a proportion of excited fluorophores do not change orientation during the short lifetime of fluoroscence (~1–10 ns). Smaller species can rotate rapidly in solution meaning that a smaller proportion of the sample remains polarised (leading to lower anisotropy). Larger fluorescent species, such as complexes, tend to have higher anisotropies as they rotate more slowly in solution. The extent of the rotational diffusion depends on the size and shape of the rotating molecule [313]. Steady-state FA assays require that the complex has a significantly increased anisotropy compared to the anisotropy of a free fluorescently-labelled complex component (Figure 6.14).

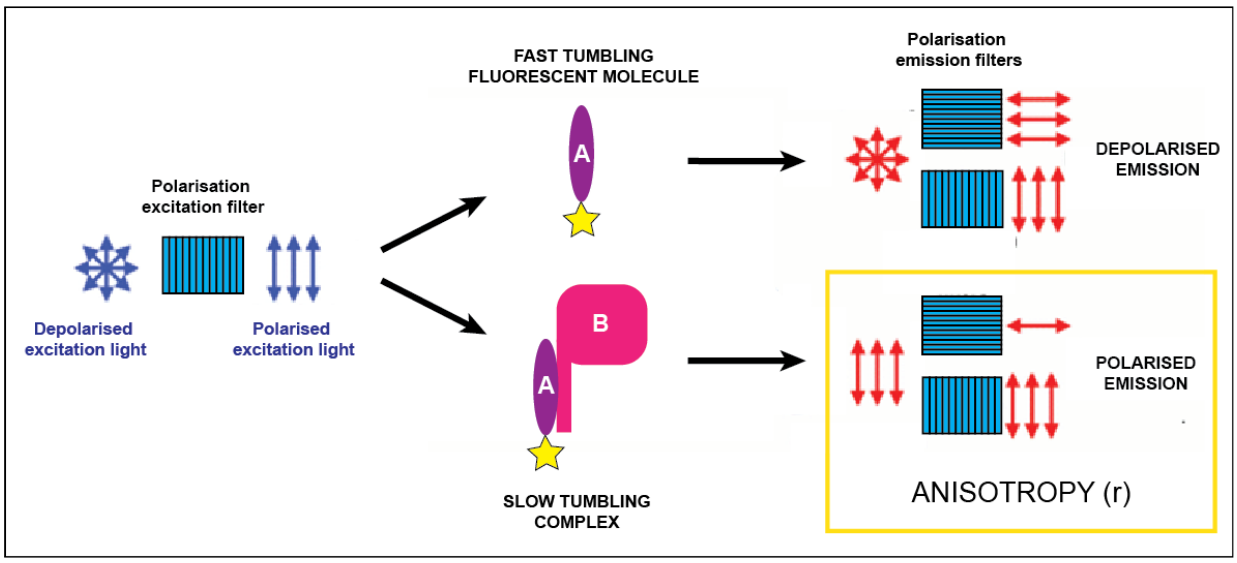


Figure 6.14. A diagram of the principle behind a fluorescence anisotropy assay. Larger complexes tumble more slowly in solution then their components so when the fluorescent tag is excited by a polarised light, complex-associated fluorophores give more polarised light on average than free protein-associated ones. Figure adapted from http://www.hi-techsci.com/techniques/anisotropy/. The fluorophore is shown as a yellow star and polarisation filters are shown as blue rectangles with black stripes. Anisotropy of the complex AB is seen in the yellow rectangle.

Assays presented in this chapter involve a specific version of the homologous competition assay where the small 4.3 kDa $Ldb1_{LID}$ peptide ligand is N-terminally tagged with a 43 kDa MBP domain and tandem LIM-domains (LIMs, which are the receptors) are fluorescently labelled with either fluorescein or GFP (Figure 6.15A). Titration of the bulky MBP-$Ldb1_{LID}$ ligand should result in the formation of larger complexes that can be detected by changes in anisotropy (Figure 6.15B).

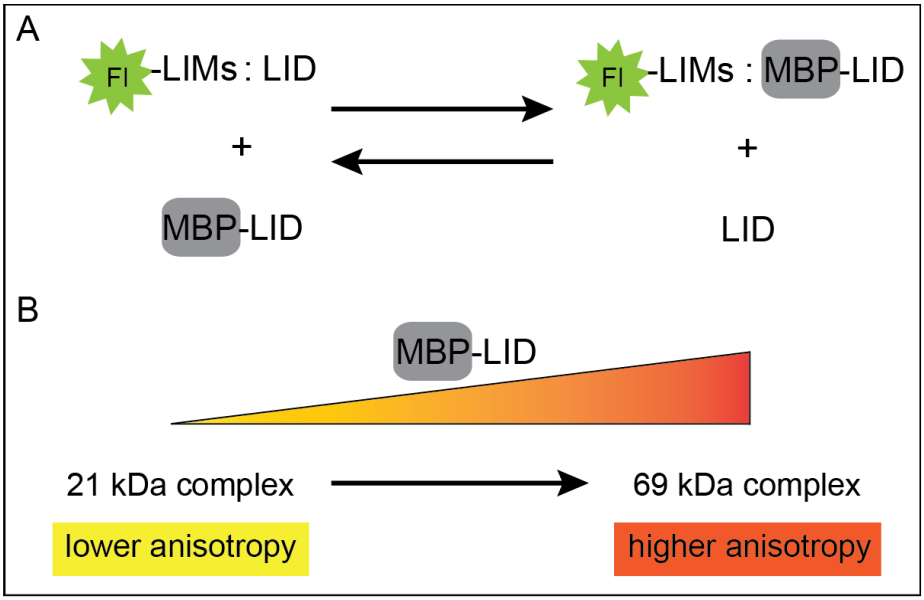


Figure 6.15. Homologous competition between different-sized LID constructs for fluorescently-labelled LIMs and the basis for an anisotropy assay. (A) An equilibrium between the Fl-LIMs:LID and Fl-LIMs:MBP-LID complexes, with colons representing protein-protein interactions. Fl-Cys-LMO4$_{LIMs}$:Ldb1$_{LID}$ is a complex obtained by Factor Xa cleavage of the respective tethered construct. Green stars represent fluorescence, whereas Fl is short for fluorescein. (B) By adding MBP-LID, the equilibrium shifts towards the formation of a larger fluorescent complex, which is expected to exhibit a higher anisotropy.

6.10.1 MBP-Ldb1$_{LID}$ and MBP

For the purpose of the experiments described in this chapter Ldb1$_{LID}$ spanning residues 331–375 was cloned into the pIH1121 vector. MBP-Ldb1$_{LID}$ and MBP-only were purified by amylose-affinity chromatography (section 2.7.3.3; Figure 6.16A). MBP-Ldb1$_{LID}$ was additionally purified by anion-exchange chromatography (section 2.7.5.2; Figure 6.16B).

Initial tests showed that the vertically polarized fluorescence intensity of a sample containing 10 nM Fl-Cys-LMO4$_{LIMs}$:Ldb1$_{LID}$ complex or the Fl-Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$ tethered construct dropped with every consecutive measurement (Figure 6.17A). In contrast, 5' fluoroscein-labelled 17 bp RNA (Dharmacon, Chicago, USA), obtained from Dr. Cuong Nguyen, and Fl-Cys-DEAF1, showed constant polarized emission intensities under the same conditions. When MBP-Ldb1$_{LID}$ was titrated into a solution containing Fl-Cys-LMO4$_{LIMs}$:Ldb1$_{LID}$ (10 nM), the anisotropy (r) increased from about 0.15 at the start of the

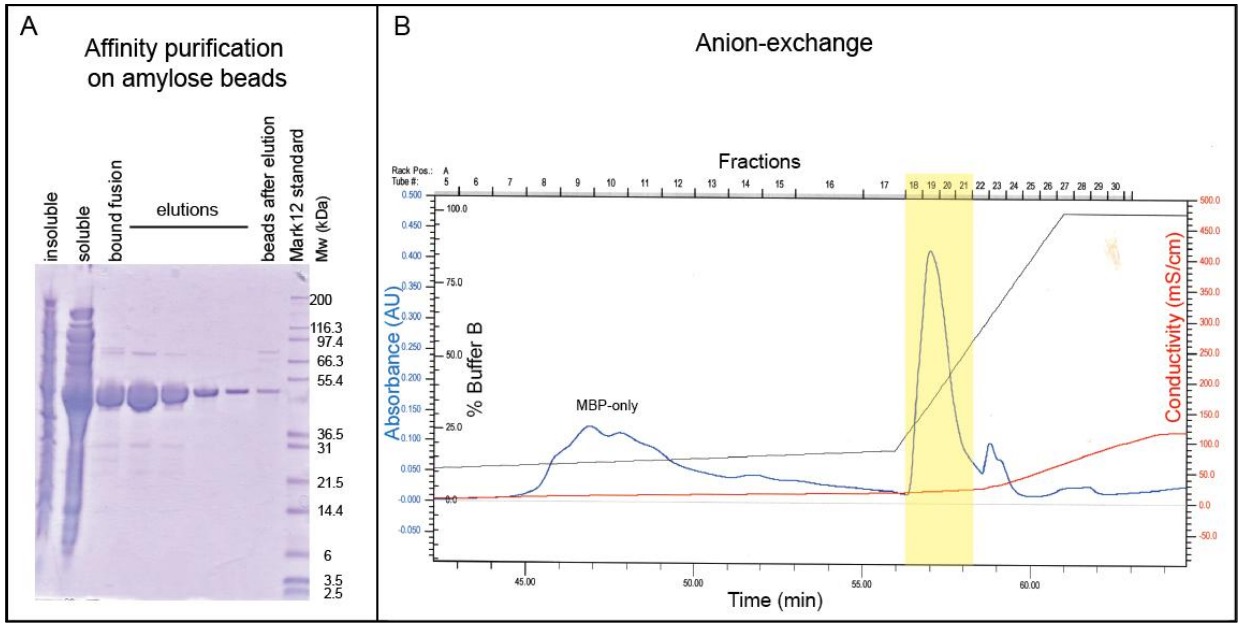


Figure 6.16. Purification of MBP-Ldb1$_{LID}$. (A) Comassie-stained SDS-PAGE reporting the affinity purification of MBP-Ldb1$_{LID}$ on amylose beads. (B) Anion-exchange chromatography of MBP-Ldb1$_{LID}$ on a UnoQ6 column in 50 mM Tris pH 8 buffer and a gradient of 0–1M NaCl. The peak containing MBP-Ldb1$_{LID}$ is marked in yellow. The grey line indicates % of buffer B, red line is conductivity and the blue line is the UV trace.

titration to >0.26 at the highest concentrations of MBP-Ldb1$_{LID}$, but displayed significant fluctuation (Figure 6.17B). However, the overall loss of fluorescence signal throughout the experiment made the anisotropy measurements unreliable.

The protocol included inversion of the cuvette after each titration step to mix the solutions. After comparing mixing techniques including no mixing, inverting the cuvette in the presence and absence of BSA, and mixing by pipetting, it was concluded that the drop of emission intensities is caused by mixing, but this can be minimized by the addition of 100 µg/mL BSA (Figure 6.17C). BSA is a commonly used to minimize non-specific binding in ELISA and other in vitro bioassays [314-316], suggesting that signal loss was due to non-specific adsorption of the fluorescent LMO4$_{LIMs}$-containing proteins to various surfaces during mixing. BSA was included in Competition assay buffer (Table 2.7) for all subsequent experiments.

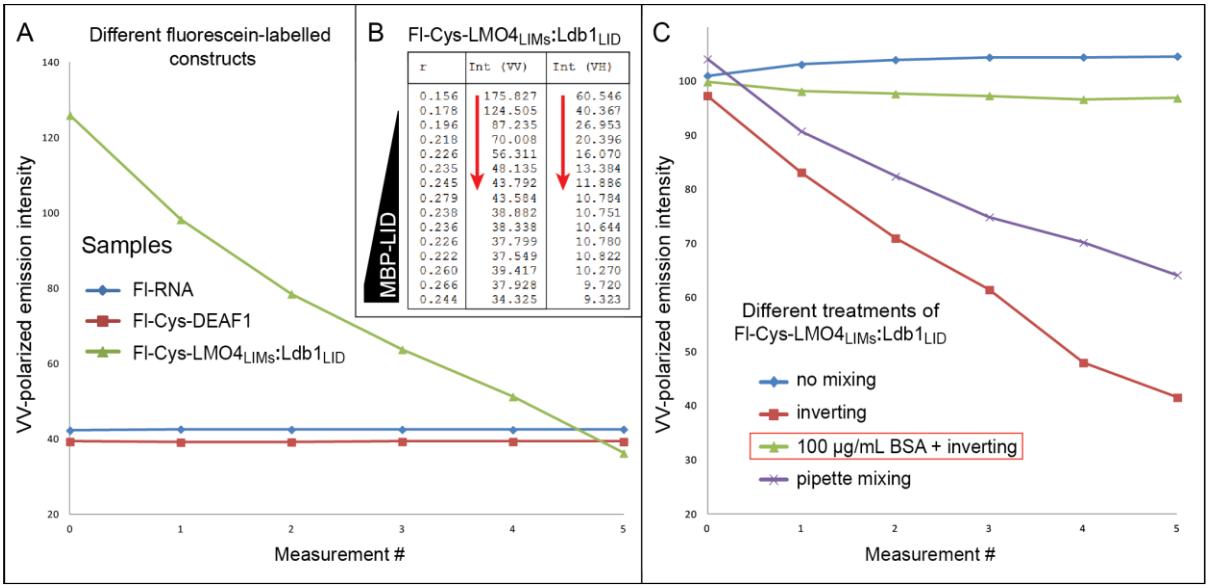


Figure 6.17. LIM-domain specific loss of fluorescence signal. (A) VV-polarized emission intensity of different fluorescent samples measured after inverting the cuvette seven times at each point. Fl-RNA is 5' fluorescein-labelled RNA. (B) Raw data for titration of MBP-Ldb1$_{LID}$ into a sample with Fl-LMO4$_{LIMs}$:Ldb1$_{LID}$ complex. Anisotropy (r) and the polarised emission light intensities (I_{vv} and I_{vh}) are shown. The last 2 points are repeated measurements. Red arrows indicate the decreasing emission intensities with increasing MBP-Ldb1$_{LID}$. (C) VV-polarized emission intensity of Fl-Cys-LMO4$_{LIMs}$:Ldb1$_{LID}$ measured after different sample treatments. The sample containing 100 µg/mL monomeric BSA in the buffer is marked.

6.10.2 FA assay results and analysis

Two experiments were carried out in which Fl-Cys-LMO4$_{LIMs}$:Ldb1$_{LID}$ (10 nM) was titrated with MBP-Ldb1$_{LID}$ (1 nM to 2 µM). The r-values were normalized (r_n, Appendix F) and plotted against the concentration of MBP-Ldb1$_{LID}$ (Figure 6.18A–B). The data was apparently monophasic and was fitted well by a 3-parameter SLogistic1 function. IC_{50} values of 24 nM (95% CI of 16–37 nM) and 31 nM (95% CI of 11–53 nM) were obtained for these two different experiments. The overlap of these data indicated that the experiment was reproducible. A control experiment with MBP-only titration was performed which did not show non-specific binding.

Assuming a number of conditions are met, IC_{50} is approximately equal to K_d for homologous competition experiments. However, one of these assumptions is that the concentration of the ligand is in excess of the receptor so that the free and total ligand concentrations are essentially the same. At the initial stages of the titration, this assumption does not hold true for this experiment because prior to the addition of any MBP-Ldb1$_{LID}$, the

concentrations of free Fl-Cys-LMO4$_{LIMs}$ and free Ldb1$_{LID}$ are the same. This phenomenon of ligand depletion can be accounted for using the Equation F.20 (Appendix F). The data from Figure 6.18A yielded a K_d of 14 ± 3 nM (the error is the reported standard error of the fit to a single data set; Figure 6.18C), but the second data set (Figure 6.18B) did not give a good fitto data. It appears that this approach could be a useful way to assess LIM-LID interactions, if problems associated with production of Fl-Cys-LMO4$_{LIMs}$:Ldb1$_{LID}$ could be resolved. However, the lack of available material did not allow further assessment of the method using FA for this thesis and other binding assays were pursued.

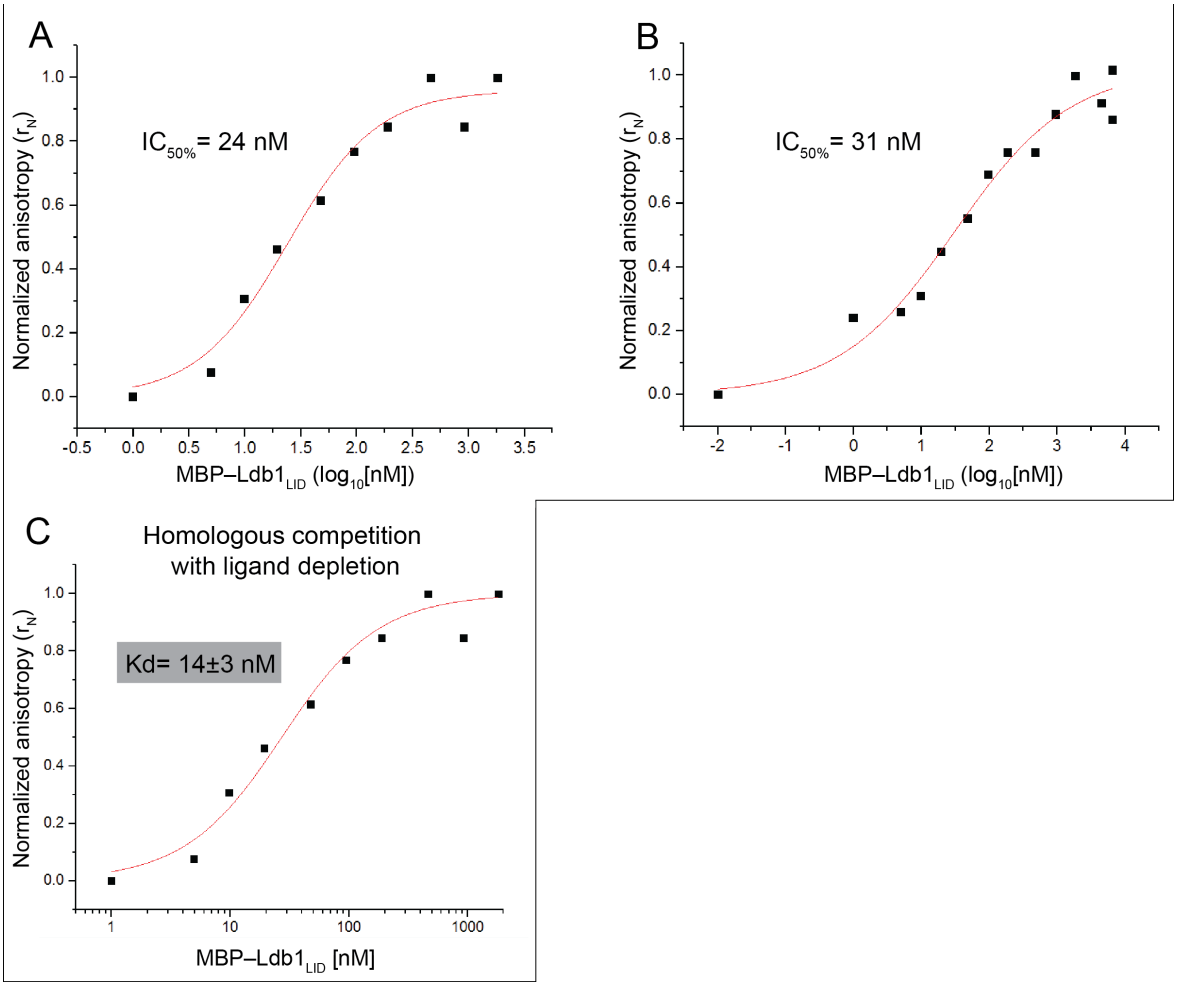


Figure 6.18. FA homogenous competition assays with the fluorescein-Cys-LMO4$_{LIMs}$:Ldb1$_{LID}$ complex. (A) and (B) Two titration experiments were fitted in Origin9.1 with a sigmoid logistic function to yield IC$_{50}$ values. (C) Data from titration shown in (A) was fitted with the Equation F.20 (Appendix F) to provide an estimate of K_d.

6.11 Summary

Although the $LMO4_{LIMs}$-$Ldb1_{LID}$ construct was initially labelled using the self-labelling enzymatic SNAP-tag with a commercial fluorescent dye, this approach was not reproducible. However, using EPL approaches, small amounts of $LMO4_{LIMs}$-$Ldb1_{LID}$ were fluorescently labelled. The yield of the purified labelled protein was relatively low but it was still useful for assessing the use of fluorescence anisotropy to assay LIM-LID interactions. The homologous competition FA assay was used to obtain an IC_{50} of 24 nM (K_d = 14±3 nM for ligand depletion) in one assay and IC_{50} of 31 nM in the second assay for the $LMO4_{LIMs}$ interaction with $Ldb1_{LID}$. The problems encountered in labelling, subsequent purification of LIMs-LID constructs and fluorescence measurements were specific to these constructs and were not encountered with Cys-DEAF1 peptide. The nonspecific adsorption of LIMs-containing samples to cuvettes and other surfaces in the FA assays was minimized by including BSA in the Competition assay buffer.

6.12 Discussion

6.12.1 Fluorescent labelling of LIMs-LID constructs

Although small amounts of several fluorescein-labelled Cys-LIMs-LID proteins were generated for use in binding studies, two key limitations of EPL labelling were noted: the long reaction times needed for Cys-LIMs-LID labelling compared to shorter peptide constructs; and, difficulties in removing the contaminating low molecular weight fluorescent species. In particular Cys-DEAF1-peptide labelling was much faster than Cys-$LMO4_{LIMs}$-$Ldb1_{LID}$ labelling, suggesting that the labelling reagents and experimental conditions were fundamentally sound, but something about the Cys-$LMO4_{LIMs}$-$Ldb1_{LID}$ construct affects the reaction efficiency. Burial of the N-terminal Cys, or interactions with hydrophobic residues of the $LMO4_{LIMs}$-$Ldb1_{LID}$ complex might account for the observed lower reaction rates. Denaturants such as guanidium-HCL (Gdn-HCL) or urea can often be used to prevent this type of problem and increase labelling efficiency [317]. Unfortunately, LIMs-LID constructs are not amenable to refolding, so such conditions could not be used.

Alternatively, zinc-coordination in the LIM domains could have contributed to poor labelling efficiency and slow reaction rates for Cys-LMO4$_{LIMs}$-Ldb1$_{LID.}$ Although only the N-terminal cysteine contains an α-amino group that reacts irreversibly with the thioester, trans-thioesterification of exposed internal cysteine residues is a reversible side-reaction in NCL (Figure 6.2) and the cause of side-products [303, 308, 318]. The zinc-coordinating cysteine residues in LMO4 might become involved in this type of side-reaction, especially if any of the reducing agents used was able to displace cysteine side chains as zinc-coordinating groups. Displacing the zinc could also lead to protein unfolding and aggregation [319], which would likely reduce accessibility of the target and further reduce labelling efficiency. At least one ZnF protein has previously been semi-synthesised by EPL, but that protein was labelled under denaturing conditions and refolded [320]. As noted above, the same approach is not feasible for Cys-LMO4$_{LIMs}$-Ldb1$_{LID}$.

Although MESNA thioesters, as used here, have been shown to rapidly label some targets [303] [304], other studies suggest that they can be slow, at least for some other targets, and that it may be possible to enhance labelling efficiency through the use of more reactive thioesters [321]. However, data from this chapter showed that although MPAA enhanced the reaction rate and improved labelling of LIMs-LID over MESNA, it resulted in higher levels of non-covalently bound dye. A compromise that might assist labelling efficiency but minimise problems with purification could be modulation of activity by lowering the temperature [309]. Similar types of contamination have been observed in the initial products from other labelling studies [322], but no purification problems were reported, so at this stage it is difficult to know how they could be avoided, or easily purified away from LIMs-LID constructs.

Alternatively, other labelling techniques that require milder conditions, without the need for potentially denaturising chemicals, could be performed. For example, some enzymatic labelling techniques such as sortase-mediated labelling (Table 6.2) could allow for highly specific labelling under physiologic conditions.

6.12.2 Possible improvements in FA assays

The preliminary FA assays that were carried out in this chapter suggest some potential improvements, if fluorescently labelled protein could be reliably generated. In order to prevent nonspecific binding of LIM-constructs to surfaces, PEG-coating of quartz cuvette walls could be a more effective alternative to BSA buffer additive [314]. A relatively small change in mass of the fluorescent species (21 kDa to 69 kDa, Figure 6.15) likely affected the reproducibility and reliability of the data. For this reason, future FA assays should be carried out using fluorescent N-terminally-labelled LID-LIMs complexes (i.e., with a cleavable tether between the C-terminus of $Ldb1_{LID}$ and the N-terminus of $LMO4_{LIMs}$), using non-fluorescent MBP-tagged LID peptides as the competitors. The displacement of Fl-LID (5 kDa to 69 kDa change in mass of fluorescent species) would likely result in a larger drop in anisotropy (a wider dynamic range of the assay), giving improved signal-to-noise data. This type of experiment is also closer in design to classical 'cold saturation' experiments, where the LID peptides take the place of labelled and cold ligands and the LIMs mimic unlabelled receptors. Data from this type of experiment may be fitted more easily with established binding equations [323]. However, rather than pursuing this approach, for this thesis it was decided to use an alternative approach to generating fluorescently tagged complexes, which is described in the following chapter.

Chapter 7 Competition assays with GFP-LIMs-LID constructs

As an alternative to fluorescence anisotropy assays, it was decided to use native gel-shift assays to investigate LIM-LID binding and determine if this approach could be used to measure binding affinities using homologous competition experiments. Given that the site-specific labelling approach described in the previous approach was problematic, GFP-labelled proteins were generated. The first part of this chapter describes the production and characterisation of these proteins, and the second part of the chapter focuses on testing these proteins in blue native PAGE (BN-PAGE) and colourless native PAGE (CN-PAGE) gel shift assays.

7.1 Protein Preparation

GFPuv-tagged $LMO4_{LIMs}$:$Ldb1_{LID}$ complex was used in blue native PAGE assays only. GFPsol-labelled complexes were used in colourless native PAGE (CN-PAGE) gel shift assays, with GFPsol-$LMO4_{LIMs}$:$Ldb1_{LID}$ being used in FA assays. MBP-tagged $Ldb1_{LID}$ constructs were generated for use in all experiments and MBP-only was used as a control in gel shift assays.

7.1.1 GFP-LIMs:LID complexes

GFPuv was created from the split-GFPuv sequence used in experiments described in Chapter 5. The inserted linker was removed by PCR method to create pET15bBE-GFPuv, pHisGB1.3C-GFPuv, and pGEX6P-GFPuv vectors (see Appendix B for sequences). Into each of these vectors $LMO4_{LIMs}$-$Ldb1_{LID}$ was subcloned to allow the expression of this protein with an N-terminal affinity tag (6×His, 6×His-GB1 from the pET15bBE-GFPuv and pHisGB1.3C-GFPuv vectors, respectively, or GST from the pGEX6P-GFPuv vector) and a GFP moiety as indicated. The affinity tags could each be removed by treatment with thrombin (pet15bBE-GFPuv) or HRV3C protease (all other vectors). Small scale overexpression and purification studies were carried out to determine which affinity tags resulted in the best overall yield and final purity. The GST-tagged protein showed the lowest level of contamination or degradation products in purification (Figure 7.1A). Thus, GST-GFPuv-$LMO4_{LIMs}$-$Ldb1_{LID}$ was produced on a larger scale. The protein was purified by GSH-affinity chromatography, followed by on-bead

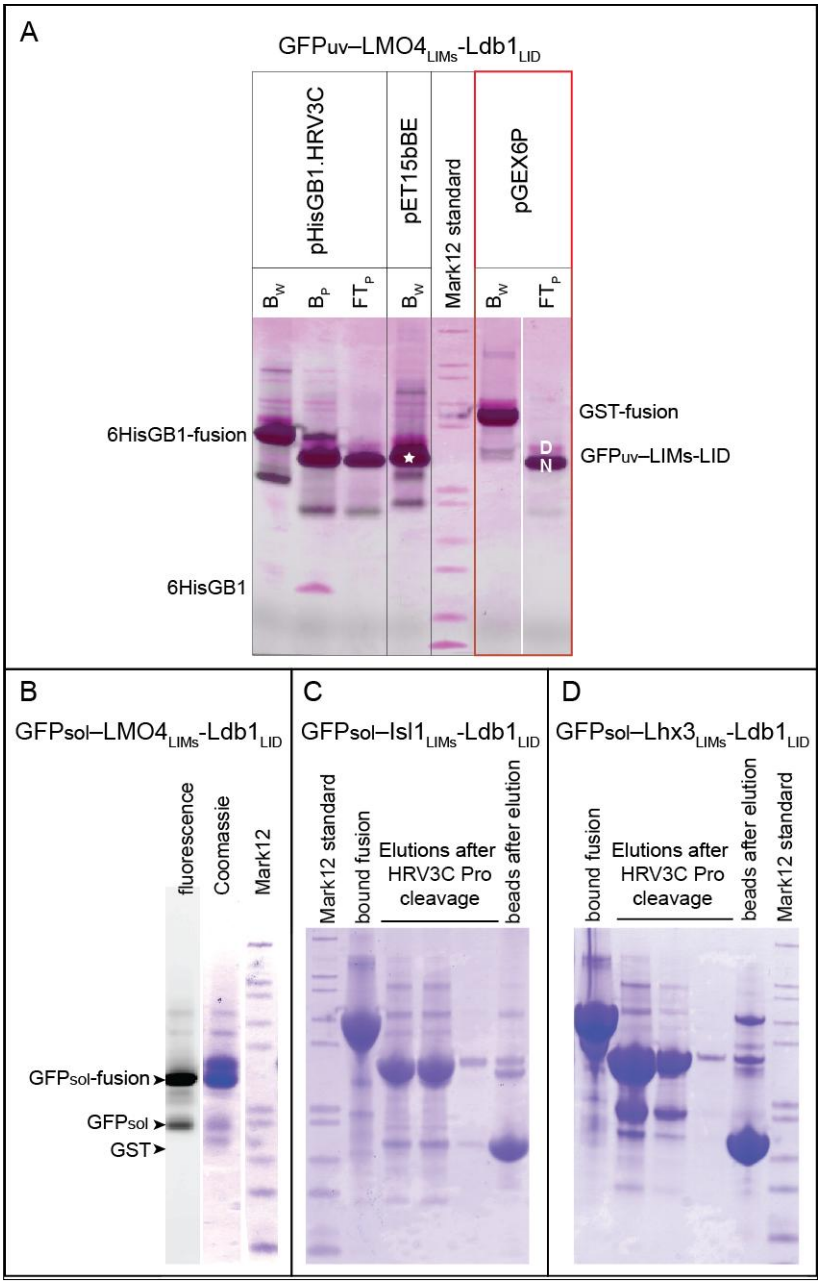


Figure 7.1. SDS-PAGE analysis of the affinity purification of GFP-LIMs-LID fusion proteins. (A) Small scale purification trials of GFPuv-$LMO4_{LIMs}$-$Ldb1_{LID}$ constructs with different affinity tags. An overlay of Coomassie-stained gel (pink) and its fluorescence scan (black). The names of corresponding vectors from which tagged constructs were expressed are shown. Note that proteins containing GFP proteins often show a double banding effect due to incomplete denaturation of GFP. The lower bands (black+pink) contain natively folded GFPuv (labelled N) while the upper bands (pink) contain the non-fluorescent denatured GFPuv (labelled D). B_W - tagged-fusion construct bound to beads; B_P - bound tagged-fusion construct cleaved with HRV3C protease to release the GFPuv-$LMO4_{LIMs}$-$Ldb1_{LID}$ from the beads; FT_P - flow-through fraction with released soluble GFPuv-$LMO4_{LIMs}$-$Ldb1_{LID}$; White star - 6×His-GFPuv-$LMO4_{LIMs}$-$Ldb1_{LID}$; Red rectangle - purification of the GST-tagged fusion construct. (B) GFPsol-$LMO4_{LIMs}$-$Ldb1_{LID}$ after affinity purification. A fluorescence scan detected GFPsol-only degradation product. (C) Purification of GFPsol-$Isl1_{LIMs}$-$Ldb1_{LID}$. (D) Purification of GFPsol-$Lhx3_{LIMs}$-$Ldb1_{LID}$.

HRV3C protease cleavage to remove the GST-tag, and elution of the GFPuv-fusion protein from the GSH-beads (Figure 7.1A).

GFPsol was created from GFPuv by mutagenesis (section 2.4.1.2) and subcloned into the pGEX6P vector to make pGEX6P-GFPsol for expression of GST-GFPuv-LIMs-LID complexes and generation of GFPsol-LIMs-LID constructs, as described above for the GFPuv versions. Three different GFPsol-LIMs-LID constructs were created: GFPsol-$LMO4_{LIMs}$-$Ldb1_{LID}$, GFPsol-$Lhx3_{LIMs}$-$Ldb1_{LID}$ and GFPsol-$Isl1_{LIMs}$-$Ldb1_{LID}$ (Figure 7.1B,C and D). Each of these constructs contained a Factor-Xa site in the linker between the LIM and LID domains for use in competition binding experiments.

A common feature of the SDS-PAGE gels featuring GFP proteins is a double banding effect. This arises from incomplete denaturation of GFP under standard protocols (heating samples in SDS-containing sample loading buffer to 80 ˚C prior to sample loading). Although this protocol is sufficient to denature most proteins, the GFP-fold is very resistant to denaturation [324]. For example, when the gel in Figure 7.1A was scanned for fluorescence, it could be seen that the lower bands contained natively fluorescent GFP.

Additional purification was carried out as required, and in all cases involved a SEC step to remove large aggregates and smaller degradation products (Figure 7.2A). In cases where there was significant contamination by GST or some degradation products close in size to the full constructs, as was the case for GFPsol-$Lhx3_{LIMs}$-$Ldb1_{LID}$ (Figure 7.1D), an additional anion-exchange chromatography step was carried out (Figure 7.2B).

Purified GFPsol-$Lhx3_{LIMs}$-$Ldb1_{LID}$ was subjected to MALLS analysis to confirm that it is monomeric and not aggregated in solution (Figure 7.3). The experimental weight average molecular weight was 47.5 kDa compared to a theoretic molecular weight of 48 kDa, indicating monomeric protein. Purified and concentrated GFPsol-LIMs-LID fusion proteins (Figure 7.4A) were diluted to 1 µM and cleaved by Factor Xa for 6 h at room temperature, or at 4 °C for 48 h to create the corresponding GFPsol-LIMs:LID complexes (note the use of a colon to indicate the cut complex versus a hyphen to indicated the uncut tethered complex). As longer incubation times showed some evidence of degradation and aggregation (Figure 7.4B), subsequent reactions were performed at room temperature for 6 h to generate the proteins used in fluoresecent gel shift assays.

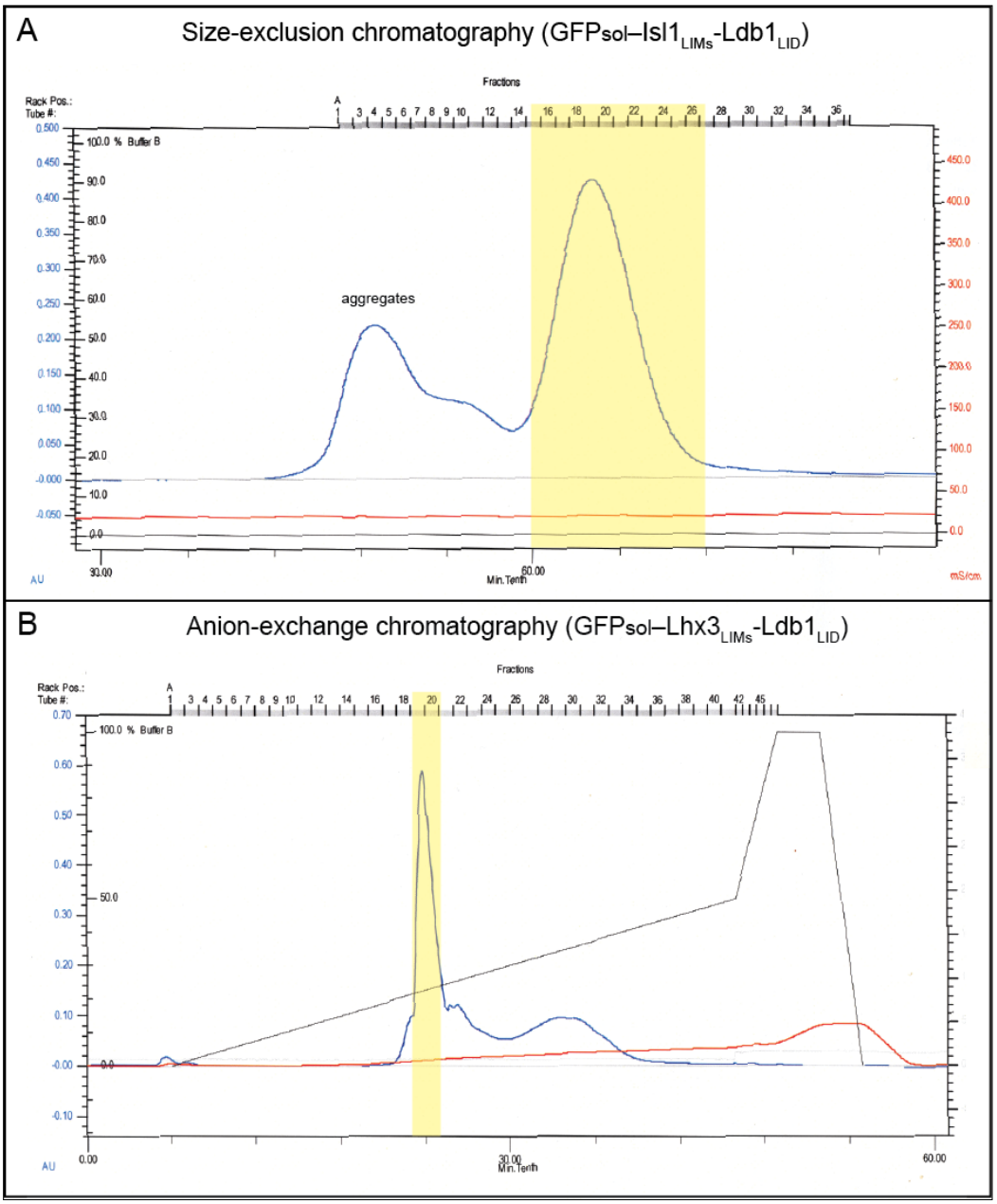


Figure 7.2. Further purification of GFP-labelled fusion proteins. (A) SEC of GFPsol-Isl1$_{LIMs}$-Ldb1$_{LID}$ using a Sephadex75 column in 20 mM Tris pH8 buffer, 150 mM NaCl and 1 mM DTT. Fractions containing non-aggregated proteins were collected (yellow rectangle). (B) Anion-exchange chromatography purification of GFPsol-Isl1$_{LIMs}$-Ldb1$_{LID}$ was carried out using a UnoQ6 column in 20 mM Tris pH 8 buffer with 0–1 M NaCl gradient and 1 mM DTT. The yellow rectangle indicates fractions containing pure protein. Blue traces in both chromatograms (A and B) show absorbance at 280 nm, grey traces show the % of buffer B and red traces show conductivity.

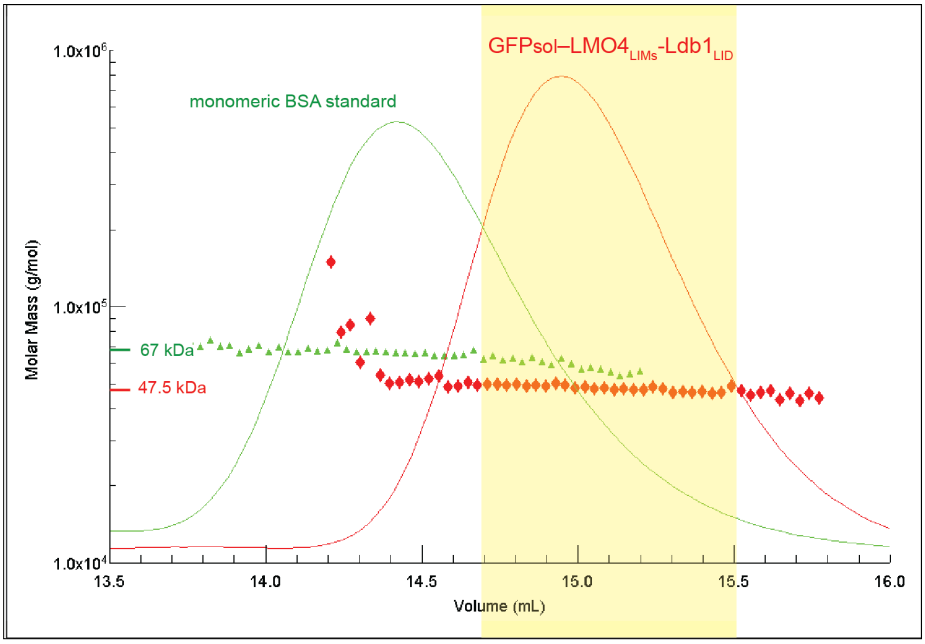


Figure 7.3. SEC-MALLS chromatogram showing analysis of an uncleaved GFPsol-LMO4$_{LIMs}$-Ldb1$_{LID}$ sample. Monomeric BSA protein was run for comparison and spectra overlayed. Yellow zone marks fractions of the aggregate-free protein that were collected.

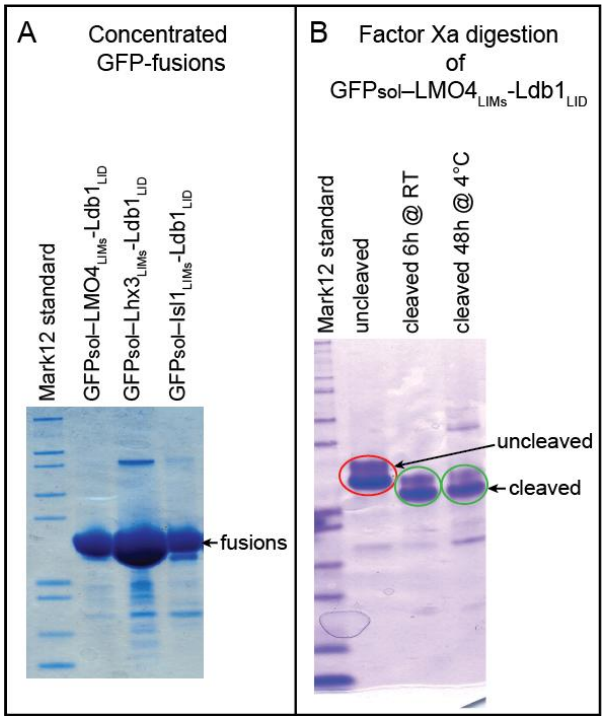


Figure 7.4. Concentrated GFPsol-fusion proteins and Factor Xa digestion of the linker. SDS-PAGE showing (A) Concentrated purified GFPsol-fusion proteins. (B) Factor Xa digestion of GFPsol-LMO4$_{LIMs}$-Ldb1$_{LID}$ for 6 h at room temperature or 2 days at 4 °C. Note that the (~5 kDa) LID peptide is typically not visible in gels due to its small size and poor staining properties.

7.2 Fluorescent protein gel shift assays

The binding of a labelled protein to an unlabelled partner protein can be detected through native (non-denaturing) PAGE, which separates bound (shifted) and unbound components. The labelled protein is held at a constant concentration while the second component is added at increasing concentrations. Binding affinities can be estimated by quantifying the band shifts at each concentration of the second component [325, 326].

The migration of the free and bound complexes in classic native PAGE depends on the charge and shape of the species involved, and only proteins and complexes with pI < pH_{buffer} migrate towards the anode [327]. In contrast, Blue Native PAGE (BN-PAGE) uses Coomassie G-250 dye in the cathode and sample loading buffers, the dye to coat (but not denature) most proteins in a sample by binding non-specifically to exposed hydrophobic surfaces. This gives the proteins a net negative charge while maintaining their native conformation, allowing visualization of protein species and their separation according to differences in size and shape at near neutral pH (e.g., Figure 7.5) [328, 329].

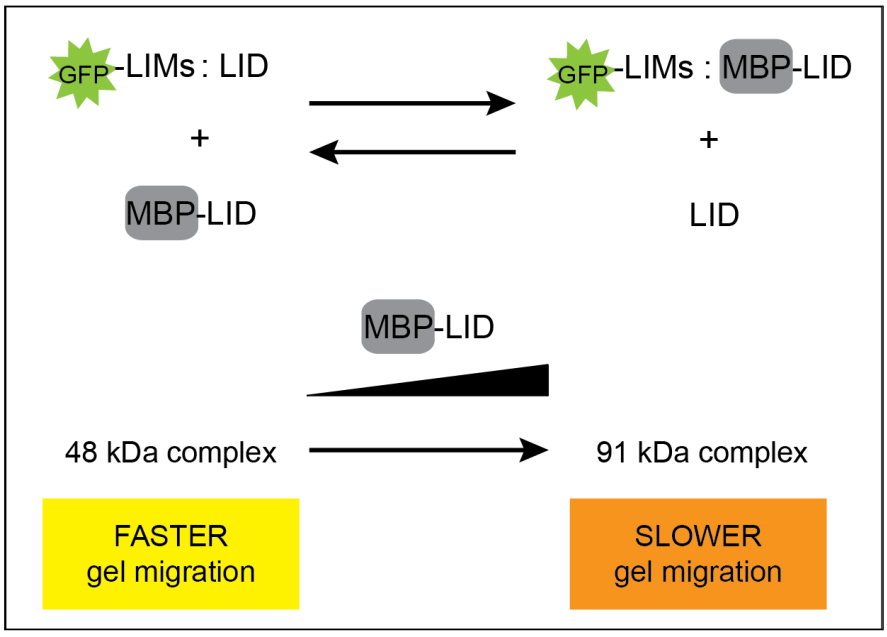


Figure 7.5. Homologous competition between different sized LID-constructs for GFP-labelled LIMs and the detection of the resulting complex. An equilibrium between the two complexes at a specific MBP-LID concentration. Colons represent protein-protein interactions in the complex. The size of the fluorescent-LIMs : LID complex dominant at low (left) or high MBP-LID concentrations (right).

Colourless PAGE (CN-PAGE) does not use the blue Coomassie dye throughout electrophoresis. Although it gives lower resolution than BN-PAGE or SDS-PAGE it omits additives that might interfere with protein binding and detection, and allows characterisation of complex formation under mildest possible conditions. BN-PAGE and CN-PAGE gel shifts were assessed to see if either was suitable for estimating LIMs-LID interaction affinities in competition style experiments.

7.2.1 BN-PAGE gel shift with the GFPuv-LMO4$_{LIMs}$:Ldb1$_{LID}$

Preliminary gel shift experiments using BN-PAGE were carried out using GFPuv-LMO4$_{LIMs}$:Ldb1$_{LID}$ and titrating in MBP-Ldb1$_{LID}$ as part of a competition assay. The addition of 1 μM MBP-Ldb1$_{LID}$ to 100 nM GFPuv-LMO4$_{LIMs}$:Ldb1$_{LID}$ resulted in a new GFPuv-LMO4$_{LIMs}$:MBP-Ldb1$_{LID}$ complex that appeared as a band well separated from GFPuv-LMO4$_{LIMs}$:Ldb1$_{LID}$. However, even though the BN-PAGE gels showed excellent resolution, they did not provide the necessary sensitivity to detect shifts at lower concentrations (≤ 50 nM GFPuv-LMO4$_{LIMs}$ with ≤ 50 nM MBP-Ldb1$_{LID}$) (Figure 7.6).

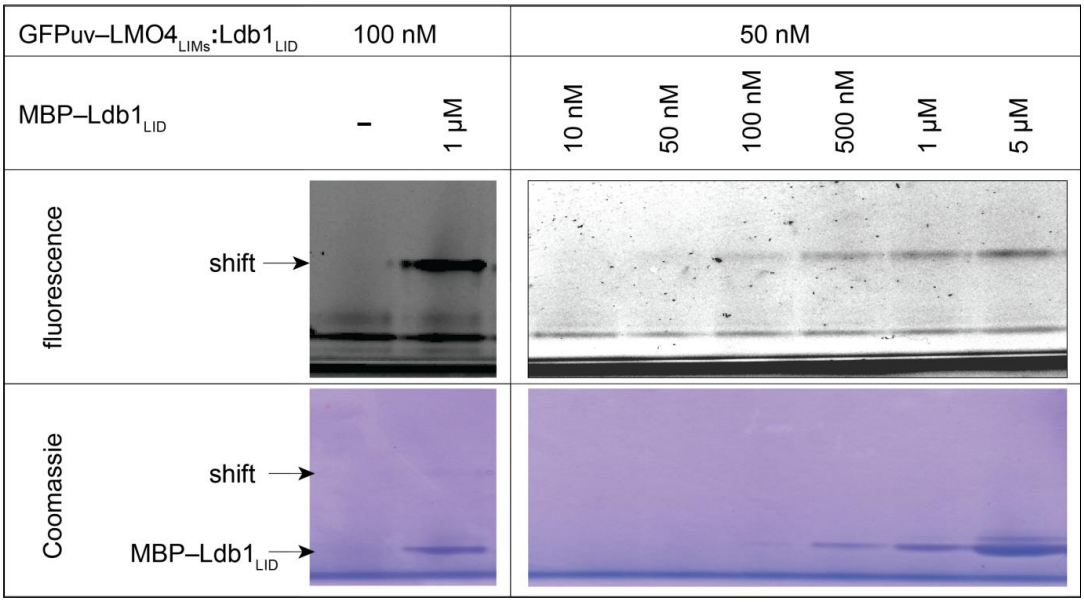


Figure 7.6. Sensitivity of the BN-PAGE shift competition assays with the GFPuv-labelled complexes. Competition binding of GFPuv-LMO4$_{LIMs}$ with MBP-Ldb1$_{LID}$ was investigated on 3-12% polyacrylamide gels using BN-PAGE. The photographs show fluorescence scanned and Coomassie stained BN-PAGE gels. Two separate gels are shown: with 100 nM GFPuv-LMO4$_{LIMs}$:Ldb1$_{LID}$ (left) and with 50 nM GFPuv-LMO4$_{LIMs}$:Ldb1$_{LID}$ (right). Shifted bands are indicated; non-shifted bands are near the buffer front at the bottom of each panel.

Since the expected K_d for the $LMO4_{LIMs}$ vs. $Ldb1_{LID}$ interaction was around 10 nM, a titration would have to cover at least an order of magnitude above and below this concentration of MBP-$Ldb1_{LID}$. Also, to minimize ligand depletion (section 6.10.2), labelled $LMO4_{LIMs}$ should be ≤ 10 nM. As this did not appear to be feasible for BN-PAGE with GFPuv-tagged attempts were made to use CN-PAGE.

7.2.2 CN-PAGE gel shifts with the GFP_{sol}-labelled complexes

GFP_{S65T} and related FP versions give more intense fluorescent signals so GFPuv-$LMO4_{LIMs}$:$Ldb1_{LID}$ was replaced with a GFPsol version for CN-PAGE gel-shift experiments. In the absence of MBP-$Ldb1_{LID}$, GFPsol-$LMO4_{LIMS}$:$Ldb1_{LID}$ appeared as multiple fluorescent bands that were not detected by Coomassie staining (Figure 7.7). These bands might correspond to GFPsol-$LMO4_{LIMs}$:$Ldb1_{LID}$, GFPsol-$LMO4_{LIMs}$, GFPsol-$LMO4_{LIMs}$-$Ldb1_{LID}$ and GFPsol-only degradation product that was noted during early stages of purification (Figure 7.1). In samples containing MBP-$Ldb1_{LID}$ a single, well-resolved, shifted band was observed (Figure 7.7). In these initial experiments, glycerol (5%) was used to help sample loading. As glycerol could potentially affect protein interactions [330], it was replaced with 5% Ficoll in subsequent experiments.

For initial experiments the samples were incubated for 15 min at room temperature prior to loading on the gel. These shifted bands had relatively low intensities compared to total fluorescence even at the highest concentrations of MBP-$Ldb1_{LID}$. To assess the optimal time of incubation, samples containing 20 nM GFPsol-$LMO4_{LIMs}$:$Ldb1_{LID}$ complex and 100 nM MBP-$Ldb1_{LID}$ were incubated at times ranging from 0–60 min before electrophoresis (Figure 7.7A). Fluorescent bands were quantified and the fraction in complex was plotted against incubation time (Figure 7.7B). These data indicated that under the conditions used the exchange between $Ldb1_{LID}$ and MBP-$Ldb1_{LID}$ competition reaches equilibrium plateau by ~60 minutes ($t_{1/2}$ ~30 min). The fraction of shifted fluorescence increased from about 30% at 15 min to about 45% at 60 min.

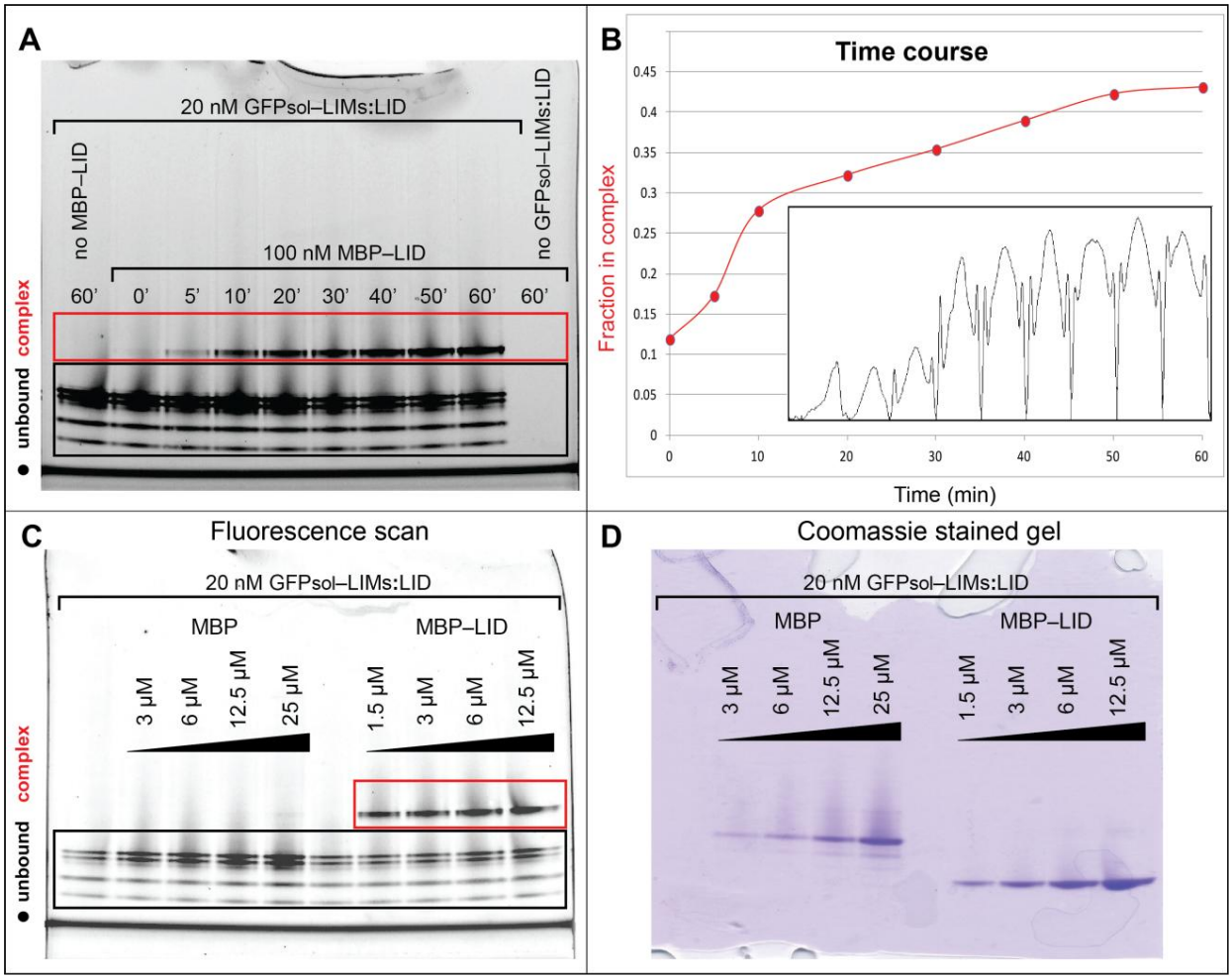


Figure 7.7. Optimization of fluorescent protein-based gel shift assays for competition binding analysis of MBP-Ldb1$_{LID}$ with GFPsol-LMO4$_{LIMs}$:Ldb1$_{LID}$. (A) Time course experiment with GFPsol-LMO4$_{LIMs}$:Ldb1$_{LID}$ (20 nM) and MBP-Ldb1$_{LID}$ (100 nM) detected by fluorescence scanning of the gel. The position of the shifted bands corresponding to GFPsol-LIMs:MBP-LID are marked with red rectangles. Bands corresponding to unbound GFPsol- LMO4$_{LIMs}$-Ldb1$_{LID}$ construct are marked with a black box. The buffer front is indicated with a black dot. (B) The boxed areas from the time course gel scan were analysed in ImageJ (see sample inset for red box frmo panel A). The fraction of the total fluorescence in each lane that makes a shifted band was plotted against time. (C) Fluorescence scan comparing the effects of MBP-Ldb1$_{LID}$ and the MBP-only samples. (D) Coomassie-stained gel from (C).

A control experiment was carried out in which 3–25 µM MBP and 1.5–12.5 µM MBP-Ldb1$_{LID}$ were each titrated into GFPsol-LMO4$_{LIMs}$:Ldb1$_{LID}$. MBP by itself does not cause a band shift, but the overall fluorescence signal appears to be higher in the presence of MBP and the signal noticeably increases for MBP concentrations above 3 µM. A similar effect was detected for MBP-Ldb1$_{LID}$ (Figure 7.7C–D). This effect may be because MBP blocks non-specific

adsorption of the LIM-domain construct to surfaces. The gel shifts gave best results when ≥ 10 nM GFPsol-LMO4$_{LIMs}$:Ldb1$_{LID}$ complex was used because this allowed detection of the shifted complex even at low nM concentrations of MBP-Ldb1$_{LID}$. Although this concentration is much lower than that required for the BN-PAGE experiments, it may still be too high for this type of analysis to be meaningful, especially using the IC_{50} as an indicator of K_d.

Nevertheless, several exploratory experiments were carried out in which MBP-Ldb1$_{LID}$ was titrated into GFPsol-LMO4$_{LIMs}$:Ldb1$_{LID}$ (Figure 7.8A and D), and the relative intensity of the shifted bands was determined (Figure 7.8B and E). These normalised data were fitted by the SLogistic1 equation to obtain estimates of IC_{50} (Figure 7.8C and F). Note that the first experiment suggests that the main binding event occurred in the early stages of the titration (Figure 7A–C). The second experiment included more points at the lower end of the concentration range. The same procedure was performed using additional titrations of MBP-Ldb1$_{LID}$ into GFPsol-Lhx3$_{LIMs}$:Ldb1$_{LID}$ and GFPsol-Isl1$_{LIMs}$:Ldb1$_{LID}$ (Figure 7.9).

The data were also fitted by the ligand depletion model to obtain estimates of K_d (Appendix F). Whereas the IC_{50} values for LMO4:Ldb1 interactions were similar (~10 nM), the data from the ligand depletion model differ more substantially and are associated with high errors (particularly for the second data set), K_d = 7 ± 2 and 0.48 ± 0.47 nM, respectively (Figure 7.10A). Values of K_d were estimated based single experiments for Lhx3$_{LIMs}$ vs. Ldb1$_{LID}$ (K_d = 30 ± 6 nM; Figure 7.10B) and for Isl1$_{LIMs}$ vs. Ldb1$_{LID}$ (K_d = 81 ± 16 nM; Figure 7.10C). If these estimates are correct, then it is possible that the method is appropriate for a subset of interactions of this class. Unfortunately it was not possible to collect additional data for these constructs due to time constraints.

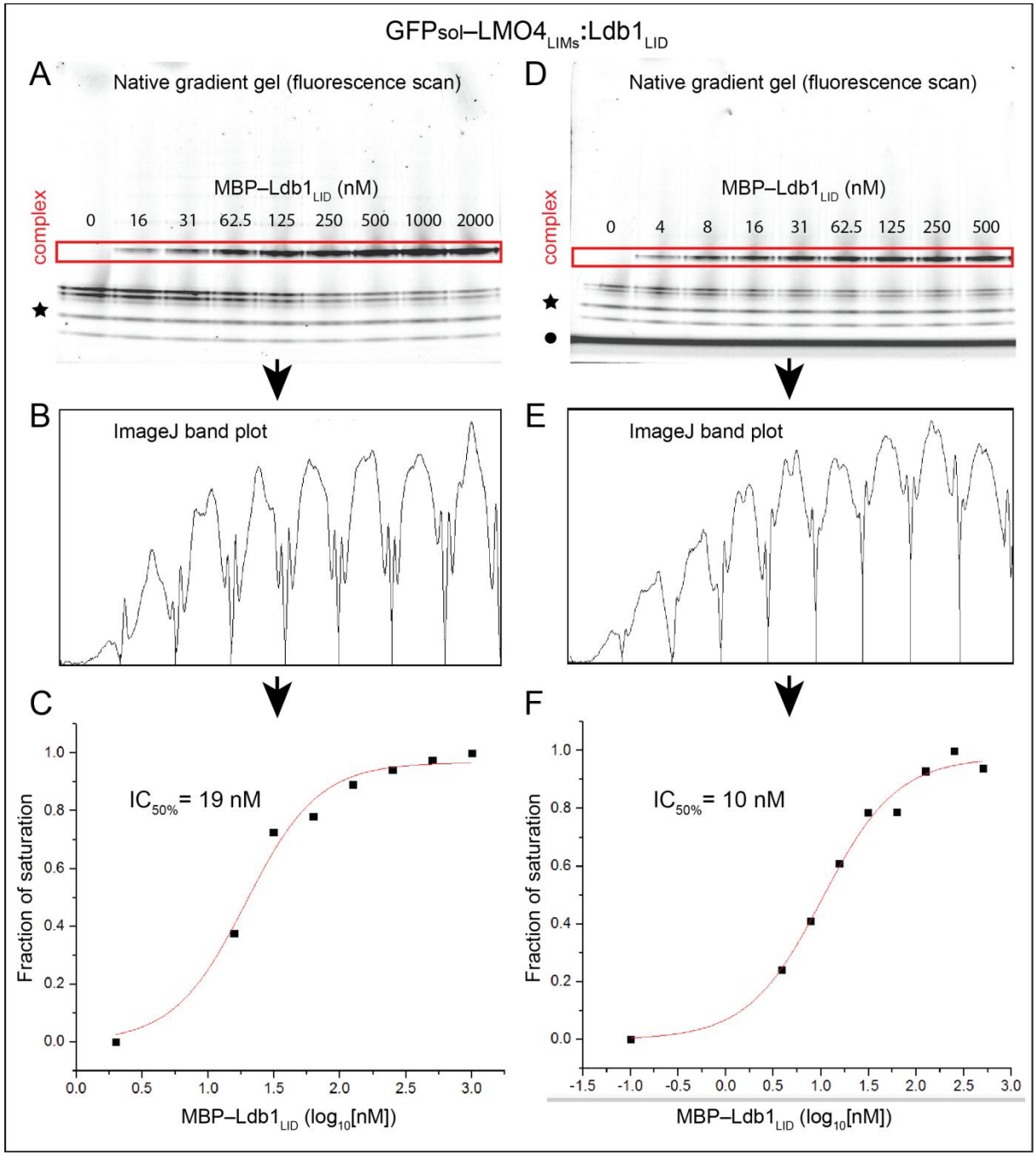


Figure 7.8. Gel shift assays with GFPsol-LMO4$_{LIMs}$:Ldb1$_{LID}$ vs. MBP-Ldb1$_{LID}$. (A and D) native gel scans show the shifted complexes (in red rectangles), characteristic additional bands (black stars) and the buffer front in (D) marked with black dot. (B and E) show ImageJ plots of band intensities. (C and F) Sigmoid logarithmic curves fitted to experimental data in Origin9.1 and the IC_{50} values obtained. The lowest MBP-Ldb1$_{LID}$ concentration was set to appropriately low values of 2 nM (C) and 0.1 nM (F), because log[0] is not mathematically defined.

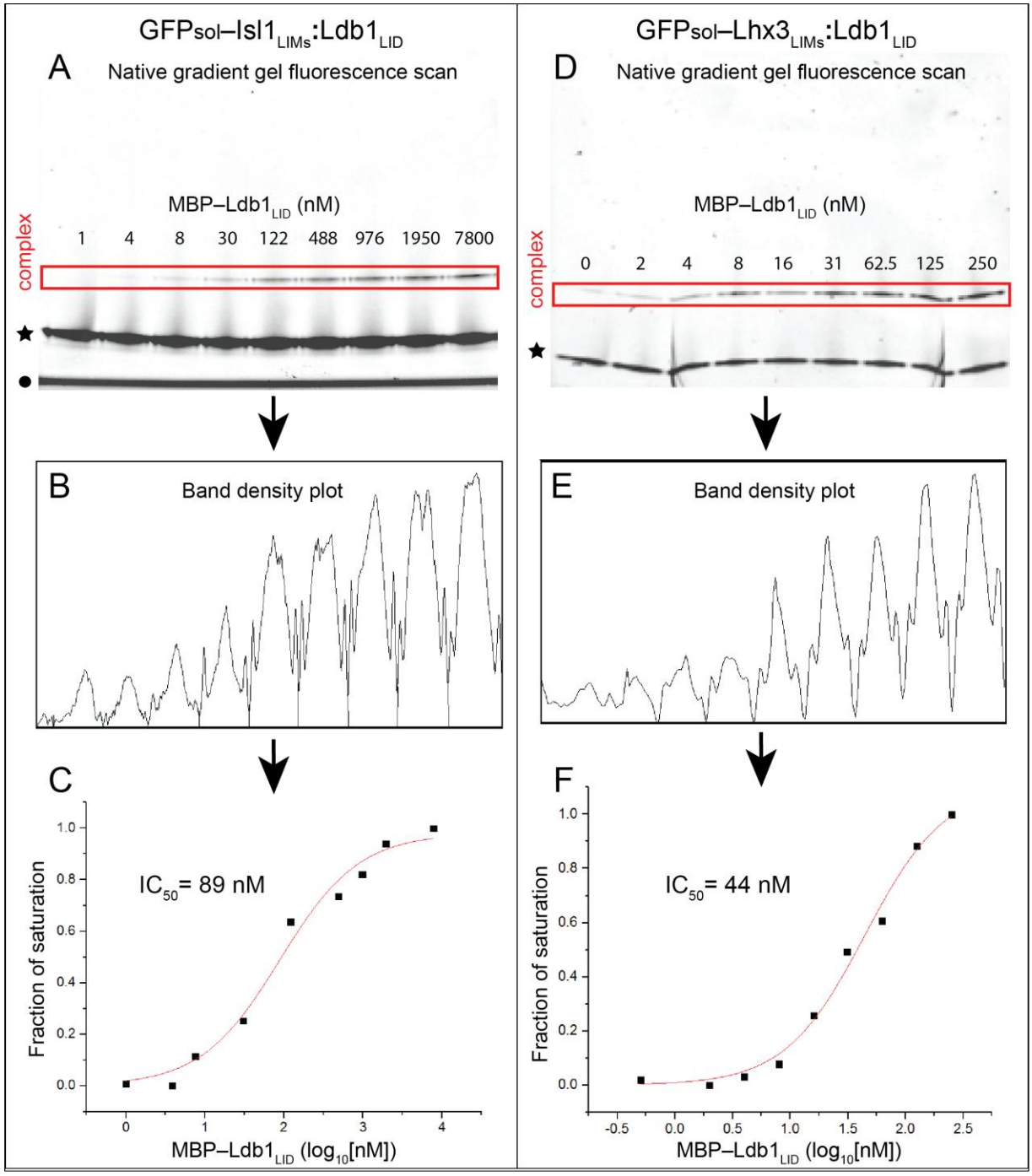


Figure 7.9. GFP gel shift assays with LIM-domains from LIM-HD proteins. (A–C) GFPsol-Isl1$_{LIMs}$:Ldb1$_{LID}$ vs. MBP-Ldb1$_{LID}$. (D–F) GFPsol-Lhx3$_{LIMs}$:Ldb1$_{LID}$ vs. MBP-Ldb1$_{LID}$. (A and D) Native gel scans show the shifted complexes (in red rectangles). Bands corresponding to non-shifted GFP-fusion constructs are marked with black stars. The buffer front in (A) is marked with a black dot. (B and E) ImageJ plots of shifted band intensities. (C and F) Sigmoid logarithmic curves fitted to experimental data in Origin9.1 and the IC_{50} values obtained. In (F), the lowest MBP-Ldb1$_{LID}$ concentration was set to appropriately low value of 0.5 nM because log[0] is not mathematically defined.

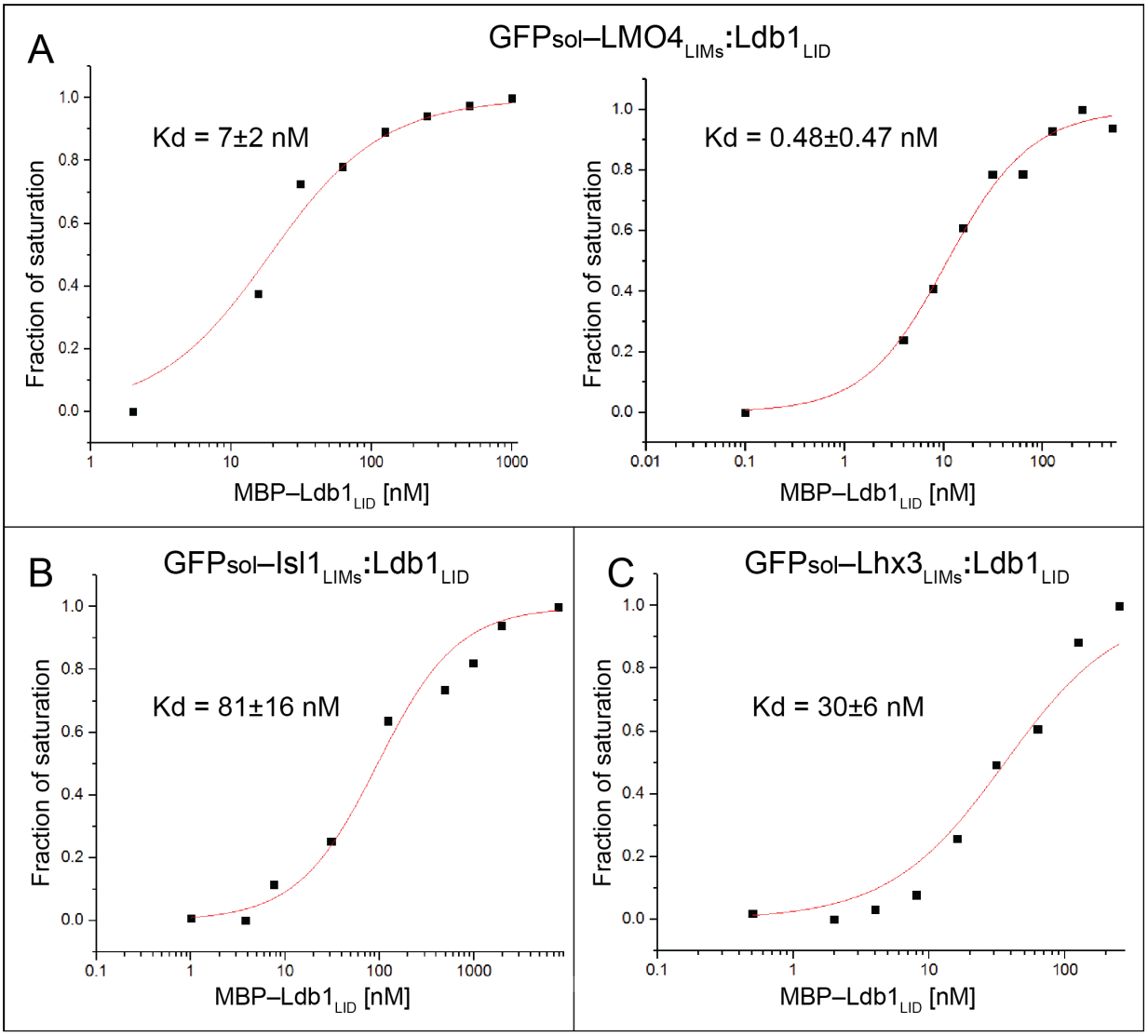


Figure 7.10. GFP gel shift assays data fitted with the derived function for homologous competitive binding with ligand depletion. Calculated *Kd* values are shown. (A) Competition assays with GFPsol-LMO4$_{LIMs}$:Ldb1$_{LID}$, (B) GFPsol-Isl1$_{LIMs}$:Ldb1$_{LID}$ and (C) GFPsol-Lhx3$_{LIMs}$:Ldb1$_{LID}$. In each case the standard error of the fit to a single data set is reported.

7.3 Fluorescent anisotropy assays with GFP-tagged proteins

In theory it should be possible to carry out FA experiments with GFP-LIM:LID complexes in the same manner as described in section 6.10 for Fl-LIM:LID complexes. Unfortunately, the intrinsic anisotropy of GFP-fusion proteins is already close to the theoretically

maximal anisotropy value of 0.4 [331, 332], meaning that change in anisotropy on binding is limiting (e.g., [333, 334]). Attempts to carry out such experiments confirmed high starting anisotropies for GFPsol-LMO4$_{LIMs}$:Ldb1$_{LID}$ (r = 0.33–0.34) and YPet-LMO4$_{LIMs}$:Ldb1$_{LID}$ (r = 0.32). The r-values did not change with titration of the MBP-Ldb1$_{LID}$ competitor into the cut complexes (not shown), indicating that the approach was impractical.

7.4 Discussion

The experiments in this chapter show that LIMs-LID fusion constructs can be generated with an N-terminal GFP-based tag, and that these proteins can be used to generate binding curves for fluorescence gel-shift-based homologous competition experiments. Similar experiments could not be found in the literature suggesting that these are the first such experiments to be reported. However, there are some limitations to these assays that would need to be overcome before they could be used to assess LIM:LID interactions for the LMO and LIM-HD families.

7.4.1 Limitations of fluorescence gel shift assays.

One of the main limitations of this gel shift assay is that the concentrations of GFP-labelled proteins required for detection are potentially too high for the accurate determination of K_d (see later for a more detailed explanation). In most previous reports of fluoresence gel shift experiments for protein:protein interactions concentrations were ≥ 0.5 μM for GFP-labelled proteins [325, 326, 333-341]. In this study it was possible to drop the concentration to ~10 nM using CN-PAGE and GFPsol-labelled proteins. By using brighter FP variants, this detection limit could potentially be lowered in future experiments. Apart from brightness of the GFP, the detection limit in native gels depends on their ability to focus protein bands during electrophoresis. It may be possible to do this using horizontal submarine polyacrylamide gels with larger wells. A modified version of BN-PAGE [342] using GFPsol-fusion constructs also hold some promise for improved resolution and sensitivity. Improved separation in clear native gel might be achieved with supercharged scGFP variants [343, 344].

An important consideration is that this approach is that it is not a true equilibrium method - complexes with fast off rates can dissociate during electrophoresis [335]. While this problem

can be at least partly overcome by incubation at low temperatures (e.g., 4 °C) to reduce off-rates, it is probably not an issue for LIM-LID interactions. Indeed the reverse is probably the case. An approximate halflife of the initial $LMO4_{LIMs}$:$Ldb1_{LID}$ complex was estimated to be about 30 minutes, which represents a moderate-to-slow off-rate. In retrospect, the concentrations at which the half-life was measured were too high and could underestimate $t_{1/2}$ of dissociation [345]. A recommendation of incubation time as 5 x $t_{1/2}$ [345] suggests that reactions probably did not reach equilibrium within the 1 h incubation period. This discrepency could explain why the shifted band from the exploratory gel shift experiments did not reach 50% of total fluorescence. For all of the binding experiments in this and the previous chapter it would be reasonable to assume that with complete equilibration, the curves are likely to shift to the left, towards low nM K_d values or even lower. As different complexes could have different $t_{1/2}$, each should be determined independently.

Based on observations that MBP appeared to enhance overall fluorescence for GFP-$LMO4_{LIMS}$-$Ldb1_{LID}$ constructs (section 7.2.2), it may be generally useful to add BSA in the buffers for any binding assay using this class of proteins.

7.4.2 A comparison of binding affinity studies

Although the fluorescence based binding studies reported in this and the previous chapter are preliminary in nature, they were compared to existing data from our laboratory for this class of interaction [18, 167]. For consistency all of the original raw data from those papers were refitted using Origin9.1 using the same equation without fixing any parameters (Table 7.1).

All of the estimates of IC_{50} are in general agreement. However, in analogy with standard cold homologous competition experiments, $IC_{50} \approx$ [starting complex] + K_d, meaning that if the concentration of the starting complex is high relative to the K_d, then the IC_{50} is not discriminatory. Aside from the possibility that data are compromised by insufficient equilibration times, ligand depletion will have occurred at low concentrations of titrant because of limits of fluorescence detection, making simple estimates of K_d from IC_{50} invalid [176]. Fitting the data for homologous competition assays with ligand depletion, should provide a better estimate of K_d, An alternative solution to the ligand depletion problem would be to carry out a series of titration experments at different concentrations of the labelled starting complex and carry out a global fit

by a single site competitive binding model. However, variability in the data obtained indicates that further optimisation of existing assays, including detection of more points in the initial stages of the titration to get more complete binding curves (as suggested in the previous section) are required. If this is not viable alternative methodological approaches will be required.

Table 7.1. A comparison of IC_{50} values of LIM-LID interactions obtained by fitting previously reported and new data with binding curves through non-linear regression analysis. (*) The original raw data was standardized and fitted by a sigmoid logarithmic curve, yielding log10(IC_{50}) and associated SE. IC_{50} values and CI 95% were calculated as in Motulsky and Christopoulos, 2003 [176]. All experiments performed in this thesis used the same (10 nM) concentration of the labelled LIMs:LID complexes. This concentration was not determined for previously reported ELISA assays.

.

Binding to $Ldb1_{LID}$	IC_{50} (nM)	CI 95% (nM)	Competitive binding assay	Reported in
$LMO4_{LIMs}$	8	3–26	ELISA	[167]*
	19	7–51	gel shift assay	this study
	10	4–27		
	24	16–37	fluorescence anisotropy	this study
	31	18–53		
$LMO2_{LIMs}$	19	7–53	ELISA	[167]*
$Lhx3_{LIMs}$	38	9–171	ELISA	[18]*
	44	16–124	gel shift assay	this study
$Isl1_{LIMs}$	192	43–856	ELISA	[18]*
	89	30–269	gel shift assay	this study

Chapter 8 Final discussion

8.1 The biological implications of LIM-HD interactions

A large portion of this thesis was focussed on trying to map interactions between LIM-HD proteins. Y2H mating arrays from chapter 3 of this thesis showed that robust LIM-mediated interactions between LIM-HD proteins exist only between Lhx3/4 and Isl1/2 proteins (Figure 8.1). Weaker interactions were identified between $Isl1_{LBD}$ and the LIM domains of Lmx1a/b, and the $Isl2_{LBD}$ and Lmx1b. BiFC experiments from Chapter 5 didn't give additional information about interactions.

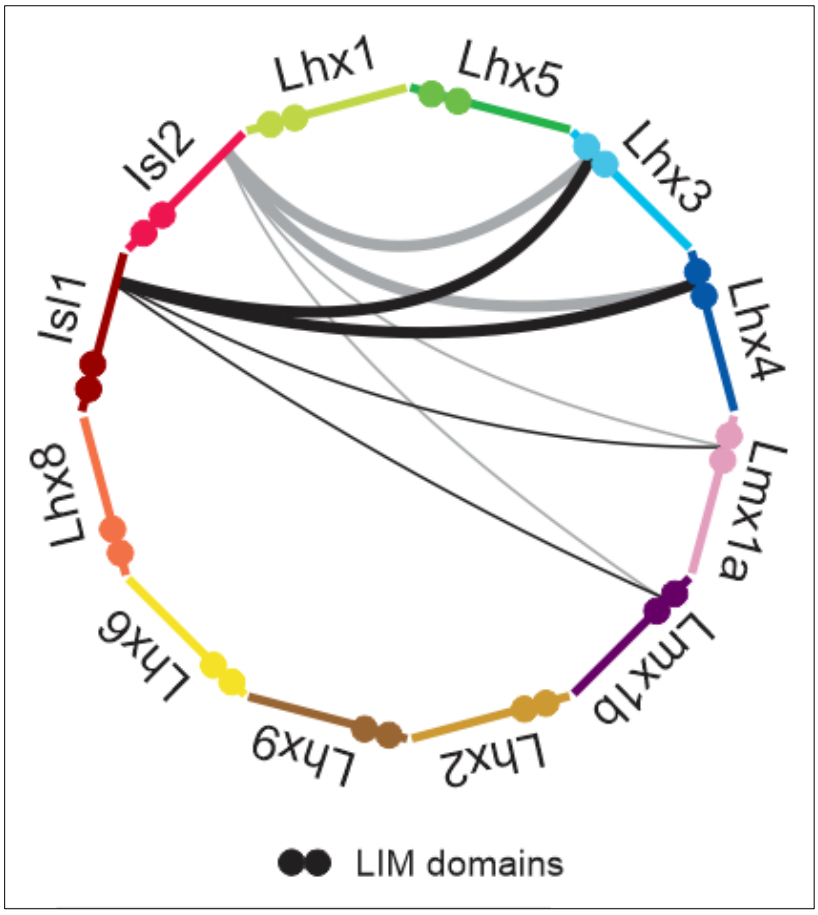


Figure 8.1. Schematic of interactions between LIM-HD proteins. Proteins are depicted as differently coloured lines, arranged in a circle. Tandem LIM domains are shown as small coloured circles. Interactions between proteins are represented by connecting lines. Black and grey lines represent interactions mediated by Isl1 and Isl2, respectively. Line width indicates apparent strength of an interaction detected by Y2H assays.

8.2 Additional evidence for Lmx1a/b vs. Isl1/2 protein interactions

Interactions between Isl1/2 proteins and Lhx3/4 proteins have been highly characterised and are established as important in various developmental processes (section 1.8). According to the Mouse Gene Expression Database Lmx1a/b and Isl1/2 proteins are co-expressed in various tissues during development (Table 8.1) [86, 191], and in the adult mouse gut [346], but it is unclear if they are co-expressed in the same cells.

Table 8.1 Co-expression of Lmx1a/b and Isl1/2 proteins in different stages of mouse development. Structures in which co-expression was detected are shown for embryonic stages E10.5–E15.5 and postnatal stages P(postnatal) and P1(postnatal day 1).

	Isl1	Isl2	
Lmx1a	hindbrain, future spinal	hindbrain	E10.5
	midbrain, hindbrain, future spinal cord	hindbrain	E11.5
	cerebellum primordium		E12.5
	cochlea, labyrinth, hypothalamus mantle layer, midbrain tegmentum	midbrain mantle layer	E14.5
	urinary bladder		E15.5
	thalamus		P
Lmx1b	midbrain, hindbrain, future spinal cord	midbrain, hindbrain, future spinal cord	E10.5
	hindbrain		E11.5
	hindbrain	midbrain, hindbrain, spinal cord	E13.5
	forelimb, hypothalamus, thalamus, superior colliculus, midbrain tegmentum, pons, medulla oblongata, spinal cord		E14.5
	thymus primordium	renal blood vessels	E15.5
	hypothalamus, thalamus, cranial ganglion, dorsal root ganglion	cranial ganglion	P
	thymus		P1

Lmx1a/b share a common ancestral gene with Lhx3/4 (Figure 1.3) and Lhx1/5. As Lhx3/4 and Lhx1/5 are more closely related than Lhx3/4 and Lmx1a/b, it is probably less surprising that Lmx1a/b_{LIMs} exhibit some Isl1/2 binding activity than the fact that Lhx1/5_{LIMs} do not. The interactions between Lmx1a/b_{LIMs} and Isl1α/2_{LBD} domains are apparently very weak compared to the interaction of Lmx1a/b_{LIMs} with the widely expressed $Ldb1_{LID}$, but it is not clear if Lmx1a/b can take part in ternary transcription factors complexes with Isl1/2. Additional experiments are required to confirm if interactions between Lmx1a/b and Isl1/2 are real and biologically relevant. It is possible that Lhx1/5 did not show interaction with Isl1/2 because some common core hydrophobic residues (positions 30, 90, 110 and 118 in Figure 1.10) are specific for Lhx1/5.

8.3 Interactions involving Lhx8

Lhx8 showed some evidence of a potentially weak interaction (some yeast growth but no colour development under routinely used selection conditions, except when extended incubation times resulted in the development of blue colour) with the LIM domains of all LIM-HD proteins (e.g., Fig 3.7). The AD-Lhx8 construct showed evidence of autoactivation under weak selection conditions (Figure 3.3), suggesting that most of the yeast growth probably stems from additive effects of autoactivation and sticky DBD-LIMs constructs. It is interesting to note that an interaction between Isl1α and Lhx8 in cholinergic cells was recently reported [347]. The authors of that paper used the results of GST-pulldown and co-IP experiments from cell lysates to suggest that Isl1α and Lhx8 form a complex similar to Isl1α and Lhx3 complex. The data presented in this thesis do not support those findings as DBD-$Lhx8_{LIMs}$ did not show any evidence of an interaction with any of full-length Isl1α, Isl2 or $Isl1_{LBD}$ (Figures 3.6 and 3.7). However, it is possible that the affinity of such an interaction lies below the limit of detection by Y2H, or that direct Lhx8:Isl1 interaction cannot be detected in yeast for various other reasons (see section 3.8.1).

8.4 Contribution of separate LIM domains to binding

In Chapter 4 it was shown that both LIM domains of mouse LIM-HD proteins are required for high affinity binding to $Ldb1_{LID}$, but that different subsets of LIM-HD proteins display higher levels of binding through their LIM1 or LIM2 domains. Note that the overall binding affinity is thought to result from negative cooperative binding of two LIM domains to $Ldb1_{LID}$, as it is likely that $K_{d\text{LIMs}} > K_{d\text{LIM1}} \cdot K_{d\text{LIM2}}$ [167].

A simple comparison of protein sequences between paralogous pairs of mouse LIM-HD proteins indicates that the dominant Ldb1-binding LIM domains are generally more conserved than the non-dominant domain, as defined by Y2H (Table 8.3). The exception here is for Lhx2/9 where there are relatively few sequence differences. Note that Lhx1/5, for which no dominant domains were identified, have the same number of sequence differences between each domain. However, at this stage it is not known if any of these sequence differences are more relevant to Ldb1 binding.

Table 8.3. The number of differences in the primary sequence of LIM1 or LIM2 domains between LIM-HD proteins in each paralogue pair in mice. Yellow fields mark domains which were shown to separately interact with $Ldb1_{LID}$.

The number of differences in the primary sequence of each domain		
	LIM1	LIM2
Isl1 vs. Isl2	8	19
Lhx3 vs. Lhx4	12	7
Lhx1 vs. Lhx5	5	5
Lmx1a vs. Lmx1b	20	10
Lhx2 vs. Lhx9	3	6
Lhx6 vs. Lhx8	17	13

Some structure and sequence-based explanations have been previously put forward to rationalise the differences between LIM1 and LIM2 dominance in terms of Ldb1-binding [162]. However, to date only one LIM2-dominant LIMs-LID structure has been determined (compared with structures for several LIM1-dominant LIM-LID structures) and a full explanation of these differences is speculative until more information is available.

Previous studies have shown that whereas $Isl1_{LIM1}$ is the preferential $Ldb1_{LID}$ binder, $Lhx3_{LIM2}$ preferentially binds both $Ldb1_{LID}$ and $Isl1_{LBD}$ [18]. In the same study strand exchange of $Isl1_{LBD}$ was observed for one molecule in the asymmetric unit of the crystal structure of $Lhx3_{LIMs}$-$Isl1_{LBD}$ (i.e., $Lhx3_{LIM2}$ binds the N-terminal half of $Isl1_{LBD}$ from one tethered complex but $Lhx3_{LIM1}$ binds the C-terminal of $Isl1_{LBD}$ from a symmetry related molecule to form a symmetry related dimer: pdb code 2RGT, chain a). Given these properties and the modular nature of LIM-peptide binding, a putative mechanism for cofactor exchange for the switch from Lhx3:Ldb1 complexes to Lhx3:Isl1:Ldb1 was proposed. In this mechanism intermediate states would involve $Ldb1_{LID}$ binding simultaneously to the $Isl1_{LIM1}$ and $Lhx3_{LIM2}$ [18]. However, there is no biological evidence to support or refute this hypothetical mechanism.

Previous Y2H experiments with *C. elegans* Lhx3/4 orthologue Ceh-14 showed that its LIM2 can independently bind the Isl1/2 orthologue Lim7, whereas LIM1 cannot [155]. This indicates that the mode of tandem LIM-domain binding could be conserved, at least for some proteins.

The strong conservation of $Ldb1_{LID}$ in metazoans implies that every residue in this sequence contributes to complex formation, in terms of making productive protein-protein interactions while preventing non-specific interactions. The $Ldb1_{LID}$ domain is mostly disordered but the charged DEDER region in the sequence has some helical propensity which is evident in the unbound state (pdb code 2LXD) [161]. The LMO2:Ldb1 (pdb code 2XJY) and Lhx3:Ldb1 complexes (pdb code 2JTN) show that this sequence can retain a degree of helicity in the bound state. This is not uncommon for proteins with internal intrinsic disorder where smaller elements of secondary structure (e.g., alpha helices) form prior to binding to function as molecular recognition elements [348, 349].

8.5 Considerations for binding assays

The yeast two-hybrid data provide some measure of relative binding affinities for similar sets of constructs, but are affected by stability of constructs in yeast, background binding events, and possible steric effects as detailed in section 3.8. The original ELISA-based homologous competition binding experiments appear to have been compromised by invalid assumptions about available concentrations of proteins and other experimental considerations as noted in sections 1.11 and 7.4.2. Thus, there is still a need for methodologies that can assay binding between LIM-HD and LMO family proteins and their peptide binding partners. The fluorescence-based binding assays described in chapters 6 and 7 have not yet resulted in high confidence binding data, but have provided useful information for design of future experiments.

Nonspecific adsorption of LIM domain constructs to surfaces in vitro leads to significant losses of LIM-domains at low concentrations. This effect can be minimized by including BSA in the buffer. Tubes, cuvettes or wells could alternatively be pre-coated with PEG or similar polymers, provided that LIM domains do not bind non-specifically to those polymers.

Future experiments that aim to determine binding affinities of LIM domains by competition assays should first determine half-lives of cleaved tethered complexes at very low complex concentrations. The incubation times should be long ($> 5 \cdot t_{1/2}$) in order for reactions to come sufficiently close to equilibrium as discussed in section 7.4.1.

Due to the use of tethered complexes, ligand depletion can affect competition assays and, depending on the assay, specific equations may need to be derived to adequately fit the data.

Finally, strong interactions require assays that are sensitive enough to detect very low concentrations of the label. Site-specific labelling of proteins with radioactive reagents could potentially provide such sensitivity.

8.6 Concluding remarks

In conclusion, this thesis has used a series of different experimental approaches to explore protein interactions between LIM-HD proteins and between LIM-HD proteins and their key binding partner Ldb1. Although many technical challenges were overcome during the course of the project, additional experiments will be required to fully characterise the networks of proteins

associated with this biologically important class of proteins. This body of work provides a good platform for such experiments.

References

1. Lelli, K.M., M. Slattery, and R.S. Mann, *Disentangling the many layers of eukaryotic transcriptional regulation.* Annu Rev Genet, 2012. **46**: p. 43-68.
2. Darrow, E.M. and B.P. Chadwick, *Boosting transcription by transcription: enhancer-associated transcripts.* Chromosome Res, 2013. **21**(6-7): p. 713-24.
3. Heinz, S., et al., *The selection and function of cell type-specific enhancers.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 144-54.
4. Allen, B.L. and D.J. Taatjes, *The Mediator complex: a central integrator of transcription.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 155-66.
5. Maeshima, K., et al., *The physical size of transcription factors is key to transcriptional regulation in chromatin domains.* J Phys Condens Matter, 2015. **27**(6): p. 064116.
6. Francesconi, M., R. Jelier, and B. Lehner, *Integrated genome-scale prediction of detrimental mutations in transcription networks.* PLoS Genet, 2011. **7**(5): p. e1002077.
7. Degnan, B.M., et al., *Early evolution of metazoan transcription factors.* Curr Opin Genet Dev, 2009. **19**(6): p. 591-9.
8. ENCODE, *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
9. Joshi, R., et al., *Functional specificity of a Hox protein mediated by the recognition of minor groove structure.* Cell, 2007. **131**(3): p. 530-43.
10. Berger, M.F., et al., *Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences.* Cell, 2008. **133**(7): p. 1266-76.
11. Noyes, M.B., et al., *Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites.* Cell, 2008. **133**(7): p. 1277-89.
12. Mann, R.S., K.M. Lelli, and R. Joshi, *Hox specificity unique roles for cofactors and collaborators.* Curr Top Dev Biol, 2009. **88**: p. 63-101.
13. Chu, S.W., et al., *Exploring the DNA-recognition potential of homeodomains.* Genome Res, 2012. **22**(10): p. 1889-98.
14. Jolma, A., et al., *DNA-binding specificities of human transcription factors.* Cell, 2013. **152**(1-2): p. 327-39.
15. Holland, P.W., H.A. Booth, and E.A. Bruford, *Classification and nomenclature of all human homeobox genes.* BMC Biol, 2007. **5**: p. 47.
16. Hiratani, I., et al., *Functional domains of the LIM homeodomain protein Xlim-1 involved in negative regulation, transactivation, and axis formation in Xenopus embryos.* Dev Biol, 2001. **229**(2): p. 456-67.
17. Yasuoka, Y., et al., *Evolutionary origins of blastoporal expression and organizer activity of the vertebrate gastrula organizer gene lhx1 and its ancient metazoan paralog lhx3.* Development, 2009. **136**(12): p. 2005-14.
18. 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. 2018-29.
19. 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. 42971-80.
20. Sloop, K.W., et al., *Differential activation of pituitary hormone genes by human Lhx3 isoforms with distinct DNA binding properties.* Mol Endocrinol, 1999. **13**(12): p. 2212-25.

21. Kimura, N., et al., *A brain region-specific gene product Lhx6.1 interacts with Ldb1 through tandem LIM-domains.* J Biochem, 1999. **126**(1): p. 180-7.
22. Ando, K., et al., *Isolation and characterization of an alternatively spliced variant of transcription factor Islet-1.* J Mol Endocrinol, 2003. **31**(3): p. 419-25.
23. Srivastava, M., et al., *Early evolution of the LIM homeobox gene family.* BMC Biol, 2010. **8**: p. 4.
24. Simmons, D.K., K. Pang, and M.Q. Martindale, *Lim homeobox genes in the Ctenophore Mnemiopsis leidyi: the evolution of neural cell type specification.* Evodevo, 2012. **3**(1): p. 2.
25. Lee, S., et al., *A regulatory network to segregate the identity of neuronal subtypes.* Dev Cell, 2008. **14**(6): p. 877-89.
26. Christensen, R.G., et al., *Recognition models to predict DNA-binding specificities of homeodomain proteins.* Bioinformatics, 2012. **28**(12): p. i84-9.
27. German, M.S., et al., *Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex.* Genes Dev, 1992. **6**(11): p. 2165-76.
28. 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. 2985-93.
29. Mochizuki, T., et al., *Xlim-1 and LIM domain binding protein 1 cooperate with various transcription factors in the regulation of the goosecoid promoter.* Dev Biol, 2000. **224**(2): p. 470-85.
30. Bridwell, J.A., et al., *Role of the LIM domains in DNA recognition by the Lhx3 neuroendocrine transcription factor.* Gene, 2001. **277**(1-2): p. 239-50.
31. Lee, S.K., et al., *Analysis of embryonic motoneuron gene regulation: derepression of general activators function in concert with enhancer factors.* Development, 2004. **131**(14): p. 3295-306.
32. Granger, A., et al., *The LIM-homeodomain proteins Isl-1 and Lhx3 act with steroidogenic factor 1 to enhance gonadotrope-specific activity of the gonadotropin-releasing hormone receptor gene promoter.* Mol Endocrinol, 2006. **20**(9): p. 2093-108.
33. Chou, S.J., et al., *Conserved regulatory elements establish the dynamic expression of Rpx/HesxI in early vertebrate development.* Dev Biol, 2006. **292**(2): p. 533-45.
34. 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. 566-77.
35. Flandin, P., et al., *Lhx6 and Lhx8 coordinately induce neuronal expression of Shh that controls the generation of interneuron progenitors.* Neuron, 2011. **70**(5): p. 939-50.
36. Guo, T., et al., *ISL1 promotes pancreatic islet cell proliferation.* PLoS One, 2011. **6**(8): p. e22387.
37. Golzio, C., et al., *ISL1 directly regulates FGF10 transcription during human cardiac outflow formation.* PLoS One, 2012. **7**(1): p. e30677.
38. 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. 287-99.
39. Ellsworth, B.S., D.L. Butts, and S.A. Camper, *Mechanisms underlying pituitary hypoplasia and failed cell specification in Lhx3-deficient mice.* Dev Biol, 2008. **313**(1): p. 118-29.
40. Davis, S.W., et al., *Molecular mechanisms of pituitary organogenesis: In search of novel regulatory genes.* Mol Cell Endocrinol, 2010. **323**(1): p. 4-19.
41. Mullen, R.D., et al., *Roles of the LHX3 and LHX4 LIM-homeodomain factors in pituitary development.* Mol Cell Endocrinol, 2007. **265-266**: p. 190-5.
42. Christiaen, L., et al., *Spatio-temporal intersection of Lhx3 and Tbx6 defines the cardiac field through synergistic activation of Mesp.* Dev Biol, 2009. **328**(2): p. 552-60.

43. Hukriede, N.A., et al., *Conserved requirement of Lim1 function for cell movements during gastrulation.* Dev Cell, 2003. **4**(1): p. 83-94.
44. Shawlot, W., et al., *Lim1 is required in both primitive streak-derived tissues and visceral endoderm for head formation in the mouse.* Development, 1999. **126**(22): p. 4925-32.
45. Bedont, J.L., et al., *Lhx1 controls terminal differentiation and circadian function of the suprachiasmatic nucleus.* Cell Rep, 2014. **7**(3): p. 609-22.
46. Cirio, M.C., et al., *Lhx1 is required for specification of the renal progenitor cell field.* PLoS One, 2011. **6**(4): p. e18858.
47. Kania, A. and T.M. Jessell, *Topographic motor projections in the limb imposed by LIM homeodomain protein regulation of ephrin-A:EphA interactions.* Neuron, 2003. **38**(4): p. 581-96.
48. Sheng, H.Z., et al., *Multistep control of pituitary organogenesis.* Science, 1997. **278**(5344): p. 1809-12.
49. Peng, G. and M. Westerfield, *Lhx5 promotes forebrain development and activates transcription of secreted Wnt antagonists.* Development, 2006. **133**(16): p. 3191-200.
50. Miquelajauregui, A., et al., *LIM-homeobox gene Lhx5 is required for normal development of Cajal-Retzius cells.* J Neurosci, 2010. **30**(31): p. 10551-62.
51. Kobayashi, A., et al., *Requirement of Lim1 for female reproductive tract development.* Development, 2004. **131**(3): p. 539-49.
52. Poche, R.A., et al., *Lim1 is essential for the correct laminar positioning of retinal horizontal cells.* J Neurosci, 2007. **27**(51): p. 14099-107.
53. Margeta, M.A., *Transcription factor Lim1 specifies horizontal cell laminar position in the retina.* J Neurosci, 2008. **28**(15): p. 3835-6.
54. Pillai, A., et al., *Lhx1 and Lhx5 maintain the inhibitory-neurotransmitter status of interneurons in the dorsal spinal cord.* Development, 2007. **134**(2): p. 357-66.
55. 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 U S A, 2007. **104**(32): p. 13182-6.
56. Failli, V., I. Bachy, and S. Retaux, *Expression of the LIM-homeodomain gene Lmx1a (dreher) during development of the mouse nervous system.* Mech Dev, 2002. **118**(1-2): p. 225-8.
57. Andersson, E., et al., *Identification of intrinsic determinants of midbrain dopamine neurons.* Cell, 2006. **124**(2): p. 393-405.
58. Simon, H.H., et al., *Midbrain dopaminergic neurons: determination of their developmental fate by transcription factors.* Ann N Y Acad Sci, 2003. **991**: p. 36-47.
59. Giraldez, F., *Regionalized organizing activity of the neural tube revealed by the regulation of lmx1 in the otic vesicle.* Dev Biol, 1998. **203**(1): p. 189-200.
60. Cheng, L., et al., *Lmx1b, Pet-1, and Nkx2.2 coordinately specify serotonergic neurotransmitter phenotype.* J Neurosci, 2003. **23**(31): p. 9961-7.
61. Ding, Y.Q., et al., *Lmx1b is essential for the development of serotonergic neurons.* Nat Neurosci, 2003. **6**(9): p. 933-8.
62. Smidt, M.P., et al., *A second independent pathway for development of mesencephalic dopaminergic neurons requires Lmx1b.* Nat Neurosci, 2000. **3**(4): p. 337-41.
63. 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.

64. Marcos-Mondejar, P., et al., *The lhx2 transcription factor controls thalamocortical axonal guidance by specific regulation of robo1 and robo2 receptors.* J Neurosci, 2012. **32**(13): p. 4372-85.
65. Yun, S., et al., *Lhx2 links the intrinsic and extrinsic factors that control optic cup formation.* Development, 2009. **136**(23): p. 3895-906.
66. Roy, A., et al., *LHX2 is necessary for the maintenance of optic identity and for the progression of optic morphogenesis.* J Neurosci, 2013. **33**(16): p. 6877-84.
67. Benveniste, R.J., et al., *Cell type-specific regulation of the Drosophila FMRF-NH2 neuropeptide gene by Apterous, a LIM homeodomain transcription factor.* Development, 1998. **125**(23): p. 4757-65.
68. Lundgren, S.E., et al., *Control of neuronal pathway selection by the Drosophila LIM homeodomain gene apterous.* Development, 1995. **121**(6): p. 1769-73.
69. Hunter, C.S. and S.J. Rhodes, *LIM-homeodomain genes in mammalian development and human disease.* Mol Biol Rep, 2005. **32**(2): p. 67-77.
70. Wilson, S.I., et al., *A molecular program for contralateral trajectory: Rig-1 control by LIM homeodomain transcription factors.* Neuron, 2008. **59**(3): p. 413-24.
71. Fogarty, M., et al., *Spatial genetic patterning of the embryonic neuroepithelium generates GABAergic interneuron diversity in the adult cortex.* J Neurosci, 2007. **27**(41): p. 10935-46.
72. Bachy, I. and S. Retaux, *GABAergic specification in the basal forebrain is controlled by the LIM-hd factor Lhx7.* Dev Biol, 2006. **291**(2): p. 218-26.
73. Mori, T., et al., *The LIM homeobox gene, L3/Lhx8, is necessary for proper development of basal forebrain cholinergic neurons.* Eur J Neurosci, 2004. **19**(12): p. 3129-41.
74. Zhao, Y., et al., *The LIM-homeobox gene Lhx8 is required for the development of many cholinergic neurons in the mouse forebrain.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 9005-10.
75. Pangas, S.A., et al., *Oogenesis requires germ cell-specific transcriptional regulators Sohlh1 and Lhx8.* Proc Natl Acad Sci U S A, 2006. **103**(21): p. 8090-5.
76. Alifragis, P., A. Liapi, and J.G. Parnavelas, *Lhx6 regulates the migration of cortical interneurons from the ventral telencephalon but does not specify their GABA phenotype.* J Neurosci, 2004. **24**(24): p. 5643-8.
77. Liodis, P., et al., *Lhx6 activity is required for the normal migration and specification of cortical interneuron subtypes.* J Neurosci, 2007. **27**(12): p. 3078-89.
78. 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. 237-49.
79. Elshatory, Y., et al., *Islet-1 controls the differentiation of retinal bipolar and cholinergic amacrine cells.* J Neurosci, 2007. **27**(46): p. 12707-20.
80. 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.
81. 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. 3931-42.
82. Lin, L., et al., *Isl1 is upstream of sonic hedgehog in a pathway required for cardiac morphogenesis.* Dev Biol, 2006. **295**(2): p. 756-63.
83. Yang, L., et al., *Isl1Cre reveals a common Bmp pathway in heart and limb development.* Development, 2006. **133**(8): p. 1575-85.

84. Nasif, S., et al., *Islet 1 specifies the identity of hypothalamic melanocortin neurons and is critical for normal food intake and adiposity in adulthood.* Proc Natl Acad Sci U S A, 2015. **112**(15): p. E1861-70.
85. Pak, W., et al., *Magnitude of binocular vision controlled by islet-2 repression of a genetic program that specifies laterality of retinal axon pathfinding.* Cell, 2004. **119**(4): p. 567-78.
86. Smith, C.M., et al., *The mouse Gene Expression Database (GXD): 2014 update.* Nucleic Acids Res, 2014. **42**(Database issue): p. D818-24.
87. Dasen, J.S. and T.M. Jessell, *Hox networks and the origins of motor neuron diversity.* Curr Top Dev Biol, 2009. **88**: p. 169-200.
88. Alaynick, W.A., T.M. Jessell, and S.L. Pfaff, *SnapShot: spinal cord development.* Cell, 2011. **146**(1): p. 178-178 e1.
89. Lu, D.C., T. Niu, and W.A. Alaynick, *Molecular and cellular development of spinal cord locomotor circuitry.* Front Mol Neurosci, 2015. **8**: p. 25.
90. Tsuchida, T., et al., *Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes.* Cell, 1994. **79**(6): p. 957-70.
91. Lumsden, A., *Neural development. A 'LIM code' for motor neurons?* Curr Biol, 1995. **5**(5): p. 491-5.
92. Bachy, I., V. Failli, and S. Retaux, *A LIM-homeodomain code for development and evolution of forebrain connectivity.* Neuroreport, 2002. **13**(2): p. A23-7.
93. Gill, G.N., *Decoding the LIM development code.* Trans Am Clin Climatol Assoc, 2003. **114**: p. 179-89.
94. Liang, X., et al., *Isl1 is required for multiple aspects of motor neuron development.* Mol Cell Neurosci, 2011. **47**(3): p. 215-22.
95. 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. 2923-32.
96. Hester, M.E., et al., *Rapid and efficient generation of functional motor neurons from human pluripotent stem cells using gene delivered transcription factor codes.* Mol Ther, 2011. **19**(10): p. 1905-12.
97. 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. 731-45.
98. 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 U S A, 2012. **109**(9): p. 3383-8.
99. Mazzoni, E.O., et al., *Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity.* Nat Neurosci, 2013. **16**(9): p. 1219-27.
100. Way, J.C. and M. Chalfie, *mec-3, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in C. elegans.* Cell, 1988. **54**(1): p. 5-16.
101. 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-11.* Nature, 1990. **344**(6269): p. 876-9.
102. 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. 879-82.
103. Krishna, S.S., I. Majumdar, and N.V. Grishin, *Structural classification of zinc fingers: survey and summary.* Nucleic Acids Res, 2003. **31**(2): p. 532-50.
104. Omichinski, J.G., et al., *NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1.* Science, 1993. **261**(5120): p. 438-46.

105. Perez-Alvarado, G.C., et al., *Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP.* Nat Struct Biol, 1994. **1**(6): p. 388-98.
106. Grishin, N.V., *Treble clef finger--a functionally diverse zinc-binding structural motif.* Nucleic Acids Res, 2001. **29**(8): p. 1703-14.
107. Matthews, J.M. and M. Sunde, *Zinc fingers--folds for many occasions.* IUBMB Life, 2002. **54**(6): p. 351-5.
108. 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. 920-31.
109. Burroughs, A.M., L.M. Iyer, and L. Aravind, *Functional diversification of the RING finger and other binuclear treble clef domains in prokaryotes and the early evolution of the ubiquitin system.* Mol Biosyst, 2011. **7**(7): p. 2261-77.
110. 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.
111. Moes, D., et al., *A LIM domain protein from tobacco involved in actin-bundling and histone gene transcription.* Mol Plant, 2013. **6**(2): p. 483-502.
112. Nishiya, N., et al., *The LIM domains of hic-5 protein recognize specific DNA fragments in a zinc-dependent manner in vitro.* Nucleic Acids Res, 1998. **26**(18): p. 4267-73.
113. Bronstein, R. and D. Segal, *Modularity of CHIP/LDB transcription complexes regulates cell differentiation.* Fly (Austin), 2011. **5**(3): p. 200-5.
114. Love, P.E., C. Warzecha, and L. Li, *Ldb1 complexes: the new master regulators of erythroid gene transcription.* Trends Genet, 2014. **30**(1): p. 1-9.
115. Ostendorff, H.P., et al., *Dynamic expression of LIM cofactors in the developing mouse neural tube.* Dev Dyn, 2006. **235**(3): p. 786-91.
116. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
117. Tuoc, T.C. and A. Stoykova, *Roles of the ubiquitin-proteosome system in neurogenesis.* Cell Cycle, 2010. **9**(16): p. 3174-80.
118. Mabb, A.M. and M.D. Ehlers, *Ubiquitination in postsynaptic function and plasticity.* Annu Rev Cell Dev Biol, 2010. **26**: p. 179-210.
119. Geng, F., S. Wenzel, and W.P. Tansey, *Ubiquitin and proteasomes in transcription.* Annu Rev Biochem, 2012. **81**: p. 177-201.
120. Durairaj, G. and P. Kaiser, *The 26S proteasome and initiation of gene transcription.* Biomolecules, 2014. **4**(3): p. 827-47.
121. Ostendorff, H.P., et al., *Ubiquitination-dependent cofactor exchange on LIM homeodomain transcription factors.* Nature, 2002. **416**(6876): p. 99-103.
122. Gungor, C., et al., *Proteasomal selection of multiprotein complexes recruited by LIM homeodomain transcription factors.* Proc Natl Acad Sci U S A, 2007. **104**(38): p. 15000-5.
123. Howard, P.W., et al., *Regulation of LIM-domain-binding 1 protein expression by ubiquitination of Lys134.* Biochem J, 2010. **429**(1): p. 127-36.
124. Chen, L., et al., *Ssdp proteins interact with the LIM-domain-binding protein Ldb1 to regulate development.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14320-5.
125. 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. 1915-25.

126. Xu, Z., et al., *Single-stranded DNA-binding proteins regulate the abundance of LIM domain and LIM domain-binding proteins.* Genes Dev, 2007. **21**(8): p. 942-55.
127. Cai, Y., et al., *Single-stranded DNA-binding proteins regulate the abundance and function of the LIM-homeodomain transcription factor LHX2 in pituitary cells.* Biochem Biophys Res Commun, 2008. **373**(2): p. 303-8.
128. Bach, I., et al., *A family of LIM domain-associated cofactors confer transcriptional synergism between LIM and Otx homeodomain proteins.* Genes Dev, 1997. **11**(11): p. 1370-80.
129. Bach, I., et al., *RLIM inhibits functional activity of LIM homeodomain transcription factors via recruitment of the histone deacetylase complex.* Nat Genet, 1999. **22**(4): p. 394-9.
130. Jurata, L.W., D.A. Kenny, and G.N. Gill, *Nuclear LIM interactor, a rhombotin and LIM homeodomain interacting protein, is expressed early in neuronal development.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11693-8.
131. Jurata, L.W. and G.N. Gill, *Functional analysis of the nuclear LIM domain interactor NLI.* Mol Cell Biol, 1997. **17**(10): p. 5688-98.
132. Jurata, L.W., S.L. Pfaff, and G.N. Gill, *The nuclear LIM domain interactor NLI mediates homo- and heterodimerization of LIM domain transcription factors.* J Biol Chem, 1998. **273**(6): p. 3152-7.
133. Agulnick, A.D., et al., *Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins.* Nature, 1996. **384**(6606): p. 270-2.
134. Morcillo, P., et al., *Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila.* Genes Dev, 1997. **11**(20): p. 2729-40.
135. Visvader, J.E., et al., *The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation.* Proc Natl Acad Sci U S A, 1997. **94**(25): p. 13707-12.
136. Breen, J.J., et al., *Interactions between LIM domains and the LIM domain-binding protein Ldb1.* J Biol Chem, 1998. **273**(8): p. 4712-7.
137. Milan, M. and S.M. Cohen, *Regulation of LIM homeodomain activity in vivo: a tetramer of dLDB and apterous confers activity and capacity for regulation by dLMO.* Mol Cell, 1999. **4**(2): p. 267-73.
138. van Meyel, D.J., et al., *Chip and apterous physically interact to form a functional complex during Drosophila development.* Mol Cell, 1999. **4**(2): p. 259-65.
139. Fernandez-Funez, P., et al., *The relative expression amounts of apterous and its co-factor dLdb/Chip are critical for dorso-ventral compartmentalization in the Drosophila wing.* EMBO J, 1998. **17**(23): p. 6846-53.
140. Cross, A.J., et al., *LIM domain binding proteins 1 and 2 have different oligomeric states.* J Mol Biol, 2010. **399**(1): p. 133-44.
141. Song, S.H., C. Hou, and A. Dean, *A positive role for NLI/Ldb1 in long-range beta-globin locus control region function.* Mol Cell, 2007. **28**(5): p. 810-22.
142. Soler, E., et al., *The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation.* Genes Dev, 2010. **24**(3): p. 277-89.
143. Deng, W., et al., *Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor.* Cell, 2012. **149**(6): p. 1233-44.
144. Stadhouders, R., et al., *Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development.* EMBO J, 2012. **31**(4): p. 986-99.
145. Schuh, A.H., et al., *ETO-2 associates with SCL in erythroid cells and megakaryocytes and provides repressor functions in erythropoiesis.* Mol Cell Biol, 2005. **25**(23): p. 10235-50.
146. Meier, N., et al., *Novel binding partners of Ldb1 are required for haematopoietic development.* Development, 2006. **133**(24): p. 4913-23.

147. 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 U S A, 2011. **108**(35): p. 14443-8.
148. Feuerstein, R., et al., *The LIM/double zinc-finger motif functions as a protein dimerization domain.* Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10655-9.
149. Arber, S. and P. Caroni, *Specificity of single LIM motifs in targeting and LIM/LIM interactions in situ.* Genes Dev, 1996. **10**(3): p. 289-300.
150. 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.
151. Deane, J.E., et al., *Letter to the Editor: (1)H, (15)N and (13)C assignments of FLIN4, an intramolecular LMO4:ldb1 complex.* J Biomol NMR, 2002. **23**(2): p. 165-6.
152. 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.
153. 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. 337-50.
154. Appel, B., et al., *Motoneuron fate specification revealed by patterned LIM homeobox gene expression in embryonic zebrafish.* Development, 1995. **121**(12): p. 4117-25.
155. Bhati, M., *Structural Characterisation of LIM Complexes That Regulate Neuronal Specification*, in *School of Molecular Biology*. 2010, University of Sydney.
156. Deane, J.E., et al., *Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex.* EMBO J, 2004. **23**(18): p. 3589-98.
157. Jeffries, C.M., et al., *Stabilization of a binary protein complex by intein-mediated cyclization.* Protein Sci, 2006. **15**(11): p. 2612-8.
158. El Omari, K., et al., *Structure of the leukemia oncogene LMO2: implications for the assembly of a hematopoietic transcription factor complex.* Blood, 2011. **117**(7): p. 2146-56.
159. Liew, C.W., et al., *(1)H, (1)(5)N and (1)(3)C assignments of an intramolecular LMO4-LIM1/CtIP complex.* Biomol NMR Assign, 2012. **6**(1): p. 31-4.
160. 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): p. e40719.
161. Dastmalchi, S., et al., *Solution structure of a tethered Lmo2(LIM2) /Ldb1(LID) complex.* Protein Sci, 2012. **21**(11): p. 1768-74.
162. 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. 21924-35.
163. Stokes, P.H., et al., *Structural basis of the interaction of the breast cancer oncogene LMO4 with the tumour suppressor CtIP/RBBP8.* J Mol Biol, 2013. **425**(7): p. 1101-10.
164. Joseph, S., et al., *The structure of an LIM-only protein 4 (LMO4) and Deformed epidermal autoregulatory factor-1 (DEAF1) complex reveals a common mode of binding to LMO4.* PLoS One, 2014. **9**(10): p. e109108.
165. Sewell, H., et al., *Conformational flexibility of the oncogenic protein LMO2 primes the formation of the multi-protein transcription complex.* Sci Rep, 2014. **4**: p. 3643.
166. Matthews, J.M. and J.R. Potts, *The tandem beta-zipper: modular binding of tandem domains and linear motifs.* FEBS Lett, 2013. **587**(8): p. 1164-71.
167. Ryan, D.P., et al., *Identification of the key LMO2-binding determinants on Ldb1.* J Mol Biol, 2006. **359**(1): p. 66-75.

168. Boeda, B., et al., *Molecular recognition of the Tes LIM2-3 domains by the actin-related protein Arp7A.* J Biol Chem, 2011. **286**(13): p. 11543-54.
169. Craig, V. 2006, University of Sydney.
170. Passner, J.M., et al., *Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex.* Nature, 1999. **397**(6721): p. 714-9.
171. Miyazono, K., et al., *Cooperative DNA-binding and sequence-recognition mechanism of aristaless and clawless.* EMBO J, 2010. **29**(9): p. 1613-23.
172. Hiraoka, J., et al., *Self-association of LIM-kinase 1 mediated by the interaction between an N-terminal LIM domain and a C-terminal kinase domain.* FEBS Lett, 1996. **399**(1-2): p. 117-21.
173. Nagata, K., et al., *The N-terminal LIM domain negatively regulates the kinase activity of LIM-kinase 1.* Biochem J, 1999. **343 Pt 1**: p. 99-105.
174. Sanchez-Garcia, I., et al., *The cysteine-rich LIM domains inhibit DNA binding by the associated homeodomain in Isl-1.* EMBO J, 1993. **12**(11): p. 4243-50.
175. Hornbeck, P., S.E. Winston, and S.A. Fuller, *Enzyme-linked immunosorbent assays (ELISA).* Curr Protoc Mol Biol, 2001. **Chapter 11**: p. Unit11 2.
176. Christopoulos, H.J.M.a.A., *Fitting Models to Biological Data using Linear and Nonlinear Regression* 2003, San Diego CA: GraphPad Software Inc.
177. Lee, N., C. Francklyn, and E.P. Hamilton, *Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter.* Proc Natl Acad Sci U S A, 1987. **84**(24): p. 8814-8.
178. Hopfner, K.P., et al., *Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3600-5.
179. Green, R., Rogers, E., *Chemical Transformation of E. coli.* Methods in Enzymology, 2013(529): p. 329-336.
180. Clontech, *Yeast Protocols Handbook*. 2008.
181. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press. p. 571-607.
182. Murphy, J.B., Kies, M.W., *Note on spectrophotometric determination of proteins in dilute solutions.* Biochimica et Biophysica Acta, 1960(45): p. 382-384.
183. Kapust, R.B., et al., *Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency.* Protein Eng, 2001. **14**(12): p. 993-1000.
184. Kapust, R.B., et al., *The P1' specificity of tobacco etch virus protease.* Biochem Biophys Res Commun, 2002. **294**(5): p. 949-55.
185. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.
186. Johnson, M., et al., *NCBI BLAST: a better web interface.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W5-9.
187. McWilliam, H., et al., *Analysis Tool Web Services from the EMBL-EBI.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W597-600.
188. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.
189. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
190. Cunningham, F., et al., *Ensembl 2015.* Nucleic Acids Res, 2015. **43**(Database issue): p. D662-9.
191. Bairoch, A., S. Cohen-Boulakia, and C. Froidevaux, *Review of the selected proceedings of the Fifth International Workshop on Data Integration in the Life Sciences 2008.* BMC Bioinformatics, 2008. **9 Suppl 8**: p. S1.
192. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.

193. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
194. Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions.* Nature, 1989. **340**(6230): p. 245-6.
195. Bruckner, A., et al., *Yeast two-hybrid, a powerful tool for systems biology.* Int J Mol Sci, 2009. **10**(6): p. 2763-88.
196. Golemis, E.A., et al., *Interaction trap/two-hybrid system to identify interacting proteins.* Curr Protoc Neurosci, 2011. **Chapter 4**: p. Unit 4 4.
197. Michnick, S.W., et al., *Protein-fragment complementation assays for large-scale analysis, functional dissection and dynamic studies of protein-protein interactions in living cells.* Methods Mol Biol, 2011. **756**: p. 395-425.
198. Stynen, B., et al., *Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system.* Microbiol Mol Biol Rev, 2012. **76**(2): p. 331-82.
199. Ruohonen, L., M.K. Aalto, and S. Keranen, *Modifications to the ADH1 promoter of Saccharomyces cerevisiae for efficient production of heterologous proteins.* J Biotechnol, 1995. **39**(3): p. 193-203.
200. Van Criekinge, W. and R. Beyaert, *Yeast Two-Hybrid: State of the Art.* Biol Proced Online, 1999. **2**: p. 1-38.
201. Serebriiskii, I.G. and E. Kotova, *Analysis of protein-protein interactions utilizing dual bait yeast two-hybrid system.* Methods Mol Biol, 2004. **261**: p. 263-96.
202. Soellick, T.R. and J.F. Uhrig, *Development of an optimized interaction-mating protocol for large-scale yeast two-hybrid analyses.* Genome Biol, 2001. **2**(12): p. RESEARCH0052.
203. Kolonin, M.G., J. Zhong, and R.L. Finley, *Interaction mating methods in two-hybrid systems.* Methods Enzymol, 2000. **328**: p. 26-46.
204. 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. 1778-9.
205. Finley, R.L., Jr. and R. Brent, *Interaction mating reveals binary and ternary connections between Drosophila cell cycle regulators.* Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12980-4.
206. Uetz, P., *Two-hybrid arrays.* Curr Opin Chem Biol, 2002. **6**(1): p. 57-62.
207. Rajagopala, S.V., et al., *The protein network of bacterial motility.* Mol Syst Biol, 2007. **3**: p. 128.
208. 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. 2691-9.
209. Uetz, P., et al., *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.* Nature, 2000. **403**(6770): p. 623-7.
210. Hauser, R., et al., *Matrix-based yeast two-hybrid screen strategies and comparison of systems.* Methods Mol Biol, 2012. **812**: p. 1-20.
211. Johnson, J.D., et al., *Transcriptional synergy between LIM-homeodomain proteins and basic helix-loop-helix proteins: the LIM2 domain determines specificity.* Mol Cell Biol, 1997. **17**(7): p. 3488-96.
212. 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. 1597-600.
213. Sloop, K.W., C.J. Dwyer, and S.J. Rhodes, *An isoform-specific inhibitory domain regulates the LHX3 LIM homeodomain factor holoprotein and the production of a functional alternate translation form.* J Biol Chem, 2001. **276**(39): p. 36311-9.
214. Thieu, K., in *School of Molecular Biosciences*. 2008, University of Sydney.

215. Bach, I., et al., *P-Lim, a LIM homeodomain factor, is expressed during pituitary organ and cell commitment and synergizes with Pit-1.* Proc Natl Acad Sci U S A, 1995. **92**(7): p. 2720-4.
216. Smirnov, M.N., et al., *Red pigment of adenine-deficient yeast Saccharomyces cerevisiae.* Biochem Biophys Res Commun, 1967. **27**(3): p. 299-304.
217. Fisher, C.R., *Enzymology of the pigmented adenine-requiring mutants of Saccharomyces and Schizosaccharomyces.* Biochem Biophys Res Commun, 1969. **34**(3): p. 306-10.
218. Sahota, A., et al., *Mutants of Saccharomyces cerevisiae deficient in adenine phosphoribosyltransferase.* Mutat Res, 1987. **180**(1): p. 81-7.
219. Weisman, L.S., R. Bacallao, and W. Wickner, *Multiple methods of visualizing the yeast vacuole permit evaluation of its morphology and inheritance during the cell cycle.* J Cell Biol, 1987. **105**(4): p. 1539-47.
220. James, P., J. Halladay, and E.A. Craig, *Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast.* Genetics, 1996. **144**(4): p. 1425-36.
221. Buckholz, R.G., et al., *Automation of yeast two-hybrid screening.* J Mol Microbiol Biotechnol, 1999. **1**(1): p. 135-40.
222. 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.
223. Estojak, J., R. Brent, and E.A. Golemis, *Correlation of two-hybrid affinity data with in vitro measurements.* Mol Cell Biol, 1995. **15**(10): p. 5820-9.
224. Fox, A.H., et al., *Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers.* EMBO J, 1999. **18**(10): p. 2812-22.
225. Liew, C.K., et al., *Solution structures of two CCHC zinc fingers from the FOG family protein U-shaped that mediate protein-protein interactions.* Structure, 2000. **8**(11): p. 1157-66.
226. Dubin, M.J., et al., *Dimerization of CtIP, a BRCA1- and CtBP-interacting protein, is mediated by an N-terminal coiled-coil motif.* J Biol Chem, 2004. **279**(26): p. 26932-8.
227. Beranger, F., et al., *Getting more from the two-hybrid system: N-terminal fusions to LexA are efficient and sensitive baits for two-hybrid studies.* Nucleic Acids Res, 1997. **25**(10): p. 2035-6.
228. Stellberger, T., et al., *Improving the yeast two-hybrid system with permutated fusions proteins: the Varicella Zoster Virus interactome.* Proteome Sci, 2010. **8**: p. 8.
229. Rubenstein, E.M. and M.C. Schmidt, *Mechanisms regulating the protein kinases of Saccharomyces cerevisiae.* Eukaryot Cell, 2007. **6**(4): p. 571-83.
230. Manning, G., et al., *Evolution of protein kinase signaling from yeast to man.* Trends Biochem Sci, 2002. **27**(10): p. 514-20.
231. Kawata, T., et al., *SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in dictyostelium.* Cell, 1997. **89**(6): p. 909-16.
232. Thien, C.B. and W.Y. Langdon, *Cbl: many adaptations to regulate protein tyrosine kinases.* Nat Rev Mol Cell Biol, 2001. **2**(4): p. 294-307.
233. Mockli, N. and D. Auerbach, *Quantitative beta-galactosidase assay suitable for high-throughput applications in the yeast two-hybrid system.* Biotechniques, 2004. **36**(5): p. 872-6.
234. Serebriiskii, I.G., et al., *A combined yeast/bacteria two-hybrid system: development and evaluation.* Mol Cell Proteomics, 2005. **4**(6): p. 819-26.
235. Chen, J., et al., *High throughput flow cytometry based yeast two-hybrid array approach for large-scale analysis of protein-protein interactions.* Cytometry A, 2012. **81**(1): p. 90-8.
236. Ruden, D.M., *Activating regions of yeast transcription factors must have both acidic and hydrophobic amino acids.* Chromosoma, 1992. **101**(5-6): p. 342-8.

237. Koegl, M. and P. Uetz, *Improving yeast two-hybrid screening systems.* Brief Funct Genomic Proteomic, 2007. **6**(4): p. 302-12.
238. Newman, J.R., E. Wolf, and P.S. Kim, *A computationally directed screen identifying interacting coiled coils from Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2000. **97**(24): p. 13203-8.
239. Hu, J.C., *A guided tour in protein interaction space: coiled coils from the yeast proteome.* Proc Natl Acad Sci U S A, 2000. **97**(24): p. 12935-6.
240. Marmorstein, R., et al., *DNA recognition by GAL4: structure of a protein-DNA complex.* Nature, 1992. **356**(6368): p. 408-14.
241. Hahn, S. and E.T. Young, *Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators.* Genetics, 2011. **189**(3): p. 705-36.
242. Ishii, J., et al., *A simple and immediate method for simultaneously evaluating expression level and plasmid maintenance in yeast.* J Biochem, 2009. **145**(6): p. 701-8.
243. Richards, F.M., *On the Enzymic Activity of Subtilisin-Modified Ribonuclease.* Proc Natl Acad Sci U S A, 1958. **44**(2): p. 162-6.
244. Ullmann, A., F. Jacob, and J. Monod, *On the subunit structure of wild-type versus complemented beta-galactosidase of Escherichia coli.* J Mol Biol, 1968. **32**(1): p. 1-13.
245. Taniuchi, H. and C.B. Anfinsen, *Simultaneous formation of two alternative enzymology active structures by complementation of two overlapping fragments of staphylococcal nuclease.* J Biol Chem, 1971. **246**(7): p. 2291-301.
246. Johnsson, N. and A. Varshavsky, *Split ubiquitin as a sensor of protein interactions in vivo.* Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10340-4.
247. Kerppola, T.K., *Visualization of molecular interactions by fluorescence complementation.* Nat Rev Mol Cell Biol, 2006. **7**(6): p. 449-56.
248. Michnick, S.W., et al., *Universal strategies in research and drug discovery based on protein-fragment complementation assays.* Nat Rev Drug Discov, 2007. **6**(7): p. 569-82.
249. Villalobos, V., S. Naik, and D. Piwnica-Worms, *Current state of imaging protein-protein interactions in vivo with genetically encoded reporters.* Annu Rev Biomed Eng, 2007. **9**: p. 321-49.
250. Shekhawat, S.S. and I. Ghosh, *Split-protein systems: beyond binary protein-protein interactions.* Curr Opin Chem Biol, 2011. **15**(6): p. 789-97.
251. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein.* Biochemistry, 1997. **36**(22): p. 6786-91.
252. Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44.
253. Magliery, T.J., et al., *Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism.* J Am Chem Soc, 2005. **127**(1): p. 146-57.
254. Kerppola, T.K., *Visualization of molecular interactions using bimolecular fluorescence complementation analysis: characteristics of protein fragment complementation.* Chem Soc Rev, 2009. **38**(10): p. 2876-86.
255. Melnik, T.N., et al., *Multi-state proteins: approach allowing experimental determination of the formation order of structure elements in the green fluorescent protein.* PLoS One, 2012. **7**(11): p. e48604.
256. Demidov, V.V., et al., *Fast complementation of split fluorescent protein triggered by DNA hybridization.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2052-6.
257. Boevink, P., et al., *In vivo protein-protein interaction studies with BiFC: conditions, cautions, and caveats.* Methods Mol Biol, 2014. **1127**: p. 81-90.

258. Crameri, A., et al., *Improved green fluorescent protein by molecular evolution using DNA shuffling.* Nat Biotechnol, 1996. **14**(3): p. 315-9.
259. Battistutta, R., A. Negro, and G. Zanotti, *Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the green fluorescent protein.* Proteins, 2000. **41**(4): p. 429-37.
260. Fukuda, H., M. Arai, and K. Kuwajima, *Folding of green fluorescent protein and the cycle3 mutant.* Biochemistry, 2000. **39**(39): p. 12025-32.
261. Remy, I. and S.W. Michnick, *A cDNA library functional screening strategy based on fluorescent protein complementation assays to identify novel components of signaling pathways.* Methods, 2004. **32**(4): p. 381-8.
262. Herrera, F., S. Goncalves, and T.F. Outeiro, *Imaging protein oligomerization in neurodegeneration using bimolecular fluorescence complementation.* Methods Enzymol, 2012. **506**: p. 157-74.
263. Park, K., et al., *A split enhanced green fluorescent protein-based reporter in yeast two-hybrid system.* Protein J, 2007. **26**(2): p. 107-16.
264. Kojima, T., et al., *Novel screening system for protein-protein interactions by bimolecular fluorescence complementation in Saccharomyces cerevisiae.* J Biosci Bioeng, 2011. **111**(4): p. 397-401.
265. Cundell, M.J. and C. Price, *The budding yeast amphiphysin complex is required for contractile actin ring (CAR) assembly and post-contraction GEF-independent accumulation of Rho1-GTP.* PLoS One, 2014. **9**(5): p. e97663.
266. Barnard, E., et al., *Development and implementation of split-GFP-based bimolecular fluorescence complementation (BiFC) assays in yeast.* Biochem Soc Trans, 2008. **36**(Pt 3): p. 479-82.
267. Shyu, Y.J. and C.D. Hu, *Fluorescence complementation: an emerging tool for biological research.* Trends Biotechnol, 2008. **26**(11): p. 622-30.
268. Kowalski, K., et al., *Characterization of the conserved interaction between GATA and FOG family proteins.* J Biol Chem, 2002. **277**(38): p. 35720-9.
269. Sung, M.K. and W.K. Huh, *Bimolecular fluorescence complementation analysis system for in vivo detection of protein-protein interaction in Saccharomyces cerevisiae.* Yeast, 2007. **24**(9): p. 767-75.
270. Sung, M.K. and W.K. Huh, *In vivo quantification of protein-protein interactions in Saccharomyces cerevisiae using bimolecular fluorescence complementation assay.* J Microbiol Methods, 2010. **83**(2): p. 194-201.
271. Zahradka, J., G.P. van Heusden, and H. Sychrova, *Yeast 14-3-3 proteins participate in the regulation of cell cation homeostasis via interaction with Nha1 alkali-metal-cation/proton antiporter.* Biochim Biophys Acta, 2012. **1820**(7): p. 849-58.
272. Sakalis, P.A., G.P. van Heusden, and P.J. Hooykaas, *Visualization of VirE2 protein translocation by the Agrobacterium type IV secretion system into host cells.* Microbiologyopen, 2014. **3**(1): p. 104-17.
273. Linke, C., et al., *Fkh1 and Fkh2 associate with Sir2 to control CLB2 transcription under normal and oxidative stress conditions.* Front Physiol, 2013. **4**: p. 173.
274. Ito, Y., M. Suzuki, and Y. Husimi, *A novel mutant of green fluorescent protein with enhanced sensitivity for microanalysis at 488 nm excitation.* Biochem Biophys Res Commun, 1999. **264**(2): p. 556-60.
275. Iizuka, R., M. Yamagishi-Shirasaki, and T. Funatsu, *Kinetic study of de novo chromophore maturation of fluorescent proteins.* Anal Biochem, 2011. **414**(2): p. 173-8.

276. Kodama, Y. and C.D. Hu, *An improved bimolecular fluorescence complementation assay with a high signal-to-noise ratio.* Biotechniques, 2010. **49**(5): p. 793-805.
277. Chen, J., et al., *A yEGFP-based reporter system for high-throughput yeast two-hybrid assay by flow cytometry.* Cytometry A, 2008. **73**(4): p. 312-20.
278. Greene, L.E., et al., *Application of GFP-labeling to study prions in yeast.* Protein Pept Lett, 2009. **16**(6): p. 635-41.
279. Cormack, B., *Green fluorescent protein as a reporter of transcription and protein localization in fungi.* Curr Opin Microbiol, 1998. **1**(4): p. 406-10.
280. Cormack, B.P., et al., *Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in Candida albicans.* Microbiology, 1997. **143 (Pt 2)**: p. 303-11.
281. Kass, G., et al., *Permeabilized mammalian cells as an experimental system for nuclear import of geminiviral karyophilic proteins and of synthetic peptides derived from their nuclear localization signal regions.* J Gen Virol, 2006. **87**(Pt 9): p. 2709-20.
282. Miyasaka, H., *The positive relationship between codon usage bias and translation initiation AUG context in Saccharomyces cerevisiae.* Yeast, 1999. **15**(8): p. 633-7.
283. Hu, C.D. and T.K. Kerppola, *Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis.* Nat Biotechnol, 2003. **21**(5): p. 539-45.
284. Parker, G.E., et al., *The homeodomain coordinates nuclear entry of the Lhx3 neuroendocrine transcription factor and association with the nuclear matrix.* J Biol Chem, 2000. **275**(31): p. 23891-8.
285. Ye, W., et al., *Karyopherins in nuclear transport of homeodomain proteins during development.* Biochim Biophys Acta, 2011. **1813**(9): p. 1654-62.
286. Kutay, U. and S. Guttinger, *Leucine-rich nuclear-export signals: born to be weak.* Trends Cell Biol, 2005. **15**(3): p. 121-4.
287. Kosugi, S., et al., *Nuclear export signal consensus sequences defined using a localization-based yeast selection system.* Traffic, 2008. **9**(12): p. 2053-62.
288. Xu, D., et al., *Sequence and structural analyses of nuclear export signals in the NESdb database.* Mol Biol Cell, 2012. **23**(18): p. 3677-93.
289. Berrade, L. and J.A. Camarero, *Expressed protein ligation: a resourceful tool to study protein structure and function.* Cell Mol Life Sci, 2009. **66**(24): p. 3909-22.
290. Modesti, M., *Fluorescent labeling of proteins.* Methods Mol Biol, 2011. **783**: p. 101-20.
291. Stephanopoulos, N. and M.B. Francis, *Choosing an effective protein bioconjugation strategy.* Nat Chem Biol, 2011. **7**(12): p. 876-84.
292. Toseland, C.P., *Fluorescent labeling and modification of proteins.* J Chem Biol, 2013. **6**(3): p. 85-95.
293. Zhang, X.F., J. Zhang, and L. Liu, *Fluorescence properties of twenty fluorescein derivatives: lifetime, quantum yield, absorption and emission spectra.* J Fluoresc, 2014. **24**(3): p. 819-26.
294. Sletten, E.M. and C.R. Bertozzi, *Bioorthogonal chemistry: fishing for selectivity in a sea of functionality.* Angew Chem Int Ed Engl, 2009. **48**(38): p. 6974-98.
295. Keppler, A., et al., *A general method for the covalent labeling of fusion proteins with small molecules in vivo.* Nat Biotechnol, 2003. **21**(1): p. 86-9.
296. Gautier, A., et al., *An engineered protein tag for multiprotein labeling in living cells.* Chem Biol, 2008. **15**(2): p. 128-36.
297. Stagge, F., et al., *SNAP-, CLIP- and Halo-tag labelling of budding yeast cells.* PLoS One, 2013. **8**(10): p. e78745.

298. NEB, *SNAP-Cell Starter Kit Instruction Manual*. 2008.
299. De Rosa, L., et al., *Semi-synthesis of labeled proteins for spectroscopic applications.* Molecules, 2013. **18**(1): p. 440-65.
300. Dawson, P.E., et al., *Synthesis of proteins by native chemical ligation.* Science, 1994. **266**(5186): p. 776-9.
301. Johnson, E.C. and S.B. Kent, *Insights into the mechanism and catalysis of the native chemical ligation reaction.* J Am Chem Soc, 2006. **128**(20): p. 6640-6.
302. Iwai, H. and A. Pluckthun, *Circular beta-lactamase: stability enhancement by cyclizing the backbone.* FEBS Lett, 1999. **459**(2): p. 166-72.
303. Gentle, I.E., D.P. De Souza, and M. Baca, *Direct production of proteins with N-terminal cysteine for site-specific conjugation.* Bioconjug Chem, 2004. **15**(3): p. 658-63.
304. Busch, G.K., et al., *Specific N-terminal protein labelling: use of FMDV 3C pro protease and native chemical ligation.* Chem Commun (Camb), 2008(29): p. 3369-71.
305. Tolbert, T.J. and C.H. Wong, *Conjugation of glycopeptide thioesters to expressed protein fragments: semisynthesis of glycosylated interleukin-2.* Methods Mol Biol, 2004. **283**: p. 255-66.
306. Gesteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press. p. 571-607.
307. Ryan, D., *Investigation of transcriptional complexes that regulate haematopoiesis and are implicated in leukemia*, in *School of Molecular Bioscience*. 2005, University of Sydney.
308. Dawson, P.E. and S.B. Kent, *Synthesis of native proteins by chemical ligation.* Annu Rev Biochem, 2000. **69**: p. 923-60.
309. Yi, L., et al., *One-pot dual-labeling of a protein by two chemoselective reactions.* Angew Chem Int Ed Engl, 2011. **50**(36): p. 8287-90.
310. Peroza, E.A. and E. Freisinger, *Tris is a non-innocent buffer during intein-mediated protein cleavage.* Protein Expr Purif, 2008. **57**(2): p. 217-25.
311. Kim, Y., et al., *Efficient site-specific labeling of proteins via cysteines.* Bioconjug Chem, 2008. **19**(3): p. 786-91.
312. Royer, C.A. and S.F. Scarlata, *Fluorescence approaches to quantifying biomolecular interactions.* Methods Enzymol, 2008. **450**: p. 79-106.
313. Lakowicz, J.R., ed. *Principles of Fluorescence Spectroscopy, Second Edition*. 1999, Kluwer Academic Plenum Publishers.
314. Rasnik, I., S.A. McKinney, and T. Ha, *Surfaces and orientations: much to FRET about?* Acc Chem Res, 2005. **38**(7): p. 542-8.
315. Xiao, Y. and S.N. Isaacs, *Enzyme-linked immunosorbent assay (ELISA) and blocking with bovine serum albumin (BSA)--not all BSAs are alike.* J Immunol Methods, 2012. **384**(1-2): p. 148-51.
316. Gibbs, J., *Effective Blocking Procedures.* ELISA Technical Bulletin - No 3, 2001.
317. Rudolph, R. and H. Lilie, *In vitro folding of inclusion body proteins.* FASEB J, 1996. **10**(1): p. 49-56.
318. Schuler, B. and L.K. Pannell, *Specific labeling of polypeptides at amino-terminal cysteine residues using Cy5-benzyl thioester.* Bioconjug Chem, 2002. **13**(5): p. 1039-43.
319. Wilkinson-White, L. and J.M. Matthews, *The PA207 peptide inhibitor of LIM-only protein 2 (Lmo2) targets Zinc Finger domains in a non-specific manner.* Protein Pept Lett, 2014. **21**(2): p. 132-9.
320. Fehr, F., et al., *Semi-synthesis and analysis of chemically modified zif268 zinc-finger domains.* ChemistryOpen, 2012. **1**(1): p. 26-32.

321. Dawson, P.E., Churchill,M. J., Ghadiri, M. R. & Kent, S. B. H. , *Modulation of Reactivity in Native Chemical Ligation through the Use of Thiol Additives.* Journal of the American Chemical Society, 1997. **119**(19).
322. Yi, L., *Development of site-specific labeling strategies for protein modification and synthesis of Rab probes*, in *Department of Chemistry*. 2012, University of Dortmund: Dortmund.
323. Swillens, S., *Interpretation of binding curves obtained with high receptor concentrations: practical aid for computer analysis.* Mol Pharmacol, 1995. **47**(6): p. 1197-203.
324. Wang, Z., et al., *Optimizing expression and purification of an ATP-binding gene gsiA from Escherichia coli k-12 by using GFP fusion.* Genet Mol Biol, 2011. **34**(4): p. 661-8.
325. Park, S.H. and R.T. Raines, *Green fluorescent protein as a signal for protein-protein interactions.* Protein Sci, 1997. **6**(11): p. 2344-9.
326. Wagstaff, K.M., et al., *Quantitative analysis of protein-protein interactions by native page/fluorimaging.* J Fluoresc, 2005. **15**(4): p. 469-73.
327. Wittig, I. and H. Schagger, *Advantages and limitations of clear-native PAGE.* Proteomics, 2005. **5**(17): p. 4338-46.
328. Schagger, H. and G. von Jagow, *Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form.* Anal Biochem, 1991. **199**(2): p. 223-31.
329. Wittig, I. and H. Schagger, *Features and applications of blue-native and clear-native electrophoresis.* Proteomics, 2008. **8**(19): p. 3974-90.
330. Farnum, M. and C. Zukoski, *Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor.* Biophys J, 1999. **76**(5): p. 2716-26.
331. Mattheyses, A.L., et al., *Fluorescence anisotropy reveals order and disorder of protein domains in the nuclear pore complex.* Biophys J, 2010. **99**(6): p. 1706-17.
332. Lakowicz, J.R., *Principles of fluorescence spectroscopy*. 1999, New York: Kluwer Academic Plenum Publishers.
333. Forwood, J.K., V. Harley, and D.A. Jans, *The C-terminal nuclear localization signal of the sex-determining region Y (SRY) high mobility group domain mediates nuclear import through importin beta 1.* J Biol Chem, 2001. **276**(49): p. 46575-82.
334. Forwood, J.K., M.H. Lam, and D.A. Jans, *Nuclear import of Creb and AP-1 transcription factors requires importin-beta 1 and Ran but is independent of importin-alpha.* Biochemistry, 2001. **40**(17): p. 5208-17.
335. Kiessig, S., et al., *Application of a green fluorescent fusion protein to study protein-protein interactions by electrophoretic methods.* Electrophoresis, 2001. **22**(7): p. 1428-35.
336. Alvisi, G., et al., *A protein kinase CK2 site flanking the nuclear targeting signal enhances nuclear transport of human cytomegalovirus ppUL44.* Traffic, 2005. **6**(11): p. 1002-13.
337. Oksayan, S., et al., *A novel nuclear trafficking module regulates the nucleocytoplasmic localization of the rabies virus interferon antagonist, P protein.* J Biol Chem, 2012. **287**(33): p. 28112-21.
338. Forwood, J.K. and D.A. Jans, *Quantitative analysis of DNA-protein interactions using double-labeled native gel electrophoresis and fluorescence-based imaging.* Electrophoresis, 2006. **27**(16): p. 3166-70.
339. Fulcher, A.J., M.M. Dias, and D.A. Jans, *Binding of p110 retinoblastoma protein inhibits nuclear import of simian virus SV40 large tumor antigen.* J Biol Chem, 2010. **285**(23): p. 17744-53.
340. Caly, L., et al., *Impaired nuclear import and viral incorporation of Vpr derived from a HIV long-term non-progressor.* Retrovirology, 2008. **5**: p. 67.

341. Kaur, G. and D.A. Jans, *Dual nuclear import mechanisms of sex determining factor SRY: intracellular Ca2+ as a switch.* FASEB J, 2011. **25**(2): p. 665-75.
342. da Cunha, E.S., C.C. Domingues, and E. de Paula, *Modified native electrophoresis protocol for the solubilization and separation of mitochondrial protein complexes.* Anal Biochem, 2011. **418**(1): p. 158-60.
343. Stafford, R.L., et al., *The molecular basis of the Caskin1 and Mint1 interaction with CASK.* J Mol Biol, 2011. **412**(1): p. 3-13.
344. McNaughton, B.R., et al., *Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins.* Proc Natl Acad Sci U S A, 2009. **106**(15): p. 6111-6.
345. Hulme, E.C. and M.A. Trevethick, *Ligand binding assays at equilibrium: validation and interpretation.* Br J Pharmacol, 2010. **161**(6): p. 1219-37.
346. Makarev, E. and M. Gorivodsky, *Islet1 and its co-factor Ldb1 are expressed in quiescent cells of mouse intestinal epithelium.* PLoS One, 2014. **9**(4): p. e95256.
347. 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.
348. Dogan, J., S. Gianni, and P. Jemth, *The binding mechanisms of intrinsically disordered proteins.* Phys Chem Chem Phys, 2014. **16**(14): p. 6323-31.
349. Mizuguchi, M., et al., *Transient alpha-helices in the disordered RPEL motifs of the serum response factor coactivator MKL1.* Sci Rep, 2014. **4**: p. 5224.
350. Kenakin, *Pharmacologic analysis of drug-receptor interaction*. 1993, Raven Press: New York.

Appendix A LIM-HD proteins: names, isoforms and homozygous knockout mice phenotypes

As the discovery of many genes predates the genome sequencing projects and the modern gene classification, different names were used for the same proteins in literature, which lead to confusion in many cases. In particular, the paralogue of Lhx6 was originally identified and named L3 (Matsumoto 1996), but was later renamed to Lhx8 (Kitanaka 1998). It was then 'rediscovered' as Lhx7 (Grigoriou 1998) and before the controversy was resolved, a newly identified paralogue of Lhx2 was named Lhx9 (Retaux 1999, Failli 2000). Identification of several alternatively spliced mRNA/protein isoforms of Lhx6 and Lhx8 additionally complicated this issue. Since no additional LIM-HD genes were identified, Lhx8 is sometimes referred to as Lhx7/8 (Hobert 2000).

Lhx and Lmx both stand for LIM-homeobox, but because of historical reasons, Lmx1 as well as Isl proteins retained their original names. LIM-HD proteins in Drosophila and Nematoda have alternative, historical names (Bach 2000).

Most isoforms of LIM-HD proteins (Table A1) are results of alternatively spliced N-terminal and/or C-terminal exons. Isl1β isoform is an exception because it is apparently a result of an alternative splice acceptor site located within the nearest downstream exon. Consequently, the beginning of exon 5 which contains the first half of $Isl1_{LBD}$ is not spliced in this isoform. Lhx2b is a validated RefSeq sequence but no data on detected protein is provided in either database. NM_027397.2 version of Isl2 was replaced on Jul 28, 2007 by the correct version NP_081673.2.

Table A.1. Protein-coding, CDS-complete protein isoforms of mouse LIM-HD proteins. Protein names from NCBI RefSeq, ENSEMBL and Uniprot databases are shown along with NCBI and EBI protein IDs and the numbers of amino acids in parallel. Double EBI IDs for some proteins indicate the same isoform translated from different transcripts. Red stars mark isoforms that are not in the NCBI Consensus CDS (CCDS) protein set. P in superscript indicates predicted isoforms (NCBI). Lhx3 isoforms from the Uniprot database have opposite names to the RefSeq entries (likely database error) and are marked in red. Lmx1b entry in Uniprot has a 23 a.a. extension of the N-terminal exon due to erroneous translation initiation and is marked with E. Some isoforms have additional alternative names in the Uniprot database and they are shown in brackets.

RefSeq protein	NCBI Protein ID	ENSEMBL protein	EBI Protein ID	UniProt protein	Number of residues
Lhx1	NP_032524.1	Lhx1-001	ENSMUSP00000018842	Lhx1	406
Lhx2a	NP_034840.1	Lhx2-001	ENSMUSP00000000253	Lhx2	406
*Lhx2b	NP_001277575.1	Lhx2-003	ENSMUSP00000114797	-	365
Lhx3a	NP_001034742.1	Lhx3-001	ENSMUSP00000028302	Lhx3b	402
*Lhx3b	NP_034841.2	Lhx3-002	ENSMUSP00000056822	Lhx3a	400
Lhx4	NP_034842.2	Lhx4-001	ENSMUSP00000027740	Lhx4	390
Lhx5	NP_032525.1	Lhx5-001	ENSMUSP00000031591	Lhx5	402
Lhx6.1	NP_032526.2	Lhx6-003	ENSMUSP00000108584	-	392
Lhx6.3	NP_001076595.1	Lhx6-001 or -201	ENSMUSP00000108585 ENSMUSP00000108587	Lhx6.1a	363
Lhx6.2	NP_001076594.1	Lhx6-007	ENSMUSP00000108591	-	377
Lhx6.4	NP_001076596.1	Lhx6-002	ENSMUSP00000108590	Lhx6.1b	348
Lhx8	NP_034843.2	Lhx8-001	ENSMUSP00000136047	Lhx8	367
*Lhx8 $X1^P$ or $X2^P$	XP_006501134.1 XP_006501135.1	Lhx8-003	ENSMUSP00000134853	-	346
Lhx9a	NP_001020736.1	Lhx9-002	ENSMUSP00000019374	Lhx9.4	330
Lhx9b	NP_034844.1	Lhx9-004 or -201	ENSMUSP00000107661 ENSMUSP00000091198	Lhx9.2 (Lhx9α)	321
Lhx9c	NP_001036042.1	Lhx9-005	ENSMUSP00000107657	Lhx9.3	397
*Lhx9 $X2^P$	XP_006529234.1	Lhx9-001	ENSMUSP00000036480	Lhx9.1 (Lhx9β)	388
Lmx1a	NP_387501.1	Lmx1a-001 or -002	ENSMUSP00000028003 ENSMUSP00000107008	Lmx1a	382
Lmx1b	NP_034855.2	Lmx1b-001	ENSMUSP00000043616	^{E}Lmx1b	372
ISL-1	NP_067434.3	Isl1-001	ENSMUSP00000044879	Isl1α	349
*ISL-1 $X1^P$	XP_006517596.1	Isl1-002	ENSMUSP00000135567	Isl1β	326
ISL-2	NP_081673.2	Isl2-001	ENSMUSP00000034869	Isl2	359

Table A.2. Phenotypes of mice with homozygous knockout mutations of LIM-HD proteins. Each column shows closely related paralogue genes.

lhx3	**Perinatal lethal** (within 24 hours of birth); failure of growth and differentiation of Rathke's pouch, lack of both anterior and intermediate lobes of the pituitary gland.	***lhx4***	**Neonatal lethal**; abnormal lung development, pituitary and decreased hormone levels.
lhx1	**Embryonic lethal**; embryos are small, fail to develop head structures anterior to rhombomere 3 in the hindbrain, lack kidneys and gonads, and show aberrant trajectories of limb motor axons; most die around embryonic day 10.	***lhx5***	**Neonatal lethal**; defective hippocampal development. Postmitotic hippocampal cells are unable to differentiate properly and migrate to correct positions, resulting in structural anomalies of the Ammon's horn and the dentate gyrus.
lmx1a	**Prenatal** to **preweaning lethality**; fewer dopaminergic neurons, neurological and skeletal abnormalities, inner ear defects, deafness	***lmx1b***	**Neonatal lethal**; exhibit various skeletal, kidney, and eye defects. Pups also fail to suckle.
lhx2	**Embryonic** to **perinatal lethal**; abnormal liver, telencephalon, olfactory bulb, basal ganglion, and eye morphology	***lhx9***	**Not lethal**; failed proliferation of the somatic cells of the genital ridge resulting in lack of discrete gonad formation, infertility in both sexes, and female-like genitalia in genetically male mice
lhx6	**Postnatal** or **preweaning lethal**; results in impaired migration as well as specification of cortical interneurons, general weakness	***lhx8***	**Neonatal lethal** with **incomplete penetrance**; cleft secondary palate. Those without cleft palate survive to adulthood. Both have decreased or absent forebrain cholinergic neurons.
isl1	**Embryonic lethal**; mice fail to develop motor neurons and die by embryonic day 11.5 with abnormal heart and pancreas development.	***isl2***	**Neonatal lethal**; motor neuron migration defects and impaired visceral motor neuron differentiation.

Appendix B Design of bacterial overexpression plasmids

pIH1119 (source plasmid)

```
tcgagctcgaacaacaacaacaataacaataacaacaacctcggg
 S  S  S  N  N  N  N  N  N  N  N  N  N  L  G

atcgagggaaggatttcagaatttggatccgaattagaattc
 I  E  G  R  I  S  E  F  G  S  E  L  E  F
```

Factor Xa cleavage site

Derivatives:

- **pIH1121** codes for a **HRV3C-cleavable** fusion linker.

```
ctcgggccatgggggctcgaagtactctttcaaggacccggatcc
 L  G  P  W  G  L  E  V  L  F  Q  G  P  G  S
```

HRV3C Pro cleavage site

- **pIH1122nc** codes for a **non-cleavable** short GS linker.

```
actaattcgagctcgggtggcggcggatccgaattagaattc
 T  N  S  S  S  G  G  G  G  S  E  L  E  F
```

pGEX-2TKE (source plasmid)

```
ggtggtggcgaccatcctccaaaatcggatctggttccgcgtggatccccgggaattcat
 G  G  G  D  H  P  P  K  S  D  L  V  P  R  G  S  P  G  I  H
```

Thrombin cleavage site

Derivative:

- **pGEX-TEVC**

```
catcctccaaaatcggatgaaaacctgtattttcagtgcggatccccgggaattcat
 H  P  P  K  S  D  E  N  L  Y  F  Q  C  G  S  P  G  I  H
```

TEV cleavage site

pSNAP-tag®(T7) (source plasmid)

```
gagatatacatatggacaaagattgcgaa...
            M  D  K  D  C  E
```

SNAP-tag

Derivative:

- **pT7-HSNAP** codes for a TEV-cleavable His-tag, N-terminal to SNAP-tag

```
...catatgcatcaccatcaccatcacgaaaacctgtattttcagggcatggacaaagattgcgaa
   H  M  H  H  H  H  H  H  E  N  L  Y  F  Q  G  M  D  K  D  C  E
```

His-tag **TEV cleavage site** **SNAP-tag**

pET15b (source vector)

```
AACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCA
GCCATATGCTCGAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA
CTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCC
CGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGAT
GAGCGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCT
TATCGATGATAAGCTGTCAAACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAA
TGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGT...
```

Derivatives:

- **pET1bBE** has a BamHI-EcoRI cloning site and lacks the downstream EcoRI site

```
AACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGC
GCGGCAGCCATATGGGATCCGCGAATTCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCC
ACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGG
AGGAACTATATCCGGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATC
CAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGC
AATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACAGACGTCAGGTGGCAC
...
```

- **pET1bBE-GFPsol** has a BglII-GFPsol-BamHI-2×STOP-EcoRI sequence inserted between BamHI and EcoRI sites.

```
...CATATGGGATCTGCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGA
CGGCGACGTTAACGGTCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGA
CCCTGAAGTTCATCTGCACTACTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTCTGACGTAC
GGTGTTCAGTGCTTCGCACGCTATCCGGATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGA
AGGTTATGTTCAGGAACGCACTATCAGCTTCAAAGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAAT
TCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAGGTATTGACTTCAAAGAAGACGGTAACATCCTG
```

GGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTA**CATTACTGCGGATAAACAAAAAAACGGCAT**
TAAAGCGAACTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATCACTACCAGCAGA
ATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAACTGTCC
AAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACG**G**
TATGGATGAGCTGTACAAG*GGATCC*TAATGAGAATTC**GGCTGCTAA**...

pGEX-6P (source vector)

caaaatcggatctggaagttctgttccaggggcccctg**ggatcc**ccg**gaattc**
K S D L E V L F Q G P L G S P E F

HRV3C Pro cleavage site

Derivative:

- **pGEX-6P-GFPsol** has a BglII-GFPsol-BamHI-2×STOP-EcoRI sequence inserted between BamHI and EcoRI sites.

...GACCATCCTCCAAAATCGGATCTGGAAGTTCTGTTCCAGGGGCCCCTG**GGATCT**GCATCTAAGGGCGAA
GAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGGTCACAAGTTCTCCGTT
TCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTACTGGCAAACTG
CCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCGGATCAC
ATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAAA
GATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTG
AAAGGTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAAT
GTGTA**CATTACTGCGGATAAACAAAAAAACGGCATTAAAGCGAACTTC**AAAATCCGTCATAACATTGAGGAT
GGCGGTGTACAGCTGGCGGATCACTACCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGAT
AACCACTACCTGTCCACCCAGTCTAAACTGTCCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTG
GAATTTGTTACTGCGGCGGGTATTACTCACG**GTATGGATGAGCTGTACAAG**GGATCCTAATGAGAATTC...

Appendix C Tethered LIMs-LID constructs with Factor Xa-cleavable linkers

C.1 Protein constructs expressed from pGEX-TEVC and pGEX-6P-GFPsol plasmids

$LMO4_{LIMs}$-$Ldb1_{LID}$

LSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQAQLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSI
PASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHLN*SGGS*IEGR*GSGG*DVMVVGEP
TLMGGEFGDEDERLITRLENTQFDAANGIDDE-

$Lhx3_{LIMs}$-$Ldb1_{LID}$

TPEIPMCAGCDQHILDRFILKALDRHWHSKCLKCSDCHVPLAERCFSRGESVYCKDDFFKRFGTKCAACQLGIPPTQ
VVRRAQDFVYHLHCFACVVCKRQLATGDEFYLMEDSRLVCKADYETAKQ*GGS*IEGR*GSGG*DVMVVGEPTLMGGEFGD
EDERLITRLENTQFDAANGIDDE-

$Isl1_{LIMs}$-$Ldb1_{LID}$

KRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFVRDGKTYCKRDYIRLYGIKCAKCSIGFS
KNDFVMRARSKVYHIECFRCVACSRQLIPGDEFALREDGLFCRADHDVVER*GGS*IEGR*GSGG*DVMVVGEPTLMGGEF
GDEDERLITRLENTQFDAANGIDDE-

C.2 Protein constructs expressed from pGEX-TEVC plasmids only

$Lhx1_{LIMs}$-$Ldb1_{LID}$

MVHCAGCKRPILDRFLLNVLDRAWHVKCVQCCECKCNLTEKCFSREGKLYCKNDFFRCFGTKCAGCAQGISPSDLVR
RARSKVFHLNCFTCMMCNKQLSTGEELYIIDENKFVCKEDYLSN*SSGGS*IEGR*GSGG*DVMVVGEPTLMGGEFGDEDE
RLITRLENTQFDAANGIDDE-

$Lhx2_{LIMs}$-$Ldb1_{LID}$

SDRAALCAGCGGKISDRYYLLAVDKQWHMRCLKCCECKLNLESELTCFSKDGSIYCKEDYYRRFSVQRCARCHLGIS
ASEMVMRARDLVYHLNCFTCTTCNKMLTTGDHFGMKDSLVYCRLHFEALLQ*GGS*IEGR*GSGG*DVMVVGEPTLMGGEF
GDEDERLITRLENTQFDAANGIDDE-

$Lhx4_{LIMs}$-$Ldb1_{LID}$

MQQIPQCAGCNQHILDKFILKVLDRHWHSSCLKCADCQMQLADRCFSRAGSVYCKEDFFKRFGTKCTACQQGIPPTQ
VVRKAQDFVYHLHCFACIICNRQLATGDEFYLMEDGRLVCKEDYETAKQ*GGS*IEGR*GSGG*DVMVVGEPTLMGGEFGD
EDERLITRLENTQFDAANGIDDE-

DNA sequences coding for similar constructs containing tandem LIM-domains of Lhx5, Lhx6, Lhx8, Lmx1a, Lmx1b and Isl2 were also cloned into pGEX-TVC vectors. $LMO4_{LIMs}$-$Ldb1_{LID}$ DNA construct was designed by Dr J. Matthews and other DNA constructs were made by Dr M. O'Connell.

All GST-fusion constructs expressed from pGEX-TEVC plasmids contained a modified TEV site in the GST-insert linker. TEV cleavage of these fusion constructs produced tethered LIMs-LID fusion constructs with N-terminal CGS sequence. All GST-fusion constructs expressed from pGEX-6P-GFPsol plasmids were cleaved with HRV3C protease on the GSH-resin and the released proteins had a GPLGS-GFPsol-GS-LIMs-LID.

Table C.1. Molecular weights and extinction coefficients of purified constructs.

Construct	Molecular weight (Da)	ε_{280} ($M^{-1}cm^{-1}$)
His_6-SNAP-$LMO4_{LIMs}$-$Ldb1_{LID}$	42148.9	42400
Cys-$LMO4_{LIMs}$-$Ldb1_{LID}$	20474	19940
GFPsol-$LMO4_{LIMs}$-$Ldb1_{LID}$	47463.3	41830
Cys-$Lhx3_{LIMs}$-$Ldb1_{LID}$	20057.7	11460
GFPsol-$Lhx3_{LIMs}$-$Ldb1_{LID}$	47047.1	33350
Cys-$Isl1_{LIMs}$-$Ldb1_{LID}$	20347	14440
GFPsol-$Isl1_{LIMs}$-$Ldb1_{LID}$	47336.4	36330
Cys-$Lhx1_{LIMs}$-$Ldb1_{LID}$	19691.4	9970
Cys-$Lhx2_{LIMs}$-$Ldb1_{LID}$	20312.1	15930
Cys-$Lhx4_{LIMs}$-$Ldb1_{LID}$	20111.7	11460
His_6-TEV	~27000	see [183, 184]
MBP-$Ldb1_{LID(331–375)}$	48275.2	71850
Cys-$DEAF1_{404-438(T435D)}$	4436.2	~0

Appendix D BiFC-related DNA and protein constructs

D.1 Split-GFP DNA constructs and BiFC plasmids

Constructs that did not contain F64L and S65T mutations that provide enhanced UV properties were named as NGFP, CGFP and NGFP-CGFP (no uv suffix). Only NGFPsol and NGFPsol-CGFP construct names contained the sol suffix.

Restriction sites:

GGATCC - BamHI
GAATTC - EcoRI
CTGCAG - PstI
AAGCTT - HindIII
AGATCT - BglII

CATATG - NdeI
CCCGGG - SmaI
CTCGAG - XhoI
GCATGC - SphI
TGA/TAA/TAG - STOP codons

D.1.1 Removing Gal4 fragments from Y2H vectors

NpGBT9 (Gal4$_{DBD}$, yADH1 promoter)

...TGTTTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATCAACTCCAAGCTTGAAGCAAGCCTCCTG
AAAGATGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGCTCCAAAGAAA
AACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCTACTCTCCCAAAACCAAAAGGTCTCCGCTGACT
AGGGCACATCTGACAGAAGTGGAATCAAGGCTAGAAAGACTGGAACAGCTATTTCTACTGATTTTTCCTCGAGAAGA
CCTTGACATGATTTTGAAAATGGATTCTTTACAGGATATAAAAGCATTGTTAACAGGATTATTTGTACAAGATAATG
TGAATAAAGATGCCGTCACAGATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGAATA
AGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACTGTATCGCCGGAATTTGGATCCCC
GGGAATTCTGCAGCCAAGCTAATTCCGGGCGAATTTCTTATGATTTATGATTTTTATTATTAA...

pW410

...TTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATCAACTCCAAGCTTGCATATGGGATCCCGGGA
ATTCTGCAGCCAAGCTAATTCCGGGCGAATTTCTTATGATTTATGATTTTTATTATTAA ...

pGAD10 (Gal4$_{DBD}$, yADH1 promoter)

...TGTTTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATCAACTCCAAGCTTTGCAAAGATGGATAA
AGCGGAATTAATTCCCGAGCCTCCAAAAAAGAAGAGAAAGGTCGAATTGGGTACCGCCGCCAATTTTAATCAAAGTG
GGAATATTGCTGATAGCTCATTGTCCTTCACTTTCACTAACAGTAGCAACGGTCCGAACCTCATAACAACTCAAACA
AATTCTCAAGCGCTTTCACAACCAATTGCCTCCTCTAACGTTCATGATAACTTCATGAATAATGAAATCACGGCTAG
TAAAATTGATGATGGTAATAATTCAAAACCACTGTCACCTGGTTGGACGGACCAAACTGCGTATAACGCGTTTGGAA
TCACTACAGGGATGTTTAATACCACTACAATGGATGATGTATATAACTATCTATTCGATGATGAAGATACCCCACCA
AACCCAAAAAAAGAGATCTCTCGAGGATCCGAATTCCAGATCTATGAATCGTAGATACTGAAAAACCCCGCAAGTTC
ACTTCAACTGTGCATCGTGCACCATCTCAATTTCTTTCATTTATACATCGTTTTGCCTTCTTTTATGTAACTATACT
CCTCTAAGTTTCAATCTTGGCCATGTAACCTCTGATCTATAGAATTTTTTAAATGACTAGAATTAATGCCCATCTTT
TTTTTGGACCTAAATTCTTCATGAAAATATATTACGAGGGCTTATTCAGAAGCTTTGGACTTCTTCGCCAGAGGTTT
GGTCAAGTCTCCAA...

pL410

...TTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATCAACTCCAAGCTTGCATATGGGATCCGAATT
CCAGATCTATGAATCGTAGATACTGAAAAACCCCGCAAGTTCACTTCAACTGTGCATCGTGCACCATCTCAATTTCT
TTCATTTATACATCGTTTTGCCTTCTTTTATGTAACTATACTCCTCTAAGTTTCAATCTTGGCCATGTAACCTCTGA
TCTATAGAATTTTTTAAATGACTAGAATTAATGCCCATCTTTTTTTTTGGACCTAAATTCTTCATGAAAATATATTAC
GAGGGCTTATTCAGAAGCTTTGGACTTCTTCGCCAGAGGTTTGG...

D.1.2 Split-GFP DNA fragments

NGFP-CGFP (This fragment was inserted in the pW410 plasmid)

CATATGGCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGG
TCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTA
CTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCG
GATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAA
AGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAG
GTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTACATT
ACTGCGGATAAACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCCTCGAGGAATTCGGTGGCGGTGGTTCCGG
TGGTGGCGGCAGCAACGGCATTAAAGCGAACTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGG
ATCACTACCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCT
AAACTGTCCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCA
CGGTATGGATGAGCTGTACAAGTAATGACTGCAG

NGFP-CGFP (This fragment was inserted in the pL410 plasmid)

CAT**ATG**GCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGG
TCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTA
CTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCG
GATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAA
AGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAG
GTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTA**CATT**
ACTGCGGATAAACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCCTCGAGGAATTCGGTGGCGGTGGTTCCGG
TGGTGGCGGCAGC**AACGGCATTAAAGCGAACTT**CAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGG
ATCACTACCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCT

AAACTGTCCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCA
CG**GTATGGATGAGCTGTACAAG**TAATGAAGATCT

NGFP$_N$

CAT**ATG**GCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGG
TCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTA
CTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCG
GATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAA
AGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAG
GTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTA**CATT**
ACTGCGGATAAACAAAAAGGTGGCGGCGGTTCTGGCGGTGGT**GGATCC**

CGFP$_N$

CATATG**AACGGCATTAAAGCGAACTT**CAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATCACTA
CCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAACTGT
CCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACG**GTATG**
GATGAGCTGTACAAGGGCGGTGGCGGTTCCGGTGGCGGTGGCTCT**GGATCC**

NGFP$_C$ (This fragment was inserted in the pW410 plasmid)

GAATTCGGCGGTGGCGGTTCCGGTGGCGGTGGCTCT**ATGGCATCTAAGGGCGAAG**AACTGTTTACGGGTGTTGTTCC
AATCCTGGTAGAGCTGGACGGCGACGTTAACGGTCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCT
ACGGCAAACTGACCCTGAAGTTCATCTGCACTACTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTC
AGCTACGGTGTTCAGTGCTTCGCACGCTATCCGGATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGA
AGGTTATGTTCAGGAACGCACTATCAGCTTCAAAGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAG
GTGATACCCTGGTTAACCGTATCGAGCTGAAAGGTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTG
GAGTACAACTATAATTCTCACAATGTGTA**CATTACTGCGGATAAACAAAAA**TGATAACTGCAG

CGFP$_C$ (This fragment was inserted in the pL410 plasmid)

GAATTCGGTGGCGGTGGTTCCGGTGGTGGCGGCAGC**AACGGCATTAAAGCGAACTT**CAAAATCCGTCATAACATTGA
GGATGGCGGTGTACAGCTGGCGGATCACTACCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATA
ACCACTACCTGTCCACCCAGTCTAAACTGTCCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTT
GTTACTGCGGCGGGTATTACTCACG**GTATGGATGAGCTGTACAAG**TAATGAAGATCT

D.1.3 BiFC plasmids

pW410-NGFP-CGFP

...CATATGGCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAA
CGGTCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCA
CTACTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTAT
CCGGATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTT
CAAAGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGA
AAGGTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTAC
ATTACTGCGGATAAACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCCTCGAGGAATTCGGTGGCGGTGGTTC
CGGTGGTGGCGGCAGCAACGGCATTAAAGCGAACTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGG
CGGATCACTACCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAG

TCTAAACTGTCCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTAC
TCACGGTATGGATGAGCTGTACAAG**TAATGA**CTGCAG...

pW410-NGFP$_N$*

* Initial construct w/o STOPs before EcoRI site.

...CATATGGCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAA
CGGTCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCA
CTACTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTAT
CCGGATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTT
CAAAGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGA
AAGGTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTAC
ATTACTGCGGATAAACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCCGGGAATTCTGCAGCCAAGCTAATTC
CGGGCGAATTTCTTATGATTTATGA...

pW410-NGFP$_N$

New construct with STOP codons between BamHI and EcoRI sites.

...CATATGGCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAA
CGGTCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCA
CTACTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTAT
CCGGATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTT
CAAAGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGA
AAGGTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTAC
ATTACTGCGGATAAACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCTGATAATAGGAATTCTGCAG...

pW410-NGFP$_C$

...CATATGGGATCCCGGGAATTCGGCGGTGGCGGTTCCGGTGGCGGTGGCTCTATGGCATCTAAGGGCGAAGAACT
GTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGGTCACAAGTTCTCCGTTTCCGGCGAAG
GTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTACTGGCAAACTGCCGGTACCTTGGCCG
ACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCGGATCACATGAAACAACATGATTTCTT
CAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAAAGATGACGGTAACTACAAAACCCGTG
CGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAGGTATTGACTTCAAAGAAGACGGTAAC
ATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTACATTACTGCGGATAAACAAAAATGATAACT
GCAG...

pW410-NLS-NGFP$_N$

...CATATGGCGGAATTAATTCCCGAGCCTCCAAAAAAGAAGAGAAAGGTCGAATTGGGTACCGCCCATATGGCATC
TAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGGTCACAAGTTCT
CCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTACTGGCAAACTG
CCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCGGATCACATGAA
ACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAAAGATGACGGTA
ACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAGGTATTGACTTC
AAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTACATTACTGCGGATAA
ACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCTGATAATAGGAATTCTGCAG...

pW410-NLS-NGFP$_C$

...CATATGGCGGAATTAATTCCCGAGCCTCCAAAAAAGAAGAGAAAGGTCGAATTGGGTACCGCCCATATGGATC
CCGGGAATTCGGCGGTGGCGGTTCCGGTGGCGGTGGCTCTATGGCATCTAAGGGCGAAGAACTGTTTACGGGTGTTG
TTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGGTCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCA
ACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTACTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCAC
TTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCGGATCACATGAAACAACATGATTTCTTCAAATCCGCTATGC
CGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAAAGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTC
GAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAGGTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAA
ACTGGAGTACAACTATAATTCTCACAATGTGTACATTACTGCGGATAAACAAAAATGATAACTGCAG...

pWAFL-NGFP$_N$ plasmids are **pW410-NGFP$_N$** plasmids with the full-length yADH1 promoter (Appendix E).

pL410-NGFP-CGFP

...CATATGGCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAA
CGGTCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCA
CTACTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTAT
CCGGATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTT
CAAAGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGA
AAGGTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTAC
ATTACTGCGGATAAACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCGAATTCGGTGGCGGTGGTTCCGGTGG
TGGCGGCAGCAACGGCATTAAAGCGAACTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATC
ACTACCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAA
CTGTCCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACGG
TATGGATGAGCTGTACAAGTAATGAAGATCT...

pL410-CGFP$_N$

...CATATGAACGGCATTAAAGCGAACTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATCA
CTACCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAAC
TGTCCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACGGT
ATGGATGAGCTGTACAAGGGCGGTGGCGGTTCCGGTGGCGGTGGCTCTGGATCCGAATTCCAGATCTATGAATCGTA
G...

pL410-CGFP$_C$

...CATATGGGATCCGAATTCGGTGGCGGTGGTTCCGGTGGTGGCGGCAGCAACGGCATTAAAGCGAACTTCAAAAT
CCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATCACTACCAGCAGAATACGCCGATTGGCGATGGCCCTG
TGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAACTGTCCAAAGACCCTAACGAAAAACGCGACCACATG
GTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACGGTATGGATGAGCTGTACAAGTAATGAAGATCT...

pL410-NLS-CGFP$_N$

...CATATGGCGGAATTAATTCCCGAGCCTCCAAAAAAGAAGAGAAAGGTCGAATTGGGTACCGCCCATATGAACGG
CATTAAAGCGAACTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATCACTACCAGCAGAATA
CGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAACTGTCCAAAGACCCT

AACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACGGTATGGATGAGCTGTA
CAAGGGCGGTGGCGGTTCCGGTGGCGGTGGCTCTGGATCCGAATTCCAGATCTATGAATCGTAG...

D.1.4 Inserting constructs into BiFC plasmids

Inserts **with STOP codons**

e.g., GGATCCconstructSTOPsGAATTC

were inserted in pW410-NGFP$_N$ and pL410-CGFP$_N$ plasmids.

Inserts **without STOP codons**

e.g., GGATCCconstructGAATTC

were inserted in pW410-NGFP$_C$ and pL410-CGFP$_C$ plasmids.

pW410-NGFP$_N$-construct

...CATATGGCATCTAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAA
CGGTCACAAGTTCTCCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCA
CTACTGGCAAACTGCCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTAT
CCGGATCACATGAAACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTT
CAAAGATGACGGTAACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGA
AAGGTATTGACTTCAAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTAC
ATTACTGCGGATAAACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCconstructSTOPsGAATTCTGCAG.
..

pW410-NGFP$_C$-construct

...CATATGGGATCCconstructGAATTCGGCGGTGGCGGTTCCGGTGGCGGTGGCTCTATGGCATCTAAGGGCGA
AGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGGTCACAAGTTCTCCGTTTCCG
GCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTACTGGCAAACTGCCGGTACCT
TGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCGGATCACATGAAACAACATGA
TTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAAAGATGACGGTAACTACAAAA
CCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAGGTATTGACTTCAAAGAAGAC
GGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTACATTACTGCGGATAAACAAAAATG
ATAACTGCAG...

pW410-NLS-NGFP$_N$-construct

...CATATGGCGGAATTAATTCCCGAGCCTCCAAAAAAGAAGAGAAAGGTCGAATTGGGTACCGCCCATATGGCATC
TAAGGGCGAAGAACTGTTTACGGGTGTTGTTCCAATCCTGGTAGAGCTGGACGGCGACGTTAACGGTCACAAGTTCT
CCGTTTCCGGCGAAGGTGAAGGCGACGCAACCTACGGCAAACTGACCCTGAAGTTCATCTGCACTACTGGCAAACTG
CCGGTACCTTGGCCGACTCTGGTAACCACTTTCAGCTACGGTGTTCAGTGCTTCGCACGCTATCCGGATCACATGAA
ACAACATGATTTCTTCAAATCCGCTATGCCGGAAGGTTATGTTCAGGAACGCACTATCAGCTTCAAAGATGACGGTA
ACTACAAAACCCGTGCGGAAGTGAAATTCGAAGGTGATACCCTGGTTAACCGTATCGAGCTGAAAGGTATTGACTTC

AAAGAAGACGGTAACATCCTGGGCCACAAACTGGAGTACAACTATAATTCTCACAATGTGTACATTACTGCGGATAA
ACAAAAAGGTGGCGGCGGTTCTGGCGGTGGTGGATCCconstructSTOPsGAATTCTGCAG...

pL410-CGFP$_N$-construct

...CATATGAACGGCATTAAAGCGAACTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATCA
CTACCAGCAGAATACGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAAC
TGTCCAAAGACCCTAACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACGGT
ATGGATGAGCTGTACAAGGGCGGTGGCGGTTCCGGTGGCGGTGGCTCTGGATCCconstructSTOPsGAATTCCAG
ATCT...

pL410-CGFP$_C$-construct

...CATATGGGATCCconstructGAATTCGGTGGCGGTGGTTCCGGTGGTGGCGGCAGCAACGGCATTAAAGCGAA
CTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATCACTACCAGCAGAATACGCCGATTGGCG
ATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAACTGTCCAAAGACCCTAACGAAAAACGC
GACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACGGTATGGATGAGCTGTACAAGTAATGAAG
ATCT...

pL410-NLS-CGFP$_N$-construct

...CATATGGCGGAATTAATTCCCGAGCCTCCAAAAAAGAAGAGAAAGGTCGAATTGGGTACCGCCCATATGAACGG
CATTAAAGCGAACTTCAAAATCCGTCATAACATTGAGGATGGCGGTGTACAGCTGGCGGATCACTACCAGCAGAATA
CGCCGATTGGCGATGGCCCTGTGCTGCTGCCGGATAACCACTACCTGTCCACCCAGTCTAAACTGTCCAAAGACCCT
AACGAAAAACGCGACCACATGGTGCTGCTGGAATTTGTTACTGCGGCGGGTATTACTCACGGTATGGATGAGCTGTA
CAAGGGCGGTGGCGGTTCCGGTGGCGGTGGCTCTGGATCCconstructSTOPsGAATTCCAGATCT...

pW410-HGFPsol-construct and **pL410-HGFPsol-construct** plasmids were created by inserting a PCR copy of the NdeI-BamHI insert from pGEX-6P-GFPsol into NdeI/BamHI cut **pW410-NGFP$_N$-construct** and **pL410-CGFP$_N$-construct** plasmids.

D.2 BiFC protein constructs

NGFP-CGFP (expressed from either plasmid)

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGSLEEFGGGGSGGGGS*NGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSK
LSKDPNEKRDHMVLLEFVTAAGITHGMDELYK-

Molecular weight: 28629.9 Da

$NGFP_N^*$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*REFCSQANSGRISYDL-

Molecular weight: 20370.7 Da

$NGFP_N$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*-

$NGFP_C$

MGSREF*GGGGSGGGGS*MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYITADKQK-

$NGFP_N$-$LMO4_{LIMs}$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*LSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQAQLGDIGTSSYTKSGMILCRNDYIR
LFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHL-

Molecular weight: 33253.6 Da

$LMO4_{LIMs}$-$NGFP_C$

MGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQAQLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACG
QSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHLNEF*GGGGSGGGGS*MASKG
EELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPDHMKQH
DFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQK
-

$Ldb1_{LID}$-$NGFP_C$

MGSDVMVVGEPTLMGGEFGDEDERLITRLENTQFDAANGIDDEEF*GGGGSGGGGS*MASKGEELFTGVVPILVELDGD
VNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTI
SFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQK-

$NGFP_N$-$Ldb1_{LID}$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*DVMVVGEPTLMGGEFGDEDERLITRLENTQF-

$NGFP_N$-Lhx3a

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD

HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*MLLEAELDCHRERPGAPGASALCTFSRTPEIPMCAGCDQHILDRFILKALDRHWHSKCLKCS
DCHVPLAERCFSRGESVYCKDDFFKRFGTKCAACQLGIPPTQVVRRAQDFVYHLHCFACVVCKRQLATGDEFYLMED
SRLVCKADYETAKQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEK
RLKKDAGRQRWGQYFRNMKRSRGSSKSDKDSIQEGQDSDAEVSFTDEPSMADMGPANGLYSSLGEPAPALGRPVGGL
GSFTLDHGGLTGPEQYRELRPGSPYGIPPSPAAPQSLPGPQPLLSSLVYPDTNLSLVPSGPPGGPPPMRVLAGNGPS
SDLSTESSSGYPDFPASPASWLDEVDHAQF-

Molecular weight: 62534.4 Da

NGFP$_N$-Lhx3$_{LIMs}$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*TPEIPMCAGCDQHILDRFILKALDRHWHSKCLKCSDCHVPLAERCFSRGESVYCKDDFFKRF
GTKCAACQLGIPPTQVVRRAQDFVYHLHCFACVVCKRQLATGDEFYLMEDSRLVCKADYETAKQ-

Lhx3a-NGFP$_C$

MGSMLLEAELDCHRERPGAPGASALCTFSRTPEIPMCAGCDQHILDRFILKALDRHWHSKCLKCSDCHVPLAERCFS
RGESVYCKDDFFKRFGTKCAACQLGIPPTQVVRRAQDFVYHLHCFACVVCKRQLATGDEFYLMEDSRLVCKADYETA
KQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEKRLKKDAGRQRWG
QYFRNMKRSRGSSKSDKDSIQEGQDSDAEVSFTDEPSMADMGPANGLYSSLGEPAPALGRPVGGLGSFTLDHGGLTG
PEQYRELRPGSPYGIPPSPAAPQSLPGPQPLLSSLVYPDTNLSLVPSGPPGGPPPMRVLAGNGPSSDLSTESSSGYP
DFPASPASWLDEVDHAQFEF*GGGGSGGGGS*MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKF
ICTTGKLPVPWPTLVTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQK-

NGFP$_N$-Isl1

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFVRD
GKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEFALREDGLFCRADHDVVERA
SLGAGDPLSPLHPARPLQMAAEPISARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVE
MTGLSPRVIRVWFQNKRCKDKKRSIMMKQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKV
LSDFALQSDIDQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA-

Isl1-NGFP$_C$

MGSMGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFVRDGKTYCKRDYIRL
YGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEFALREDGLFCRADHDVVERASLGAGDPLSPLH
PARPLQMAAEPISARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVW
FQNKRCKDKKRSIMMKQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDIDQ
PAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEAEF*GGGGSGGGGS*MASKGEELFTGVVPILVELDG
DVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERT
ISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQK-

NLS-NGFP$_N$

MAELIPEPPKKKRKVELGTAHMASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPV
PWPTLVTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQK*GGGGSGGGGS*-

NLS-NGFP$_N$-LMO4$_{LIMs}$

MAELIPEPPKKKRKVELGTAHMASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPV
PWPTLVTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQK*GGGGSGGGGS*LSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQAQLG
DIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSL
FCEHDRPTALINGHL-

NLS-NGFP$_C$

MAELIPEPPKKKRKVELGTAHMGSREF*GGGGSGGGGS*MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATY
GKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEG
DTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQK-

NGFPsol$_N$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT**LT**YGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*-

NGFPsol$_C$

MGSREF*GGGGSGGGGS*MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTT**LT**YGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYITADKQK-

NGFPsol-CGFP (expressed from either plasmid)

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT**LT**YGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGSLEEFGGGGSGGGGS*NGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSK
LSKDPNEKRDHMVLLEFVTAAGITHGMDELYK-

NGFPsol$_N$-LMO4$_{LIMs}$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTT**LT**YGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*LSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQAQLGDIGTSSYTKSGMILCRNDYIR
LFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHL-

NGFP$_N$-LMO2$_{LIMs}$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*LLTCGGCQQNIGDRYFLKAIDQYWHEDCLSCDLCGCRLGEVGRRLYYKLGRKLCRRDYLRLF
GQDGLCASCDKRIRAYEMTMRVKDKVYHLECFKCAACQKHFCVGDRYLLINSDIVCEQDIYEWTKINGII-

NGFP$_N$-USH$_{F1}$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT

ADKQK*GGGGSGGGGS*PDTQAQPEVAEPLLKPARFMCLPCGIAFSSPSTLEAHQAYYCSHRIKDTDEAGSDKSGAGGS
GATAGDAAGLTGGSTEPPA-

NGFP$_N$-USH$_{F9}$

MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFARYPD
HMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIT
ADKQK*GGGGSGGGGS*GNSSSVAAAAAAAAEVMKKYCSTCDISFNYVKTYLAHKQFYCKNKPIRPEASDSPSPNHLGG
GVAVGLGIGGLVGGHGQQKN-

CGFP$_N$

MNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYK*GGGGSGGGGSGS*EFQIYES-

Molecular weight: 10758.8 Da

CGFP$_C$

MGSEF*GGGGSGGGGS*NGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVL
LEFVTAAGITHGMDELYK-

CGFP$_N$-Ldb1$_{LID}$

MNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYK*GGGGSGGGGSGS*DVMVVGEPTLMGGEFGDEDERLITRLENTQF-

Molecular weight: 13342.8 Da

Ldb1$_{LID}$-CGFP$_C$

MGSDVMVVGEPTLMGGEFGDEDERLITRLENTQFDAANGIDDEEF*GGGGSGGGGS*NGIKANFKIRHNIEDGGVQLAD
HYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK-

CGFP$_N$-LMO4$_{LIMs}$

MNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYK*GGGGSGGGGSGS*LSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQAQLGDIGTSSYTKSGMILCRNDYI
RLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHL-

LMO4$_{LIMs}$-CGFP$_C$

MGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQAQLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACG
QSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHLNEF*GGGGSGGGGS*NGIKA
NFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK-

CGFP$_N$-Isl1

MNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYK*GGGGSGGGGSGS*MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFVR
DGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEFALREDGLFCRADHDVVER
ASLGAGDPLSPLHPARPLQMAAEPISARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLV

EMTGLSPRVIRVWFQNKRCKDKKRSIMMKQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWK
VLSDFALQSDIDQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA-

Molecular weight: 48879.6 Da

CGFP$_N$-Isl1$_{LBD}$

MNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYK*GGGGSGGGGSGS*GTPMVAASPERHDGGLQANPVEVQSYQPPW-

Isl1-CGFP$_C$

MGSMGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQYLDESCTCFVRDGKTYCKRDYIRL
YGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEFALREDGLFCRADHDVVERASLGAGDPLSPLH
PARPLQMAAEPISARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPRPDALMKEQLVEMTGLSPRVIRVW
FQNKRCKDKKRSIMMKQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPVEVQSYQPPWKVLSDFALQSDIDQ
PAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEAEF*GGGGSGGGGS*NGIKANFKIRHNIEDGGVQLA
DHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK-

CGFP$_N$-Lhx3

MNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYK*GGGGSGGGGSGS*MLLEAELDCHRERPGAPGASALCTFSRTPEIPMCAGCDQHILDRFILKALDRHWHSKCLKC
SDCHVPLAERCFSRGESVYCKDDFFKRFGTKCAACQLGIPPTQVVRRAQDFVYHLHCFACVVCKRQLATGDEFYLME
DSRLVCKADYETAKQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKE
KRLKKDAGRQRWGQYFRNMKRSRGSSKSDKDSIQEGQDSDAEVSFTDEPSMADMGPANGLYSSLGEPAPALGRPVGG
LGSFTLDHGGLTGPEQYRELRPGSPYGIPPSPAAPQSLPGPQPLLSSLVYPDTNLSLVPSGPPGGPPPMRVLAGNGP
SSDLSTESSSGYPDFPASPASWLDEVDHAQF-

Lhx3-CGFP$_C$

MGSMLLEAELDCHRERPGAPGASALCTFSRTPEIPMCAGCDQHILDRFILKALDRHWHSKCLKCSDCHVPLAERCFS
RGESVYCKDDFFKRFGTKCAACQLGIPPTQVVRRAQDFVYHLHCFACVVCKRQLATGDEFYLMEDSRLVCKADYETA
KQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQLSSETGLDMRVVQVWFQNRRAKEKRLKKDAGRQRWG
QYFRNMKRSRGSSKSDKDSIQEGQDSDAEVSFTDEPSMADMGPANGLYSSLGEPAPALGRPVGGLGSFTLDHGGLTG
PEQYRELRPGSPYGIPPSPAAPQSLPGPQPLLSSLVYPDTNLSLVPSGPPGGPPPMRVLAGNGPSSDLSTESSSGYP
DFPASPASWLDEVDHAQFEF*GGGGSGGGGS*NGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQS
KLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK-

NLS-CGFP$_N$

MAELIPEPPKKKRKVELGTAHMNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNE
KRDHMVLLEFVTAAGITHGMDELYK*GGGGSGGGGSGS*EFQIYES-

NLS-CGFP$_N$-Ldb1$_{LID}$

MAELIPEPPKKKRKVELGTAHMNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNE
KRDHMVLLEFVTAAGITHGMDELYK*GGGGSGGGGSGS*DVMVVGEPTLMGGEFGDEDERLITRLENTQF-

CGFP$_N$-PNR$_{NF}$

MNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITHGMD
ELYK*GGGGSGGGGSGS*EGRECVNCGAISTPLWRRDGTGHYLCNACGLYHKMNGMNRPLIKPSKRL-

His$_6$-GFPsol-Lhx3a

MRGSHHHHHHGLVPRGSASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEK
RDHMVLLEFVTAAGITHGMDELYK*GS*MLLEAELDCHRERPGAPGASALCTFSRTPEIPMCAGCDQHILDRFILKALD
RHWHSKCLKCSDCHVPLAERCFSRGESVYCKDDFFKRFGTKCAACQLGIPPTQVVRRAQDFVYHLHCFACVVCKRQL
ATGDEFYLMEDSRLVCKADYETAKQREAEATAKRPRTTITAKQLETLKSAYNTSPKPARHVREQLSSETGLDMRVVQ
VWFQNRRAKEKRLKKDAGRQRWGQYFRNMKRSRGSSKSDKDSIQEGQDSDAEVSFTDEPSMADMGPANGLYSSLGEP
APALGRPVGGLGSFTLDHGGLTGPEQYRELRPGSPYGIPPSPAAPQSLPGPQPLLSSLVYPDTNLSLVPSGPPGGPP
PMRVLAGNGPSSDLSTESSSGYPDFPASPASWLDEVDHAQF-

Molecular weight: 72776.0 Da

Theoretical pI: 6.82

His$_6$-GFPsol-Isl1

MRGSHHHHHHGLVPRGSASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTFSYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL
GHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEK
RDHMVLLEFVTAAGITHGMDELYK*GS*MGDMGDPPKKKRLISLCVGCGNQIHDQYILRVSPDLEWHAACLKCAECNQY
LDESCTCFVRDGKTYCKRDYIRLYGIKCAKCSIGFSKNDFVMRARSKVYHIECFRCVACSRQLIPGDEFALREDGLF
CRADHDVVERASLGAGDPLSPLHPARPLQMAAEPISARQPALRPHVHKQPEKTTRVRTVLNEKQLHTLRTCYAANPR
PDALMKEQLVEMTGLSPRVIRVWFQNKRCKDKKRSIMMKQLQQQQPNDKTNIQGMTGTPMVAASPERHDGGLQANPV
EVQSYQPPWKVLSDFALQSDIDQPAFQQLVNFSEGGPGSNSTGSEVASMSSQLPDTPNSMVASPIEA-

Molecular weight: 67802.0 Da

Theoretical pI: 7.82

Linkers are in italic font and molecular weight was shown only for proteins that were subjected to Western blot.

Appendix E ADH1 promoters in BiFC plasmids

pW410 (with the **short yADH1 promoter**, 411 bp)

GCATGCAACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCCGTTGTTGTCTCACCATATCCGCAATGACAAAAAAA
ATGATGGAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTGTTCCAGAGCTGATGAGG
GGTATCTTCGAACACACGAAACTTTTTCCTTCCTTCATTCACGCACACTACTCTCTAATGAGCAACGGTATACGGCC
TTCCTTCCAGTTACTTGAATTTGAAATAAAAAAAGTTTGCCGCTTTGCTATCAAGTATAAATAGACCTGCAATTATT
AATCTTTTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCTTCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTATAC
CAAGCATACAATCAACTCC**AAGCTT**GCATATG**GGATCC**C**GGGAATTCTG**CAGCCAAGCTAATTCCGGGCGAATTTCT
TATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAA**GTGACTCTTAGGTT**
TTAAAACGAAAATTCTTGTTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGGTCGC
TCTTATTGACCACACCTCTACCGGCATGC

pWAFL (with **full-length yADH1 promoter**, 1482 bp)

GCATGCCTGCAGGTCGAGATCCGGGATCGAAGAAATGATGGTAAATGAAATAGGAAATCAAGGAGCATGAAGGCAAA
AGACAAATATAAGGGTCGAACGAAAAATAAAGTGAAAAGTGTTGATATGATGTATTTGGCTTTGCGGCGCCGAAAAA
ACGAGTTTACGCAATTGCACAATCATGCTGACTCTGTGGCGGACCCGCGCTCTTGCCGGCCCGGCGATAACGCTGGG
CGTGAGGCTGTGCCCGGCGGAGTTTTTTGCGCCTGCATTTTCCAAGGTTTACCCTGCGCTAAGGGGCGAGATTGGAG
AAGCAATAAGAATGCCGGTTGGGGTTGCGATGATGACGACCACGACAACTGGTGTCATTATTTAAGTTGCCGAAAGA
ACCTGAGTGCATTTGCAACATGAGTATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGTGGTGCG
AACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATAACCGCTAGAGTACTTTGAAGAGGAAACAGCAATAGG
GTTGCTACCAGTATAAATAGACAGGTACATACAACACTGGAAATGGTTGTCTGTTTGAGTACGCTTTCAATTCATTT
GGGTGTGCACTTTATTATGTTACAATATGGAAGGGAACTTTACACTTCTCCTATGCACATATATTAATTAAAGTCCA
ATGCTAGTAGAGAAGGGGGGTAACACCCCTCCGCGCTCTTTTCCGATTTTTTTCTAAACCGTGGAATATTTCGGATA
TCCTTTTGTTGTTTCCGGGTGTACAATATGGACTTCCTCTTTTCTGGCAACCAAACCCATACATCGGGATTCCTATA
ATACCTTCGTTGGTCTCCCTAACATGTAGGTGGCGGAGGGGAGATATACAATAGAACAGATACCAGACAAGACATAA
TGGGCTAAACAAGACTACACCAATTACACTGCCTCATTGATGGTGGTACATAACGAACTAATACTGTAGCCCTAGAC
TTGATAGCCATCATCATATCGAAGTTTCACTACCCTTTTTCCATTTGCCATCTATTGAAGTAATAATAGGCGCATGC
AACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCCGTTGTTGTCTCACCATATCCGCAATGACAAAAAAAATGATG
GAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTGTTCCAGAGCTGATGAGGGGTATC
TTCGAACACACGAAACTTTTTCCTTCCTTCATTCACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTT
CCAGTTACTTGAATTTGAAATAAAAAAAGTTTGCCGCTTTGCTATCAAGTATAAATAGACCTGCAATTATTAATCTT
TTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCTTCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCA
TACAATCAACTCAAGCTTGCATATGGGATCCCGGGAATTCTGCAGCCAAGCTAATTCCGGGCGAATTTCTTATGATT

pL410 (with the **short yADH1 promoter**, 411 b)

GCATGCAACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCCGTTGTTGTCTCACCATATCCGCAATGACAAAAAAA
ATGATGGAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTGTTCCAGAGCTGATGAGG
GGTATCTTCGAACACACGAAACTTTTTCCTTCCTTCATTCACGCACACTACTCTCTAATGAGCAACGGTATACGGCC
TTCCTTCCAGTTACTTGAATTTGAAATAAAAAAAGTTTGCCGCTTTGCTATCAAGTATAAATAGACCTGCAATTATT

AATCTTTTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCTTCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTATAC
CAAGCATACAATCAACTCCAAGCTTGCATATGGGATCCGAATTCCAGATCTATGAATCGTAGATACTGAAAAACCCC
GCAAGTT

pLAFL (with **full-length yADH1 promoter**, 1482 bp)

GCATGCCTGCAGGTCGAGATCCGGGATCGAAGAAATGATGGTAAATGAAATAGGAAATCAAGGAGCATGAAGGCAAA
AGACAAATATAAGGGTCGAACGAAAAATAAAGTGAAAAGTGTTGATATGATGTATTTGGCTTTGCGGCGCCGAAAAA
ACGAGTTTACGCAATTGCACAATCATGCTGACTCTGTGGCGGACCCGCGCTCTTGCCGGCCCGGCGATAACGCTGGG
CGTGAGGCTGTGCCCGGCGGAGTTTTTTGCGCCTGCATTTTCCAAGGTTTACCCTGCGCTAAGGGGCGAGATTGGAG
AAGCAATAAGAATGCCGGTTGGGGTTGCGATGATGACGACCACGACAACTGGTGTCATTATTTAAGTTGCCGAAAGA
ACCTGAGTGCATTTGCAACATGAGTATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGTGGTGCG
AACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATAACCGCTAGAGTACTTTGAAGAGGAAACAGCAATAGG
GTTGCTACCAGTATAAATAGACAGGTACATACAACACTGGAAATGGTTGTCTGTTTGAGTACGCTTTCAATTCATTT
GGGTGTGCACTTTATTATGTTACAATATGGAAGGGAACTTTACACTTCTCCTATGCACATATATTAATTAAAGTCCA
ATGCTAGTAGAGAAGGGGGGTAACACCCCTCCGCGCTCTTTTCCGATTTTTTTCTAAACCGTGGAATATTTCGGATA
TCCTTTTGTTGTTTCCGGGTGTACAATATGGACTTCCTCTTTTCTGGCAACCAAACCCATACATCGGGATTCCTATA
ATACCTTCGTTGGTCTCCCTAACATGTAGGTGGCGGAGGGGAGATATACAATAGAACAGATACCAGACAAGACATAA
TGGGCTAAACAAGACTACACCAATTACACTGCCTCATTGATGGTGGTACATAACGAACTAATACTGTAGCCCTAGAC
TTGATAGCCATCATCATATCGAAGTTTCACTACCCTTTTTCCATTTGCCATCTATTGAAGTAATAATAGGCGCATGC
AACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCCGTTGTTGTCTCACCATATCCGCAATGACAAAAAAAATGATG
GAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTGTTCCAGAGCTGATGAGGGGTATC
TTCGAACACACGAAACTTTTTCCTTCCTTCATTCACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTT
CCAGTTACTTGAATTTGAAATAAAAAAAGTTTGCCGCTTTGCTATCAAGTATAAATAGACCTGCAATTATTAATCTT
TTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCTTCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCA
TACAATCAACTCAAGCTTGCATATGGGATCCGAATTCCAGATCTATGAATCGTAGATACTGAAAAACCCCGCAAGTT

GGATCC - BamHI

GAATTC - EcoRI

CTGCAG - PstI

AAGCTT - HindIII

AGATCT - BglII

CATATG - NdeI

CCCGGG - SmaI

CTCGAG - XhoI

GCATGC - SphI

TGA/TAA/TAG - STOP codons

Appendix F Calculations and curve fitting of data from binding affinity assays

In the competition binding assays described in this thesis, the cleaved $Ldb1_{LID}$ peptide (4.7 kDa) should be displaced by the larger MBP-$Ldb1_{LID}$ construct (48.3 kDa) with an assumed identical binding affinity. In this case there is a fluorescent 'receptor' (A= Fl-Cys-$LMO4_{LIMs}$, Table F.1) and two homologous non-fluorescent 'ligands' (B=$Ldb1_{LID}$ and B'=MBP-$Ldb1_{LID}$) as opposed to standard FA competition formats where a smaller fluorescent ligand is displaced from the larger receptor protein resulting in reduced anisotropy. Therefore, instead of two fluorescent species, this assay contains three (free A and two complexes AB and AB'), which makes calculations of binding affinity complex. Note that the total concentrations of A and B are also equal ($[A]_T=[B]_T$).

Table F.1. Definition of terms.

Term	Definition
A	Fluorescently labelled LIMs construct
B	$Ldb1_{LID}$
B'	MBP-$Ldb1_{LID}$
AB	Fluorescent-LIMs : $Ldb1_{LID}$ complex
AB'	Fluorescent-LIMs : MBP-$Ldb1_{LID}$ complex
[A], [B] and [B']	Concentrations of relevant unbound (free) species
[AB] and [AB']	Concentrations of relevant complexes
$[A]_T$, $[B]_T$ and $[B']_T$	Total concentrations. Sums of concentrations of a free species and relevant complex(es) it takes part in. $[A]_T=[A]+[AB]+[AB']$, $[B]_T=[B]+[AB]$ and $[B']_T=[B']+[AB']$
r_{obs}	Observed (measured) anisotropy
r_A, r_{AB} and $r_{AB'}$	Anisotropy values of theoretical samples with 100% A, AB or AB' respectively
r_{max} and r_{min}	Maximal and minimal observed anisotropy for any specific titration (binding assay).
I_{shift}	Intensity of a shifted complex band in protein gel shift assays
I_{ns}	Intensity of a non-shifted band(s) in protein gel shift assays
I_T	Total fluorescence intensity in a gel lane. $I_T = I_{shift} + I_{ns}$

In this analysis it is assumed that the $Ldb1_{LID}$ peptide binds along the labelled LIMs construct without significantly changing shape and size of the labelled construct upon dissociation. Therefore, it can be assumed that the relevant anisotropy value of a sample with 100% of labelled-LIMs monomer (r_A) or with a 100% labelled-LIMs:LID complex (r_{AB}) are approximately the same.

$$r_A \approx r_{AB} \qquad \text{Equation F.1}$$

The observed anisotropy (r_{obs}) is a sum of anisotropies from the relevant labelled species,

$$r_{obs} = r_A \cdot f(A) + r_{AB'} \cdot f(AB') + r_{AB} \cdot f(AB) \qquad \text{Equation F.2}$$

By replacing the r_A with r_{AB} and some rearranging the following is obtained,

$$r_{obs} = r_{AB} \cdot (f(A) + f(AB)) + r_{AB'} \cdot f(AB') \qquad \text{Equation F.3}$$

and since $$f(A) + f(AB) + f(AB') = 1 \qquad \text{Equation F.4}$$

$f(A) + f(AB)$ can be substituted to obtain

$$r_{obs} = r_{AB} \cdot (1 - f(AB')) + r_{AB'} \cdot f(AB') \qquad \text{Equation F.5}$$

which can be further rearranged to

$$r_{obs} = r_{AB} + f(AB') \cdot (r_{AB'} - r_{AB}) \qquad \text{Equation F.6}$$

and knowing that $r_{AB'} = r_{max}$ and $r_{AB} = r_{min}$, a normalized response r_n can be calculated

$$r_n = (r_{obs} - r_{min}) / (r_{max} - r_{min}) \qquad \text{Equation F.7}$$

Since $$r_n = f(AB') = [AB']/A_T \qquad \text{Equation F.8}$$

r_n values can be used to plot the data.

In gel shift experiments, each lane contains bands with the free labelled monomer and/or the complex(es) containing it, whereas the band intensities are proportional to amounts of labelled monomer in each band. Even though the reaction is not at equilibrium during electrophoresis, it is assumed that for strong interactions (with low off rates) and at a low temperature (4° C in this case), the ratio of band intensities in each lane is a good approximation to a ratio of labelled species in the relevant sample. If the bands in a gradient gel do not have a significant smear, this assumption is likely to be true.

A ratio of the shifted band intensity (I_{shift}) vs. the sum of shifted and non-shifted (I_{ns}) band intensities equals the fraction of labelled protein that is in the complex:

$$I_{shift}/(I_{shift}+I_{ns}) = [AB']/[A]_T \qquad \text{Equation F.9}$$

and if the following is defined

$$I_{total} = I_{shift} + I_{ns} \quad \text{Equation F.10}$$

$$I_{shift}/I_T = [AB']/[A]_T \quad \text{Equation F.11}$$

Since I_T calculations were not always reliable, and knowing that multiplying a function by a constant would not affect the IC_{50} estimate from the non-linear regression analysis [176], I_{shift} instead of I_{shift}/I_T values could be easily entered when the function is fitted, provided that the slope (k) parameter is not fixed.

F.1 Fitting the data for homologous competition binding with ligand depletion

A solution to the ligand depletion problem can be obtained [323, 350]. When two different constructs B = $Ldb1_{LID}$ and B' = MBP-$Ldb1_{LID}$ with the same binding affinity K_d for a fluorescently labelled protein (A) are in competition and total concentrations of A and B are the same and constant, assuming an equilibrium and no non-specific binding, the following equations can be worked out:

$$[AB'] = [A]_T \cdot [B']/([B] + [B'] + Kd) \quad \text{Equation F.12}$$

and the experimentally obtained responses (Y) to a titration of $\mathbf{[B']_T}$ can be normalized,

$$(Y - Ymin)/(Ymax - Ymin) = \mathbf{Y_N} \quad \text{Equation F.13}$$

As $\mathbf{Y_N} = [AB']/[A]_T$ Equation F.14

using the F.12, the following can be obtained

$$\mathbf{Y_N} = [B']/([B] + [B'] + Kd) \quad \text{Equation F.15}$$

At equilibrium, the fractions of free vs. total protein are the same for B and B', therefore

$$[B] = [B'] \cdot [B]_T/\mathbf{[B']_T} \quad \text{Equation F.16}$$

and $[B'] = \mathbf{[B']_T} - [AB']$ Equation F.17

so the following can be derived:

$$\mathbf{Y_N} = (\mathbf{[B']_T} - [AB'])/((\mathbf{[B']_T} - [AB']) \cdot ([A]_T/\mathbf{[B']_T} + 1) + Kd) \quad \text{Equation F.18}$$

and after exchanging $[AB'] = [A]_T \cdot \mathbf{Y_N}$ the following is obtained

$$([A]_T^2/\mathbf{[B']_T} + [A]_T) \cdot \mathbf{Y_N}^2 - (Kd + 2 \cdot [A]_T + \mathbf{[B']_T}) \cdot \mathbf{Y_N} + \mathbf{[B']_T} = 0 \quad \text{Equation F.19}$$

Consequently, the solution to this quadratic equation gives

$$\mathbf{Y_N} = (- b - ((b)^2 - 4 \cdot a \cdot c)^{1/2})/(2 \cdot a) \qquad \text{Equation F.20}$$

where

$$a = [A]_T^2/\mathbf{[B']_T} + [A]_T \qquad \text{Equation F.21}$$

$$b = - (Kd + 2 \cdot [A]_T + \mathbf{[B']_T}) \qquad \text{Equation F.22}$$

$$c = \mathbf{[B']_T} \qquad \text{Equation F.23}$$

By entering values for $[A]_T$ and fitting this formula to experimental data for $\mathbf{[B']_T}$ and Y ($\mathbf{Y_N}$) in the Origin9.1 program, K_d values were determined.