Methods to identify novel α-helical peptides that bind to LMO4

Gabrielle Cathryn McClymont

A thesis submitted in fulfilment of the requirements for the degree of Master of Philosophy (Science)

School of Life and Environmental Sciences

The University of Sydney

September 2018

Table of Contents

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Jacqui Matthews for all your help, advice, guidance and encouragement. Working and learning in your lab has been an incredible experience and I am enormously grateful to you for taking me on as a student. I would also like to thank Ann Kwan my auxiliary supervisor and Flyp for all the time and effort spent teaching me as well as your patience, wisdom and support.

I am grateful to Mario for teaching me to use the plate fluorimeter, to Phil and David for helping me with the flow cytometer to Dr Tim Newsome for helping me with the fluorescent microscopy. To Caroline and David for helping me analyse the flow cytometry data and Melanie for helping me analyse the fluorescence microscopy data. To Jason for providing the methods to create super competent BL-21(DE3) cells and his transformation method. To Jacqui, Ann, Lorna, Ngaio, Janine and James, thank you very much for helping to edit my thesis. To the rest of the Matthews lab thank you for your support, friendship and patience.

Thank you to the Mackay/Matthews lab your friendship, wisdom and encouragement has made this year absolutely fantastic. A special thanks to the endless sugar suppliers: James, Vania and Taylor, Janine for the hugs and Ngaio for the smiles and advice an. To my fellow students thank you for your companionship.

To my friends and family outside the lab thank you for all your support, encouragement and faith in me. In particular I would like to thank my mum Rosalie for your endless faith in me, without you I would not have made it this far.

Abstract

Protein-protein interactions (PPIs) play an essential role in regulating cellular processes. Although long regarded as undruggable, PPIs have recently become attractive targets for drug design. One family of proteins that functions entirely through PPIs is the LIM domain-only (LMO) family. LMO protein 4 (LMO4) is a transcriptional co-regulatory protein that contributes to varied developmental processes as well as the pathogenesis of breast cancer, although the mechanism or mechanisms involved are only partially known. The development of inhibitors of LMO4 could form useful research tools for understanding these mechanisms of action and might also form the basis of new breast cancer therapies. Peptide binders can form an excellent starting point for designing inhibitors of PPIs. Previously, the Matthews Laboratory has worked with peptide inhibitors of LMO4 that were based on natural binding partners of LMO4. Although it was possible to generate peptides that bound LMO4 with higher affinity, these β-strand peptides lacked specificity for LMO4.

The hypotheses underlying this thesis is that α-helical peptides can be more specific than β-strand peptides, and that it should be possible to generate α-helical peptide binders of LMO4, using natural α-helices as templates. The aims for this thesis were to further develop two methods to identify novel α-helical peptides that bind to LMO4: a split EGFP system providing a high through-put method for the initial screening of a library of α-helical peptides; and, a Yeast Two Hybrid Competition Assay (Y2HCA) as an orthogonal method to validate hits and provide an assessment of the relative binding affinities of the α-helical peptides. For the Y2HCA work, several new constructs were investigated to span a range of different binding affinities. A new construct was added to a series of tethered LMO4-CtIP$_{LID}$ constructs to provide some relative assessment of the affinities of weakly binding peptides. Two other constructs were not suitable for this purpose but do appear to stabilise LMO4 to increase sensitivity for detecting LMO4-peptide interactions by this method.

The split EGFP complementation system previously designed during my Honours year was further developed with positive and negative controls. A construct to screen a library of α-helical peptides for affinity to LMO4$_{LIM1}$ was also developed. Naturally occurring α-helices were considered for their suitability as templates, and three - BAD, HIV-1 GP41 and PUMA - were tested in the split EGFP system. Attempts were made to introduce an mCherry

expression control to the split EGFP system. Chemical transformation into BL-21(DE3) was optimised to allow efficient expression of the library. To facilitate library screening plasmid DNA extraction protocols from BL-21 (DE3) cells were also developed. This thesis provides mechanisms to identify α-helical peptides that bind to LMO4 and provides the ground work to make a high-throughput library screen possible.

Abbreviations

3-AT	3-amino-1,2,4-triazole	**MQW**	MilliQ® water
CtIP	C-terminal Binding Protein-interacting protein	**OD_{600nm}**	Optical density at 600 nm
DMSO	Dimethyl sulfoxide	**OD_{700nm}**	Optical density at 700 nm
E. coli	*Escherichia coli*	**PBS**	Phophate-buffered Saline
EDTA	Ethylenediamine tetra-acetic acid	**PAGE**	Polyacrylamide gel electrophoresis
EGFP	Enhanced Green Fluorescent Protein	**PCR**	Polymerase chain reaction
FACS	Fluorescence Activated Cell Sorting	**PEG**	Polyethylene glycol
GFP	Green Fluorescent Protein	**PMSF**	Phenylmethanesulfonyl fluoride
HRV-3C	Human Rhinovirus 3C	**PPI**	Protein-protein interaction
IPTG	Isopropyl β-D-thiogalactopyranoside	**SD**	Synthetic dextrose
LB	Luria-Bertani broth	**SDS**	Sodium dodecyl sulphate
LDB1	LIM domain binding protein 1	**TAE**	Tris/acetate/EDTA
LID	LIM interaction domain	**Tris**	Tris(2-carboxymethyl)aminomethane
LIM	Lin1 Isl1 Mec3		
LIM-HD	LIM-homeodomain	**X-α-gal**	5-bromo-4-chloro-3 indolyl-α-D-galactopyranoside
LMO	LIM only protein	**YPD**	Yeast peptone dextrose
LMO4	LIM only protein 4	**Y2HCA**	Yeast 2 Hybrid Competition Assay

Chapter 1: Introduction

Protein-protein interactions (PPIs) play an essential role in regulating and controlling cellular processes. PPIs involve direct contact between two or more proteins, where the binding can be mediated by short sequences or larger folded domains. PPIs allow the formation of protein complexes that have important functional roles in biology. Manipulating PPIs with inhibitors and/or agonists can be a possible method for treatment of disease, which means that some PPIs have become targets for drug design. Inhibitors and agonists can also be used as tools to better understand the role of certain PPIs within cells. One protein that is thought to function entirely through PPIs is LIM-only protein 4 (LMO4), which is reported to bind a variety of protein partners and regulate numerous cellular processes (1, 2).

1.1 LMO4

1.1.1 Function

Members of the LIM domain-only (LMO) family of proteins are generally thought of as transcriptional co-regulators that mediate their functions solely via PPIs (3-5). There are four LMO proteins in mammals, LMO1–4, which play crucial roles in cell fate determination, cell growth and differentiation (1).

LMO4 is a negative regulator of differentiation (6) and a positive regulator of cell proliferation (7-9) in the mammary epithelium. It plays an essential role in organ development such as neural tube closure and inner ear development, especially in the cochlea (10-12). LMO4 acts at different stages of the cell cycle, likely by affecting the transcriptional programs or by interacting with proteins that control the cell cycle (1). There are many mechanisms through which LMO4 may produce transcriptional effects. LMO4 may repress transcription through the recruitment of co-repressor proteins such as CtIP/RBBP8 (C-terminal Binding Protein-interacting protein/retinoblastoma binding protein 8)(13) and MTA1 (metastasis-associated protein 1) (14). In contrast, it may activate transcription by recruiting GATA factors (15) and/or preventing HDAC2 from binding to chromatin (9).

1.1.1 Structure

LMO4 essentially contains only two tandem LIM domains (LIM1 and LIM2; Figure 1.1A). Each domain is approximately 60 residues in length and contains two sequential zinc coordinating modules (Figure 1.1A). Each module comprises two β-hairpins followed by an α-helix of up to eight residues (4). In LMO proteins, the two LIM domains are separated by a short two-residue linker (4).

All well-characterized LMO4-binding proteins bind to their target via a LIM interaction domain (LID). These binding proteins are LIM domain-binding protein-1 (LDB1), CtIP and Deformed epidermal autoregulatory factor-1 (Deaf-1). However, there are many additional potential binding partners that appear to confer roles to LMO4 in a variety of regulatory pathways. These partners include the basic loop-helix-loop proteins, nescient helix-loop-helix 1 (16) and neurogenin-2 (17), estrogen receptor α (14), neogenin (18), TGF-β family protein BMP7, HDAC1 and 2 and MTA1 (2, 9).

It is difficult to express and purify isolated recombinant LMO4, as it is highly insoluble and prone to aggregation. To overcome this problem, the LIM domains of LMO4 can be tethered via a flexible linker sequence to the LID of a binding partner (19, 20) (Figure 1.1B). Using this strategy, the structures of LMO4 in complex with each of LDB1, CtIP or Deaf-1 have been determined (Figure 1.1B-D) and the interactions well characterized (2, 21, 22). The tethering strategy works well for these LMO4-LID complexes because the N- and C-termini of LMO4 lie close to the C- and N- termini, respectively, of the LID within the complex (Figure 1.1C and D). This arrangement and the validity of the tethering approach were confirmed through several methods. For example, a protease cleavage site was placed in the linker of a tethered LMO4-LDB1 complex. ^{15}N-HSQC spectra of the intact and cut complexes were identical, apart from peaks arising from residues at the cut site, indicating that the topology of the complex was not dependent on the intact linker (20).

LDB1, CtIP, and Deaf-1 each target the same face on the LIM domains of LMO4 and therefore bind in a mutually exclusive manner (22) (Figure 1.1E and F). Each partner contains a ~30-residue LID that share little sequence identity; however, they have higher similarity in terms of amino acid composition. The LIDs contain two linear binding motifs that form an extended interface across both LIM domains, forming a series of β-strands that interact with β-hairpins

in the LIM domains (e.g. Figure 1.1D) (2, 19, 22-24). The linear binding motifs are specific for LIM1 or LIM2 (Figure 1.1C and D) (19). The formation of β-strands at the interfaces means that backbone-backbone hydrogen bonds and extensive hydrophobic contacts are a feature of binding at both LIM domains. However, mutagenic studies have demonstrated that LIM1-partner interactions tend to be focused on a single binding hotspot, with a bulky hydrophobic sidechain residue (or two residues in the case of CtIP) buried in a hydrophobic binding pocket between the two Zn-coordinating modules in LMO4-LIM1 (22, 24). LIM2-partner interactions appear to have a more extended binding interface, featuring surface interactions of hydrophobic sidechains. These differences, along with a small number of electrostatic interactions, are thought to direct the orientation and specificity of binding (19, 21, 23).

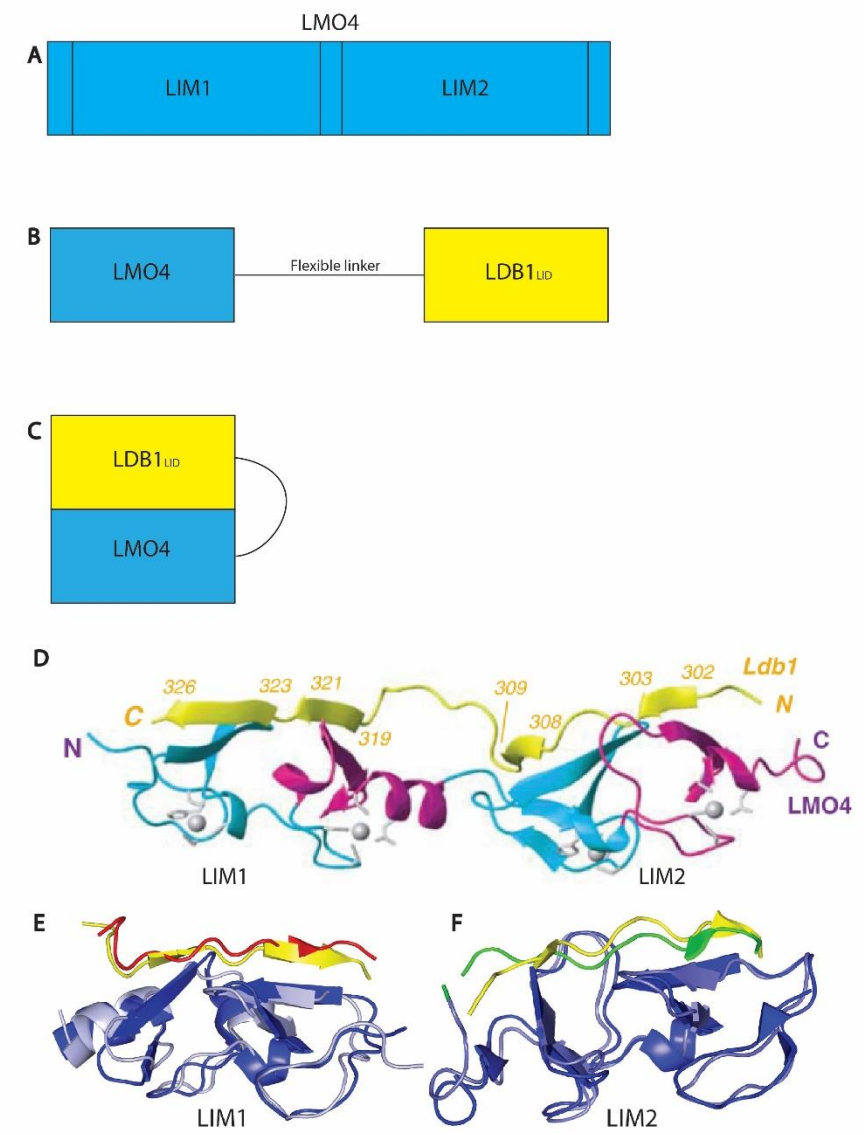


Figure 1.1: Structure of LMO4 and LMO4-LID complexes. (A) Schematic of LMO4 showing LIM domains. (B) Tethered LMO4 and LDB1 construct (C) tethered LMO4 and LDB1 with LMO4 binding to LDB1. (D) Structure of a tethered LMO4-LDB1 complex; LMO4 Zn-binding modules are alternating cyan and magenta with LDB1 in yellow. The numbering shows elements of the secondary structure of LDB1 (as determined by the 1D). Figure taken from Deane et al (19). (E) Structure of LDB1 and CtIP in complex with LIM1 of LMO4, LDB1 is yellow, CtIP is red, LIM1 in complex with LDB1 is blue, LIM1 in complex with CtIP is light blue. (F) Structure of LDB1 and Deaf-1 in complex with LIM2 of LMO4, LDB1 is yellow, Deaf-1 is green, LIM2 in complex with $LDB1_{LID}$ is blue, LIM2 in complex with $Deaf\text{-}1_{LID}$ is light blue. Images were generated from PyMOL using the coordinates of PDB codes 2L4Z (LMO4-CtIP)(22), 2MBV (LMO4-Deaf-1)(2) and 1RUT (LMO4-LDB1)(19), respectively.

1.1.2 LMO4 in breast cancer

LMO4 is expressed at high levels during pregnancy, where it plays key roles in remodelling of mammary tissue (6, 9, 25). It is also overexpressed in over half of breast cancers, where levels of expression positively correlate with tumour progression and poor prognosis (1, 6-8, 25-27). The exact mechanisms of LMO4 function in tumourigenesis are unknown, but changes to the regulation of proliferation, differentiation and apoptosis are common to all types of LMO4 positive breast cancers (7, 13, 28). Trastuzumab is an effective antibody drug for HER2 positive breast cancer, but resistance to this drug has been recently been associated with the overexpression of LMO4 (27). CtIP and BRCA1 are tumour suppressor proteins (29-32). One hypothesis is that the overexpression of LMO4 in cancer tissue may sequester CtIP to prevent CtIP from fulfilling its roles in DNA repair (where it interacts with BRCA1) and thereby contribute to a decrease in tumour suppression activity (13, 22). Another possible role of LMO4 in tumorigenesis is as a co-activator of transforming growth factor β (TGF-β) signalling (33). Thus, overexpression of LMO4 could increase the proliferation of epithelial cells by amplifying TGF-β signalling (33).

1.1.3 Inhibition of LMO4

Forced overexpression of LMO4 in mouse and human mammary epithelium contributes to mammary tumorigenesis by promoting cell proliferation and motility (7, 27). Indeed, the levels of LMO4 overexpression correlate with proliferation rates (8). In contrast, down-regulation of LMO4 using RNAi was shown to significantly reduce the proliferation, migration and invasion of both human and mouse breast cancer cell lines (8, 25, 33).

Given these properties, the inhibition of LMO4 is a potential strategy for both the investigation of the mechanism of action of LMO4 and for the development of diagnostics and therapeutics for breast cancer. The inhibition of specific LMO4-associated PPIs with small molecules or peptides, rather than broad based knockdown strategies (such as with RNAi) has the potential to allow a subset of regulatory pathways to be blocked, which may provide an advantage for research (to probe specific pathways) or therapy (to minimise side-effects).

1.1.4 Inhibition of LMO4's PPIs using peptides

Previously the Matthews laboratory focused on developing β-strand peptide inhibitors of LMO4, based on the natural binding partners of LMO4. Both specificity and affinity are vital components of inhibitor design. However, all LMO and LIM-HD family proteins contain LIM domains that are recognised by LDB1. Ideally an inhibitor should specifically target LMO4 and not other family members, to minimize off-target affects that might lead to toxicity, while maintaining reasonable levels of binding affinity that are required for efficacy.

Attempts were made to increase the binding affinity and specificity of LDB1 sequences using random mutagenesis around binding hotspots, coupled with selection by phage display (J. Duncan, R. van den Hoven, unpublished). In other work, collaborators at the University of North Carolina used a structure-based *in silico* method to design variants of LDB1 that were optimised for binding to LMO4, which were tested in yeast two hybrid (Y2H) assays in the laboratory (P. Stokes & M. Petitmangin, unpublished). Lastly, chimeric peptides generated by fusing the LIM1-binding LID region of CtIP and the LIM2-binding LID region of Deaf-1 were tested for binding (P. Stokes, unpublished). Overall, those peptides with improved binding properties tended to display increased binding affinity but decreased specificity, or decreased affinity with increased specificity.

1.1.5 Converting from β-strand to α-helical inhibitors

It is likely the lack of specificity is linked to the properties of β-strands. In particular, most of the stabilising features of the LMO4-peptide interfaces are intrinsically non-specific, formed by backbone-backbone hydrogen bonds and hydrophobic sidechain interactions. In contrast, α-helical peptides generally interact with partners using sidechains. Moreover, α-helical peptides have been engineered to have increased specificity and affinity in other systems (34, 35)

The hypothesis that underlies this thesis is that helical, rather than β-strand, peptide inhibitors of LMO4 could provide increased specificity. There are several factors that point to the feasibility of this hypothesis. Interaction surfaces on proteins tend to be plastic, able to adapt to bind structurally different targets. LIM domains, including those of LMO4 are

intrinsically flexible making them able to accommodate different binding partners (2, 21), and are made up of small elements of secondary structure joined by loops. These features enable LMO4 to adopt altered conformations when binding to different partners (19). For example, LMO4 adopts a slightly different conformation when bound to CtIP compared with LDB1 (Figure 1.1E), accommodating two sidechains in the cleft between the two Zn-modules in $LMO4_{LIM1}$ in the case of CtIP binding, rather than one for LDB1 (22). Thus it is reasonable that the specific sidechain interactions of an α-helix, while different to that of the natural β-strand LID, are likely to be accomodated by LMO4. The binding face on LMO4 that β-strands interact with is very broad and thus would need little remodelling to accommodate an α-helix rather than a β-strand.

1.2 Initial Screening system for library of α helical peptides: EGFP Complementation

The general strategy chosen for identification of helical peptide binders was to develop a GFP-based complementation system to screen a library of potential candidates and identify those that bind to LMO4. The levels of green fluorescence produced by a reconstituted split GFP will be used to identify if binding has occurred and to provide a measure of the affinity of the interaction.

1.2.1 GFP

Green Fluorescent Protein (GFP) was originally discovered in *Aequorea* jellyfish (36). The 11-stranded β-barrel structure of GFP surrounds an α-helix that forms an autocatalytic fluorophore (Figure 1.4). Residues S65, Y66 and G67 mature through a series of cyclisation, dehydration, and oxidation reactions to become a functional fluorophore only when the protein is properly folded (37-39).

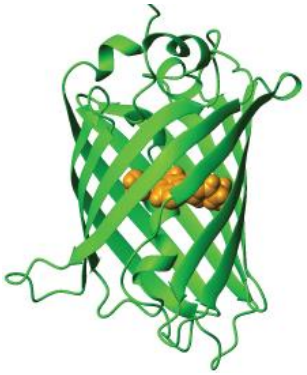


Figure 1.4: Structure of GFP. β-Barrel structure in green with fluorophore in yellow. Image was generated from PyMOL using PDB 2YOG (40).

GFP has been engineered and optimised for a variety of assays. Three common mutations (F99S, M153T, and V163A - the 'cycle 3' mutations) enhance protein folding, thereby increasing whole cell fluorescence during expression without affecting levels of transcription and translation in bacteria (41, 42). Two other mutations, S65T and F64L, were discovered to make EGFP (Enhanced GFP) (43). The excitation wavelength of EGFP was increased from 395 nm to 488 nm with a concomitant 35-fold increase in fluorescence intensity compared to wild type GFP (43). EGFP also shows enhanced folding and solubility (43). Both sets of mutations are combined in 'folding reporter' GFP, which has intense whole cell fluorescence due to increased native-like folding, solubility, and improved spectral characteristics (42, 44, 45). This folding reporter GFP is often used in *in vivo* assays.

1.2.2 Split EGFP Systems

EGFP has been adapted for use in split EGFP systems, which are used to report PPIs (46). The split EGFP systems are a type of protein-fragment complementation assay (PCA) (47). In this type of assay two proteins of interest are fused to two fragments of a reporter protein, where each of the isolated fragments has no activity (46-49). Interactions between the proteins of interest bring the fragments of the reporter gene together, allowing the fragments to associate and fold, thus reconstituting the activity of the reporter (Figure 1.5) (46, 47). PCAs can utilize many different proteins as reporter proteins (47). Those that use fluorescent reporter proteins such as GFP are referred to as bimolecular fluorescence complementation assays (BiFCs) (49). In this study, EGFP split in the loops between β-sheets 7 and 8 was utilized as a reporter (50).

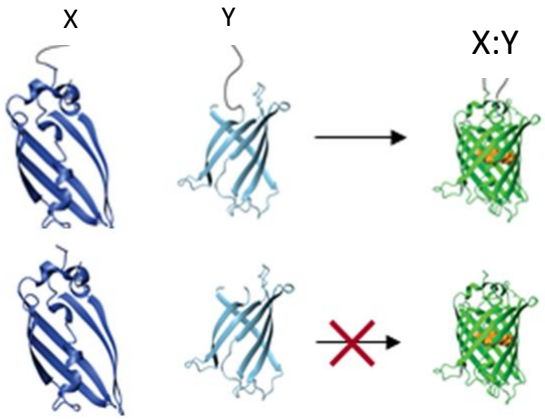


Figure 1.5: Split EGFP System. If the proteins of interest X and Y come together the EGFP molecule assembles and fluorescence occurs. Without the interaction the fragments of EGFP remain separate and no fluorescence occurs. Figure supplied by Jacqui Matthews.

1.2.3 Split EGFP system to find α-helical peptide inhibitors for LMO4

Tethered LMO4-LDB1 complexes have previously been adapted for use in a split GFP system in the Matthews laboratory. A prototype system using GFP was tested many years ago (M. Swanton, unpublished) but not further developed, and a research student developed a BiFC assay for use in yeast cells (51). As part of my Honours research project I designed a split EGFP system for use in identifying potential α-helical peptides that bind LMO4 (Figure 1.6). EGFP is

split as above with one segment tethered to LMO4 and the other to an α-helical peptide. If the α-helix binds to LMO4 then the two parts of EGFP will be brought together and auto-fluoresce. Due to the solubility problems associated with recombinant LMO4, this protein and the α-helical peptide need to be tethered (Section 1.1.1). The aim was to create a library of α-helical peptides in the split EGFP system (Figure 1.6) and overexpress these constructs in *E. coli* cells. The cells would then be screened by fluorescence activated cell sorting (FACS). This high throughput method depends on the fluorescence properties of cells (or markers of cellular proteins) and sorting of cells to isolate those with high levels of EGFP fluorescence, thereby selecting peptides with increased affinity for LMO4.

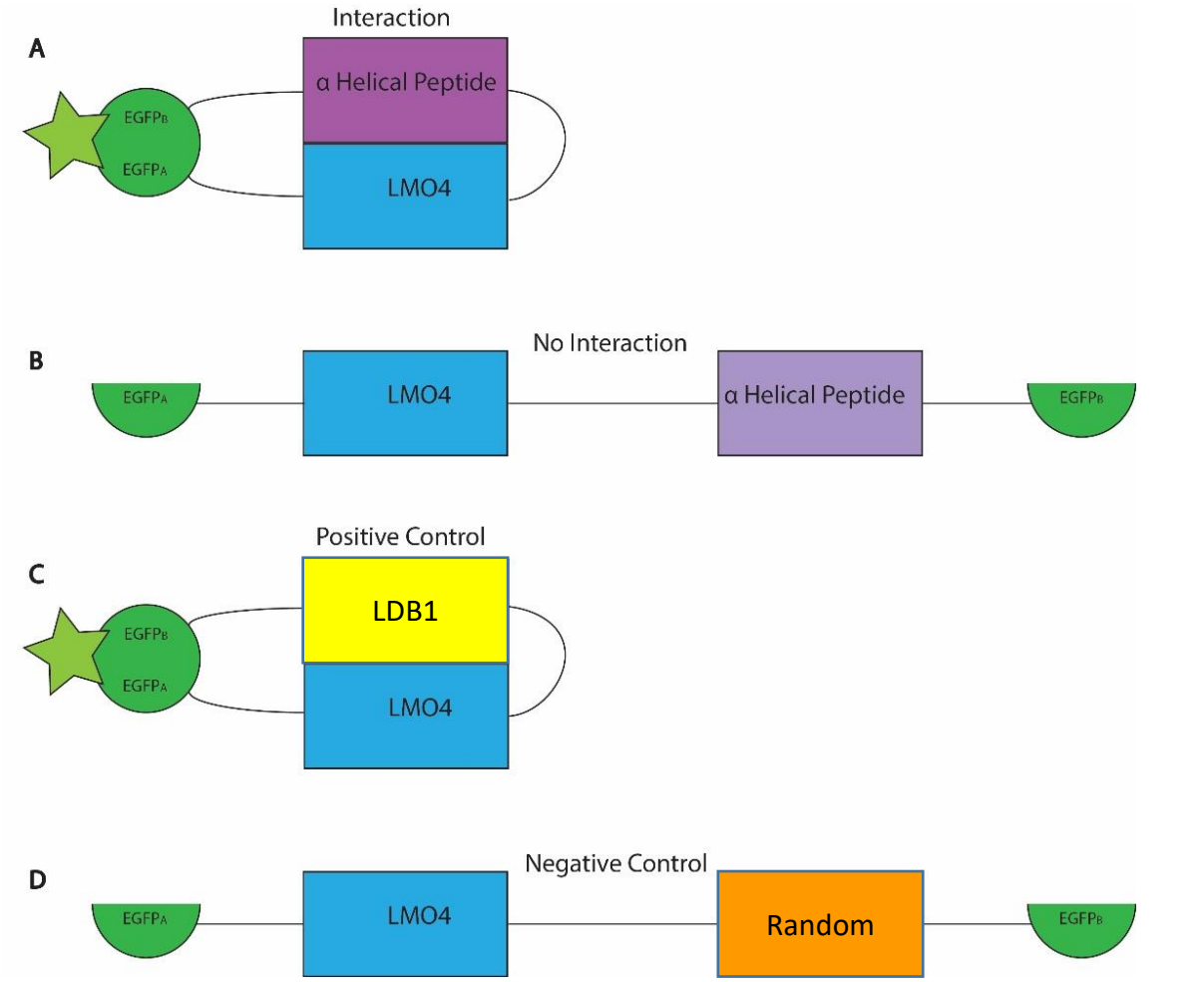


Figure 1.6: Schematic showing general design of the split EGFP system. Star = fluorescence (A) an α-helical peptide that binds to LMO4 bringing the two fragments of EGFP together to produce fluorescence (B) non-binding peptides should not fluoresce. (C) In the positive control LMO4 is tethered to LDB1 and binding between the two molecules brings the two fragments of EGFP together to produce fluorescence. (D) In the negative control LMO4 is tethered to a random peptide and should not bind or produce fluorescence.

1.2.4 Previously created constructs for split EGFP screening

As part of my Honours research project, tight binding positive controls that produce fluorescence (Positive 36 and 45) and non-binding negative controls that do not fluoresce (Negative 36 and 45) were created (Figure 1.7). The positive controls (Figure 1.7B and D) expressed high levels of fluorescence, and the negative controls (Figure 1.7C and E) showed very low levels of fluorescence. These properties were determined by the analysis of the soluble fraction by bulk fluorimetry (see Section 3.1) of BL-21 (DE3) *E. coli* cells expressing the constructs.

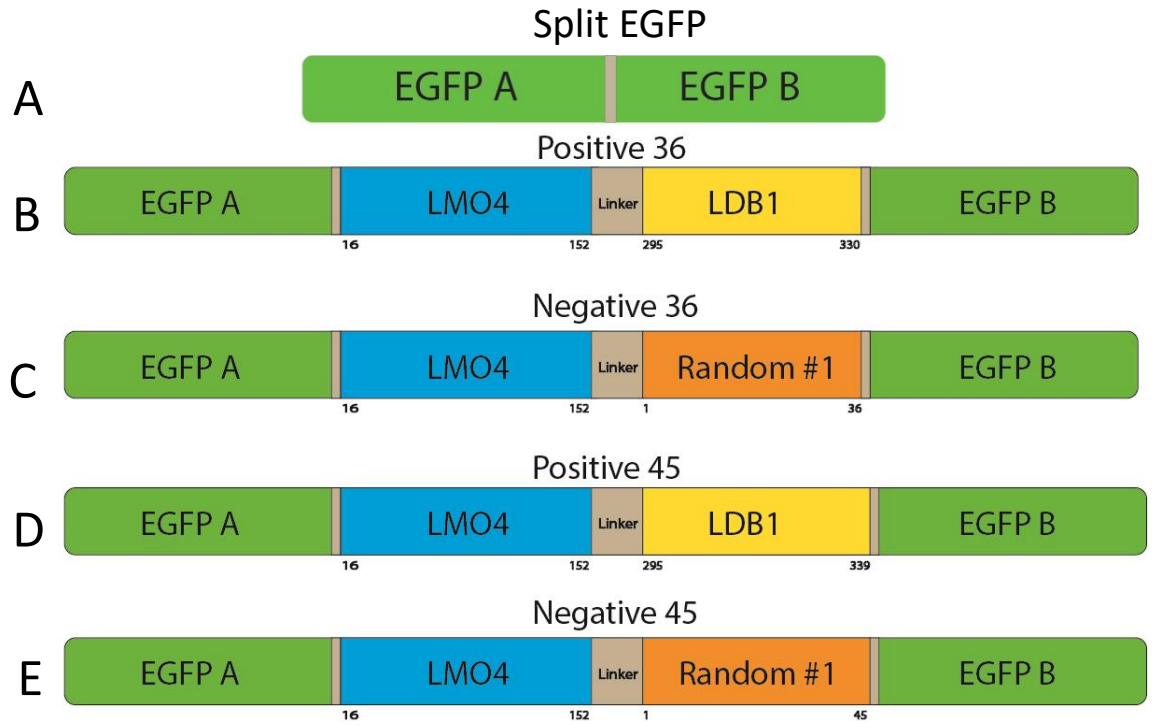


Figure 1.7: Split EGFP and split EGFP controls. (A) Empty split EGFP. For B-E the N-terminal fragment (EGFP-A) is tethered to LMO4, and the EGFP C terminal fragment (EGFP-B) is tethered to (A) a 36-residue fragment of LDB1 (residues 295-330) and is termed Positive 36 (B) a 36-residue Random peptide #1 and is termed Negative 36 (C) a 45-residue fragment of LDB1 (residues 295-339) and is termed Positive 45 (D) a Random peptide #1 45 residues (the first 36 residues are the same as Negative 36 with nine additional random residues).

Figure 1.7 is the schematic for the key spit EGFP constructs used in this research. For split EGFP alone (Figure 1.7A) there is a 6-residue flexible linker between the fragments of EGFP.

In the remaining constructs (Figure 1.7 B-E) a 3-residue linker separates each of EGFP-A and LMO4, and LDB1/Random #1 and EGFP-B, whereas an 18-residue flexible linker that contains an HRV-3C protease site separates LMO4 and LDB1/Random #1. This linker has been used with a good level of success in several related projects involving LIM-LID constructs (see section 1.1.1 for more details). The HRV-3C protease site is not required for the research described in this thesis, but it may be useful future.

1.2.5 mCherry

The original red fluorescent protein was initially discovered in the reef coral *Discosoma* (52). This protein was then substantially engineered to become dsRed and eventually mCherry (53-55). mCherry (Figure 1.8) is a red fluorescent protein that matures quickly, is photostable and monomeric; it has been widely used in biochemistry and cell biology as a fluorescent tag (55). mCherry is often used in combination with EGFP (56) because their excitation and emission spectra (Figure 1.9) are very different allowing both to be measured without interference from the other.

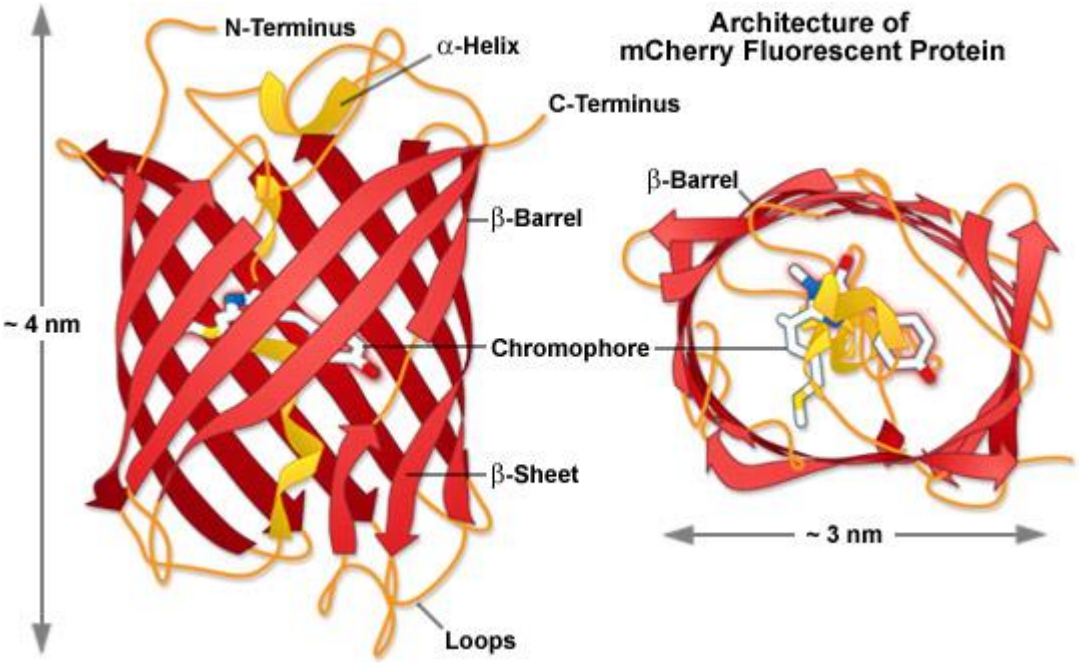


Figure 1.8: Structure of mCherry. mCherry shares a core β-barrel structure (red) with GFP, but folding and autofluoresence results in a variant fluorophore (yellow) that has different fluorescent properties. Figure taken from (57).

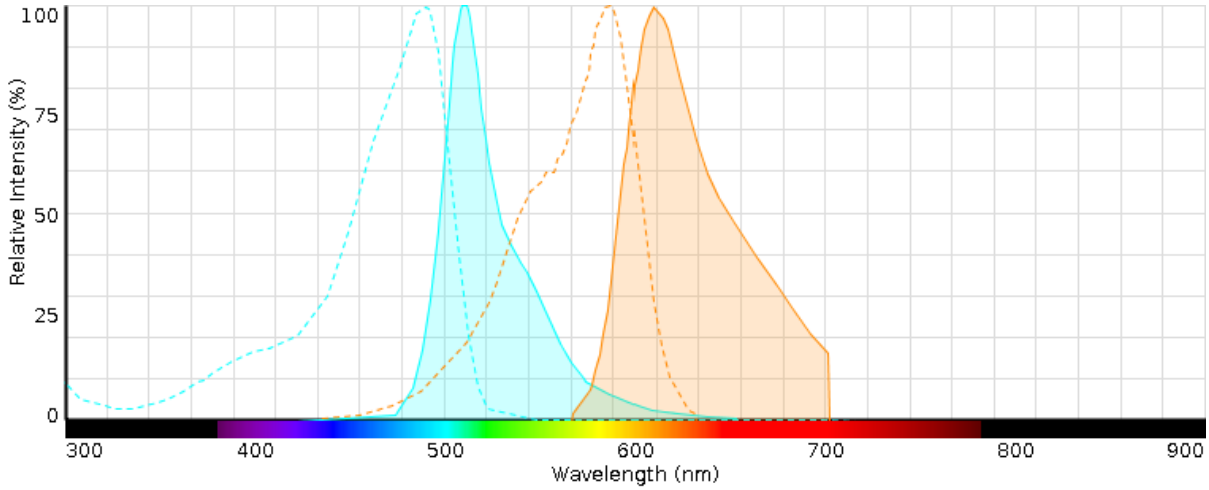


Figure 1.9: Excitation and emission spectra for EGFP and mCherry. The dotted lines for EGFP (blue) and mCherry (orange) show the excitation spectra and the shaded areas report the emission spectra. This graph was created in ThermoFisher Scientific Fluorescence Spectra Viewer.

1.3 α-helix in the split EGFP context

Isolated α-helical peptides are often intrinsically disordered and form helical structure only upon binding. Folding upon binding was previously shown to be enhanced by tethering α-helical peptide sequences to another protein or subdomain (58-60). This tethering approach is incorporated into the construct design for our split EGFP reporter (Figure 1.7).

Most studies that have generated α-helical libraries have been based on a naturally occurring helical sequence that has intrinsic affinity for a target (or targets)(61-63). There is no known α-helix that naturally binds to LMO4 that can be used as a template, but it was decided to select existing α-helices from other sources for the library design. The criteria for selection were that the α-helices must be well characterised, and have been used previously as a template in a peptide library and/or been successfully modulated. Peptides that were successfully used in other systems were considered more likely to be successful in similar experiments than untested template sequences.

1.3.1 BCL-2 family proteins

The B-cell lymphoma (BCL-2) family of proteins contains numerous well-characterised and studied α-helices. In particular the BH3 domains within these proteins are α-helices that

mediate interactions between partners. BCL-2 and its homologues (Figure 1.10) regulate apoptosis through protein-protein interactions. This family includes both pro-apoptotic and pro-survival proteins (Figure 1.10) and governs mitochondrial outer membrane permeabilisation. Interactions between these proteins determine whether cells undergo apoptosis (through mitochondrial permeablisation) or continue to survive.

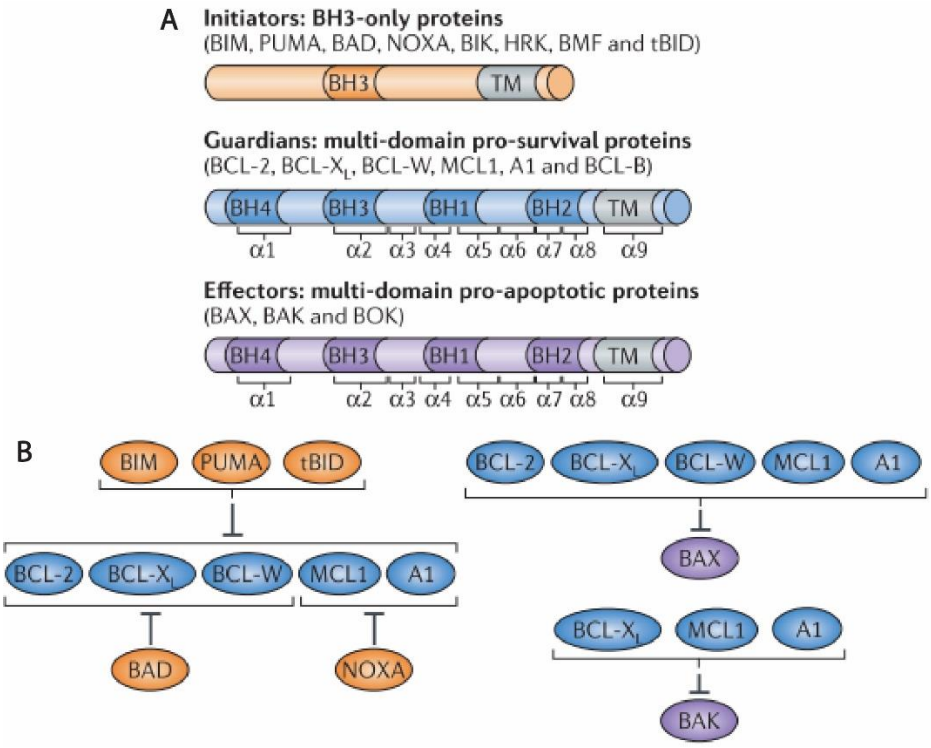


Figure 1.10: BCL-2 family of proteins. (A) BCL-2 family of proteins and protein domains. (B) BH3 α-helix mediated interactions between BCL-2 family members. Taken from (64).

BCL-2 and its homologues show sequence homology in four helical domains BH1–4 (65) that form a helix bundle. The multidomain BCL-2 proteins interact with different subsets of BH3-containing proteins, with varying affinity, to achieve pro-apoptotic or pro-survival outcomes (66, 67). In BH3-only proteins, the BH3 domain is disordered but forms an amphipathic α-helix upon binding to a common groove on multidomain BCL-2 proteins (64, 68) (Figure 1.11). BH3-only proteins BIM and PUMA bind with high affinity to all pro-survival proteins, causing

inhibition to promote apoptosis (Figure 1.10B). The other BH3-only proteins (BIK, BAD, BMF and HRK) interact only with BCL-2 and BCL-X_L with low binding affinities to weakly promote apoptosis (69).

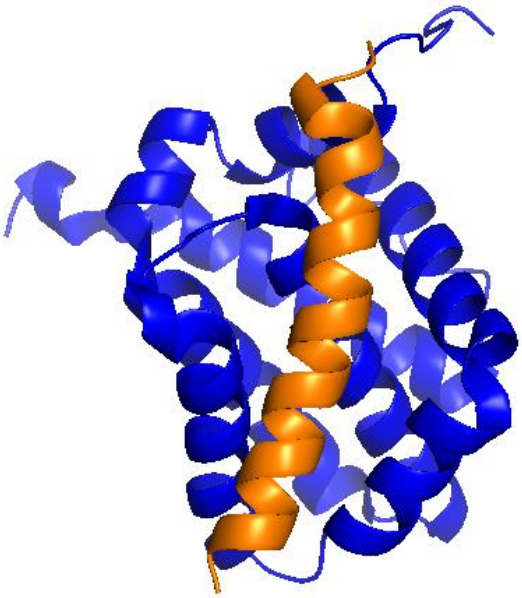


Figure 1.11: BH3 domain of Puma interacting with the major hydrophobic grove of BCL-2 protein (MCL-1). BH3 α-helix of Puma (orange) interacting with Mcl-1 (blue). Image generated in PyMOL using PDB 2ROC (70).

1.3.1.1 Developing novel α-helices based on the BH3 domain

The BH3 domain is an amphipathic α-helix containing the sequence motif A-X-X-X-A-X-X-A-B-C-X-A; where A is a hydrophobic residue, B is a smaller residue (often glycine) and C is aspartate or glutamate (71). When bound, the A and/or the B residues in the motif form the hydrophobic side of the helix that contacts partner proteins (71, 72). Many small molecule BH3 mimics and synthetic BH3-derived peptides have been generated that inhibit PPIs in the BCL-2 system (73). For example, BH3-based α-helical peptides were generated to bind with increased selectivity to various BCL-2 proteins (34). Both positive steps to improve affinity and negative steps to improve specificity were included in the design process. These helices inhibited native PPIs in fluorescence polarization and fluorescence-based Y2H assays (34, 72). Using this approach, the investigators were able to create a novel α-helical peptide inhibitor for BCL-X_L(34).

1.3.2 Other α-helices

Another naturally occurring α-helix that has been used in peptide libraries and/or has been successfully modulated to improve affinity is Son of Sevenless (SOS), which interacts with Ras (small GTPase) via an α-helix called αH (74). A rational structure-based iterative design process was deployed to improve the affinity of αH for Ras. The resultant peptide was used as the basis for design of a small molecule that mimicked the peptide structure (74).

α-helical peptides that bind to VEGF receptors (VEGFR) have been taken from the N-terminus of vascular endothelial growth factor (VEGF) and Vammin (a form of VEGF found in snake venom)(35). In silico approaches were used to develop peptide libraries designed to have improved specificity and affinity to VEGFRs. The peptide library was created by solid-phase synthesis and screened for binding to VEGFRs. A peptide was identified that bound to VEGFR-1 with greater affinity than VEGF (35). A similar in silico and solid phase synthesis selection approach was used to examine VEGF and receptor Flt-1 as well as p53 and MDM2 (75).

Other helical peptide candidates are an α-helix from the p53 tumour suppressor that binds to E3 ubiquitin-protein ligase (MDM2) (76). An α-helix from Avian Pancreatic Polypeptide that binds to a small DNA protein PPBR4, where helices of increased binding affinity were generated and selected using phage display (59). Finally, an α-helix from HIV-1 envelope protein GP41. Using rational design, the C-terminal α-helical region of GP41 was used in the development of the drug enfuvirtide, which binds to the different region of GP41 and blocks HIV-1 entry into cells (63, 77).

1.3.3 Using an α-helix as a template

One of the benefits of using α-helices whose structures have been solved is the ability to know which residues form the natural binding interface (Figure 1.12), and solely manipulate these residues to create a library of variants that could bind to LMO4. This is typical of the design of α-helical libraries (74-76). The approach maintains the propensity to form an α-helix whilst modulating binding properties and minimizes the size of the library as only a subset of residues need to be randomised.

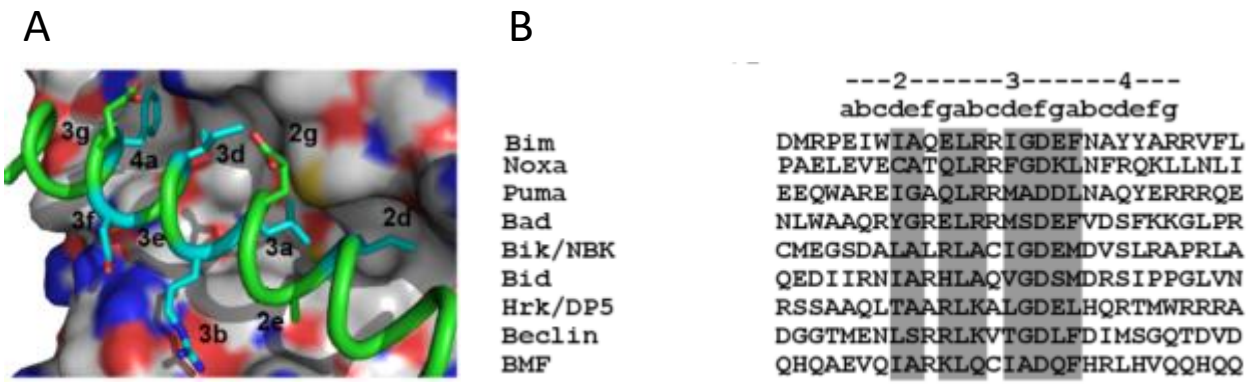


Figure 1.12: The BH3 α-helices showing interacting residues. (A) BH3 α-helix of BIM interacting with BCL-2. The interacting residues are labelled in terms of their position in the peptide sequence. (B) Amino acid sequence of BH3 domains in BCL-2 family proteins. Of the 30 residues shown, those assigned numbers and letters correspond to the main part of the helix are labelled in panel A. Figure taken from (61).

1.3.4 α-helical mimetics

α-Helical mimetics are compounds that mimic the structural configuration of α-helices but utilise various non-natural chemistries to stabilise the α-helix (78). Such mimetics may be highly helical and can be very resistant to proteolysis *in vivo*. At this stage of the project such mimetics are not useful because they cannot be readily incorporated into the EGFP reporter constructs. However, α helical mimetics are showing promise as good drug candidates, and if suitable α-helical peptide inhibitors of LMO4 can be identified, conversion into mimetics could be a viable step towards the development of drugs.

1.4 Yeast two hybrid competition assay (Y2HCA)

1.4.1 Yeast two hybrid (Y2H) assay

Yeast two-hybrid assays are used to determine if two proteins, termed the bait and the prey, will interact (79). Y2H assays have been used previously to assess possible binding partners and potential LID-based inhibitors for LMO4 (2, 19, 22, 23). For example, the prey LID protein can be fused to the transcriptional activation domain of the transcription factor GAL4, with the bait LIM protein fused to the DNA binding domain of GAL4 (Figure 1.13). An interaction is detected by activation of reporter genes, which are usually associated with yeast growth under nutritional selection conditions and/or the production of colour. Although Y2H assays are able give a qualitative assessment of binding, it can be difficult to assess the relative affinity of an interaction, and the assay cannot distinguish binding affinities that lie above or below the threshold of detection.

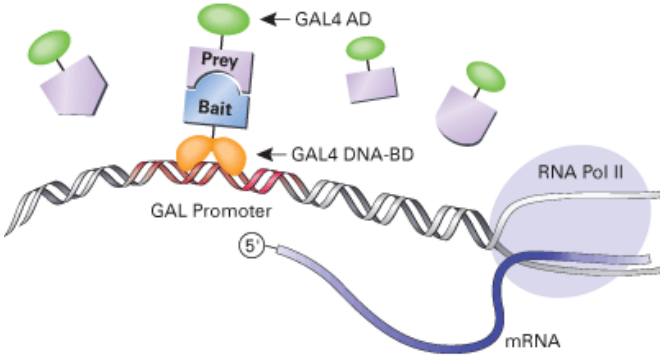


Figure 1.13: Schematic of the Y2H system. AD: Activation domain, DNA-BD: DNA binding domain. Figure from Clontech Matchmaker® Gold Yeast Two Hybrid System User Manual.

1.4.2 Yeast two hybrid competition assay

Yeast two hybrid competition assays are an extension of traditional Y2H assays that provide a means to better discern relative binding affinities for similar binding partners. In this system, instead of LMO4 alone, a tethered LMO4–LID construct is used as bait. The LID that is being assessed must 'compete off' the tethered LID peptide before it can bind LMO4 (M) (Figure 1.14).

This approach was used previously to investigate the relative strength of an intramolecular interaction in Isl1 (80). $Isl1_{LID}$ tethered to $ISL1_{LIM1+2}$ was fused to the Gal4 activation domain, whereas $LDB1_{LID}$ peptides were fused to the Gal4 binding domain (80). The $Isl1_{LID}$ and $Isl1_{LIM1+2}$ interaction was too weak to be detected in standard Y2H assays but could inhibit LDB1 binding in the context of the tethered LIM-LID complex (80).

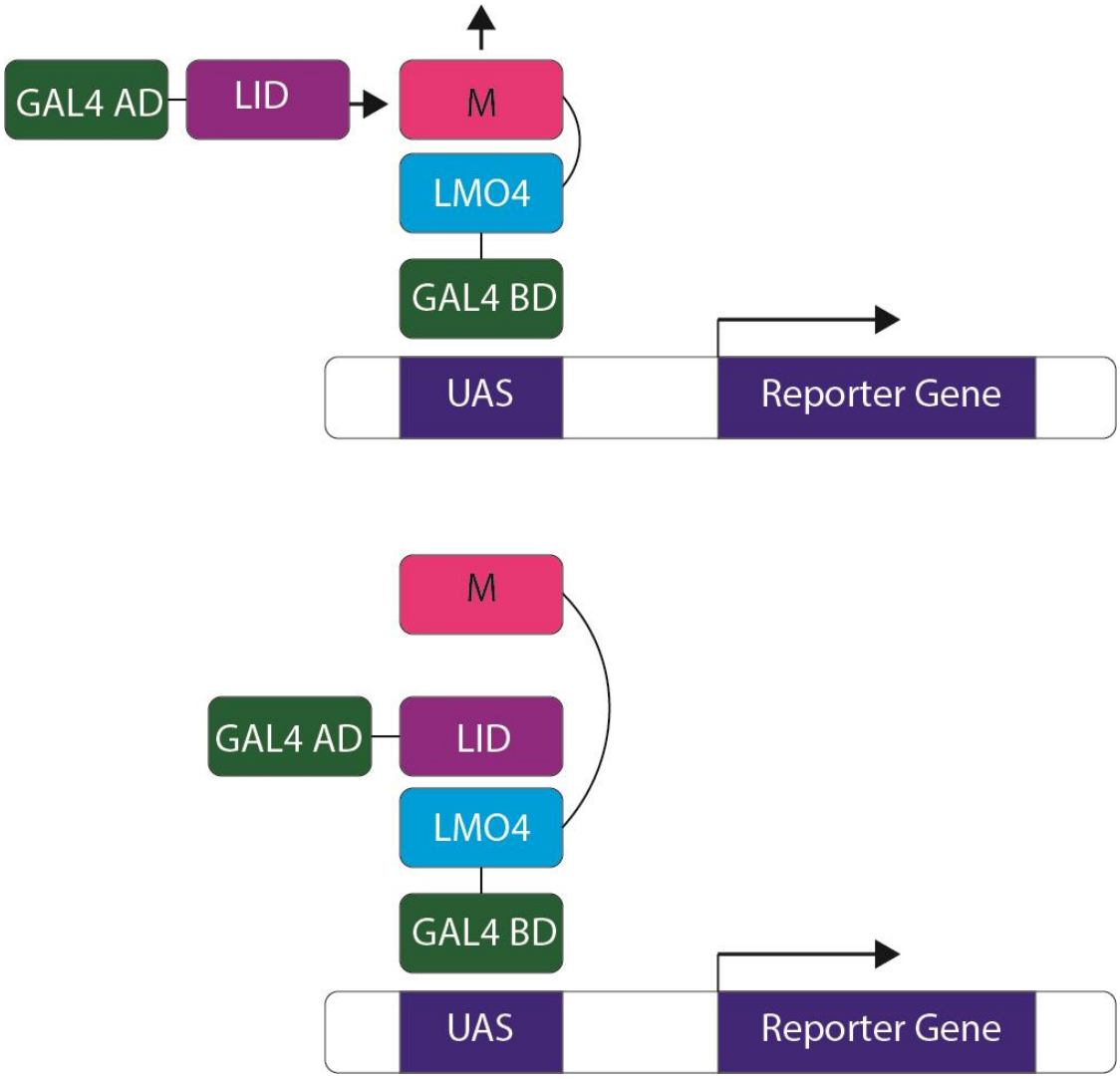


Figure 1.14: Schematic of the yeast two hybrid competition assay. In order for the LID to bind to LMO4 it must first compete off the tethered peptide (M). Then the two parts of the GAL4 are brought together to activate transcription of the reporter genes allowing the yeast to grow on nutritionally selective media, and produce a blue colour in the presence of X-α-gal.

1.4.3 Previous Development of Yeast 2 Hybrid Competition Assay for LMO4 binders

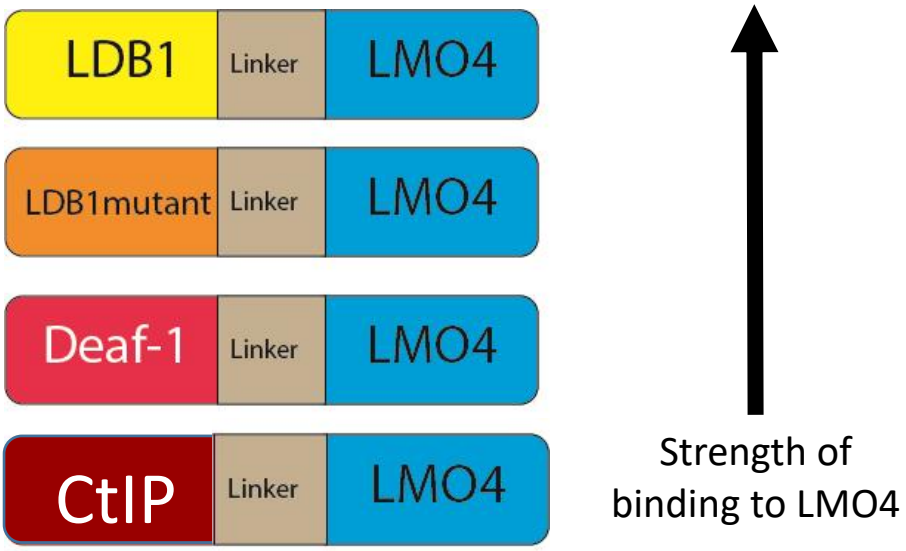


Figure 1.15: Tethered LMO4 constructs previously used in the Y2HCA

The relative binding affinities for LMO4 of for test LIDs (Section 1.1.4) were assessed by other members of the Matthews laboratory using both Y2H and Y2HCA systems. Most of these LIDs bind very well in the standard Y2H assay but can't be easily ranked in terms of binding affinities. In the Y2HCA (initially developed by P. Stokes and M. Petitmangin, unpublished data) a range of tethered constructs were utilised in which $CtIP_{LID}$ was the weakest binding partner (Figure 1.15). It was the only construct to be effectively competed off by LDB1 (positive control). The test LIDs could not compete with tethered CtIP for binding to LMO4 so could not be further assessed in the Y2CA in that form. During my Honours project research CtIP mutant constructs were designed to bind to LMO4 more weakly than CtIP and thus be capable of being competed off by weak LIDs. These constructs were cloned and preliminary experiments carried out, and are reported in Chapter 3 (Section 3.1) as an introduction to that part of the work carried out for this thesis.

1.5 Aims of thesis

The key aims of this thesis were to further develop and refine two methods: a split EGFP system providing a high throughput method for initial screening of an α-helical peptide library; and Y2HCA, an orthogonal method for validating hits and providing an assessment of the relative binding affinities of α-helical peptides in a moderate throughput manner.

Chapter 3 describes efforts to test constructs for Y2HCA that were designed during my Honours year, and to incorporate new designs to try and expand the usefulness of the Y2HCA for testing weakly binding peptides.

Chapter 4 describes efforts to further develop a split EGFP system with positive and negative controls that can detect potential LMO4 α-helical binders. Studies described in this thesis build on research that was initiated during my Honours year and focuses on generating capacity to detect helical elements that bind to a single LIM domain in LMO4, and introduces a new mCherry element for assessing cellular levels of expression.

Chapter 5 describes efforts to design the α-helical library for the split EGFP complementation system. It also describes efforts to optimise chemical transformation protocols and plasmid DNA extraction from cells to aid in the screening of the library.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals and Reagents

Table 2.1 shows all chemicals and reagents used throughout this project. Solutions were prepared in MilliQ® water (MQW) (18.2 MΩ.cm Resistivity).

Table 2.1: Chemicals and reagents used in this project

Chemical/Reagent	**Supplier**
2-log DNA ladder	New England Bio Labs (Beverly, MA)
2-β-mercaptoethanol	Sigma-Aldrich (Castle Hill, NSW)
3-Amino-1,2,4-triazole (3-AT)	Sigma-Aldrich (Castle Hill, NSW)
5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-gal)	Astral Scientific (Caringbah, NSW)
Adenine (A)	Sigma-Aldrich (Castle Hill, NSW)
Agarose	Affymetrix (Santa Clara, CA)
Ampicillin	Sigma-Aldrich (Castle Hill, NSW)
Bacterial peptone	Affymetrix (Santa Clara, CA)
Bacteriological agar	Affymetrix (Santa Clara, CA)
$CaCl_2$	Ajax Finechem (Taren Point, NSW)
Chloramphenicol	Amresco (Solon, OH)
Complete protease inhibitor cocktail tablets	Roche Applied Science (Castle Hill, NSW)
Complete supplement mixture (CSM) – ade-his-leu-trp	MP Biomedicals, LLC (Solon, OH)
Deoxyribonucleotide triphosphates (dNTPs)	Roche Applied Science (Castle Hill, NSW)
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich (Castle Hill, NSW)
Ethanol	Chem-Supply (Gillman, SA)
Ethylenediaminetetraacetic acid (EDTA)	Ajax Finechem (Taren Point, NSW)

Glucose	Univar (Auburn, NSW)
Glycerol	Chem-Supply (Gillman, SA)
HCl	Univar (Auburn, NSW)
Histidine (H)	Sigma-Aldrich (Castle Hill, NSW)
HydraGreen™ Safe DNA Dye	ACTGene Inc. (Piscataway, NJ)
Isopropyl β-D-thiogalactopyranoside (IPTG)	Progen (Darra, Qld)
KCl	Ajax Finechem (Taren Point, NSW)
KH_2PO_4	Chem-Supply (Gillman, SA)
Lithium acetate (LiAc)	Ajax Finechem (Taren Point, NSW)
Methanol	Chem-Supply (Gillman, SA)
$MgCl_2$	Ajax Finechem (Taren Point, NSW)
Mg_2SO_4	Ajax Finechem (Taren Point, NSW)
Monomeric bovine serum albumin (BSA)	New England Bio Labs (Beverly, MA)
Na_2HPO_4	Ajax Finechem (Taren Point, NSW)
NH_4Cl	Ajax Finechem (Taren Point, NSW)
NaCl	Chem-Supply (Gillman, SA)
NaOH	Chem-Supply (Gillman, SA)
Paraformaldehyde	Sigma-Aldrich (Castle Hill, NSW)
Phenylmethylsulfonyl fluoride (PMSF)	Sigma-Aldrich (Castle Hill, NSW)
Polyethylene glycol, MW 4000 (PEG)	Hampton Research (Aliso Viejo, CA)
Salmon sperm DNA	Sigma-Aldrich (Castle Hill, NSW)
Tris(hydroxymethyl)aminomethane (Tris)	Chem-Supply (Gillman, SA)
Tryptone	Amyl media (Dandenong, VIC)
Yeast extract	Amyl media (Dandenong, VIC)
Yeast nitrogen base	Difco (Detroit, MI)

2.1.2 Enzymes

All enzymes were used with the buffers provided by the manufacturer according to the manufacturer's instructions.

Table 2.2: Enzymes used in this project

Enzyme	Enzyme Commission Number (EC)	Manufacturer
BamHI	EC 3.1.21.4	New England BioLabs (Beverly, MA)
EcoRI	EC 3.1.21.4	New England BioLabs (Beverly, MA)
NdeI	EC 3.1.21.4	New England BioLabs (Beverly, MA)
HindIII	EC 3.1.21.4	New England BioLabs (Beverly, MA)
Phusion Hot Start DNA polymerase	EC 2.7.7.7	Jason Low (University of Sydney, Sydney)

2.1.3 Media

A full list of media used in this project can be found in Table 2.3. All media were prepared in MilliQ® water. For solid media, agarose was added to a concentration of 1.5% w/v. All media were sterilised by autoclaving before use, with the exception of glucose which was filter-sterilised and added to autoclaved media.

Table 2.3 Media

Media	**Components**
Luria-Bertani broth (LB)	1% w/v peptone, 0.5% w/v yeast extract, 0.5% w/v NaCl
Yeast peptone dextrose (YPD)	2% w/v tryptone, 1% w/v yeast extract, 2% w/v glucose, 0.2% w/v adenine
Synthetic dextrose (SD)	0.65% w/v yeast nitrogen base, 0.4% w/v glucose, 0.016% w/v CSM – ade-his-leu-trp,
SOB	2% w/v Tryptone, 0.5% w/v Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$
SOC	2% w/v Tryptone, 0.5% w/v Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10 mM $MgSO_4$, 20mM Glucose
YT	2% w/v Tryptone, 0.5% w/v Yeast Extract, 0.1 M NaCl

Additional nutrients were added to SD media as required

2.1.4 Oligonucleotides and plasmids

Single-stranded oligonucleotides used for cloning were synthesised by IDT (Coralville, IA). See Appendix A for a full list of sequences used.

The yeast plasmids used throughout this project were pGAD10 (*LEU2, amp*^{R}) (BD Bioscienes Clontech, CA), and pGBT9 (*TRP1, amp*^{R}*p*) (M. Crossley, University of Sydney). pGBT9 is a modified version of NpGBT9 (Clontech, Mountain View, CA) in which the MCS has been modified such that the *BamHI* and *EcoRI* sites have been reversed.

Bacterial expression plasmids used for split EGFP system were produced using a modified pET-15b vector (Novagen, Darmstadt, DEU) termed pET-15be (T. Ohashi, Duke University Medical Centre, NC, USA). The vector contains the sequence for an N-terminal His-tag, a thrombin cleavage site (LVPRGS) sequence and a multiple cloning site (MCS) spanning 12 amino acids (Figure 2.4).

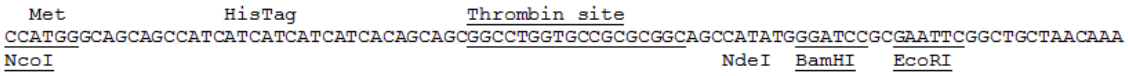


Figure 2.4: Multiple cloning site of pET15$_{BE}$.

Bacterial expression plasmids used for split EGFP and mCherry was pETDuet -1 vector (Novagen, Darmstadt, DEU). pETDuet -1 is designed for the coexpression of two target genes. The vector contains two multiple cloning sites, each of which is preceded by a T7 promoter, lac operator and ribosome binding site.

2.1.5 Yeast and Bacterial Strains

Table 2.5 lists the yeast and bacterial strains used throughout this project, and their genotypes. The *E. coli* strain DH5α was used for propagating plasmid DNA, whereas recombinant protein overexpression was conducted in the *E. coli* strains BL-21(DE3), BL-21 (DE3) Gold and Rosetta™ 2(DE3) pLysS (hereafter referred to as Rosetta 2). Bacterial strains were maintained in sterile LB broth or LB agar. *Saccharomyces cerevisiae* yeast strains were maintained in sterile liquid YPD or YPD agar.

Table 2.5 Organism strains

Organism Strain	Genotype
DH5α cells (Betheseda Research Laboratories, Gaithersburg, MD, USA)	*supE44, ΔlacU169,[Φ80lacZΔMI5], hsdR17, recA1, endA1, gyrA96, thi-1, relA1;*
BL-21 (DE3) (Integrated Sciences, Willoughby, NSW)	*F- ompT, hsdSB(rB–, mB–),gal, dcm, (DE3)*
BL-21 (DE3) Gold (Agilent, CA, USA)	*F– ompT hsdS(rB– mB–) dcm+ Tetr gal λ(DE3) endA Hte*
Rosetta™2 (Merck, Dermstadt, Germany)	*F– ompT, hsdS$_B$(r$_B^-$ m$_B^-$), gal, dcm,RARE2 (Cam^R), (DE3)*
AH109 (BD Bioscienes Clontech, CA)	*MATa, trp1-901, leu2-3, 112, ura3-52, his3 200, gal4Δ, gal80Δ, LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3, GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2 URA3::MEL1$_{UAS}$-MEL1$_{TATA}$-LacZ MEL1*

2.2 Cloning of CtIP mutant constructs and split EGFP constructs

The following inserts were generated by PCR amplification

- M3-LMO4
- M668A-LMO4
- D669A-LMO4
- V670A-LMO4

- LDB1 318
- LDB1 318 Mutant
- LDB1 315
- LDB1 315 Mutant
- mCherry
- BAD
- GP41
- PUMA

PCR amplification of all inserts was carried out using Phusion DNA polymerase (1 U) in 20 mM Tris (pH 8.8), 10 mM KCl, 10 mM $(NH_4)SO_4$, 2 mM $MgSO_4$, 1% (v/v) TritonX-100, 1 mg/mL BSA, 0.1 mM dNTPs, 400 nM primers and 50 ng of template DNA. The PCR cycle involved denaturation at 95 °C for one min followed by 30 cycles of denaturation (95 °C for 1 min), annealing (55 °C for 1 min) and extension (72 °C for 1 min) and a further extension step at 72 °C for 10 min. PCR products were run on a 2% agarose gel in TAE buffer (Tris-acetate buffer; 40 mM Tris, 40 mM acetic acid, 1 mM EDTA), stained with ethidium bromide (167 µl/L) or hydra green(25 µl/L), the bands excised and the DNA purified using a gel-extraction kit according to the manufacturer's instructions (Bioline, Alexandria, NSW). Overlap extension PCR (81) was used to fuse the fragments and amplify the fused DNA. Initially the reaction was prepared as above, without primers but including purified PCR products. Five cycles of denaturation (95 °C for 1.5 min), annealing (55 °C for 50 s) and extension (72 °C for 1 min), were carried out, plus a final extension step at 72 °C for 10 min. An aliquot of this 5-cycle reaction was added to a new PCR reaction, including end-primers, followed by a round of conventional 30-cycle PCR to amplify the full-length fragment.

The PCR inserts were cloned into the target vector (pGBT9, pET15be, or pETDuet) using Gibson Assembly (82). Fragments were amplified by PCR with primers that add a 30 bp overlap with the flanking regions of the destination vector, and the products were run on a 2% agarose gel, bands excised and purified. A Gibson Assembly reaction was set up combining this insert with vector linearized by a restriction digest with *BamHI* and *EcoRI* for pGBT9 and pET15be or *NdeI* and *XhoI* for pETDUET, at a 1:3 and 1:6 vector-to-insert molar ratio. The reaction mixture was made up to 5 µL with MQW and 5 µL of Gibson Assembly® Master Mix

(T4 DNA ligase, T5 exonuclease, Phusion hot start DNA polymerase and recommended buffer) was added. The reaction mixture was incubated at 50 °C for 1 h before being DH5α cells were transformed.

The presence of the correct insert was assessed by restriction digestion with *HindIII* for pGBT9 constructs, *HindIII* and *NdeI* for pET15be constructs. *BamHI and HindIII* for pETDuet MCS 1 and *HindIII* and *AvrII* for pETDuet MCS 2. All constructs were confirmed by DNA sequencing conducted by the Australian Genome Research Facility (Westmead, NSW).

2.3 Experimental methods for Yeast

2.3.1 Preparation of competent yeast

Yeast (AH109) were grown in rich YPD media at 30 °C overnight, with shaking at 200 rpm. Cells were then pelleted by centrifugation at room temperature, 5 min, 1000 × g. Cell pellets were resuspended in 25 mL sterile MQW and combined, before being centrifuged again at 1000 × g for 5 min. Supernatant was removed, and cells resuspended in buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 100 mM LiAc (pH 7.5)) in preparation for transformations.

2.3.2 Transformation of competent yeast

A 100 µL aliquot of competent yeast was added to plasmid DNA (approximately 200 ng of pGBT9 and pGAD10 plasmid) and 0.1 mg carrier salmon sperm DNA. Sterile PEG/LiAc solution (600 µL; 40% w / v PEG 4000, 100 mM LiAc, 10 mM Tris, 1 mM EDTA, pH 8.0) was added and the solution vortexed at low speed for 5 s. This mixture was then incubated for 30 min at 30 °C, shaking at 200 rpm, following which DMSO (70 µL) was added. Cells were then heat shocked at 42 °C for 15 min, with gentle mixing every 5 min. Cells were then chilled on ice for 2 min. The mixture was then centrifuged at 12000 x g for 5 s in a microfuge, and the supernatant removed. The cell pellet was resuspended in resuspension buffer (300 µL; 10 mM Tris (pH 8.0), 1 mM EDTA). This mixture was inoculated onto selective media (SD-LW, SD media with no leucine or tryptophan supplementation), for co- transformations. Plates were incubated for 72 h at 30 °C.

2.3.3 Yeast 2 Hybrid Spot Tests

Transformed yeast were grown overnight with shaking in rich SD media (1 mL, 30 °C, 200 rpm). Cell cultures were then normalised using the optical density of the solution at 600 nm with a 1 cm pathlength (OD600). Two 1-in-10 serial dilutions were prepared from the normalised cell suspension, resulting in solutions with OD600s of 0.2, 0.02 and 0.002. Aliquots of each sample were then spotted onto plates at each selection condition, as well as growth control plates made from SD-LW. Selection conditions used included, from least stringent to most stringent:

- SD-LW-H+0.5 mM 3-AT – SD-LW-H supplemented with 0.5 mM 3-AT.
- SD-LW-H+5 mM 3-AT – SD-LW-H supplemented with 5 mM 3-AT
- SD-LW-H-A – SD-LW with no histidine and no adenine supplementation.

Exact components of these media are listed in Appendix D.

2. 4 Production of split EGFP proteins

2.4.1 Super Competent BL-21 (DE3) Cells

BL-21 (DE3) cells were grown in 5 ml of YT media at 37 °C overnight, with shaking at 200 rpm. Cells were added to 100ml YT media at 37 °C, 200 rpm for 2 h. Cells were then pelleted by centrifugation at 5 min, 3000 × g at 4 °C. Cell pellets were resuspended in 30 mL sterile TFBI (30 mM KAcetate, 50 mM $MnCl_2$, 100 mM KCl, 10 mM $CaCl_2$, 15% Glycerol (pH 5.8)) and incubated at 4°C for 1.5 h. Cells were centrifuged again at 1000 × g for 15 min, 4 °C. Supernatant was removed, and cells resuspended in TFBII (10 mM NaMOPS (pH 7.0), 75 mM $CaCl_2$, 10 mM KCl, 15% Glycerol).

2.4.2 Bacterial transformations

2.4.2.1 Method for DH5α and Rosetta 2 transformations

Plasmid DNA (50–500 ng) was added to KCM buffer (10–50 µL; 0.1 M KCl, 30 mM $CaCl_2$, 50 mM $MgCl_2$). Resuspended competent *E. coli* cells were then added (10–50 µL). After 30 min at 4 °C, this mixture was then heat shocked at 42 °C for 90 s. LB media (200 µL)

was then added and the solution incubated for 1 h with shaking (37 °C, 150 rpm), before the solution was spread onto LB plates containing appropriate selection conditions.

2.4.2.2 Methods for super competent BL-21 (DE3) cells transformation.

Table 2.5 Methods for super competent BL-21 (DE3) transformation

Method No.	Procedure
1	50-300 ng DNA in 20-50 µl super competent cells, 5-10 min on ice, 42 °C heat shock for 90 s, 30 s on ice, 1 h recovery with LB, SOC or SOB media at 37 °C, 200 rpm. Plated onto LB plates containing Ampicillin (100 µg/mL).
2	50-300 ng DNA in 20-50 µl super competent cells, 30 min on ice, 42 °C heat shock for 60 s, 30 min on ice, 1.5 h recovery with LB, SOC or SOB media at 37 °C, 200 rpm. Plated onto LB plates containing Ampicillin (100 µg/mL).
3	0.005-300 ng DNA in 20-50 µl super competent cells, 30 min on ice, 42 °C heat shock for 60 s, 2 min on ice, 1 h recovery with LB, SOC or SOB media at 37 °C, 200 rpm. Plated onto LB plates containing Ampicillin (100 µg/mL).

The exact amount of DNA, volume of cells and recovery media are noted in Appendix A. Method 1 with 20 µL cells, 200ng DNA and LB recovery media is the method used for all BL-21(DE3) and BL-21 (DE3) Gold transformations unless explicitly stated.

2.4.3 Protein Overexpression

BL-21(DE3) cells were transformed by the spEGFP fusion construct plasmids. Single colonies were used to inoculate 5ml of LB broth supplemented with 100 µg/mL ampicillin. The inoculum was grown with shaking overnight (37 °C, 200 rpm). These cultures were used to inoculate fresh media (10 ml) at a concentration corresponding to an OD600 of 0.05. The culture was grown at 37 °C, with shaking at 150 rpm until an OD600 of 0.6–0.8 was reached. At this point, IPTG was added to a concentration of 1 mM to induce protein overexpression. Cultures were then allowed to continue growing according to Table 2.6. Cells were harvested through centrifugation at 5000 × g for 5 min.

Table 2.6: Growth Temperatures and Times

Temperature (^{o}C)	Growth time after induction
25	Overnight (12 h)
20	Overnight (16 h)

2.5 Fluorescence

2.5.1 Cell Fixing

After the protein overexpression protocol cells were fixed using paraformaldehyde. Cells were resuspended in PBS (10 ml) (Table 2.7) then centrifuged at 5000 × g for 5 min, resuspended in 4% Paraformaldehyde and incubated in the dark at room temperature for 10 min. Cells were washed in PBS (10 ml) before being finally resuspended in PBS (1 ml).

Table 2.7 Buffers

Buffer	**Components**
Phosphate-buffered saline (PBS)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4
4% Paraformaldehyde	4% w/v paraformaldehyde in PBS
Flow Cytometer Buffer	0.5 mM EDTA in PBS

2.5.2 Plate Fluorimetry

Fluorescence data were collected using the Tecan infinite m1000 plate reader (Tecan Austria GmbH). Whole cells samples were prepared at an OD600 of 0.3 in PBS. Cells were read in a Costar® black flat bottom 96 well plate. Plate was read at 4 °C, with cells shaken prior to fluorescence read. Data were recorded using excitation wavelengths of 478-90nm and emission wavelengths of 504-16 nm for EGFP, or excitation wavelengths of 581-93 nm and emission wavelengths of 604-16 nm for mCherry, and 400 v gain setting. The data were captured using i-control software.

2.5.3 Fluorescence Spectroscopy

An Olympus BX51 Microscope (Olympus, New York) with a reflected fluorescence system was used to image cells by fluorescent microscopy. Cells were resuspended in PBS buffer at an OD600 of 0.3. Other components included a Mercury Burner (U-RFL-T), and F- view monochrome fluorescence camera with FITC (495 nm/515 nm [#31001])) Chroma filter. Micrographs were captured using AnalySIS LS Starter (Olympus Soft Imaging Systems, ver. 2.8).

2.5.4 Flow Cytometry

Flow cytometry data were collected using a BD Accuri C6 plus desktop flow cytometer (BD Biosciences, California) and BD Accuri software. Samples of 1 x 10^6 fixed cells in flow cytometer buffer were run through the flow cytometer at room temperature. The filters used had an excitation range of 470-90 nm and emission wavelengths of 500-20 nm for EGFP, or an excitation range of 570-90 nm and emission wavelengths of 600-20 nm for mCherry. Flow Cytometry data were analysed using FlowJo V10 (FlowJo LLC, Oregon).

Chapter 3: Yeast Two Hybrid Competition Assay

Previously, as part of an Honours research project (G. McClymont, Honours Thesis in Biochemistry, The University of Sydney, 2016) Y2HCA experiments with a range of LID-LMO4 tethered constructs (Section 1.3.2) had been developed to assess the relative binding affinities of potential peptide inhibitors of LMO4. In the broader context of this thesis the Y2HCA acts as an orthogonal validation tool to confirm that peptides identified by the split EGFP complementation assay are LMO4 binders, as well as to provide a qualitative assessment of their binding strength. It is expected that the hits from an initial screen will be very weak binders, so the Y2HCA needs to be sufficiently sensitive to detect and assess differences between weak binders.

The aim of my Honours project was to expand the range of Y2HCA tethered constructs to include weaker LIDs tethered to LMO4 that would create a more sensitive assay capable of detecting and assessing the binding of weakly interacting peptides. The weakly binding LIDs chosen were based on $CtIP_{LID}$ (itself the naturally occurring LID with the weakest known binding affinity for LMO4) with mutations to further weaken the LID-LIM interaction. I generated the triple alanine mutations of CtIP, M1 and M2 (Figure 3.1), M1 encompasses the mutations M668A, D669A and V670A. M2 encompasses the mutations T671A, V672A and I673A. These were two of a series of triple alanine mutants designed to probe the LMO4-binding interface of $CtIP_{LID}$ (22, 83). Preliminary data suggested the resultant peptides bound LMO4 so weakly that they were easily displaced by all LIDs tested. However, the results were inconclusive as the experiment was only conducted once and there were issues with the controls affecting the validity of that single experiment.

Figure 3.1: CtIP triple alanine mutant constructs M1 and M2. M1 and M2 are tethered at the C-terminus via an 11-residue flexible linker to the N-terminus of LMO4. These constructs are both sub-cloned into the plasmid pGBT9.

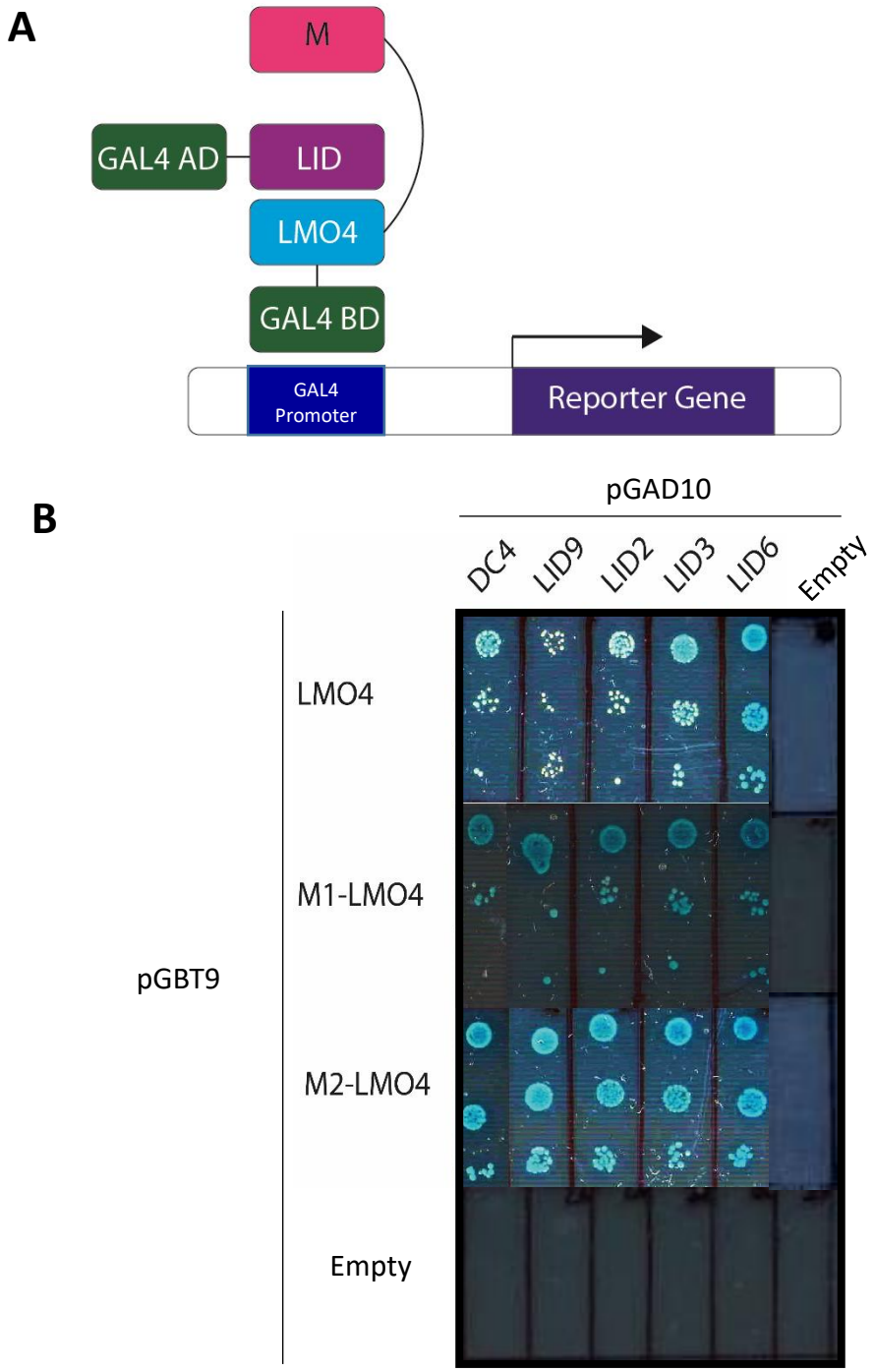


Figure 3.2 Y2HCA for M1-LMO4 and M2-LMO4 against test LIDs. (A) Y2HCA schematic (B) AH109 Yeast was transformed with pGBT9/pGAD10 plasmids containing the indicated constructs and spotted onto High stringency selection (SD-LWHA+X-α-gal) agar as serial (1:10) dilutions (top to bottom). LMO4 in pGBT9 against the test LIDs in pGAD10 are positive controls and the empty plasmids in both pGBT9 and pGAD10 are negative controls.

To begin research for this thesis, the Y2HCA experiments with the previously generated constructs (Figure 3.1) were repeated (Figure 3.2). The top panel in Figure 3.12B reports the direct interaction between LMO4 and the test LID. Robust yeast growth and blue colour indicate a strong interaction for with the test LIDs except LID9, which showed less yeast

growth. The top panel in Figure 3.2 shows a set of positive controls for this experiment as LMO4 was previously demonstrated to bind to all the test LIDs (M. Petitmangin, unpublished data). The subsequent panels report the ability of the same test LIDs to bind M1-LMO4 and M2-LMO4, respectively. In these cases the test LID must successfully compete with M1 and M2 for binding to LMO4. For M1-LMO4 there is a solid blue circle reporting robust yeast growth and α-galactosidase activity for the first dilution spot indicating that M1 has been displaced by LMO4. There are a few spots present for the next two dilution spots. M2 was also displaced by LMO4 as evidenced by the solid blue circles for the first dilution point (same as M1) but then an almost complete circle of growth for the 2nd dilution spot and spots for the final dilution spot. It was noted that compared to LMO4 alone, and to a lesser extent M1-LMO4, M2-LMO4 shows a greater level of blue yeast. However, all the test LIDs appeared to bind equally well to M1-LMO4 or to M2-LMO4. These observations indicate that the M1 and M2 mutant constructs bind LMO4 too weakly to be useful in the Y2HCA as LIDs are required that bind LMO4 weakly enough to be displaced by some test LIDs but strong enough to resist others. Since neither M1 no M2 fulfil the requirements for this system, new constructs were needed. The development and testing of such alternative constructs follows.

3.1 Designing the new CtIP mutants

M1 and M2 are consecutive triple alanine mutants of CtIP where the mutations lie in the region of CtIP (664-675) that contacts the hydrophobic binding pocket in $LMO4_{LIM1}$. These were previously shown to abrogate binding in Y2H assays (22). As these mutants bound LMO4 too weakly to be of use in the Y2HCA, new mutants were chosen for testing. The first is a triple alanine mutation D674A/T675A/K676A (M3) that was previously shown to have a smaller effect on LMO4-$CtIP_{LID}$ binding (22, 83) A further three single mutants (M668A, D669A, and V670A) were designed that should have a smaller effect on LMO4-binding than the original triple mutants (M1: M668A/D669A/V670A and M2 T671A/V672A/I673A and M2).

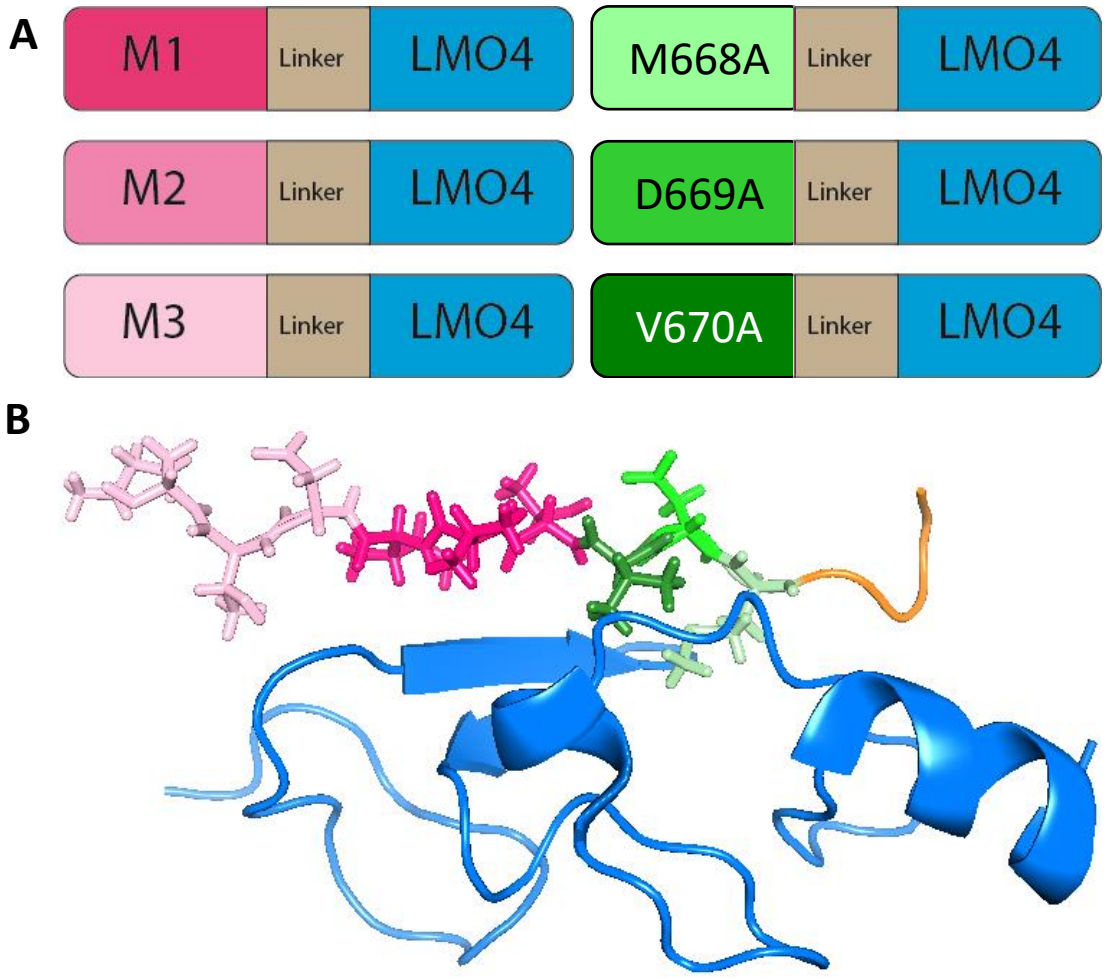


Figure 3.3: CtIP Mutant Constructs. (A) Schematic representation of previous and new CtIP mutant constructs tethered to LMO4 by a flexible 11-residue linker. (B) Structure showing LMO4 (blue, ribbon structure) binding to CtIP (sticks and ribbon structure) with residues mutated to alanine M3 (light pink), residues mutated in M1 (hot pink), residues mutated in M2 (green) as indicated. Residues that were mutated in the single alanine versions of M1 were V670A (dark green), D669A (bright green) and M668A (light green). The remainder of the CtIP binding region is shown in orange. Image generated in PyMOL from PDB 2L4Z (22).

The triple alanine mutant M3 was sub-cloned into a pGBT9 plasmid tethered to LMO4 (Figure 3.3 A) using Gibson Assembly (Section 2.2). The single alanine mutants (Figure 3.3A) were generated and cloned into pGBT9 plasmids using mutagenic primers, overlap extension and Gibson Assembly (Section 2.2).

3.2 M3, M668A, D669A and V670A in Y2HCA

The new M3 (Figure 3.5), M668A, D669A and V670A constructs (Figure 3.4) were tested in the Y2HCA.

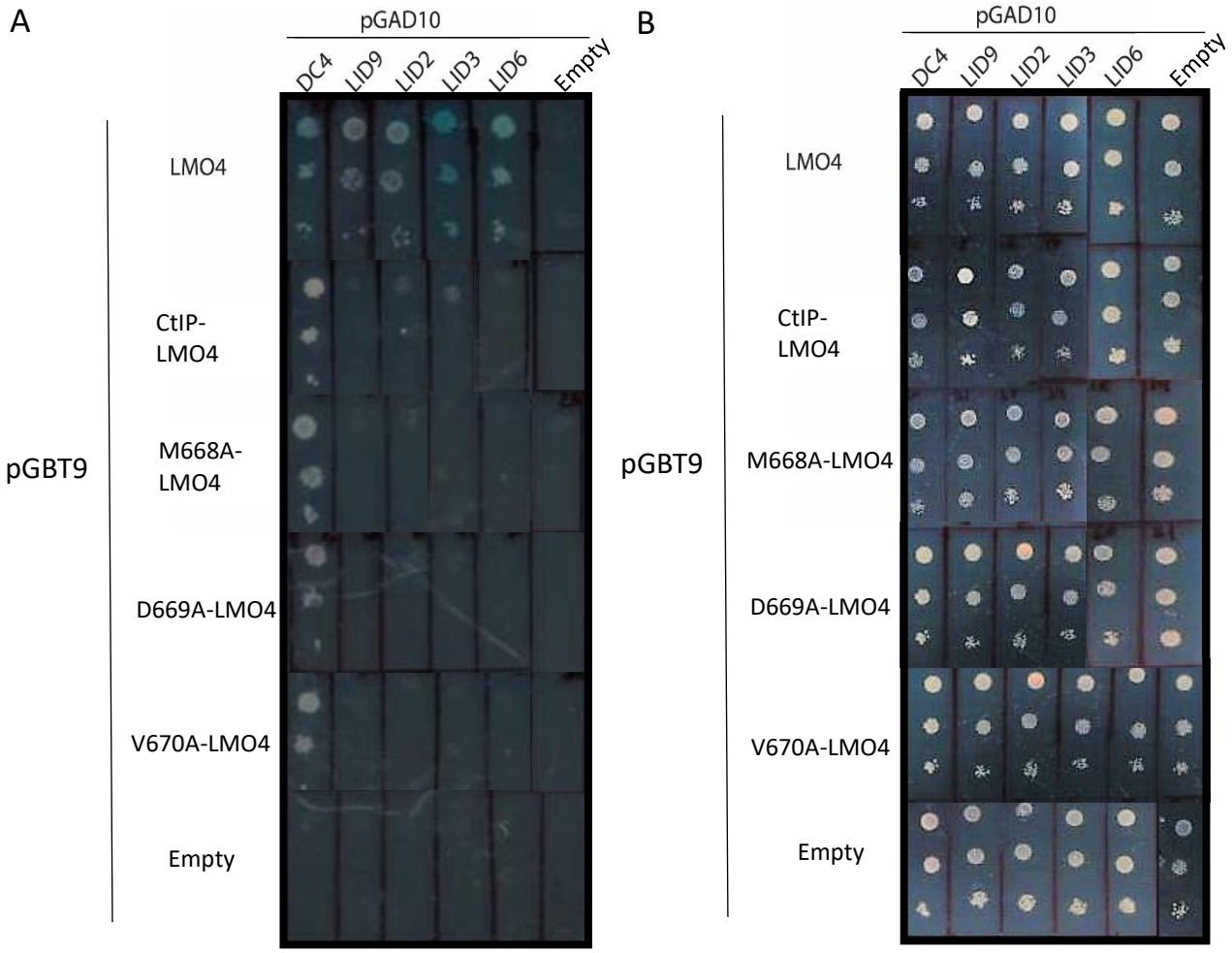


Figure 3.4 Y2HCA for CtIP-LMO4, M668A-LMO4, D669A-LMO4 and V670A-LMO4 against test LIDs. AH109 Yeast was transformed with pGBT9/pGAD10 plasmids containing the indicated constructs and spotted onto low stringency selection (SD-LWH+0.5 mM 3AT+X-α-gal) agar as serial (1:10) dilutions (top to bottom). (B) Transformation controls comprise AH109 Yeast transformed with pGBT9/pGAD10 plasmids containing the indicated constructs and spotted onto control media (SD-LW) agar as serial (1:10) dilutions (top to bottom).

The data for M668A-LMO4, D669A-LMO4 and V670A-LMO4 (Figure 3.4A) showed that these constructs were indistinguishable from CtIP-LMO4. In all experiments yeast growth was present at all three dilutions for DC4 with the reporter constructs, but no significant amounts of yeast growth were seen for the other LIDS. These data suggest that DC4, but no other test LID, was able to displace all four $CtIP_{LID}$ variants in the tethered constructs. The lack of blue colour suggests a relatively weak interaction as only the one of the reporter genes (selecting

for yeast growth) is being activated, while that for α-galactosidase (to produce blue colour in the presence of α-X-gal) is not being activated. As the single alanine mutants behaved the same as CtIP in this assay they are not useful additions to the Y2HCA, as they are unlikely to provide additional information about relative binding affinities.

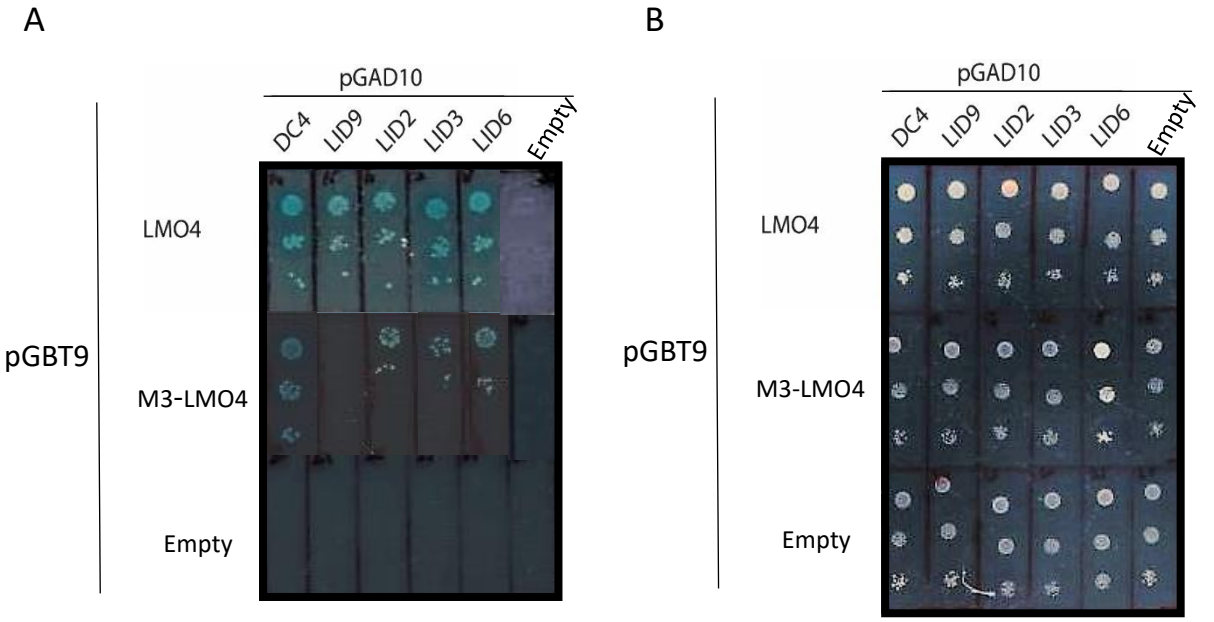


Figure 3.5 Y2HCA for M3-LMO4 against test LIDs. (A) AH109 Yeast was transformed with pGBT9/pGAD10 plasmids containing the indicated constructs and spotted onto High stringency selection (SD-LWHA+X-α-gal) agar as serial (1:10) dilutions (top to bottom). LMO4 in pGBT9 against the test LIDs is a positive control and the empty plasmids in both pGBT9 and pGAD10 are negative controls. (B) Transformation controls, comprising AH109 Yeast transformed by pGBT9/pGAD10 plasmids containing the indicated constructs and spotted onto control media (SD-LW) agar as serial (1:10) dilutions (top to bottom).

The data for M3-LMO4 (Figure 3.5A middle panel) showed robust yeast growth and blue colour for interactions with DC4, and less growth for LID2, LID3 and LID6. No growth was observed for interaction with LID9, suggesting that only this LID was unable to displace M3 to bind LMO4. The transformation controls (Figure 3.5B) indicated that lack of yeast growth on the interaction plate (Figure 3.5A) was not caused by errors in co-transformation. These data indicate that M3-LMO4 is a useful addition to the Y2HCA, as it can distinguish the relative binding affinities of weak LIDs.

3.3 Chapter Discussion

The previous CtIP mutants from my Honours research (M1-LMO4 and M2-LMO4) were tested in the Y2HCA, where they were displaced by all the test LIDs. The more robust yeast growth seen for M2-LMO4, and to a lesser extent M1-LMO4, compared to LMO4 (Figure 3.2) could represent a stronger interaction. However, Y2H experiments can be affected by factors such as stability of the constructs used. As M2-LMO4 shows more robust yeast growth with all the tests LIDs than does LMO4 alone, there may a stabilising effect associated with M2 and M1 that makes it easier for the LID to bind to LMO4. That is, a weak interaction between LMO4 and M1/M2 may be sufficient to prevent the aggregation of LMO4 making more protein available to bind the test LID in the assay. These fusion constructs could be useful for detecting the interaction of very weak interactions with LMO4 in Y2H.

Four more $CtIP_{LID}$ mutant constructs (M3-LMO4**,** M668A-LMO4, D669A-LMO4 and V670A-LMO4) were designed and tested in the Y2HCA. The triple alanine mutant was displaced by all test LIDs except one and thus could provide some estimation of relative binding affinities. The single alanine mutants were too similar to $CtIP_{LID}$, and bound too strongly to LMO4, to be displaced by the test LIDs. In the single alanine experiments there were small differences in colour production by the yeast, with some yeast growth appearing white rather than blue. The small differences in colour production alone (Figure 3.4) are not a good indicator of absolute binding affinity, as illustrated by the fact that interactions with LMO4 only (Figure 3.4A, top panel) show highest levels of blue colour for LID3 in this experiment, but the Y2HCA assay indicates that DC4 is the tightest binder, which is supported by biophysical experiments (M. Petitmangin and P. Stokes, unpublished data) to assess relative binding affinity.

This work added a new tethered construct (M3-LMO4) to the Y2HCA repertoire to help distinguish between the relative binding affinities of test LIDs. Figure 3.6 shows the new range of tethered constructs that can be used to validate and assess the relative binding affinity of an α-helix identified by the high-throughput EGFP complementation screen (Chapter 4). It may still be desirable to add additional constructs that help to distinguish weaker binders (Figure 3.6). Given that most of the triple alanine mutants were too weak

(Figure 3.2) and single alanine mutants (Figure 3.5) were too strong, double alanine mutants of $CtIP_{LID}$ might form useful CtIP variants that can be displaced by some, but not all, test LIDs.

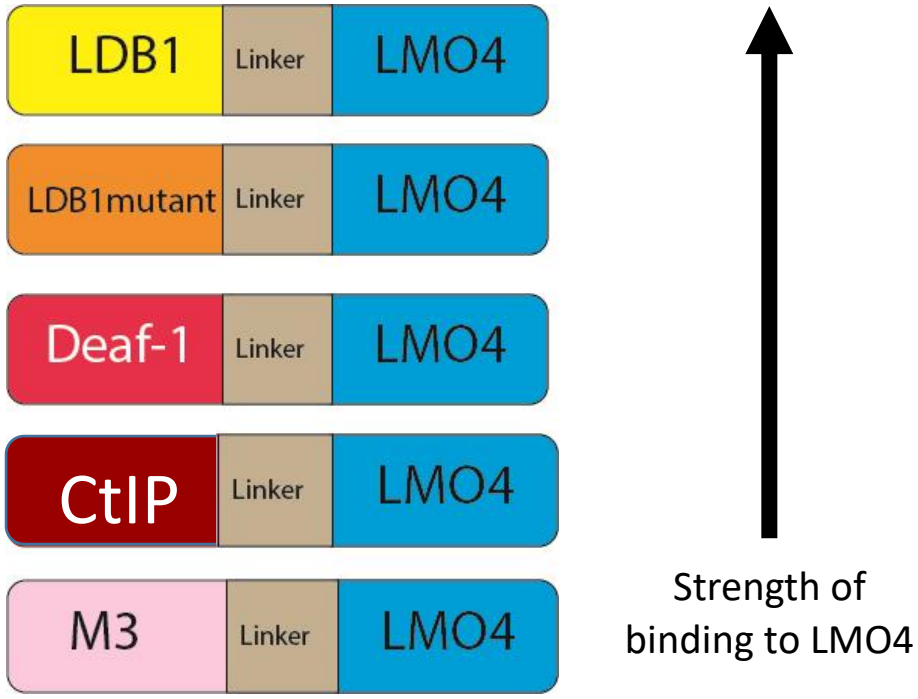


Figure 3.6: New range of tethered complexes to be used in Y2HCA

Chapter 4: High through-put EGFP complementation system

This chapter continues the development of a split EGFP complementation system to identify α-helical peptides that bind to LMO4 (Section 1.2.3). Previously, as part of an Honours research project (G. McClymont, Honours Thesis in Biochemistry, The University of Sydney, 2017) constructs for use as positive controls based on the tightly-binding natural binding partner LDB1 were developed and produced high levels of fluorescence. Constructs for use as negative controls that were created using a random peptide, the negative controls, were largely insoluble and produced very low levels of fluorescence. The controls were validated by the quantification of their whole cell fluorescence when overexpressed in *E. coli*. The system was further developed to be capable of screening peptides that vary only in their ability to bind $LMO4_{LIM1}$. An additional fluorescent protein (mCherry) was added to the system in an attempt to provide an expression control.

4.1 Previous work on a high through-put split EGFP complementation system

Prior to the start of this project, four control constructs, Positive 36, Positive 45, Negative 36 and Negative 45 (Figure 4.1) were created and use to transform BL-21 (DE3) *E. coli* cells for expression (84) (Section 1.2.4.1). Fluorescence was assessed using bulk fluorescence for the soluble fractions (Figure 4.2) and by eye for whole cells in ambient light (Figure 4.3).

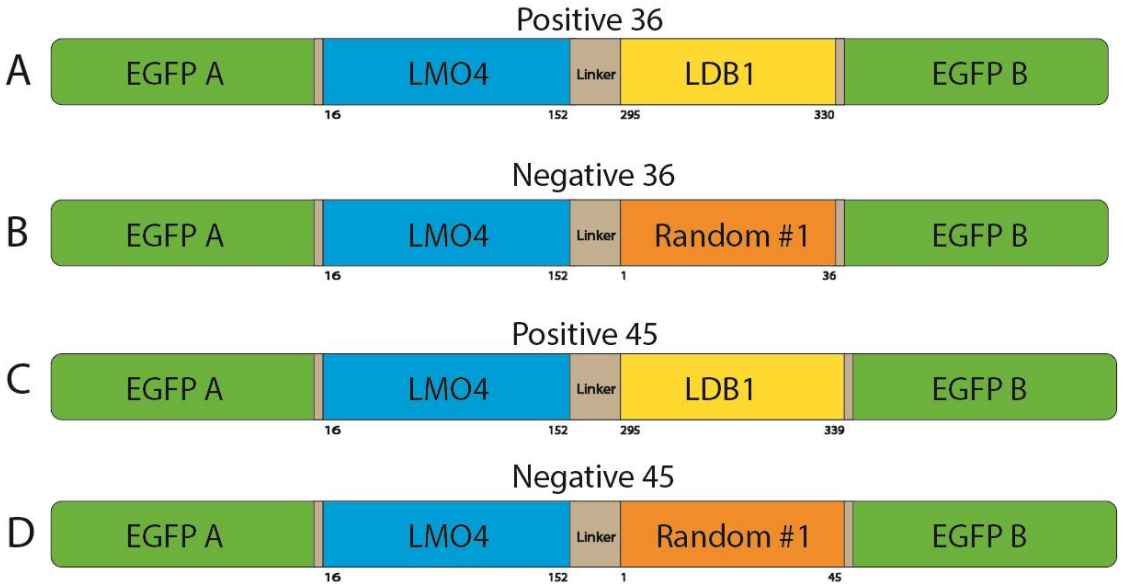


Figure 4.1 Split EGFP Control Constructs. The N-terminal fragment (EGFP-A) is tethered to LMO4, and the EGFP C terminal fragment (EGFP-B) is tethered to (A) a 36-residue fragment of LDB1 (residues 295-330) and is termed Positive 36; (B) a 36-residue Random peptide #1 and is termed Negative 36; (C) a 45-residue fragment of LDB1 (residues 295-339) and is termed Positive 45; and, (D) Random peptide #1 45 residues (the first 36 residues are the same as Negative 36 with nine additional random residues).

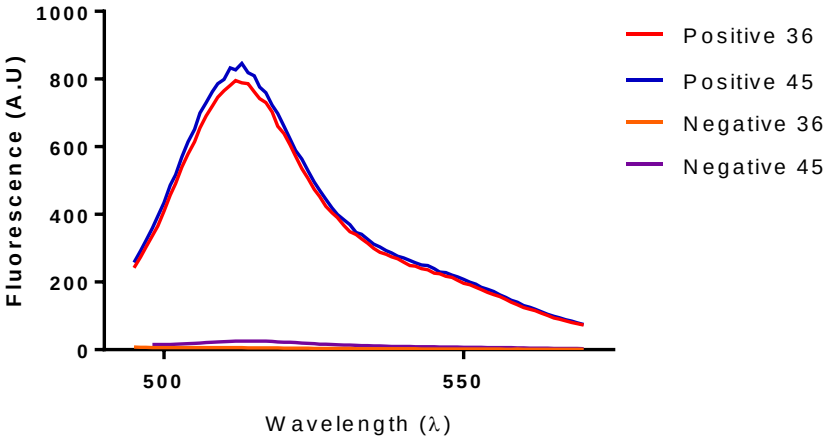


Figure 4.2: Fluorescence emission spectra of spEGFP control constructs. Positive 36, Positive 45, Negative 36 and Negative 45 were overexpressed in BL-21 (DE3) cells, 18 h 20 °C, the soluble fractions obtained, and fluorescence emission spectra measured. Fluorescence intensity at 400 v (excitation at 480 nm and emission spectra from 495–570 nm), was normalised for cell density by dividing the intensity by absorbance at 600 nm. The maximum emission peak wavelength of EGFP is 510 nm.

The normalised fluorescence spectra for Positive 36 and Positive 45 were essentially identical, with a maximum fluorescence at 510 nm, which is consistent with the theoretical maximum

for EGFP (Figure 2.2). In comparison, the soluble fractions of the negative controls, Negative 36 and Negative 45 had very low levels of fluorescence. These data demonstrated that Positive 36 and Positive 45 may by useful fluorescing positive controls, and that Negative 36 and Negative 45 should be suitable non-fluorescing negative controls.

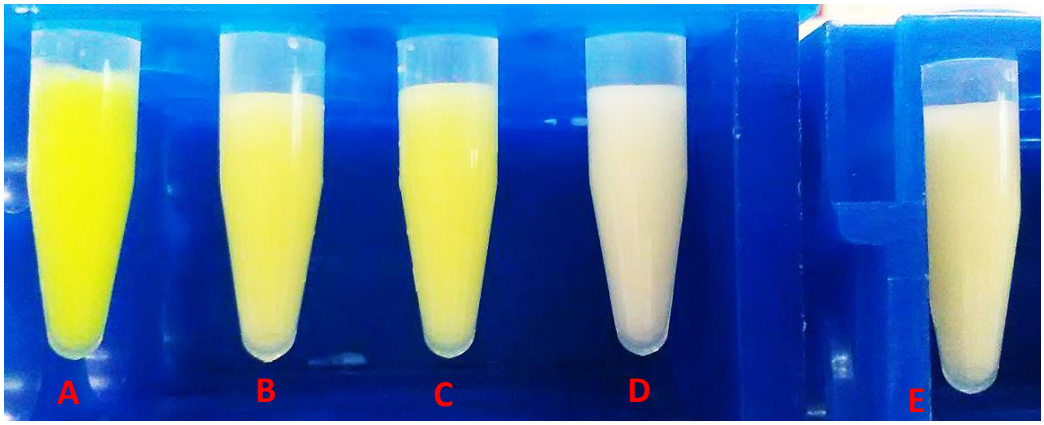


Figure 4.3: Ambient light image of resuspended whole cells. Resuspended whole cell samples of (A) split EGFP, (B) Positive 36, (C) Positive 45, (D) Negative 36 and (E) Negative 45, grown and induced in *E. coli* at 20 °C. Samples were normalised for a cell density of 1 at OD_{600nm}.

As the intention was to use whole cell fluorescence to screen for peptides binders, whole cell expression samples of these constructs were examined (Figure 4.3). The green colour visible in the whole cell resuspensions of Positive 36 and Positive 45 under ambient light conditions is a good indication that the fluorescence would be detectable in whole cell samples. The lack of green colour in the Negative 36 and Negative 45 samples was similarly a good indication of no or low fluorescence. By ambient light there appeared to be a difference between Negative 36 (Figure 4.3D) and Negative 45 (Figure 4.3E). However, according to Figure 4.2 there was no significant difference between them. These potential discrepancies were not resolved during my Honours project work.

4.2 Whole cell fluorescence analysis of previous constructs

For this thesis a first step was to reassess the whole cell fluorescence of the different constructs. This was carried out using a fluorescence plate reader, which offered several advantages. Firstly, the plate reader allows samples to be shaken before measurement. This is important for whole cells as they can quickly settle to the bottom of the well/cuvette and thus affect the reading. Even if settling occurs the sample is measured by an excitation beam

from the top of the well and the emitted light is reflected back through the body of the sample (rather than across a cuvette) minimising errors in measurement. The temperature control of the fluorometric plate reader enables rapid cooling to 4 °C, which is the desired temperature for preserving cells resuspended in PBS. Thirdly, the plate reader allows rapid assessment of 96 samples all under the same experimental conditions.

The previous experiments (Figures 4.2 and 4.3) were carried out using BL-21 (DE3) cells at 20 °C. In order to find the optimal *E. coli* strain and temperature for expression of the spEGFP-LMO4 constructs for whole cell fluorescence measurements, two cell strains and temperatures were trialled (Figure 4.4). Lower expression temperatures have been previously implicated in promoting LIM domain solubility (20) and the choice of cell type can affect levels of expression (85).

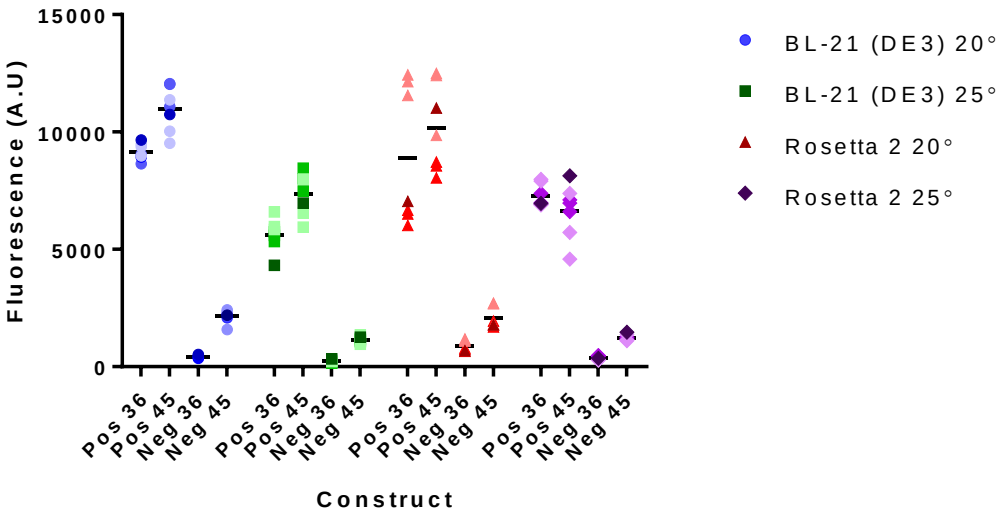


Figure 4.4: Fluorescence at 507 nm of *E. coli* cells overexpressing the LMO4 constructs. Blue circles represent BL-21 (DE3) cells at grown at 20 °C for 18 h. Green squares represent BL-21 (DE3) cells at grown at 25°c for 18 h. Red triangles represent Rosetta 2 cells at grown at 20 °C for 18 h. Purple diamonds represent Rosetta 2 cells at grown at 25 °C for 18 h. The shades of colour indicate biological replicates (N = 3), with 1 or 3 technical replicates. Bars represent the mean values of all replicates. Data were obtained using a fluorescence plate reader, excitation 482-454 nm and emission 504-517nm (Section 2.4.2). Samples were normalised for a cell density of 0.3 at OD_{600nm}.

In terms of cell type, for each construct/temperature combination, there was little overall difference in fluorescence measured. The only exception was for increased variation for one

the positive controls at 20 °C, where the fluorescence of one biological replicate was relatively high and two others relatively low. Based on these results, the BL-21 (DE3) cells were used in subsequent experiments. In terms of temperature, essentially all constructs showed higher levels of fluorescence when expressed at 20 °C rather than at 25 °C, and the overall difference in fluorescence between the negative and positive controls was higher at 20 °C.

Ideally, the combination of expression conditions and constructs should be where the negative control has the least fluorescence (background fluorescence) and the positive control has the highest fluorescence, creating a large dynamic range with which to detect improved fluorescence. Negative 36 gave the lowest fluorescence across all strains and conditions and the corresponding Positive 36 had a significantly higher fluorescence than Negative 36 for all conditions. Although the fluorescence of Positive 45 was the highest for all constructs, the levels of Negative 45 fluorescence meant that the difference of fluorescence between the pairs of negative and positive constructs is equivalent. The 36-residue constructs were considered a better option because they showed less negative control/background fluorescence. Note that the 295-330 region of LDB1 used in the Positive 36 construct is sufficient for binding to LMO4, as previously determined by Y2H and assessment of the LMO4-LDB1 structure (19). The longer peptide LDB1 Positive 45 (LDB1 295-339) contains an additional 9-residue acidic region that was at one point defined as being part of $LDB1_{LID}$, and appears to confer increased solubility LIM-LID constructs. In summary, the 36-residue variants, expressed in BL-21 (DE3) *E. coli* cells at 20 °C appear to be optimal for further development.

4.3 Validating whole cell fluorescence

Two orthogonal methods were used to validate the fluorescence in the BL-21 (DE3) cells. The first was flow cytometry, which is important as the goal is to screen libraries using FACS. Flow cytometry lacks the cell sorting capability of FACS but has the same analytical capacity and was a more convenient method for testing at this stage of the project. Although not necessary for analysis, the cells were fixed with paraformaldehyde (see section 2.1.3) prior to analysis, to prevent possible bacterial contamination of mammalian cells on the in-house instrument. The 36-residue constructs were subjected to flow cytometry alone, and together in known

ratios, to establish if different populations could be identified and separated using lasers and filters common to Flow cytometry and FACS.

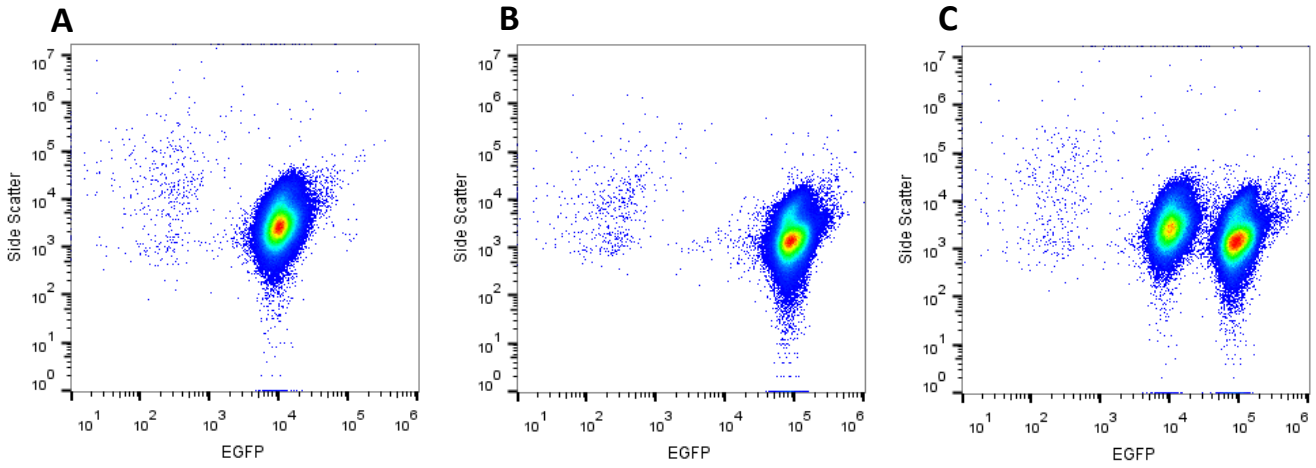


Figure 4.5: EGFP fluorescence of BL-21 (DE3) cells with Positive 36 and Negative 36. BL-21 (DE3) cells at grown at 20 °C for 18 h overexpressing either Positive 36 or Negative 36. Cells were fixed in PBS with 4% paraformaldehyde. In each experiment 2×10^5 cells were counted. (A) Negative 36 (B) Positive 36 (C) 1:1 ratio of positive 36 and negative 36. Figures were created using FlowJo.

The pseudocolour plots in Figure 4.5 use side-scatter (Y-axis) to assess cell size and EGFP fluorescence (X-axis) to characterise populations of cells. Panels A and B show Negative 36 and Positive 36, respectively. Each produces a population of cells that is fluorescent, with Positive 36 being significantly (~10-fold) more fluorescent than Negative 36. The small scatter of signals at ~10^2 on the X-axis represents a small number of cells (less than 8% of the total population) that are not expressing EGFP. In Figure 4.5 C, equal amounts of cells expressing each construct were detected, showing the ability of the flow cytometer to distinguish these populations. These data suggest that these constructs are appropriate for a FACS system.

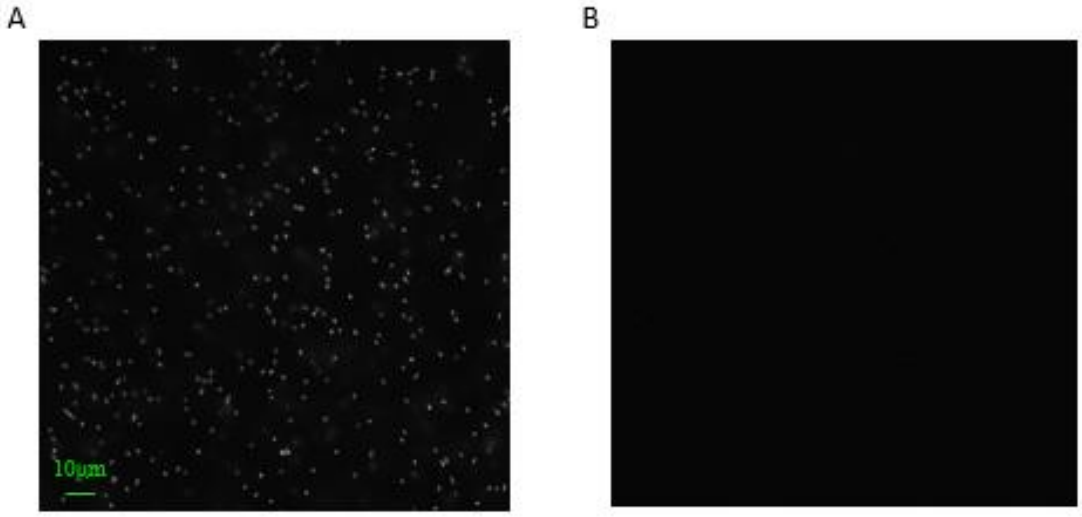


Figure 4.6: Fluorescence microscopy of Positive 36 and Negative 36. BL-21(DE3) *E. coli* cells overexpressing constructs (A) Positive 36 and (B) Negative 36 were for 18 h at 20 °C. Live cells were resuspended in PBS, normalised for cell density of 0.3 OD_{600nm} and visualised by fluorescent microscopy for 1.4 s. The scale bar is 10 µm.

Fluorescence microscopy was also used to validate the whole cell fluorescence. Figure 4.6A shows Positive 36 expressing cells. The many white dots represent fluorescent cells. Conversely, for Negative 36 (Figure 4.6B) no fluorescent cells can could be seen. This corroborates the data from fluorimetry and flow cytometry (Figures 4.4 and 4.5) demonstrating that the split EGFP system produces fluorescence that is visible in whole *E. coli* cells upon the protein-protein interaction facilitated interaction of the two split EGFP sections shown in Figure 4.1.

4.3 New split EGFP constructs

LMO4 is modular, made up of two tandem LIM domains (section 1.2.1), each of which binds to a region of LDB1. Thus, in terms of screening the library it should be possible to screen against each LIM domain separately in the context of a larger interaction interface. This is beneficial for two reasons. First it decreases the size of the region in the peptide binding target. This is beneficial as it should require shorter peptides for randomisation and screening, which reduces the size of the library. Second, by leaving one binding module of LDB1 intact, we maintain some ability to bind LMO4. Only a small positive binding event from the α-helical peptide will be required to increase fluorescence and enable detection. In contrast, a truly naïve library would most likely be populated with weak binders and it would be challenging

to find hits strong enough to detect. We decided to begin with targeting the LIM1 region of LMO4 ($LMO4_{LIM1}$), which required some modifications to the construct design.

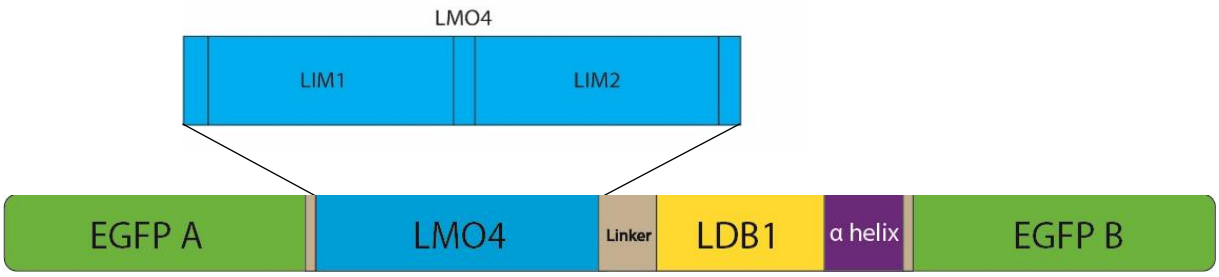


Figure 4.7 Split EGFP constructs for screening α helices against LMO4 LIM1. The N-terminal fragment (EGFP-A) is tethered to LMO4, and the EGFP C-terminal fragment (EGFP-B) is tethered to LDB1 LIM2 binding region and the α-helical peptide.

In this design the LIM2 binding portion of LDB1 (Figure 4.7) is maintained to bind $LMO4_{LIM2}$ and bring $LMO4_{LIM1}$ close to the introduced α-helix. However, an important consideration was at what point LDB1 should switch to the α-helix? The structure of LDB1 bound to LMO4, along with a series of other LIM-LID complexes and mutagenic studies, indicates that residues 299-311 of LDB1 contact LIM2, residues 318-327 contact LIM1, and residues ~312-317 form a spacer region between the two binding modules (19-21). Interestingly, residues 316-318 show some helical characteristics which was reported in terms of the secondary structure assessments from the NMR data (19) (Figure 4.8B) that might serve a helix-capping function, facilitating the folding of introduced helices. Thus, it was decided to test the replacement of two lengths of LDB1, from 315 (eliminating most of the spacer) and from 318 (leaving the putative helix capping residues in place). Note that even if the extra residues in the latter version don't promote helical structure, they could allow some flexibility between the native β-strand and the introduced helix, thus eliminating steric hindrance and promoting helix formation. Random peptide #2 and #3 are different random peptides generated using Sequence Manipulation Suite (86). The use of 3 different random peptides in the constructs so far is to ensure that data for the constructs with random peptides are not biased by a specific randomly generated sequence.

If LDB1-$LMO4_{LIM2}$ binding is too high in the context of these constructs, background fluorescence could be too high to detect positive binding helices. To control for this possibility,

a V303A mutation was made to the LIM2 binding module of LDB1, as this mutation had been previously shown to decrease the affinity of LDB1 for $LMO4_{LIM2}$ (19).

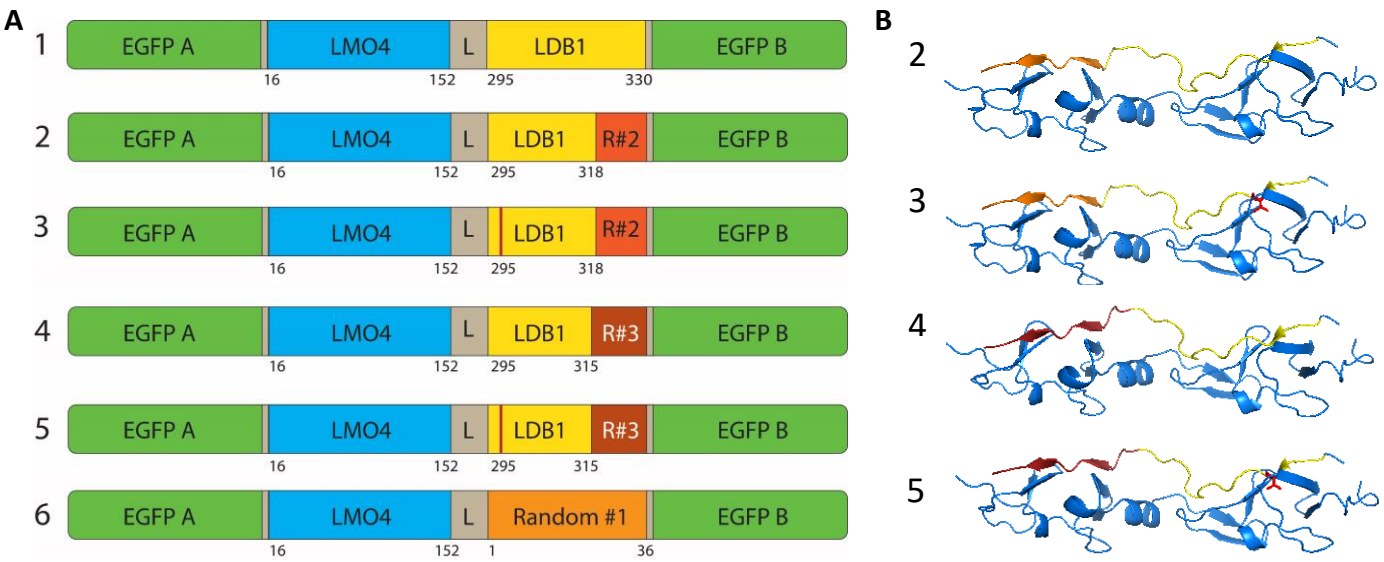


Figure 4.8 New Split EGFP Constructs. (A) The N-terminal fragment (EGFP-A) is tethered to LMO4, the EGFP C-terminal fragment (EGFP-B) is tethered to (1) Positive 36 (from Figure 4.1) (2) LDB1 295-318 and Random peptide #2 (R#2), and is termed LDB1 318; (3) LDB1 V303A 295-318 and Random peptide #2 (R#2), and is termed LDB1 318 mutant; (4) LDB1 295-315 and Random peptide #3 (R#3), and is termed LDB1 315; (5) LDB1 V303A 295-315 and Random peptide #3 (R#3), and is termed LDB1 315 mutant; (6) Negative 36 (from Figure 4.1). A short flexible linker (L) connects LMO4 to 1-6. (B) Cartoon structures of new constructs 2-5. LMO4 is blue, LDB1 is yellow, Random peptide #2 is orange, Random peptide #3 is burnt orange and mutation V303A is red (linker is not shown). Images were generated in PyMOL from PDB 1RUT (21).

4.4 Fluorescence analysis of new constructs

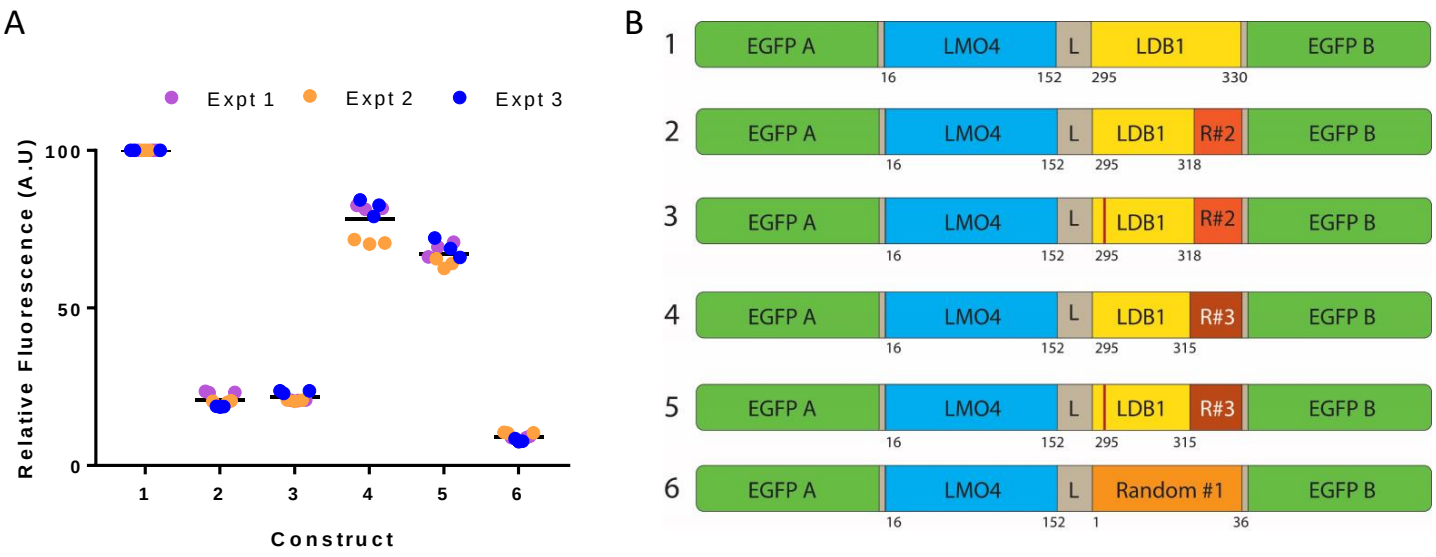


Figure 4.9 Whole cell fluorescence at 507 nm of *E. coli* cells expressing the new constructs. (A) Relative Fluorescence of Constructs 1-6 as compared to Construct 1. (N = 3 with 3 technical replicates in each experiment). Samples were normalised to a cell density of 0.3 at $0D_{600nm}$ (B) Reference schematics of constructs expressed.

In Figure 4.9A the bulk fluorescence properties of each sample were expressed relative to that of Construct 1 in each experiment to control for experimental variance of fluorescence. Constructs 2 and 3 showed almost identical levels of fluorescence that were just above that of the negative control, suggesting that the longer 318 variants could be suitable constructs for screening $LMO4_{LIM1}$-binding peptides. Constructs 4 and 5, however, showed levels of fluorescence that were much closer to the positive control. For the 318 constructs the V303A mutation had no measurable effect, and for the 315 constructs the mutant appears to have slightly reduced fluorescence.

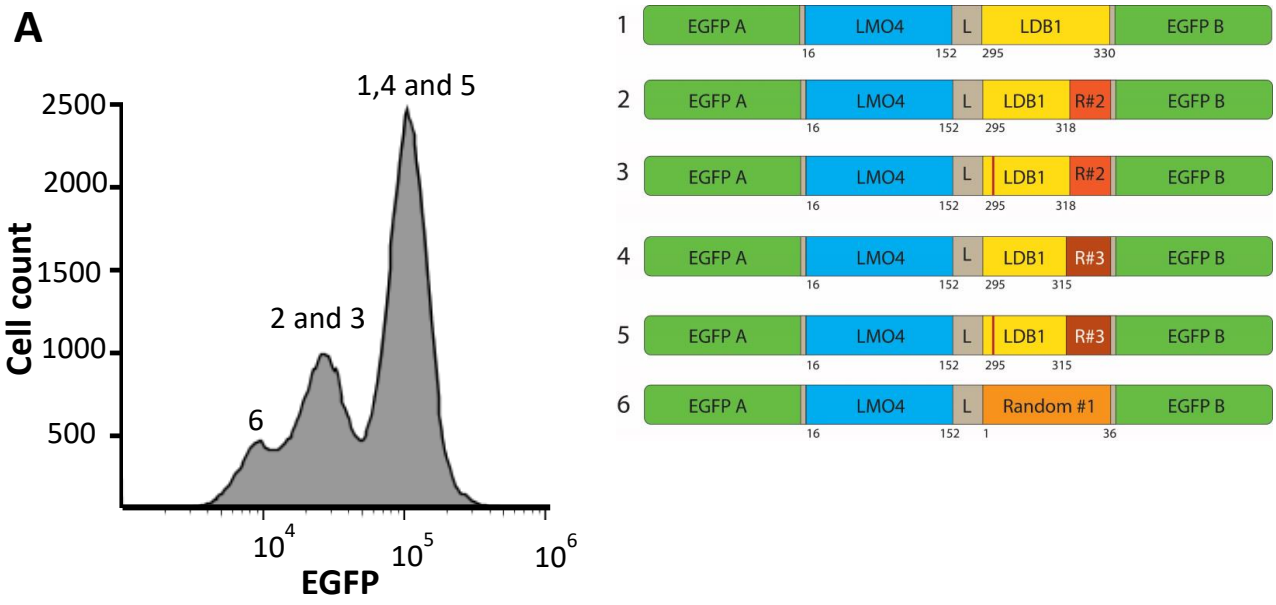


Figure 4.10: Flow cytometry histogram of 6 split EGFP constructs. (A) BL-21 (DE3) cells grown at 20 °C for 18 h overexpressing each of the constructs 1-6. Cells were fixed in PBS with 4% paraformaldehyde. Samples containing 2 × 10^5 cells expressing each of the constructs 1-6 were combined and subjected to flow cytometry. The construct expression contributing to the peaks are labelled and are based on data for each individual construct. Fluorescence histogram created with FlowJo (B) Reference schematics of constructs expressed.

The samples were also subjected to Flow Cytometry, as individual samples (not shown) and mixtures of all samples (same number of cells from each sample; Figure 4.10A). Under these conditions, the 315 constructs and Positive 36 are indistinguishable. The 318 constructs form the middle peak, which can be distinguished from the small peak from Negative 36. No significant differences were detected between the mutant and wild-type constructs.

The small increase in fluorescence for the 318 constructs can be assumed to arise from the binding of $LMO4_{LIM2}$ to LDB1 in the constructs. It is more observable in the flow cytometry experiments (Figure 4.10) than the bulk fluorescence experiments (Figure 4.9) because of the log scale in the representation of the Flow cytometry data.

Based on these data, the construct chosen for further screening for α-helical peptides against $LMO4_{LIM1}$ was Construct 2, LDB1 318. It was decided to try some test α-helices in the system to see if non-specific binding of sequences is a common problem or a rare artefact (Chapter 5). A means to control for overall expression levels is described in the next section.

4.5 Introduction to mCherry

It was decided to introduce a second fluorescent protein under the control of the same promoter as an expression reporter (Section 1.2.4). mCherry was chosen for this purpose because its fluorescence properties are reportedly distinguishable from those of EGFP (56). A ratio of green to red fluorescence should be a more accurate measure of binding affinity than EGFP fluorescence only as it normalises for expression levels in the cell. The introduction of mCherry made it necessary to switch from a pET15b vector to a pETDuet vector (Figure 4.11).

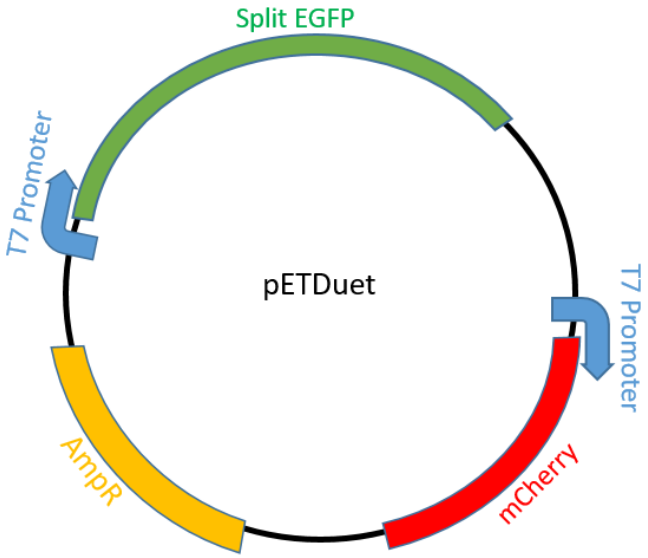


Figure 4.11: pETDuet with split EGFP and mCherry. The split EGFP is shown in the first multiple cloning site and mCherry is in the second multiple cloning site. Both are under the control of a T7 promoter.

pETDuet has two multiple cloning sites, each under the control of a T7 promoter, allowing the simultaneous expression of two different proteins. The split EGFP construct was cloned into multiple cloning site one, whilst mCherry was introduced into multiple cloning site two.

4.6 Testing mCherry

Attempts to clone all the split EGFP constructs into pETDuet with mCherry were made However, this cloning was difficult due to the low copy number of pETDuet vectors, and the need to use restriction enzymes with different efficiencies. At the end of the time available for this project, three constructs had been successfully cloned: mCherry only, as well as spEGFP and LDB1 318 with mCherry (Figure 4.12).

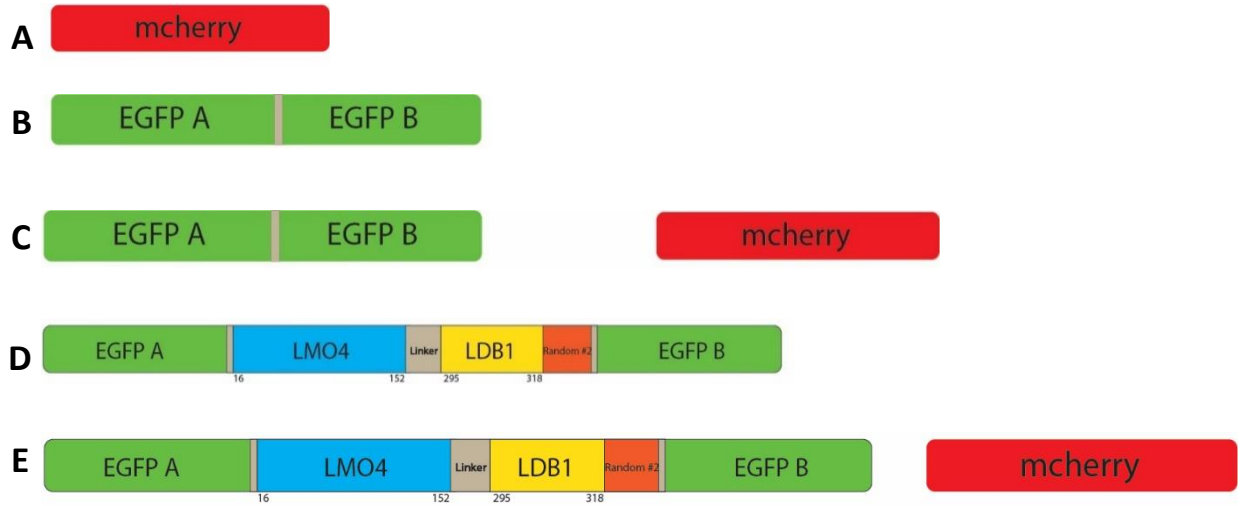


Figure 4.12: Schematic representations of split EGFP and mCherry constructs. (A) New construct, mCherry alone termed mCherry (B) Pre-existing split EGFP alone, termed spEGFP, (C) New construct, split EGFP and mCherry, termed spEGFP + mCherry. (D) Pre-existing LDB1 318, (E) LDB1 318 and mCherry, termed LDB1 318 + mCherry.

BL-21 (DE3) cells were transformed by constructs A-E and expression induced (Section 2.4). The EGFP and mCherry fluorescence of constructs A-E (Figure 4.12) were measured using plate fluorimetry (Figure 4.13) and flow cytometry (Figure 4.17). The emission wavelength for mCherry is 600 nm, which is the same wavelength traditionally used for measuring the optical density of cells. The cell density of cells expressing mCherry can be underestimated by up to 10% using OD_{600nm}, but measuring all cell densities at 700 nm eliminates this problem (87).

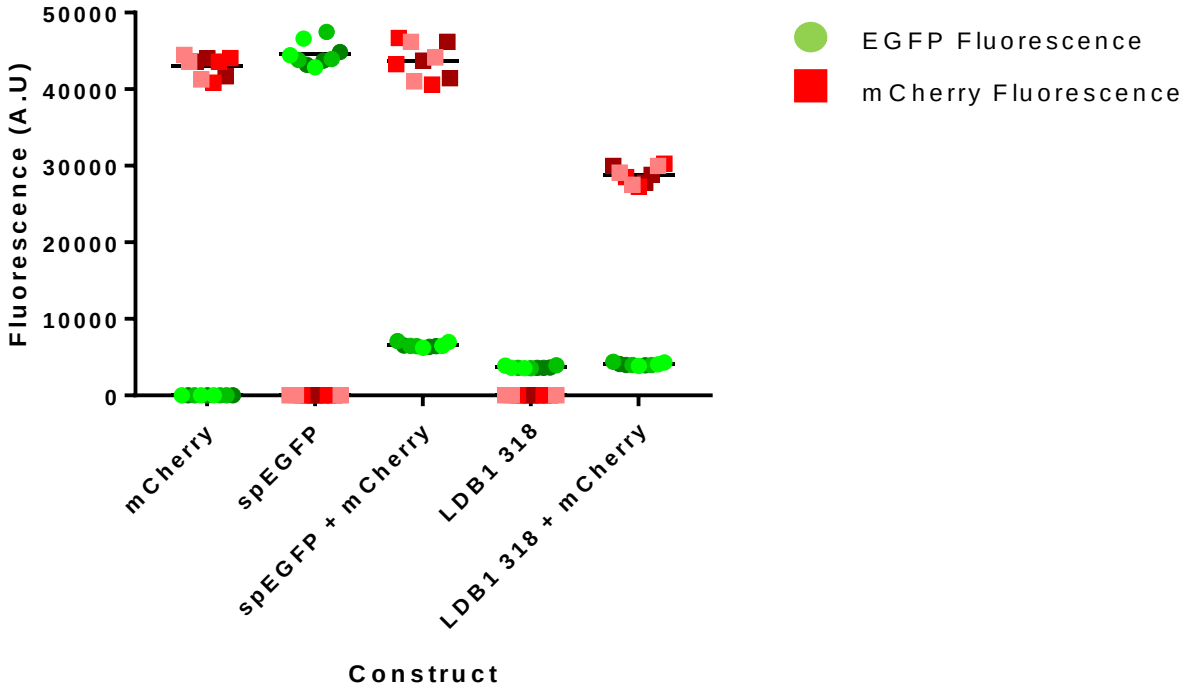


Figure 4.13: Whole cell EGFP and mCherry fluorescence of *E. coli* cells expressing the new mCherry constructs. Fluorescence of constructs as labelled (N = 3 with 3 technical replicates in each experiment). Samples were normalised to a cell density of 0.2 at OD_{700nm}. Fluorescence of EGFP was measured by excitation at 484 nm and emission at 507 nm. Fluorescence of mCherry as measured by excitation at 587 nm and emission at 610 nm.

As expected, cells that did not contain an mCherry construct (spEGFP and LDB1 318) showed no fluorescence at 600 nm, and cells that did not contain EGFP (mCherry) showed no fluorescence at 507 nm (Figure 4.13). Levels of mCherry fluorescence in the cells expressing the mCherry and spEGFP + mCherry constructs were the same, but considerably lower (~65%) in the cells expressing the LDB1 318 + mCherry construct. If mCherry is a good control for overall expression levels, this data indicates that the overall cellular expression levels in LDB1 318 + mCherry expressing cells are lower than that of the other two constructs. However, there was a significant difference in EGFP fluorescence between the cells expressing spEGFP only and spEGFP + mCherry constructs, whereby spEGFP + mCherry was decreased to ~15% of the spEGFP only. In contrast, the level of EGFP signal is the same for cells expressing LDB1 318 and LDB1 318 + mCherry. The level of mCherry fluorescence in cells expressing mCherry only, and in spEGFP + mCherry are also the same.

If we did assume that mCherry is a good expression reporter for this system, we would compare the ratios of EGFP:mCherry fluorescence and use that as a reporter of binding. At this stage we can only compare cells expressing spEGFP + mCherry and LDB1 318 + mCherry,

and those data indicate no significant change, despite EGFP being a positive control and LDB1 318 being a negative control (Figure 4.14).

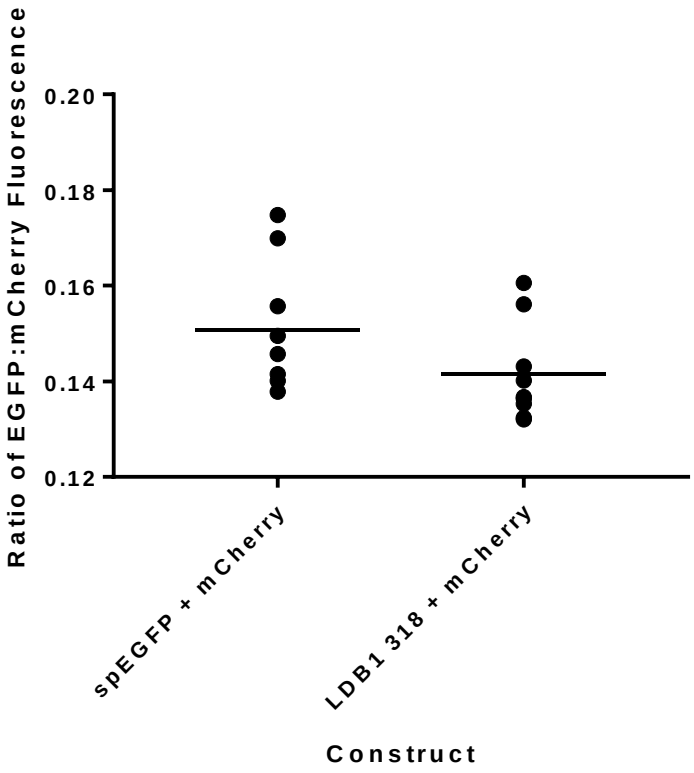


Figure 4.14: Ratio of EGFP: mCherry Fluoresence for spEGFP + mCherry and LDB1 318 + mCherry.

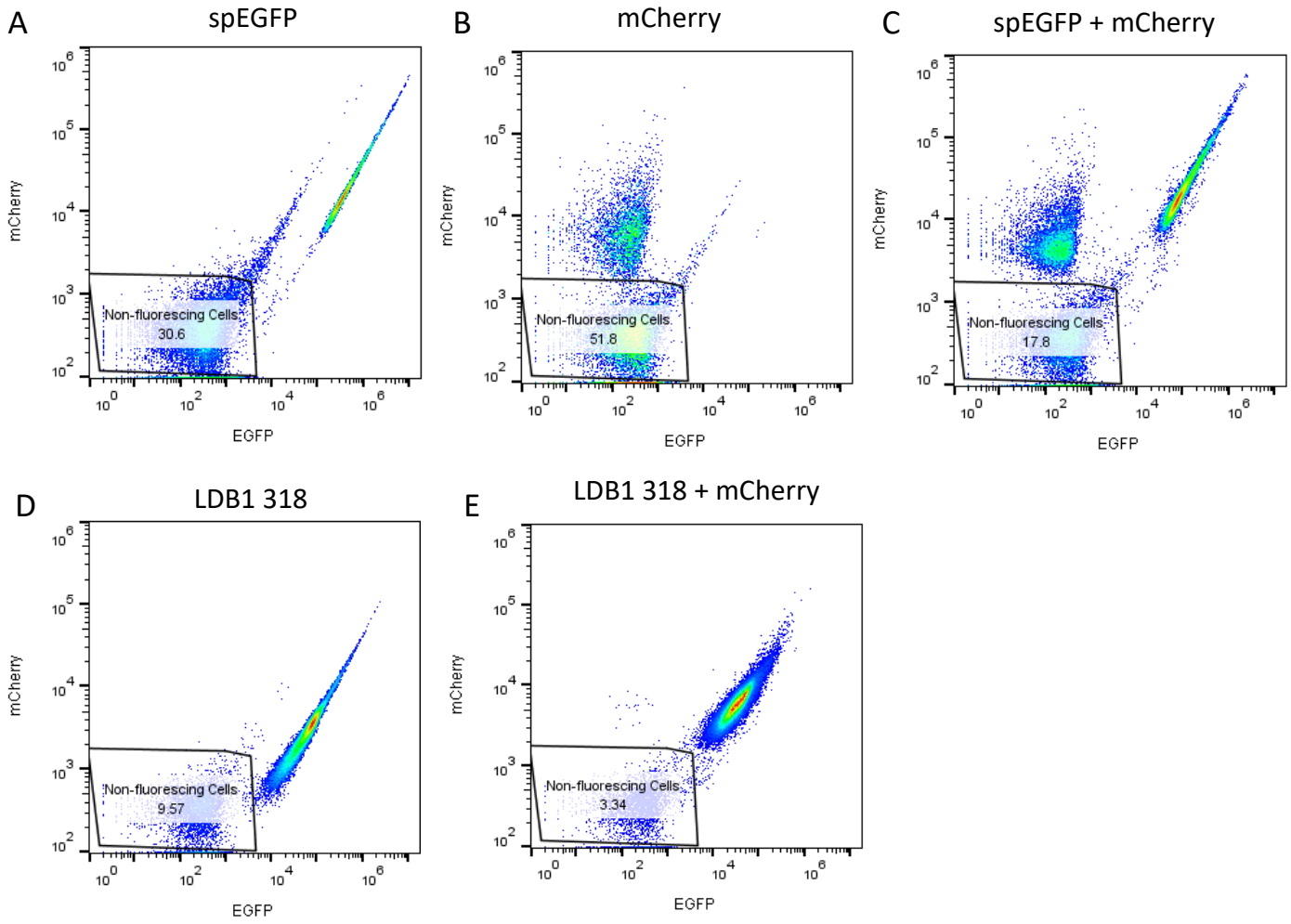


Figure 4.15: Pseudo colour dot plots measuring EGFP and mCherry fluorescence. BL-21 (DE3) cells at grown at 20 °C for 18 h overexpressing constructs A-E from Figure 4.11. Cells were fixed in PBS with 4% paraformaldehyde. In each experiment 1 × 10^5 cells were counted. Fluorescence of EGFP was measured by excitation at 480-90 nm and emission at 500-10 nm. Fluorescence of mCherry as measured by excitation at 580-90 nm and emission at 600-10 nm. (A) spEGFP (31% non-fluorescing), (B) mCherry (52% non-fluorescing), (C) spEGFP + mCherry (18% non-fluorescing), (D) LDB1 318 (10% non-fluorescing) and (E) LDB1 318 + mCherry (3% non-fluorescing). Images generated in FlowJo.

One of the replicates from each experiment in Figure 4.13 was analysed using flow cytometry (Figure 4.15). These data indicated a large range in the sample populations that did not express fluorescent protein (Figure 4.15 boxed). In particular, for spEGFP, mCherry and spEGFP + mCherry (Figure 4.15A-C) 20-50% of cells did not express fluorescent protein. Variation in levels of fluorescent protein expression was seen to a much smaller extent in previous experiments (Figure 4.5). For further analysis (Figure 4.16) the non-fluorescing cells

identified here were excluded, which meant that while equal amounts of cells were not being compared, relative differences in cell populations can still be observed.

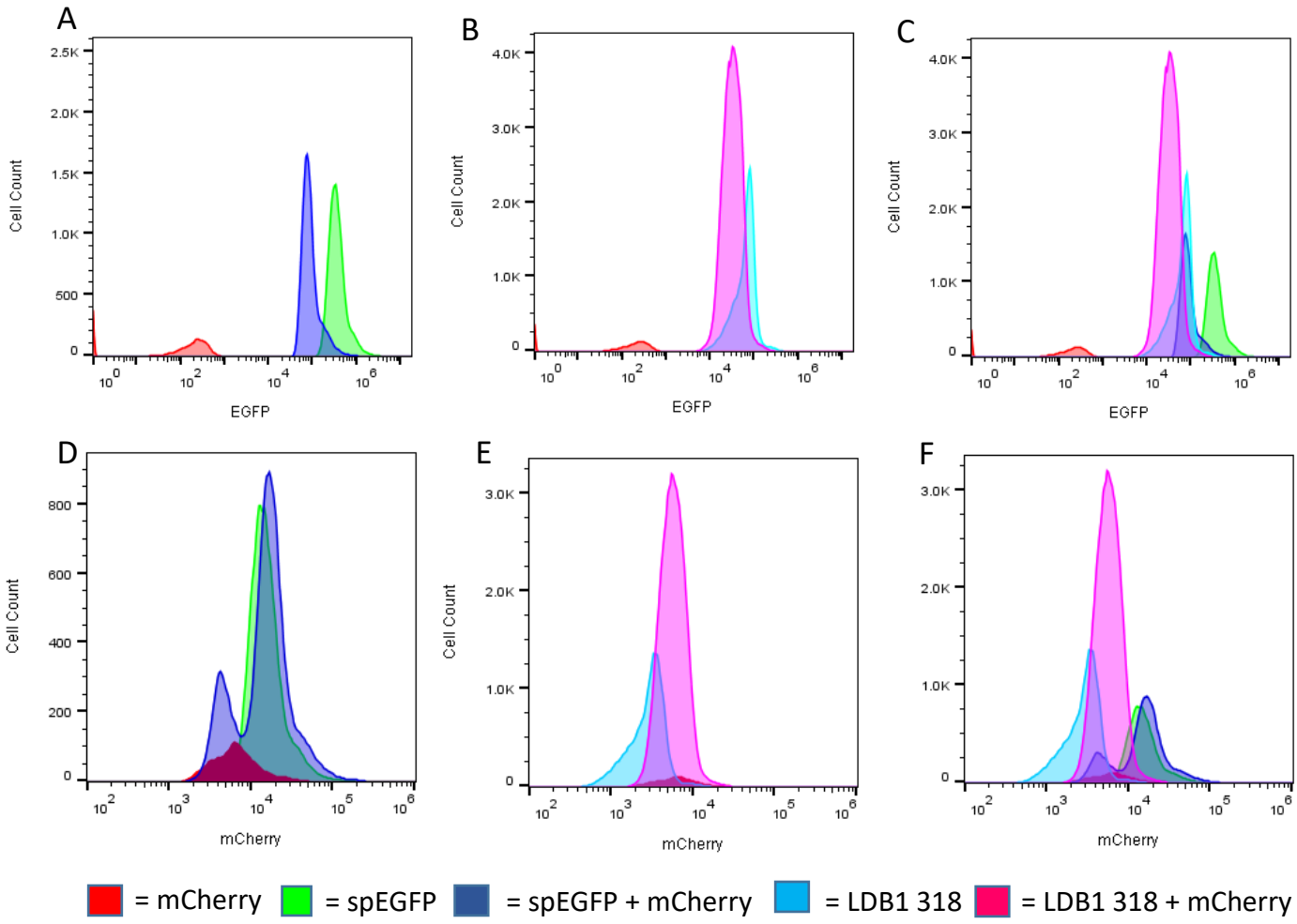


Figure 4.16: Flow cytometry histograms measuring EGFP and mCherry Fluorescence. BL-21 (DE3) cells at grown at 20 °C for 18 h overexpressing constructs A-E from Figure 4.11. Cells were fixed in PBS with 4% paraformaldehyde. In each experiment 1×10^5 cells were counted. Fluorescence of EGFP was measured by excitation at 480-90 nm and emission at 500-10 nm. Fluorescence of mCherry as measured by excitation at 580-90 nm and emission at 600-10 nm Fluorescence histograms were generated and overlayed using FlowJo. The constructs shown are spEGFP (green), mCherry (red), spEGFP + mCherry (dark blue), LDB1 318 (cyan) and LDB1 318 + mCherry (pink). (A) spEGFP, mCherry and spEGFP + mCherrry (B) mCherry, LDB1 318 and LDB1 318 + mCherry (C) spEGFP, mCherry, spEGFP + mCherrry, LDB1 318 and LDB1 318 + mCherry . (D) spEGFP, mCherry and spEGFP + mCherry (E) mCherry, LDB1 318 and LDB1 318 + mCherry (F) spEGFP, mCherry, spEGFP + mCherry, LDB1 318 and LDB1 318 + mCherry. Histogram colours are true to the key in the non-overlapped regions.

Cells expressing spEGFP (Figure 4.16A) had the highest level of EGFP fluorescence. There was a dramatic decrease (~10x) in EGFP fluorescence when mCherry was also being expressed (Figure 4.16A). A much smaller reduction (~4 x) of EGFP fluorescence upon mCherry addition was observed between cells expressing LDB1 318 and LDB1 318 + mCherry constructs (Figure 4.16B).

There was an obvious problem associated with mCherry expression (Figure 4.16D-F), as all the samples, regardless of whether or not they carried an mCherry expression construct or had undetectable expression by plate fluorimetry (Figure.12), had a mCherry fluorescence peak. Unfortunately, the setup used to collect the data in Figure 4.16 does not appear to be optimised for simultaneous EGFP and mCherry fluorescence in *E. coli*, so there is no reliable way to accurately quantify mCherry fluorescence from this data.

4.7 Conclusion and Future Directions

In this chapter a split EGFP complementation system for use in *E. coli* cells was modified to be suitable for screening a library of α-helical peptides against $LMO4_{LIM1}$. Four constructs (LDB1 318, LDB1 318 mutant, LDB1 315 and LDB1 315 mutant) were tested. The LDB1 318 construct was selected for library screening because it showed the lowest levels of background EGFP fluorescence as measured by plate fluorimetry and flow cytometry, and provided a large dynamic range for selection of binders. LDB1 318 may also help induce helicity through the inclusion of LDB1 residues that have helical properties. The LDB1 315 and LDB1 315 mutant construct had unexpectedly levels of high fluorescence as it was assumed that constructs with three extra LDB1 residues might increase the binding of the 318 constructs as compared to the 315 constructs. The two most plausible explanations for this observation are as follows. First, the LDB1 315 constructs had a higher level of expression within the cell than LDB1 318 and thus the cells had higher fluorescence due not to increased binding but rather to increased copy number. Second, the sequence in Random #3, the part of LDB1 315 that comes into contact with $LMO4_{LIM1}$, could be biased towards binding $LMO4_{LIM1}$. The sequence of Random #3 is KANTMTLMPIPAVYR, where the red highlighted residues are all hydrophobic. The red hydrophobic region could bind to the hydrophobic peptide-binding surface of $LMO4_{LIM1}$ when brought into proximity by LDB1 binding to $LMO4_{LIM2}$. Although this could be high affinity binding, it is unlikely to represent specific binding, as was shown previously for

the scrambled $ISL1_{LID}$ binding to the LIM domains of LHX3 (80). In contrast, the sequence of Random#2 in LDB1 318 is AKQLATAAISMC, which is shorter, has fewer bulky hydrophobic residues and a shorter stretch of hydrophobic residues than Random#3. By testing some α-helices in the LDB1 318 construct we can better assess if the LDB1 315 high result is a common problem or rare artefact. Chapter 5 continues with the design of α-helical libraries and testing binding of template α-helices in place of the Random peptide sequences.

Three constructs with fluorescent protein mCherry were tested in an attempt to create an expression control. EGFP and mCherry fluorescence was measured via plate fluorimetry and flow cytometry. For both methods it was observed that the EGFP signal decreased (to varying degrees) with the addition of mCherry**.** A possible explanation for this phenomenon is that the cells may be unable to express spEGFP and mCherry in equal amounts, despite the use of the bicistronic expression system, and/or there may be a maximum capacity for overexpression and correct folding of these proteins that favours mCherry at higher levels of expression. Attempts were made to assess the same cells used for plate fluorimetry and flow cytometry on SDS-PAGE to visualise the expression levels of EGFP and mCherry. However, the fixing of the cells for flow cytometry meant their migration on SPS-PAGE was affected, and no distinct protein bands were able to be seen.

The mCherry levels measured by flow cytometry were similar for all samples regardless of the presence of a mCherry protein (Figure 4.16D-F). The cause of this may be a bleed-through of the strong fluorescence of EGFP or intrinsic *E. coli* proteins that emit at around 610 nm. However, it is plausible that in a more sophisticated Flow Cytometer or FACS machine with more appropriate filter sets for excitation and emission this type of bleed-through could be substantially reduced or eliminated.

From the limited number of constructs tested in Figures 4.13-15 mCherry does not appear to be a useful expression control the split EGFP system because of variable expression levels of the two fluorescent proteins when used in combination. However, additional constructs with combinations such as those shown in Figure 4.17 should be tested before it is decided which construct design to use in screening the peptide library. In addition, unfixed cells should be analysed by SDS-PAGE to assess the relative levels of mCherry and EGFP expression.

Figure 4.17: Additional split EGFP and mCherry constructs. (A) Positive 36 and mCherry, termed Positive 36 + mCherry. (B) LDB1 315 and mCherry, termed LDB1 315 + mCherry, (C) New construct, split EGFP and mCherry, termed spEGFP + mCherry. (D) Pre-existing LDB1 318, (E) LDB1 318 and mCherry, termed LDB1 318 + mCherry.

While it may be suitable to look for an alternative expression control, the split EGFP system can operate without one. It may be appropriate to select a larger pool of fluorescing cells by FACS and to select from smaller libraries than might otherwise be the case. However, validation will be even more important, so that the Y2HCA (Chapter 3) and other binding assays will be particularly useful.

Chapter 5: Library Design

In Chapter 4 I developed a split EGFP complementation system to screen peptides for binding to $LMO4_{LIM1}$ in the context of a split EGFP tethered LMO4-LDB1 construct. The LDB1 318 construct was selected for screening. The next step was to test naturally occurring α-helices in this system that could be potentially used as the template α-helix for the library. This Chapter looks at the process by which three candidate α-helices were chosen, and the result of adding these template peptides to the split EGFP system.

To ensure the full diversity of a large library is represented in the cells being screened we need a very efficient transformation method. This Chapter recounts experiments to optimise the chemical transformation method to maximise transformational efficiency.

It is intended that FACS will be used to isolate cells with higher levels of fluorescence, and the plasmid DNA from the selected cells will be sequenced to identify candidate binders. This Chapter describes efforts to optimise high quality plasmid extraction from *E. coli* cells that had been previously optimised for overexpression of proteins.

5.1 α-helix template candidates and initial selection criteria

A preliminary set of 15 pre-existing α-helices (Table 5.1) were selected by mining the literature for naturally occurring α-helices for which structures had been solved. Specifically, α-helices were identified that were involved in protein-protein interactions and had previously been successfully manipulated through protein engineering. These criteria increase the likelihood of the helix being sufficiently malleable in the context of a library. The α-helix should be able to bind the hydrophobic groove on LMO4 (Section 1.1.1). Thus, amphipathic helices, which have hydrophobic and polar residues segregated on opposite sides (88-90), were selected for consideration as they would provide a hydrophobic binding face but have properties that promote solubility.

Table 5.1: Candidate template α-helices. Bold residues where present represent the key binding region to its natural binding partner while the red residues are the key binding residues (when known).

Source Protein Name	Sequence	Ref
Bim[1]	DMRP**EIWIAQELRRIGDEFNAYYAR**RVFL	(61, 91)
Noxa[1]	PAEL**EVECATQLRRFGDKLNFRQKL**LNLI	(61)
Puma[1]	EEQ**WAREIGAQLRRIADDLNAQYERR**RQE	(61, 92)
Bad[1]	NLWA**AQRYGRELRRMSDEFVDSFKKG**LPR	(61, 93)
Bik/NBK[1]	CMEG**SDALALRLACIGDEMDVSLRA**PRLA	(61)
Bid[1]	QEDI**IRNIARHLAQVGDSMDRSIPPG**LVN	(61)
HRK/DP5[1]	RSSA**AQLTAARLKALGDELHQRTMWR**RRA	(61, 94)
Beclin[1]	DGGT**MENLSRRLKVTGDLFDIMSGQT**DVD	(61, 95)
BMF[1]	QHQA**EVQIARKLQCIADQFHRLHVQQ**HQQ	(61)
VEGF	**EVVKFMDVYQRSY**	(35, 96)
Vammin	**VRPFLEVHERSA**	(35, 96)
p53	SQ**ETFSDWKLLPE**N	(76, 96)
Avian Pancreatic Polypeptide	**VEDLIRFYDNLQQYLNVV**	(58, 97)
HIV-1 Envelope protein GP41	**MEWDREINNYTSLIHSCIEESQNQQEKNEQELL**	(77, 98)
SOS- αH	F**FGIYLTNILKTEEG**	(74, 99)

[1]Bim, Noxa, Puma, Bad, Bik/NBK, BID, HRK/DP5, Beclin and BMF are BH3 α-helices.

It was decided to test three helices from this set in the first instance. The set (Table 5.1) was more closely examined for length and amino acid composition to select the peptides that would be best-suited to binding to LMO4.

In terms of length, it was considered desirable to span the whole peptide-binding surface of $LMO4_{LIM1}$ and to allow some flexibility in binding. It was estimated that this would correspond to 5-6 turns of helix, so the lower limit in length was set at 24 residues. This excludes VEGF, Vammin, p53, Avian Pancreatic Polypeptide and SOS-αH from Table 5.1.

In terms of amino acid composition, it was decided to exclude those with cysteine residues. Cysteines form disulphide bonds, which can be a problem when expressing proteins if disulphide bonds form incorrectly and cause misfolding. This criterion excludes Noxa, Bik/NBK

and BMF. As discussed in Chapter 4 (Section 4.3) the final residues of LBD1 318 appear to adopt a helix-like conformation that could enhance helical formation. Those residues are Glu-Asp-Glu. It was thought reasonable to look for naturally occurring helices that contain Glu, or a similar residue, towards the beginning of the binding region to promote helical formation. Candidate helices Bim, Puma, Bad, HRK/DP5, Beclin, GP41 all fit this criterion. To further narrow down the selection, simple modelling was carried out to gauge how those helices might interact with $LMO4_{LIM1}$.

As it is difficult to estimate levels of plasticity of LMO4 and account for changes in conformation to accommodate peptide binding, a very simple approach was used to model potential interactions. The structures of the helices were manually overlaid on the structure of LMO4 bound to LDB1 in PyMOL. General features such as hydrophobicity and potential to make electrostatic interactions (and not make unfavourable interactions) were assessed. To create the overlays with a reasonable register, a Glu, Gln, Asp or Asn in the helix was aligned with Glu318 in LDB1. The rest of the helix was placed along the hydrophobic grove of $LMO4_{LIM1}$.

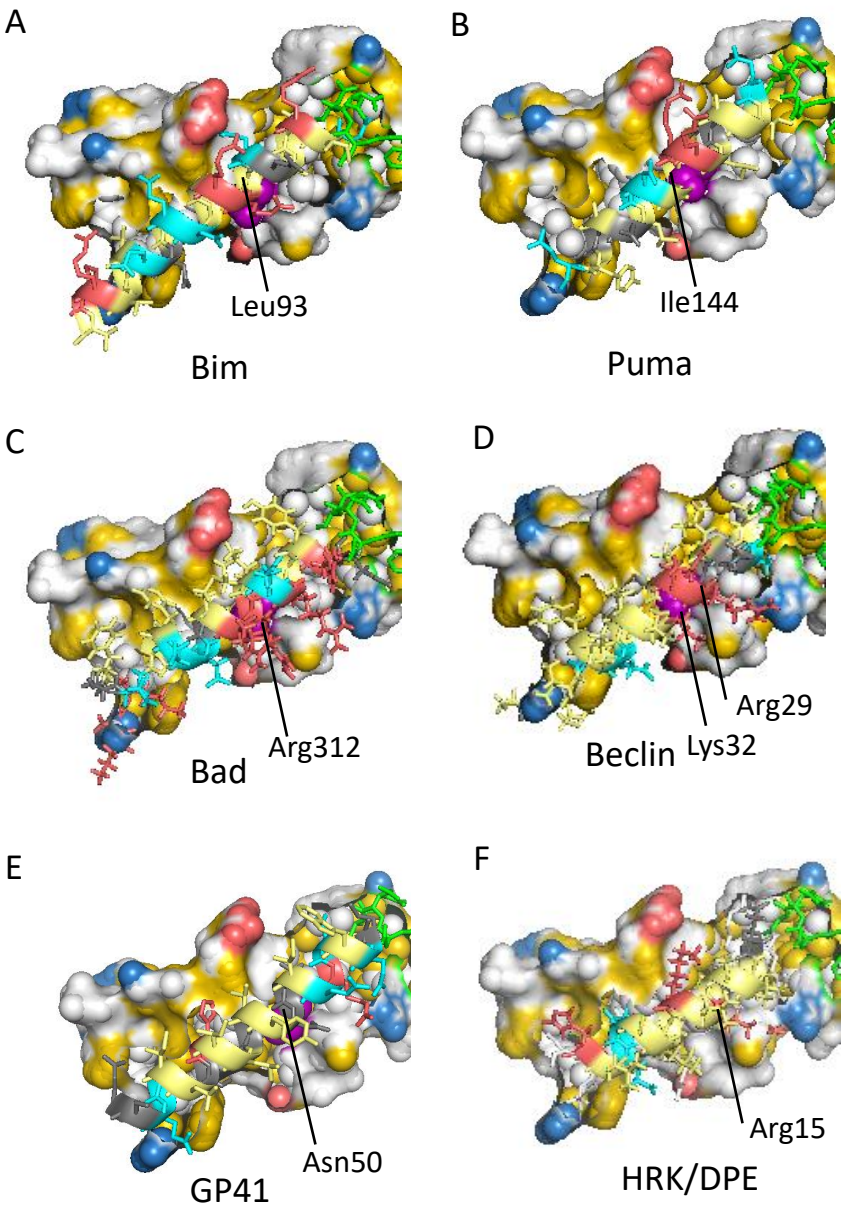


Figure 5.2 LMO4$_{LIM1}$ and candidate α-helix. The α-helix from (A) Bim, (B) Puma, (C) Bad, (D) Beclin, (E) GP41 and (F) Beclin was overlayed onto LMO4$_{LIM1}$. LMO4 is shown in surface representation with YRB colouring: blue is negative, red is positive and yellow is hydrophobic. LDB1 (316-318) is shown in green sticks. The hydrophobic binding pocket, where LDB1-Ile321 binds, is in purple. The α-helices are shown sticks and cartoon, the hydrophobic residues are in pale yellow, positively charged residues red, negatively charged residues in blue and polar but uncharged residues in grey. The images were generated in PyMOL using PDB files 1RUT (21), 1PQ1 (91), 2VOF (92), 1G5J (93), 2PON (95), 1IEX (98) and 2L58 (94).

From the PyMOL overlays (Figure 5.2) the residues in contact with the hydrophobic binding groove of LMO4 were assessed residues around the hydrophobic binding pocket (labelled

residues). Using this method of alignment, it can be seen that the hydrophobic face of BIM α-helix fits along the $LMO4_{LIM1}$ hydrophobic groove (Figure 5.2A). Only hydrophobic and non-charged residues look to be in contact with LMO4. Leu93 occupies a position similar to that of the key binding residue in LDB1, Ile321. Bim BH3 α-Helix was considered a reasonable candidate.

The hydrophobic face of the PUMA α-helix fits along the $LMO4_{LIM1}$ hydrophobic groove in a similar fashion (Figure 5.2B). In this case, Ile144 fits in the hydrophobic binding pocket of $LMO4_{LIM1}$, there were no potential problems indicated by the overlay. PUMA was considered a very strong candidate template helix.

The hydrophobic face of BAD α-helix fits along the $LMO4_{LIM1}$ hydrophobic groove in a similar fashion, but with one notable exception (Figure 5.2C). The residue in contact with the hydrophobic binding pocket is Arg312, which carries a positive charge. However, as the long sidechain of Arg has a hydrophobic component, it can engage in hydrophobic interactions, provided the positively charged end is surface exposed or complemented with a negative charge. Note that Arg323, in the key LMO4-binding region of LDB1, appears to make extensive hydrophobic interactions with LMO4. In addition, residue Leu311 in BAD, which lies next to Arg312 in the sequence could occupy the hydrophobic binding pocket on LMO4. There is precedent for rearrangements at this site as $CtIP_{LID}$ uses two residues to bind this site rather than one (Chapter 1.1.1). BAD remained a candidate.

The Beclin α-helix fits along the $LMO4_{LIM1}$ hydrophobic groove with hydrophobic and hydrophilic residues along the LMO4 interface (Figure 5.2D). Lys32 lies in the hydrophobic binding pocket of $LMO4_{LIM1}$. The other residue that is close the pocket is Arg29 (a turn before Lys32). If only one of these residues was positively charged residue, it could probably have been incorporated as explained above for Arg312 in BAD. However, the presence of two positively charged residues in close proximity to the hydrophobic binding pocket is likely too much positive charge to accommodate binding. This rules Beclin out as a test α-helix.

The GP41 α-helix fits along the $LMO4_{LIM1}$ hydrophobic groove with hydrophobic and non-charged residues along the LMO4 interface (Figure 5.2E). Asn50 lies in the hydrophobic binding pocket of $LMO4_{LIM1}$. Although there is a polar amino acid nearby, residues are

hydrophobic and small rearrangements could accommodate binding, so GP41 remains a candidate.

Using the standard alignment, the HRK/DP5 α-helix would contact $LMO4_{LIM1}$ with the hydrophilic polar/charged side facing LMO4 and the hydrophobic side on the surface (Figure 5.2F) which is not desirable. Although other registers might promote binding of the hydrophobic surface, the other registers would not work as well with the helical residues 316-18 of LDB1. HRK/DO5 was eliminated from consideration.

At this stage there were four candidate peptides BIM, PUMA, BAD and GP41. This set of proteins is dominated by BH3-only helices (BIM, PUMA and BAD). It was considered desirable to test varied of α-helices, and as GP41 is the only non-BH3 candidate remaining it was selected. Of the three BH3 α-helicies PUMA and BIM are very similar to one another, they have 38% sequence identity and 55% sequence similarity. BAD only has 21% sequence identity with PUMA and 34% similarity, BIM and BAD have 24% sequence identity and 41% similarity. As PUMA and BAD are the most different BH3s, PUMA and BAD were chosen. The three candidates selected to test in the split EGFP system were GP41, PUMA and BAD

5.1 Testing BAD, GP41 and PUMA in the split EGFP system

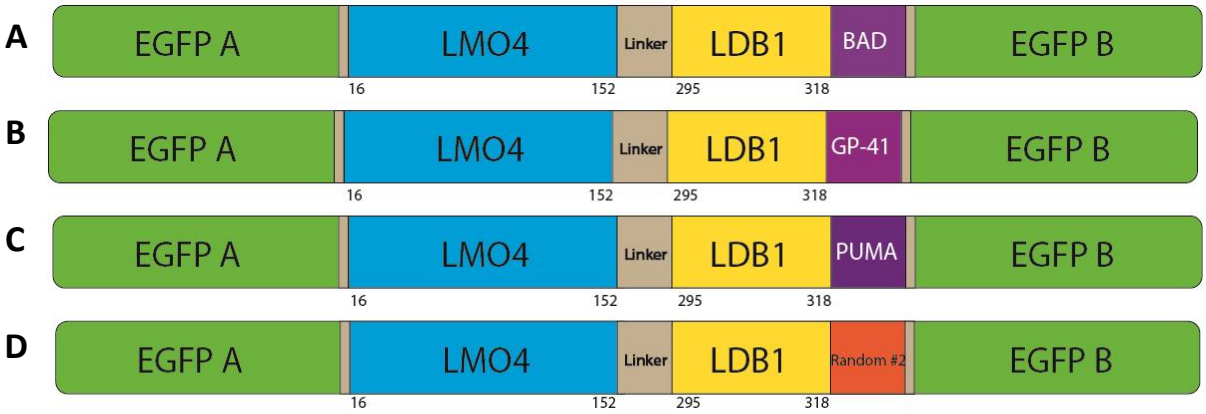


Figure 5.3 Schematic representations of BAD, GP-41 and PUMA in the split EGFP complementation system. The N-terminal fragment (EGFP-A) is tethered to LMO4, and the EGFP C-terminal fragment (EGFP-B) is tethered to (A) LDB1 295-318 and BAD; termed BAD; (B) LDB1 295-318 and GP41; termed GP41; (C) LDB1 295-318 and PUMA; termed PUMA. (D) Pre-existing construct LDB1 318.

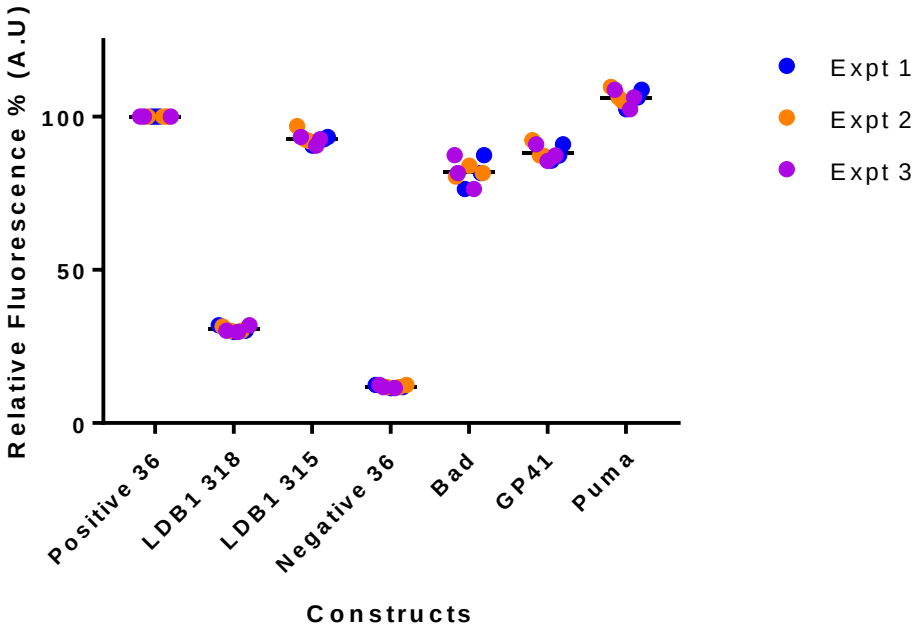


Figure 5.4 Whole cell fluorescence at 507 nm of *E. coli* cells expressing the new constructs. BL-21 (DE3) cells at grown at 20 °C for 18 h overexpressing constructs as labelled. Cells were fixed in PBS with 4% paraformaldehyde. Relative fluorescence of constructs as labelled compared to Positive 36 (N = 3 with three technical replicates in each experiment). Samples were normalised to a cell density of 0.3 at $0D_{600nm}$.

The three test α-helices BAD, PUMA and GP41 were cloned into the LDB1 318 construct, replacing residues 319 and onwards in LDB1 with the α-helix sequence (Figure 5.3). The new constructs were tested for levels of fluorescence by plate fluorimetry and flow cytometry, and compared to a range of controls and previously analysed fluorescence constructs. The three new constructs BAD, GP41 and PUMA all displayed much higher levels of fluorescence than LDB1 318 when analysed in bulk solution (Figure 5.4). The levels were similar to that of LDB1 315. BAD and GP41 showed levels of relative fluorescence just below that of Positive 36 (81 ± 6 % and 88 ± 3 %, respectively) whereas PUMA displayed a level of fluorescence (105 ± 4 %) close to that of Positive 36.

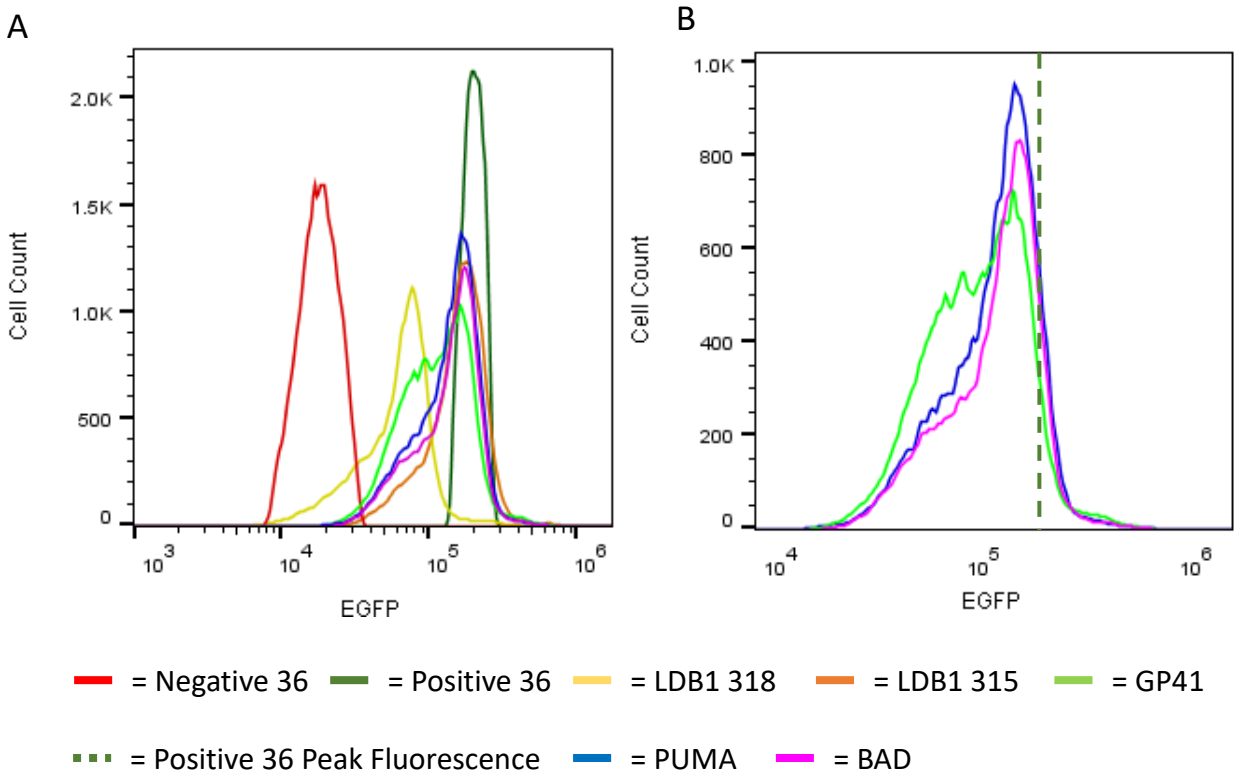


Figure 5.5: Flow cytometry EGFP fluorescence of *E. coli* cells expressing the new constructs. For each sample 5 x 10^4 cells were counted. Fluorescence of EGFP was measured by excitation at 480-90 nm and emission at 500-10 nm. Images were generated and overlayed using FlowJo. The constructs shown are Positive 36 (green line), Negative 36 (red), LDB1 318 (yellow), LDB1 315 (orange), BAD (pink), GP41 (Cyan), PUMA (dark blue), Positive 36 EGFP peak (dark green) (A) Positive 36, Negative 36, LDB1 318, LDB1 315, BAD, GP41 and PUMA. (B) BAD, GP41, PUMA and Positive 36 peak.

One of the replicates for each construct in Figure 5.4 was analysed by flow cytometry to provide some preliminary data about the populations of cells. Only fluorescent cells were

used to make the histograms in Figure 5.4; excluding the ~10% non-fluorescent cells from each sample. The results for Positive 36, LDB1 315, LDB1 318 and Negative 36 in Figure 5.5A were consistent with the previous flow cytometry results (Figure 4.10) and the plate fluorimetry data (Figure 5.4). Negative 36 showed a population of cells with low levels of expression, and Positive 36 showed a single population of highly fluorescent cells (Figure 5.5A). LDB1 315 was slightly less fluorescent than Positive 36 and LDB1 318 was less fluorescent than Positive 36 and LDB1 315, and more fluorescent than Negative 36. Unlike the plate fluorimetry data (Figure 5.4), the new constructs showed peak fluorescence populations that were similar to each other and all a bit lower than that for Positive 36 (Figure 5.4). GP41 had a more significant shoulder in this histogram than the other constructs, representing more populations of cells with lower levels of fluorescence. This comparison illustrates the depth of information provided by flow cytometry as it registers every cell whereas the plate reader provides average values.

5.3 Optimising Chemical Transformation

Any library will comprise a set of plasmids with varied sequences at the target positions that must be transform BL-21 (DE3) cells. Depending on the number of different sequences in the library, this must be a very efficient transformation step to ensure screening of the complete library. The standard protocol for bacterial transformation in the laboratory is a chemical transformation protocol using thawed aliquots of flash-frozen competent cells. This method was selected for its high efficiency in the laboratory compared to other methods for chemical transformation that vary slightly (100). However, it is well suited for transforming individual plasmids, but is not sufficiently effective for transforming a library. The aim of this section was to optimise the transformation procedure for the cells and types of plasmids being used (101).

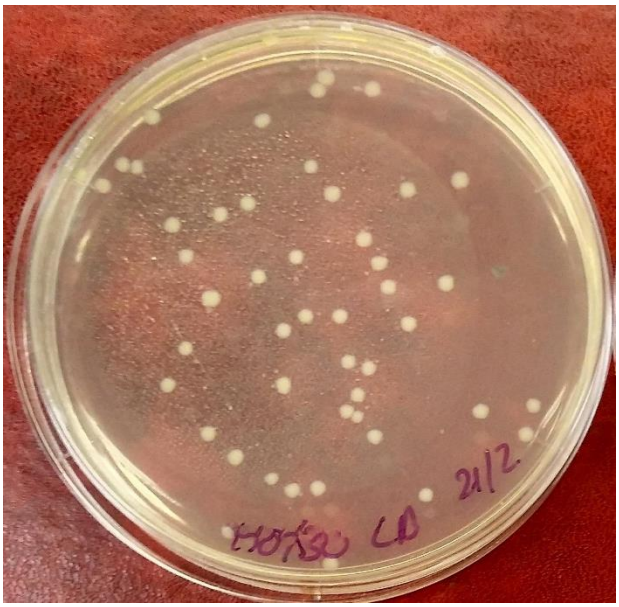


Figure 5.6: Representative original transformation plate. LDB1 315 DNA (200 ng) was added to 20 µl of chemically supercompetent BL-21 (DE3) cells, incubated on ice for 5 min, heat shocked for 90 s at 42 °C, left on ice for 30 s, recovered in LB media for 1 h and plated onto LB agar.

Figure 5.6 shows the typical outcome of the standard chemical transformation protocol approximately 50 colonies representing the same number of successfully transformed cells with approximately 3 x 10^{10} plasmids in 200 ng DNA. To transform a library a much more efficient method is required. To optimise the protocol I varied three different parameters in the standard method (Figure 5.7) and four different recovery media (Figure 5.8). In each case cells were made super-competent by using freshly cultured BL-21 (DE3) cells according to Section 2.4.2, without the freezing, storage and thawing that is part of the standard protocol. Over 100 transformations were performed varying the time parameters according to Table 5.7 and the recovery media according to Table 5.8, each unique combination was repeated three times, where each replicate used each a different plasmid (Positive 36, LDB1 318 and Negative 36). The full data from these experiments is provided in Appendix A). In this section representative plates are shown.

Table 5.7: Times for each of the components in the heat shock protocol.

Method	Resting Ice Time (min)	Heat Shock Time (s)	Ice time (min)	Recovery Time (h)
1	5-10	90	0.5	1
2	30	90	30	1.5-2
3	30	60	2	1-1.5

Method 1 is the standard protocol.

These three methods represent three common sets of parameters used in the Mackay/Matthews lab. Method 1 (the standard protocol) is from Dr Philippa Stokes, Method 2 from Dr Neil Robertson and Method 3 from Dr Jason Low.

Table 5.8: Recovery Media and Components.

Recovery Media	Components
LB	1% w/v Tryptone, 0.5% w/v Yeast Extract, 0.17 M NaCl
SOB	2% w/v Tryptone, 0.5% w/v Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$
SOC	2% w/v Tryptone, 0.5% w/v Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM Glucose
YT	2% w/v Tryptone, 0.5% w/v Yeast Extract, 0.1 M NaCl

LB is the current recovery media

5.3.1 Results of Transformation Optimisation Experiments

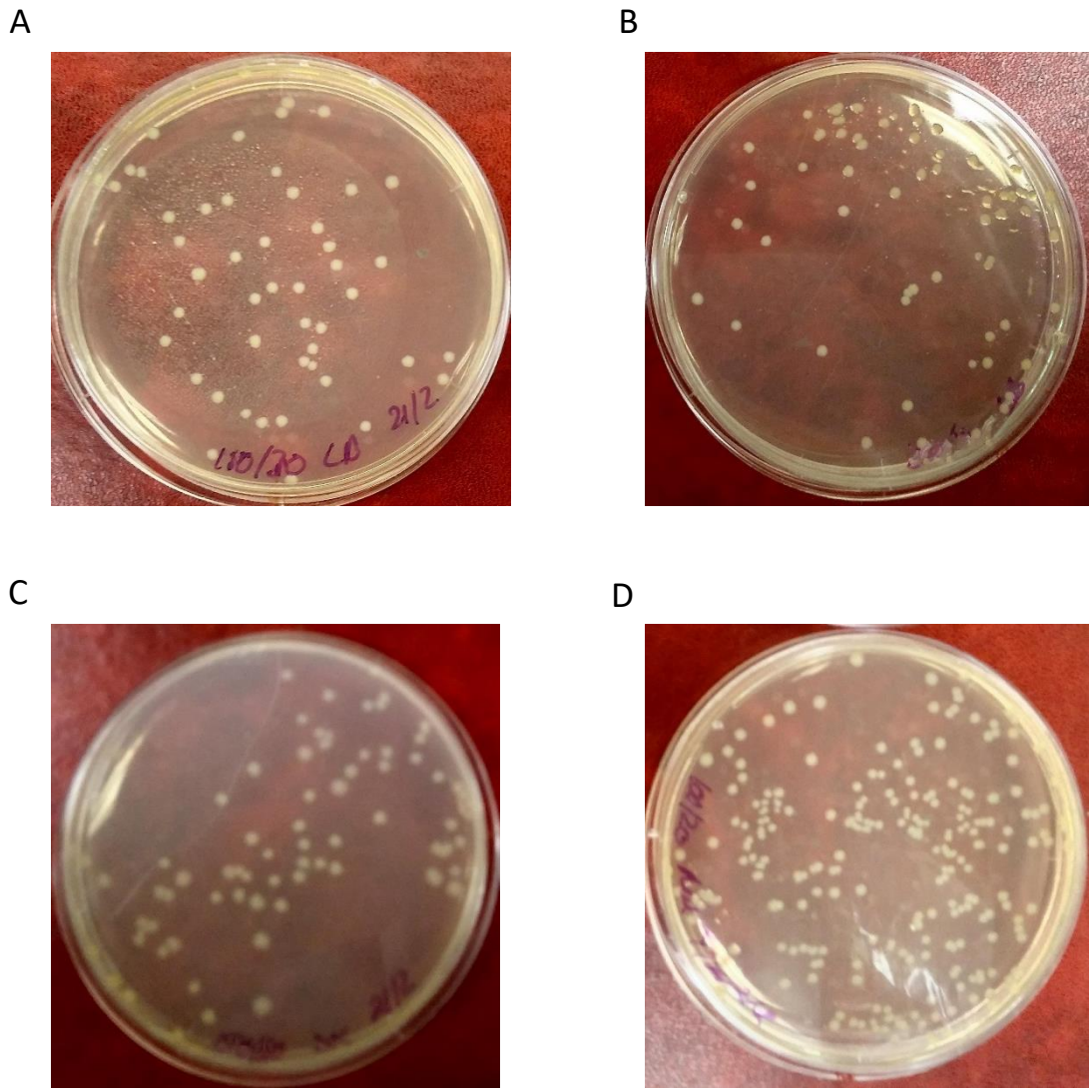


Figure 5.9: Representative transformation plate for Method 1. LDB1 318 DNA (100 ng) was added to 20 µl of chemically supercompetent BL-21 (DE3) cells, incubated on ice for 10 min, heat shocked for 90 s at 42 °C, left on ice for 30 s, recovered for 1 h in (A) LB media (B) SOC media (C) SOB media and (D) YT media. Cells were plated onto LB agar.

Varying the recovery media for the original method revealed an approximate three-fold increase in colony number with YT media (150 colonies) (Figure 5.9). SOC, SOB and LB recovery media all have approximately 50 colonies.

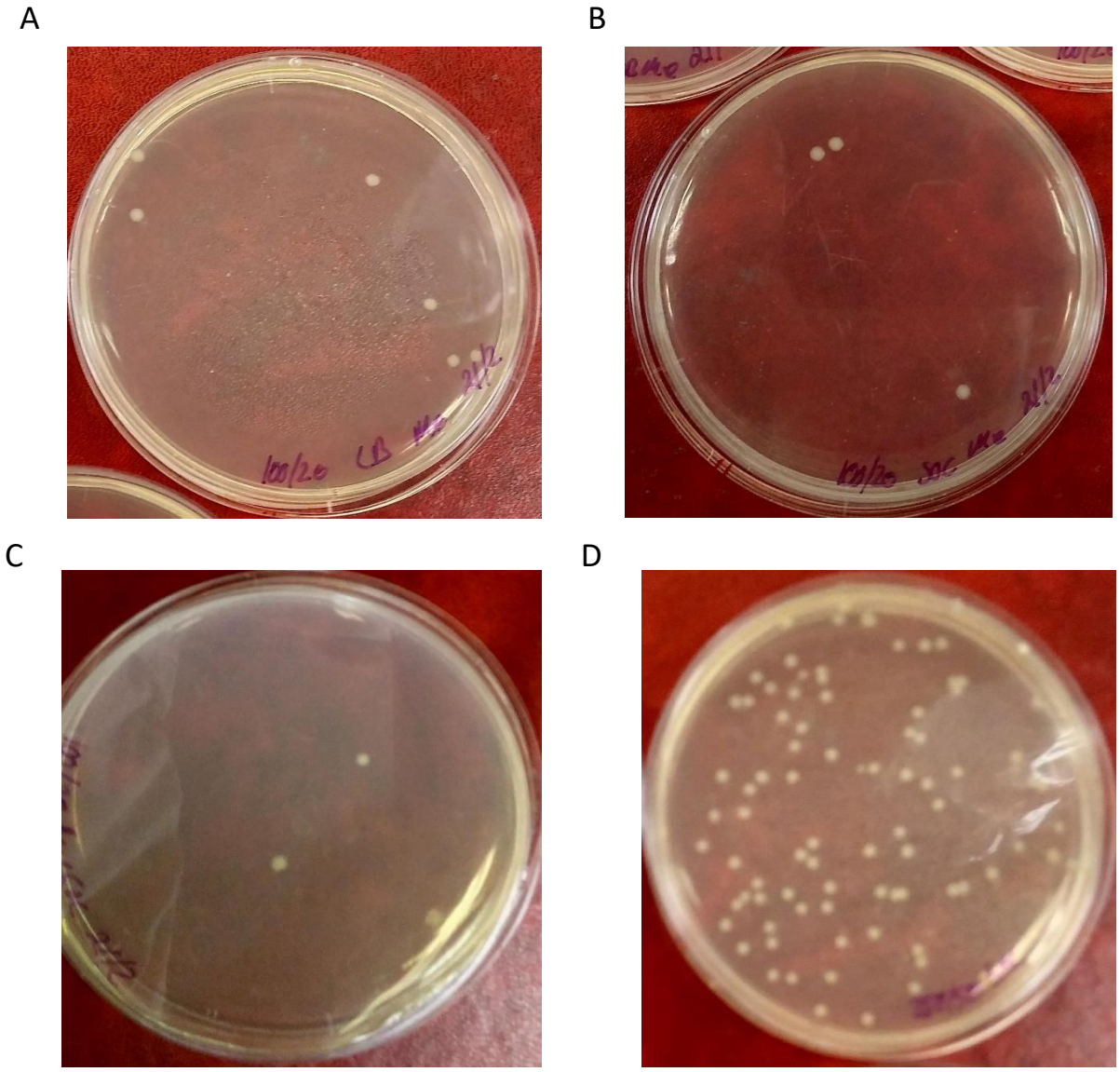


Figure 5.10: Representative transformation plates for Method 2. LDB1 318 DNA (100 ng) was added to 20 µl of chemically supercompetent BL-21 (DE3) cells, incubated on ice for 30 min, heat shocked for 90 s at 42 °C, left on ice for 30 min, recovered for 1.5 h in (A) LB media (B) SOC media (C) SOB media and (D) YT media. Cells were plated onto LB agar.

For Method 2, YT media again gave the highest number of colonies, ~50 (Figure 5.7). LB, SOC and SOB resulted in less than 10 colonies per transformation. All the plates here have fewer colonies than the corresponding plates in Figure 5.9, suggesting that Method 2 is not an improvement on the original method.

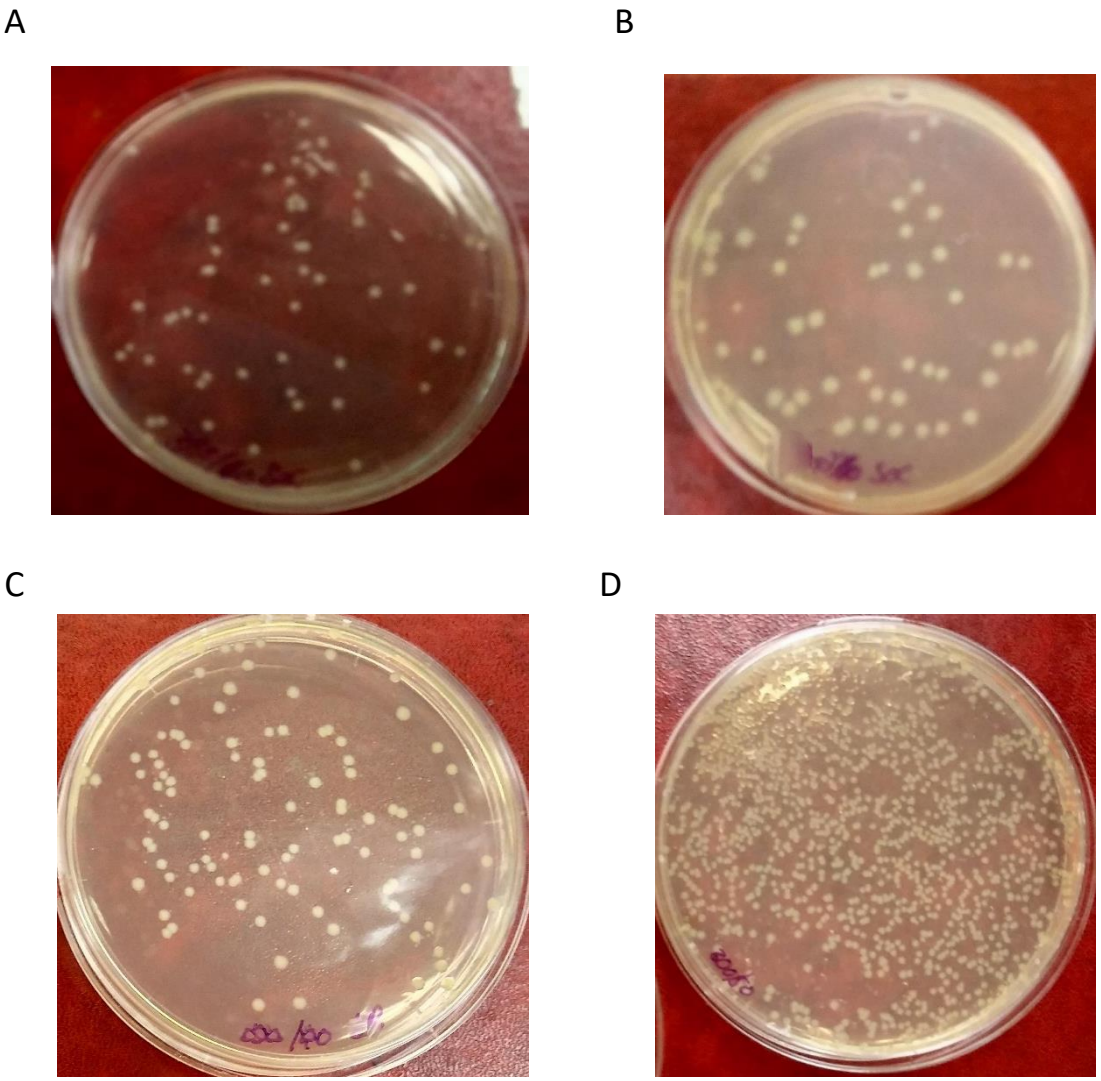


Figure 5.11: Representative transformation pate for Method 3. LDB1 318 DNA (100 ng) was added to 20 µl of chemically supercompetent BL-21 (DE3) cells, incubated on ice for 30 min, heat shocked for 60 s at 42 °C, left on ice for 2 min, recovered for 1 h in (A) LB media (B) SOC media (C) SOB media and (D) YT media. Cells were plated onto LB agar.

Method 3 resulted in approximately 50 colonies for LB, SOC and SOB and 1000 colonies for YT (Figure 5.11). For Methods 1 and 3 recovery in YT media resulted in the largest number of colonies, which was substantially higher for Method 3.

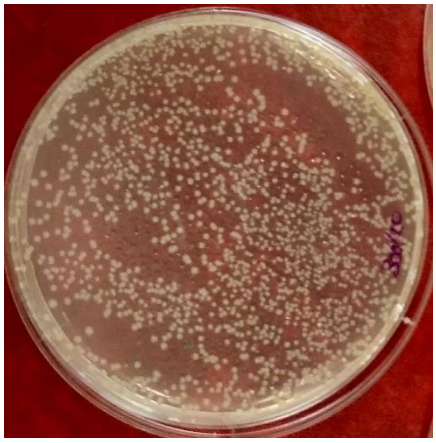
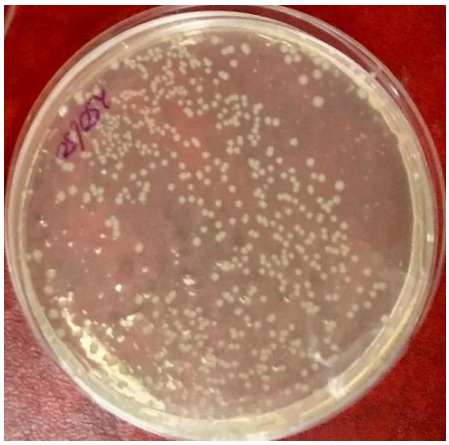
A  B 

Figure 5.12: YT transformation plates. Supercompetent BL-21 (DE3) cells, incubated on ice for 30 min, heat shocked for 60 s at 42 °C, left on ice for 2 min, recovered for 1 h in (A) LDB1 318 DNA (300 ng) was added to 30 µl cells (B) LDB1 318 DNA (50 ng) was added to 20 µl cells.

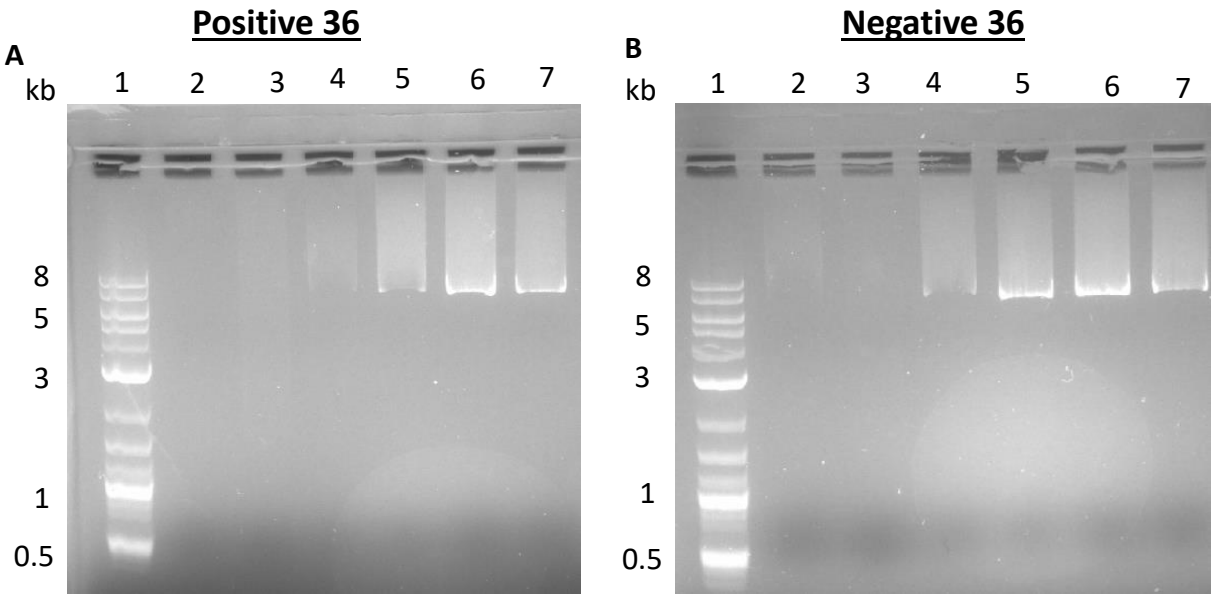
Further attempts were made to increase the transformational efficiency of Method 3 with YT media by increasing the volume of cells being transformed and varying the DNA (ng)-to-cells (µl) ratio. Increasing the cell volume or this ratio did not increase the colony number (Figure 5.12 and Appendix A). For example, 300 ng DNA in 30 µl cells yielded ~1000 colonies (Figure 5.12A). In contrast, decreasing the amount of DNA to 50 ng decreased the colony number by approximately half (Figure 5.12B), and smaller amounts of DNA further decreased the colony number (Appendix A).

Overall, these experiments indicate that Method 3, with recovery in YT media using 100 ng DNA and 20 µl cells is optimal for these cells. This corresponds to ~1000 colonies from 1.4 x 10^{10} plasmids from a single transformation experiment, or ~10^4 cfu (colony forming units)/µg DNA

5.5 Plasmid Extraction Optimisation

The sequence of the LMO4-peptides will be identified by DNA-sequencing of plasmids extracted from highly fluorescing cells isolated by FACS. The cells used for expression of the split EGFP system, BL-21(DE3), belong to a strain that was optimised for protein expression rather than plasmid transformation, amplification and DNA extraction. A variant strain, BL-21(DE3) Gold, possesses an inactivated endonuclease I gene, *endA*, that can degrade plasmid DNA isolated by most miniprep procedures (102).

I sought to compare plasmid extracted from BL-21 (DE3) and BL-21 (DE3) Gold cells and compared these to our standard cell line for plasmid amplification and extraction, DH5α cells. I also compared two commercially available miniprep kit used to extract the DNA in use in our Laboratory, from Bioline and from Qiagen. Cells were transformed with Positive 36 and Negative 36 plasmids, minipreps carried out according to the Manufacturers' instructions (including all optional wash steps) and the resultant preparations analysed by agarose gel electrophoresis (Figure 5.13).



1. Ladder
2. DH5α Bioline
3. BL-21 (DE3) Bioline
4. BL-21 (DE3) Gold Bioline
5. DH5α Qiagen
6. BL-21 (DE3) Qiagen
7. BL-21 (DE3) Gold Qiagen

Figure 5.13. Comparison of plasmid extraction efficacies for mini-prep kits. 500 ng of plasmid (as estimated by spectrophotometry) were run on 1% Agarose gels at 110 v, 400 mA for 40 min and visualised with hydra green.

A correctly sized band on an Agarose gel is a good indication of the presence and quality of the plasmid. For the plasmids prepared using the Bioline kit there were no DNA bands visible except for a very faint one for Negative 36 (Figure 5.13B, lane 4). For plasmids prepared using the Qiagen kit bands of the correct size (6.5 kb) were observed. These data indicate that the Qiagen kit is more appropriate for extracting plasmid DNA from cells. The cell type used did not appear to influence if a DNA band was observed. The yields of the plasmids as estimated by spectrophotometry, along with measures of purity as estimated by the ratios of absorbance at 260/280 (to check for protein contamination) and 230/260 (to check for

carbohydrate contamination), appearance of bands (Figure 5.13) and success in DNA sequencing (Section 2.2) are reported in Tables 5.14/15.

Table 5.14: Plasmid yields for Positive 36.

Construct	Positive 36					
Cell Type	DH5α		BL-21 (DE3)		BL-21 (DE3) Gold	
Plasmid Extraction Kit	Bioline	Qiagen	Bioline	Qiagen	Bioline	Qiagen
Avg Yield (ng)	98	90	95	127	47	94
Avg A260/280	1.8	2.1	1.8	1.9	1.9	1.9
Avg A230/260	1.1	2.3	1.3	2.3	1.0	2.3
Gel Band	no	yes	no	yes	No	yes
Sequencing Success	no	yes	no	yes	No	yes

Preparations were carried out in triplicate (biological replicates) apart from DNA sequencing (one sample only).

Table 5.15: Plasmid yields for Negative 36.

Construct	Negative 36					
Cell Type	DH5α		BL-21 (DE3)		BL-21 (DE3) Gold	
Plasmid Extraction Kit	Bioline	Qiagen	Bioline	Qiagen	Bioline	Qiagen
Avg Yield (ng)	86	89	105	131	78	130
Avg A260/280	1.8	2.0	1.8	1.8	1.8	1.9
Avg A230/260	0.9	2.3	1.0	2.3	1.2	2.2
Gel Band	no	yes	no	yes	yes	yes
Sequencing Success	no	yes	no	yes	no	yes

Preparations were carried out in triplicate (biological replicates) apart from DNA sequencing (one sample only)

The yields estimated by spectrophotometry (absorbance at 260 nm) were generally higher for the Qiagen rather than the Bioline preparations, although there was some variation (Tables

5.14/15). Pure DNA preparations are expected to have an A260/280 ratio of ~1.8, with lower values indicating contamination with protein (103). None of the reported values suggest this problem. The A230/260 ratio has an expected range of 2.0-2 for pure samples with lower values being a marker of carbohydrate and/or phenol contamination (103). The values for the Qiagen kit preparations fall inside or just outside this range, but those for the Bioline preparation were much lower. Carbohydrate and/or phenol contamination may be the reason DNA bands were not seen in agarose gels and sequencing is of poor quality.A Qiagen kit will be used to extract the library plasmids out of the sorted cells. The data does not show the BL-21(DE3) Gold strain having any improvements in DNA extraction over the standard BL-21(DE3) strain.

5.6 Conclusion and Future Directions

In this Chapter naturally occurring α-helices were identified from the literature and examined for their use as templates in the split EGFP selection screen on the basis of length, amino acid composition and potential to form interactions with LMO4. BAD, GP41 and PUMA were selected to be tested and cloned into the split EGFP vectors, which were expressed in cells and fluorescence measured by plate fluorimetry and flow cytometry. There were some differences observed between the fluorescence data obtained by these methods. The plate fluorimetry indicated that the GP41, BAD constructs had levels of fluorescence that were just below that of the positive control, whereas that of PUMA was equivalent to that in Positive 36 (Figure 5.3). However, the flow cytometry data indicated that all three helical constructs resulted in levels of fluorescence that were close to that of LDB1 315, but less than Positive 36 (Figure 5.4). Note that only one set of flow cytometer experiments was completed, so it should be repeated for reliability. However, it is clear from these studies that plate fluorimetry reports average values for the whole cell population, which can be misleading, especially if there are varied numbers of non-fluorescing cells, whereas flow cytometry reports data for each individual cell and provides information about different populations of cells.

As discussed for LDB1 315 (Chapter 4.4), the data for BAD, GP41 and PUMA could indicate that those helices bind non-specifically to LMO4. Indeed, they were selected as having a hydrophobic face that had the potential to interact with LMO4 (Chapter 1). It is not possible to tell from these data if the interactions are specific or not. Rather it will be necessary to express and purify the BAD, GP41 and PUMA constructs for structural characterisation. This

would include circular dichroism studies to look for differences in helicity compared with LMO4-LID complexes, as well as size exclusion chromatography and light scattering to assess oligomeric state, and NMR and/or X-ray crystallography studies to determine if the helices do bind as intended or exhibit non-specific binding.

Regardless, these high values for BAD, GP41 and PUMA mean that there will be a high baseline from which increases in fluorescence would be measured, which may not provide sufficient dynamic range from improved binding. Alternate options include exploration of different mutations in the LDB1 $LMO4_{LIM2}$ binding region to weaken that interaction and thus lower the overall baseline fluorescence. It may be possible to dispense with the $LMO4_{LIM2}$ and corresponding region of LDB1 (295-318) altogether.

The protocols of chemical transformation and DNA extraction were optimised for maximal efficiency for potential screening. The improved protocol for transformation is to incubate 100 ng of plasmid DNA in 20 µl of super competent BL-21 (DE3) cells on ice for 30 min followed by heat shock at 42 °C for 1 min, resting on ice for 2 min and recovery in YT media for 1 h. This method resulted in 1000 colonies on a standard 10-cm agar plate, which is a greater than 20-fold improvement over the original protocol. Much of the improvement in chemical transformation efficiency was due to changing the recovery media to YT (Figures 5.10-15). Further experiments would need to be conducted to know which media components (Table 3) are responsible, but perhaps YT sits in the sweet spot for these cells, being a very rich media with high levels of Tryptone and enough salt to help the bacteria grow (more than SOB/SOC) but not so much that it impacts growth (less than LB). However, despite the improvements, the efficiency of transformation may still be insufficient for library transformations. Electroporation, which utilises and electric current to permeabilise cells, and/or the use of commercially prepared component cells should be investigated to further increase transformation efficiency.

For DNA extraction, it was not necessary to switch from our BL-21 (DE3) cells to another cell line that is optimised for plasmid recovery. This result was surprising, as BL-21(DE3) cells are usually considered to be poor candidates for plasmid extraction. It is possible that our strain has changed over time, but it would also be worth confirming if we are using BL-21(DE3) cells, or if they are mislabelled BL-21(DE3)Gold cells. This could be checked by testing for tetracycline resistance as BL-21(DE3) Gold cells are resistant to this antibiotic whereas BL-

21(DE3) cells are not. Regardless, when those cells were used in combination with the Qiagen mini-prep kit sufficient high quality DNA was produced for DNA-sequencing (Table 4 and 5).

Future Directions

Specific inhibitors of LMO4 that allow LMO4 activity to be modulated would provide tools with which to probe the functions of LMO4 and LMO4-interacting proteins. Understanding the function of LMO4 in different cells types will allow greater understanding of LMO4 in human biology, particularly in terms of development and the onset and/or progression of breast cancer. Inhibitory peptides themselves are unlikely to be useful therapeutics, but they could be used to establish proof of principle that inhibiting LMO4 is a useful strategy for therapy. The peptides would, however, be useful reagents to aid the development of small molecule drug inhibitors that could be used to treat breast cancer. In particular α-helical peptides could be converted into α-helical mimetics - compounds that use non-natural chemistry to stabilise the helical structure that are highly resistant to proteolysis. α-Helical mimetics are showing good promise as drug candidates (63, 73, 104). The approach outlined in this thesis for finding specific α-helical inhibitors of β-strand peptide binding domain may be useful for finding inhibitors for other β-peptide binding proteins that would have broad uses in medical research, biotechnology and agriculture.

The aim of this thesis was to develop methods to identify α-helical peptides from a library of α-helical peptides that bind to LMO4. The split EGFP complementation system was further developed to target $LMO4_{LIM1}$ through the incorporation of the LDB1 318 construct. This construct was subsequently tested with three naturally occurring template α-helices. The fluorescence results for results for PUMA, BAD and GP41 by flow cytometry and plate fluorimetry reveal a background fluorescence that is too high for FACS to identify tighter binding peptides. It may be that these helices are binding to LMO4 as intended, which needs to be assessed by structural and biophysical characterisation as outlined in the previous chapter (Section 5.2).

Regardless, for the split EGFP system to work we need to reduce the background fluorescence as discussed in the previous Chapter (Section 5.2), perhaps by introducing destabilising mutations or dispensing with the $LMO4_{LIM2}$ binding components. The next stage is to design an initial test library. In library design we want to focus on small smart libraries to maximise the chance of finding peptides with increased affinity to LMO4.

Traditional large libraries use either use a naive completely randomised sequence or a template peptide and then perform saturation mutagenesis to create extremely large libraries of peptides. Saturation mutagenesis libraries have a very low probability of finding peptides that bind better to the target, as approximately 30% of the library is significantly worse at binding the target (105). Given this low efficiency it makes sense to use a small smart libraries (106). A small smart library uses structural information to identify which residues are most likely to be involved in binding and to specifically mutate these residues to amino acids that could improve binding (107). Small changes and substitutions of one amino acid for another can have great effects, both positive and negative, on binding affinity (108). Although we don't yet have structures of the template helices bound to LMO4, we can use information from the structures of LMO4 in complex with LDB1, CtIP-1 and Deaf-1 to provide information about what kind of residues would be useful in which locations. In particular, we can target the hydrophobic binding pocket of $LMO4_{LIM1}$ by making libraries biased towards hydrophobic residues in likely binding positions. We have already identified the hydrophobic faces of the amphipathic helices as the likely LMO4-interaction surfaces (Figure 5.2) and intend to select a small number of sites for targeted mutagenesis. For example, the sequence of PUMA is **E**I**G**A**QLR**R**IADDL**NAQYERR**RQE where **bold** indicates residues in the main binding region for its natural target BCL-XL and red indicates residues directly involved in binding. From Figure 5.2 we predict that the Ile114 (blue) will bind in the hydrophobic binding pocket (**E**I**G**A**QLR**R**IADDL**NAQYERR**RQE). We would initially select residues for mutation that are predicted to lie in or around this pocket. Because of the 3-dimensional structure of helices, these will be separated in sequence and are indicated in pink, EIGAQ**L**RR**IA**DD**L**NAQYERRRQE.

Whereas full randomisation of sites could use NNN codons (equimolar amounts of A, C, G and T at each position), it is more common to use restricted codons (NNS or NNK where S is G or C, and K is G or T) to reduce library size and even out representation of amino acids. While all three schemes encode every amino acid, they also include stop codons that will cause a fraction of peptides to be truncated. However, other selective codon schemes (Figure 5.10) encode subsets of the amino acids (109) and can avoid stop codons altogether.

Table 1: Partially degenerate codon schemes.

Codon	Description	Amino Acids	Stop codons
NTN	Hydrophobic	M,F,L,I,V	None
NTT	Hydrophobic	F, I L, V	None
RVK	Charged Hydrophobic	D, E, F, H, I, K, N, R, S, T	TAA
TDK	Hydrophobic	C, F, L, W, Y	TAG
NNT	Mixed	A, D, G, H, I, L, N, P, R, S, T, V	None
NNC	15 aa	A, D, F, G, H, I, L, N, P, R, S, T, V, Y,	None
NDT	12aa	C, D, F, G, H, I, L, N, R, S,V,Y	None
NAN	Charged larger side chains	D,E,H,K,N,Q,Y	TAA, TAG
NCN	Smaller side chains	A, P,S,T	None
RST	Small side chains	A,G,S,T	None
VVC	Hydrophilic	A,D,F,H,N,P,R,S,T	None
DVT	Hydrophilic	A,C,D,G,N,S,T,Y	None
NNW	Charged, hydrophobic	D,E,F,H,I,K,L,N,Q,V,Y	TAA

Equimolar mixtures. N - all nucleotides; D - A/G/T; K - G/T; R - A/G; S - G/C; V - A/C/G; W - A/T.

As we want to focus on hydrophobic residues, we can focus on codon schemes that result in all or predominantly hydrophobic residues. We can also eliminate schemes that contain one or more stop codons. Four possible schemes NTN, NTT, NNT and NDT could be suitable. NTN and NTT code for five and four amino acids, respectively, whereas NNT and NDT encode 15 and 12 amino acids, respectively. We prefer a smaller rather than a large initial library so would likely use NTT or NTN. As the latter includes an extra hydrophobic amino acid (Met), it is a likely candidate.

Attempts were made to use mCherry to control for expression levels in the cells. Due to cloning difficulties only three constructs with mCherry were tested. In order to properly asses if mCherry can be useful as an expression control a complete set of constructs will need to be examined.

In conclusion, thesis has made significant improvements to the split EGFP complementation system and Y2HCA validation systems. The Y2HCA has one additional construct (M3-LMO4) to help qualitatively asses the relative binding affinities of weak peptides and another to stabilise LMO4 and better visualise weak binders (M2-LMO4). The split EGFP system now has positive and negative controls and can incorporate α-helical peptides that have been designed to bind $LMO4_{LIM1}$. Improvements were made to a chemical transformation and a method to extract plasmid DNA from BL-21(DE3) cells was established.

References

1. Matthews, J. M., Lester, K., Joseph, S., and Curtis, D. J. (2013) LIM-domain-only proteins in cancer. *Nat. Rev. Cancer* **13**, 111-122
2. Joseph, S., Kwan, A. H., Stokes, P. H., Mackay, J. P., Cubeddu, L., and Matthews, J. M. (2014) 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* **9**, 13
3. Sang, M. X., Ma, L., Sang, M. J., Zhou, X. L., Gao, W., and Geng, C. Z. (2014) LIM-domain-only proteins: multifunctional nuclear transcription coregulators that interacts with diverse proteins. *Mol. Biol. Rep.* **41**, 1067-1073
4. Matthews, J. M., Bhati, M., Lehtomaki, E., Mansfield, R. E., Cubeddu, L., and Mackay, J. P. (2009) It Takes Two to Tango: The Structure and Function of LIM, RING, PHD and MYND Domains. *Curr. Pharm. Des.* **15**, 3681-3696
5. Bach, I. (2000) The LIM domain: regulation by association. *Mech. Dev.* **91**, 5-17
6. Visvader, J. E., Venter, D., Hahm, K., Santamaria, M., Sum, E. Y. M., O'Reilly, L., White, D., Williams, R., Armes, J., and Lindeman, G. J. (2001) The LIM domain gene LMO4 inhibits differentiation of mammary epithelial cells in vitro and is overexpressed in breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14452-14457
7. Sum, E. Y. M., Segara, D., Duscio, B., Bath, M. L., Field, A. S., Sutherland, R. L., Lindeman, G. J., and Visvader, J. E. (2005) Overexpression of LMO4 induces mammary hyperplasia, promotes cell invasion, and is a predictor of poor outcome in breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7659-7664
8. Montanez-Wiscovich, M. E., Shelton, M. D., Seachrist, D. D., Lozada, K. L., Johnson, E., Miedler, J. D., Abdul-Karim, F. W., Visvader, J. E., and Keri, R. A. (2010) Aberrant expression of LMO4 induces centrosome amplification and mitotic spindle abnormalities in breast cancer cells. *J. Pathol.* **222**, 271-281
9. Wang, N., Lin, K. K., Lu, Z., Lam, K. S., Newton, R., Xu, X., Yu, Z., Gill, G. N., and Andersen, B. (2007) The LIM-only factor LMO4 regulates expression of the BMP7 gene through an HDAC2-dependent mechanism, and controls cell proliferation and apoptosis of mammary epithelial cells. *Oncogene* **26**, 6431-6441
10. Hahm, K. M., Sum, E. Y. M., Fujiwara, Y., Lindeman, G. J., Visvader, J. E., and Orkin, S. H. (2004) Defective neural tube closure and anteroposterior patterning in mice lacking the LIM protein LMO4 or its interacting partner deaf-1. *Mol. Cell. Biol.* **24**, 2074-2082
11. Tse, E., Smith, A. J. H., Hunt, S., Lavenir, I., Forster, A., Warren, A. J., Grutz, G., Foroni, L., Carlton, M. B. L., Colledge, W. H., Boehm, T., and Rabbitts, T. H. (2004) Null mutation of the Lmo4 gene or a combined null mutation of the Lmo1/Lmo3 genes causes perinatal lethality, and Lmo4 controls neural tube development in mice. *Mol. Cell. Biol.* **24**, 2063-2073
12. Deng, M., Luo, X.-j., Pan, L., Yang, H., Xie, X., Liang, G., Huang, L., Hu, F., Kiernan, A. E., and Gan, L. (2014) LMO4 Functions As a Negative Regulator of Sensory Organ Formation in the Mammalian Cochlea. *J. Neurosci* **34**, 10072-10077
13. Sum, E. Y. M., Peng, B., Yu, X., Chen, J. J., Byrne, J., Lindeman, G. J., and Visvader, J. E. (2002) The LIM domain protein LMO4 interacts with the cofactor CtIP and the tumor suppressor BRCA1 and inhibits BRCA1 activity. *J. Biol. Chem.* **277**, 7849-7856
14. Singh, R. R., Barnes, C. J., Talukder, A. H., Fuqua, S. A. W., and Kumar, R. (2005) Negative regulation of estrogen receptor alpha transactivation functions by LIM domain only 4 protein. *Cancer Res.* **65**, 10594-10601
15. Setogawa, T., Shinozaki-Yabana, S., Masuda, T., Matsuura, K., and Akiyama, T. (2006) The tumor suppressor LKB1 induces p21 expression in collaboration with LMO4, GATA-6, and Ldb1. *Biochem. Biophys. Res. Commun.* **343**, 1186-1190

16. Manetopoulos, C., Hansson, A., Karlsson, J., Jonsson, J. I., and Axelson, H. (2003) The Lim-only protein LMO4 modulates the transcriptional activity of HEN1. *Biochem. Biophys. Res. Commun.* **307**, 891-899
17. Asprer, J. S. T., Lee, B., Wu, C. S., Vadakkan, T., Dickinson, M. E., Lu, H. C., and Lee, S. K. (2011) LMO4 functions as a co-activator of neurogenin 2 in the developing cortex. *Development* **138**, 2823-2832
18. Schaffar, G., Taniguchi, J., Brodbeck, T., Meyer, A. H., Schmidt, M., Yamashita, T., and Mueller, B. K. (2008) LIM-only protein 4 interacts directly with the repulsive guidance molecule A receptor Neogenin. *J. Neurochem.* **107**, 418-431
19. Deane, J. E., Ryan, D. P., Sunde, M., Maher, M. J., Guss, J. M., Visvader, J. E., and Matthews, J. M. (2004) Tandem LIM domains provide synergistic binding in the LMO4 : Ldb1 complex. *Embo J.* **23**, 3589-3598
20. Deane, J. E., Sum, E., Mackay, J. P., Lindeman, G. J., Visvader, J. E., and Matthews, J. M. (2001) Design, production and characterization of FLIN2 and FLIN4: the engineering of intramolecular ldb1 : LMO complexes. *Protein Eng.* **14**, 493-499
21. Deane, J. E., Mackay, J. P., Kwan, A. H. Y., Sum, E. Y. M., Visvader, J. E., and Matthews, J. M. (2003) Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4. *Embo J.* **22**, 2224-2233
22. Stokes, P. H., Liew, C. W., Kwan, A. H., Foo, P., Barker, H. E., Djamirze, A., O'Reilly, V., Visvader, J. E., Mackay, J. P., and Matthews, J. M. (2013) Structural Basis of the Interaction of the Breast Cancer Oncogene LMO4 with the Tumour Suppressor CtIP/RBBP8. *J. Mol. Biol.* **425**, 1101-1110
23. Cubeddu, L., Joseph, S., Richard, D. J., and Matthews, J. M. (2012) Contribution of DEAF1 Structural Domains to the Interaction with the Breast Cancer Oncogene LMO4. *PLoS One* **7**, 8
24. Matthews, J. M., and Potts, J. R. (2013) The tandem beta-zipper: Modular binding of tandem domains and linear motifs. *Febs Letters* **587**, 1164-1171
25. Wang, N., Kudryavtseva, E., Ch'en, I. L., McCormick, J., Sugihara, T. M., Ruiz, R., and Andersen, B. (2004) Expression of an engrailed-LMO4 fusion protein in mammary epithelial cells inhibits mammary gland development in mice. *Oncogene* **23**, 1507-1513
26. Wittlin, S., Sum, E. Y. M., Jonas, N. K., Lindeman, G. J., and Visvader, J. E. (2003) Two promoters within the human LMO4 gene contribute to its overexpression in breast cancer cells. *Genomics* **82**, 280-287
27. Ding, K., Wu, Z., Li, X., Sheng, Y., Wang, X., and Tan, S. (2018) LMO4 mediates trastuzumab resistance in HER2 positive breast cancer cells. *Am J Cancer Res* **8**, 594-609
28. Montanez-Wiscovich, M. E., Seachrist, D. D., Landis, M. D., Visvader, J., Andersen, B., and Keri, R. A. (2009) LMO4 is an essential mediator of ErbB2/HER2/Neu-induced breast cancer cell cycle progression. *Oncogene* **28**, 3608-3618
29. Chen, L. C., Nievera, C. J., Lee, A. Y. L., and Wu, X. H. (2008) Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair. *J. Biol. Chem.* **283**, 7713-7720
30. Chen, P. L., Liu, F., Cai, S. N., Lin, X. Q., Li, A. H., Chen, Y. M., Gu, B. N., Lee, E., and Lee, W. H. (2005) Inactivation of CtIP leads to early embryonic lethality mediated by G(1) restraint and to tumorigenesis haploid insufficiency. *Mol. Cell. Biol.* **25**, 3535-3542
31. Murphy, C. G., and Moynahan, M. E. (2010) BRCA Gene Structure and Function in Tumor Suppression A Repair-Centric Perspective. *Cancer J.* **16**, 39-47
32. Yu, X., Wu, L. J. C., Bowcock, A. M., Aronheim, A., and Baer, R. (1998) The C-terminal (BRCT) domains of BRCA1 interact in vivo with CtIP, a protein implicated in the CtBP pathway of transcriptional repression. *J. Biol. Chem.* **273**, 25388-25392
33. Lu, Z., Lam, K. S., Wang, N., Xu, X., Cortes, M., and Andersen, B. (2006) LMO4 can interact with Smad proteins and modulate transforming growth factor-beta signaling in epithelial cells. *Oncogene* **25**, 2920-2930

34. Dutta, S., Ryan, J., Chen, T. S., Kougentakis, C., Letai, A., and Keating, A. E. (2015) Potent and Specific Peptide Inhibitors of Human Pro-Survival Protein Bcl-x(L). *J. Mol. Biol.* **427**, 1241-1253
35. Garcia-Aranda, M. I., Gonzalez-Lopez, S., Santiveri, C. M., Gagey-Eilstein, N., Reille-Seroussi, M., Martin-Martinez, M., Inguimbert, N., Vidal, M., Garcia-Lopez, M. T., Jimenez, M. A., Gonzalez-Muniz, R., and de Vega, M. J. P. (2013) Helical peptides from VEGF and Vammin hotspots for modulating the VEGF-VEGFR interaction. *Org. Biomol. Chem.* **11**, 1896-1905
36. Shimomura, O., Johnson, F. H., and Saiga, Y. (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from luminous hydromedusan, aequorea. *J. Cell. Physio* **59**, 223-239
37. Reid, B. G., and Flynn, G. C. (1997) Chromophore formation in green fluorescent protein. *Biochemistry* **36**, 6786-6791
38. Tsien, R. Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509-544
39. Wachter, R. M., King, B. A., Heim, R., Kallio, K., Tsien, R. Y., Boxer, S. G., and Remington, S. J. (1997) Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein. *Biochemistry* **36**, 9759-9765
40. Royant, A., and Noirclerc-Savoye, M. (2011) Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein. *Journal of structural biology* **174**, 385-390
41. Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. C. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnol* **14**, 315-319
42. Battistutta, R., Negro, A., and Zanotti, G. (2000) Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the Green Fluorescent Protein. *Proteins* **41**, 429-437
43. Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33-38
44. Ito, Y., Suzuki, M., and Husimi, Y. (1999) A novel mutant of green fluorescent protein with enhanced sensitivity for microanalysis at 488 nm excitation. *Biochem. Biophys. Res. Commun.* **264**, 556-560
45. Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**, 691-695
46. Magliery, T. J., Wilson, C. G. M., Pan, W. L., Mishler, D., Ghosh, I., Hamilton, A. D., and Regan, L. (2005) Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: Scope and mechanism. *J. Am. Chem. Soc.* **127**, 146-157
47. Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I., and Stefan, E. (2007) Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat. Rev. Drug Discov.* **6**, 569-582
48. Johnsson, N., and Varshavsky, A. (1994) Split ubiquitin as a sensor of protein interactions in-vivo. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10340-10344
49. Ghosh, I., Hamilton, A. D., and Regan, L. (2000) Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *J. Am. Chem. Soc* **122**, 5658-5659
50. Kerppola, T. K. (2006) Visualization of molecular interactions by fluorescence complementation. *Nat. Rev. Mol. Cell Biol* **7**, 449-456
51. Nisevic, I. (2016) Detection and analysis of LIM domain-mediated interactions between transcription factors In *School of Life and Environmental Sciences* Vol. Doctor of Philosophy, University of Sydney, Sydney
52. Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L., and Lukyanov, S. A. (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **17**, 969-973
53. Bevis, B. J., and Glick, B. S. (2002) Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). *Nat. Biotechnol.* **20**, 83-87

54. Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11984-11989
55. Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp red fluorescent protein. *Nat. Biotechnol.* **22**, 1567-1572
56. Wu, B., Chen, Y., and Müller, J. D. (2009) Fluorescence Fluctuation Spectroscopy of mCherry in Living Cells. *Biophys. J* **96**, 2391-2404
57. Piston, D. W., Campbell, R. E., Day, R. N., and Davidson, M. W. Anthozoa Fluorescent Proteins. In *Education in Microscopy and Digital Imaging* Zeiss
58. Chin, J. W., and Schepartz, A. (2001) Concerted evolution of structure and function in a miniature protein. *J. Am. Chem. Soc.* **123**, 2929-2930
59. Chin, J. W., Grotzfeld, R. M., Fabian, M. A., and Schepartz, A. (2001) Methodology for optimizing functional miniature proteins based on avian pancreatic polypeptide using phage display. *Bioorg. Med. Chem. Lett.* **11**, 1501-1505
60. Chin, J. W., and Schepartz, A. (2001) Design and evolution of a miniature bcl-2 binding protein. *Angew. Chem.-Int. Edit.* **40**, 3806-+
61. Dutta, S., Gulla, S., Chen, T. S., Fire, E., Grant, R. A., and Keating, A. E. (2010) Determinants of BH3 Binding Specificity for Mcl-1 versus Bcl-x(L). *J. Mol. Biol.* **398**, 747-762
62. Groß, A., Hashimoto, C., Sticht, H., and Eichler, J. (2016) Synthetic Peptides as Protein Mimics. *Front Bioeng Biotechnol* **3**
63. Araghi, R. R., and Keating, A. E. (2016) Designing helical peptide inhibitors of protein-protein interactions. *Curr. Opin. Struct. Biol.* **39**, 27-38
64. Czabotar, P. E., Lessene, G., Strasser, A., and Adams, J. M. (2014) Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* **15**, 49-63
65. Hunter, J. J., Bond, B. L., and Parslow, T. G. (1996) Functional dissection of the human Bc12 protein: sequence requirements for inhibition of apoptosis. *Mol. Cell. Biol.* **16**, 877-883
66. Cory, S., Huang, D. C. S., and Adams, J. M. (2003) The Bcl-2 family: roles in cell survival and oncogenesis. *Oncogene* **22**, 8590-8607
67. Gross, A., McDonnell, J. M., and Korsmeyer, S. J. (1999) BCL-2 family members and the mitochondria in apoptosis. *Genes Dev* **13**, 1899-1911
68. Petros, A. M., Olejniczak, E. T., and Fesik, S. W. (2004) Structural biology of the Bcl-2 family of proteins. *Biochim. Biophys. Acta-Mol. Cell Res.* **1644**, 83-94
69. Chen, L., Willis, S. N., Wei, A., Smith, B. J., Fletcher, J. I., Hinds, M. G., Colman, P. M., Day, C. L., Adams, J. M., and Huang, D. C. S. (2005) Differential Targeting of Prosurvival Bcl-2 Proteins by Their BH3-Only Ligands Allows Complementary Apoptotic Function. *Mol. Cell* **17**, 393-403
70. Day, C. L., Smits, C., Fan, F. C., Lee, E. F., Fairlie, W. D., and Hinds, M. G. (2008) Structure of the BH3 domains from the p53-inducible BH3-only proteins Noxa and Puma in complex with Mcl-1. *J Mol Biol* **380**, 958-971
71. Wang, K., Gross, A., Waksman, G., and Korsmeyer, S. J. (1998) Mutagenesis of the BH3 Domain of BAX Identifies Residues Critical for Dimerization and Killing. *Mol. Cell. Biol.* **18**, 6083-6089
72. Yin, X.-M., and Dong, Z. P. D. (2009) *Essentials of apoptosis: a guide for basic and clinical research* Vol. 2nd, Humana Press, New York, N.Y
73. Delbridge, A. R., and Strasser, A. (2015) The BCL-2 protein family, BH3-mimetics and cancer therapy. *Cell Death Differ* **22**, 1071-1080
74. Patgiri, A., Yadav, K. K., Arora, P. S., and Bar-Sagi, D. (2011) An orthosteric inhibitor of the Ras-Sos interaction. *Nat. Chem. Biol.* **7**, 585-587

75. Bonache, M. A., Balsera, B., Lopez-Mendez, B., Millet, O., Brancaccio, D., Gomez-Monterrey, I., Carotenuto, A., Pavone, L. M., Reille-Seroussi, M., Gagey-Eilstein, N., Vidal, M., de la Torre-Martinez, R., Fernandez-Carvajal, A., Ferrer-Montiel, A., Garcia-Lopez, M. T., Martin-Martinez, M., de Vega, M. J. P., and Gonzalez-Muniz, R. (2014) De Novo Designed Library of Linear Helical Peptides: An Exploratory Tool in the Discovery of Protein-Protein Interaction Modulators. *ACS Comb. Sci.* **16**, 250-258
76. Henchey, L. K., Porter, J. R., Ghosh, I., and Arora, P. S. (2010) High Specificity in Protein Recognition by Hydrogen-Bond-Surrogate alpha-Helices: Selective Inhibition of the p53/MDM2 Complex. *ChemBioChem* **11**, 2104-2107
77. Matthews, T., Salgo, M., Greenberg, M., Chung, J., DeMasi, R., and Bolognesi, D. (2004) Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nat. Rev. Drug Discov.* **3**, 215-225
78. Shaginian, A., Whitby, L. R., Hong, S., Hwang, I., Farooqi, B., Searcey, M., Chen, J. C., Vogt, P. K., and Boger, D. L. (2009) Design, Synthesis, and Evaluation of an alpha-Helix Mimetic Library Targeting Protein - Protein Interactions. *J. Am. Chem. Soc.* **131**, 5564-5572
79. Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246
80. Gadd, M. S., Jacques, D. A., Nisevic, I., Craig, V. J., Kwan, A. H., Guss, J. M., and Matthews, J. M. (2013) A Structural Basis for the Regulation of the LIM-Homeodomain Protein Islet 1 (Isl1) by Intra- and Intermolecular Interactions. *J. Biol. Chem.* **288**, 12
81. Bryksin, A. V., and Matsumura, I. (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* **48**, 463-465
82. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* **6**, 343-345
83. Mitchell, V. (2006) Protein-protein interactions betwenn LMOs and their partners, Ldb1 and CtIP. In *School of Molecular and Microbial Biosciences* Vol. Bachelor of Science in Molecular Biotechnology (Honours), The Univeristy of Sydney, Sydney
84. McClymont, G. (2016) Redesigning a Peptide:Protein Interface. In *School of Life and Environmental Sciences* Vol. Bachelor of Science (Advanced) (Honours), University of Sydney, Sydney
85. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**, 172
86. Stothard, P. (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences.
87. Hecht, A., Endy, D., Salit, M., and Munson, M. S. (2016) When Wavelengths Collide: Bias in Cell Abundance Measurements Due to Expressed Fluorescent Proteins. *ACS Synth. Biol.* **5**, 1024-1027
88. Segrest, J. P., De Loof, H., Dohlman, J. G., Brouillette, C. G., and Anantharamaiah, G. M. (1990) Amphipathic helix motif: Classes and properties. *Proteins* **8**, 103-117
89. Drin, G., and Antonny, B. (2010) Amphipathic helices and membrane curvature. *FEBS Letters* **584**, 1840-1847
90. Drin, G., Casella, J. F., Gautier, R., Boehmer, T., Schwartz, T. U., and Antonny, B. (2007) A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14**, 138-146
91. Liu, X., Dai, S., Zhu, Y., Marrack, P., and Kappler, J. W. (2003) The structure of a Bcl-xL/Bim fragment complex: implications for Bim function. *Immunity* **19**, 341-352
92. Smits, C., Czabotar, P. E., Hinds, M. G., and Day, C. L. (2008) Structural Plasticity Underpins Promiscuous Binding of the Prosurvival Protein A1. *Structure* **16**, 818-829
93. Petros, A. M., Nettesheim, D. G., Wang, Y., Olejniczak, E. T., Meadows, R. P., Mack, J., Swift, K., Matayoshi, E. D., Zhang, H., Thompson, C. B., and Fesik, S. W. (2000) Rationale for Bcl-

xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein science : a publication of the Protein Society* **9**, 2528-2534
94. Barrera-Vilarmau, S., Obregon, P., and de Alba, E. (2011) Intrinsic order and disorder in the bcl-2 member harakiri: insights into its proapoptotic activity. *PLoS One* **6**, e21413
95. Feng, W., Huang, S., Wu, H., and Zhang, M. (2007) Molecular basis of Bcl-xL's target recognition versatility revealed by the structure of Bcl-xL in complex with the BH3 domain of Beclin-1. *J Mol Biol* **372**, 223-235
96. Bonache, M. Á., Balsera, B., López-Méndez, B., Millet, O., Brancaccio, D., Gómez-Monterrey, I., Carotenuto, A., Pavone, L. M., Reille-Seroussi, M., Gagey-Eilstein, N., Vidal, M., de la Torre-Martínez, R., Fernández-Carvajal, A., Ferrer-Montiel, A., García-López, M. T., Martín-Martínez, M., de Vega, M. J. P., and González-Muñiz, R. (2014) De Novo Designed Library of Linear Helical Peptides: An Exploratory Tool in the Discovery of Protein–Protein Interaction Modulators. *ACS Comb. Sci.* **16**, 250-258
97. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P., and Wu, C. W. (1981) X-ray analysis (1. 4-Å resolution) of avian pancreatic polypeptide: Small globular protein hormone. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 4175-4179
98. Lu, M., Stoller, M. O., Wang, S., Liu, J., Fagan, M. B., and Nunberg, J. H. (2001) Structural and functional analysis of interhelical interactions in the human immunodeficiency virus type 1 gp41 envelope glycoprotein by alanine-scanning mutagenesis. *Journal of virology* **75**, 11146-11156
99. Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D., and Kuriyan, J. (1998) The structural basis of the activation of Ras by Sos. *Nature* **394**, 337-343
100. Yoshida, N., and Sato, M. (2009) Plasmid uptake by bacteria: a comparison of methods and efficiencies. *Appl Microbiol Biotechnol* **83**, 791-798
101. Chan, W.-T., Verma, Chandra S., Lane, David P., and Gan, Samuel K.-E. (2013) A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. *Bioscience Reports* **33**, e00086
102. Technologies, A. BL21-Gold(DE3) Competent Cells Instruction Manual. Agligent Technologies
103. (2009) 260/280 and 160/230 Ratios. In *Technical Bulletin NanoDrop Spectrophotometer*, Thermo Fisher Scientific, Delaware, USA
104. Bullock, B. N., Jochim, A. L., and Arora, P. S. (2011) Assessing Helical Protein Interfaces for Inhibitor Design. *J. Am. Chem. Soc.* **133**, 14220-14223
105. Bershtein, S., and Tawfik, D. S. (2008) Ohno's Model Revisited: Measuring the Frequency of Potentially Adaptive Mutations under Various Mutational Drifts. *Mol. Biol. Evol.* **25**, 2311-2318
106. Goldsmith, M., and Tawfik, D. S. (2012) Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **22**, 406-412
107. Goldsmith, M., and Tawfik, D. S. (2012) Directed enzyme evolution: beyond the low-hanging fruit. *Curr Opin Struct Biol* **22**, 406-412
108. Zelezetsky, I., and Tossi, A. (2006) Alpha-helical antimicrobial peptides—Using a sequence template to guide structure–activity relationship studies. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1758**, 1436-1449
109. Currin, A., Swainston, N., Day, P. J., and Kell, D. B. (2015) Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **44**, 1172-1239

Appendix A

Method	DNA (ng)	Cells (µl)	Recovery media	Colonies
3	100	20	YT	1000
3	300	30	YT	1000
3	150	20	YT	1000
3	50	20	YT	750
3	250	50	YT	750
1	250	50	YT	250
1	300	20	YT	200
1	150	20	YT	200
1	250	50	LB	155
2	250	50	YT	150
2	300	20	YT	150
1	100	20	YT	150
3	250	50	LB	150
3	250	50	SOC	150
3	144	20	LB	150
1	250	50	SOC	150
3	240	50	SOB	125
2	150	20	YT	120
3	200	20	LB	115
2	100	20	YT	100
3	300	20	LB	100
3	200	40	LB	100
1	200	40	LB	100
1	300	20	LB	100
3	250	20	LB	90
1	250	20	LB	75
3	250	50	SOB	70
3	100	20	LB	60
3	300	60	LB	60
3	100	20	SOB	60
1	250	50	SOB	60
3	150	20	SOB	60
3	150	30	LB	50
3	100	20	SOC	50
3	150	30	SOC	50
3	200	40	SOC	50
3	300	60	SOC	50
1	200	20	LB	50
1	144	20	LB	50
1	100	20	SOB	50
1	100	20	SOC	50

1	100	20	LB	50
1	150	30	LB	50
1	300	60	LB	50
1	150	30	SOC	50
1	200	40	SOC	50
1	300	60	SOC	50
1	150	30	SOB	50
1	200	40	SOB	50
1	300	60	SOB	50
3	100	20	SOB	50
3	150	30	SOB	50
3	200	40	SOB	50
3	300	60	SOB	50
1	50	20	LB	42
3	50	20	LB	37
1	100	20	LB	20
1	100	20	SOC	14
1	100	20	SOB	12
2	100	20	LB	6
2	100	20	SOC	3
2	100	20	SOB	3
3	5	20	LB	1
3	0.5	20	LB	0
3	0.05	20	LB	0
3	0.005	20	LB	0

Full data for chemical transformation optimisation experiments. Methods 1-3 are described in Figure 5.11 and Section 2.4.5. Methods 1-3 describe a transformation protocol which have a range of possible amounts of DNA, cell volumes and recovery media. The exact DNA amount, cell volume and recovery media used in each transformation is indicated in the table. The times and other steps indicated in the Methods were not varied. Data are arranged in descending order of colonies obtained.

Appendix B – Protein Sequences

Protein sequences of constructs generated in the thesis

LDB1 318

MGSSHHHHHHSSGLVPRGSHMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQKGGGGSGGGGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQA
QLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYIN
GSLFCEHDRPTALINGHLNSGGSGGLEVLFQGPGGSGGSSQVPDVMVVGEPTLMGGEFGDEDERLITRLENTQFD
AANGIDDEEFGGGGSGGGGSNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNE
KRDHMVLLEFVTAAGITHGMDELYK

LDB1 318 mutant

MGSSHHHHHHSSGLVPRGSHMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQKGGGGSGGGGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQA
QLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYIN
GSLFCEHDRPTALINGHLNSGGSGGLEVLFQGPGGSGGSSQVPDVMVVGEPTLMGGEFGDEDERLITRLENTQFD
AANGIDDEEFGGGGSGGGGSNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNE
KRDHMVLLEFVTAAGITHGMDELYK

LDB1 315

MGSSHHHHHHSSGLVPRGSHMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQKGGGGSGGGGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQA
QLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYIN
GSLFCEHDRPTALINGHLNSGGSGGLEVLFQGPGGSGGSSQVPDVMVVGEPTLMGGEFGDEDERLITRLENTQFE
FGGGGSGGGGSNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLE
FVTAAGITHGMDELYK

LDB1 315 mutant

MGSSHHHHHHSSGLVPRGSHMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQKGGGGSGGGGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQA
QLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYIN
GSLFCEHDRPTALINGHLNSGGSGGLEVLFQGPGGSGGSSQVPDVMVVGEPTLMGGEFGDEDERLITRLENTQFE
FGGGGSGGGGSNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLE
FVTAAGITHGMDELYK

mCherry

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSP
QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSD
GPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKL
DITSHNEDYTIVEQYERAEGRHSTGGMDELYK

PUMA

MGSSHHHHHHSSGLVPRGSHMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQKGGGGSGGGGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQA
QLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYIN
GSLFCEHDRPTALINGHLNSGGSGGLEVLFQGPGGGSSTLYCQFPITIVEQSRQCLHNSYAWPRRIFTDVVMCLGK
KGSTWLTEFGGGGSGGGGSNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEK
RDHMVLLEFVTAAGITHGMDELYK

BAD

MGSSHHHHHHSSGLVPRGSHMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQKGGGGSGGGGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQA
QLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYIN
GSLFCEHDRPTALINGHLNSGGSGGLEVLFQGPGGGSSTLYCQFPITIVEQSRQCLHNSYAWPRRIFTDVVMCLEFG
GGGSGGGGSNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFV
TAAGITHGMDELYK

GP41

MGSSHHHHHHSSGLVPRGSHMSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEYNYNSHNVYITADKQKGGGGSGGGGSLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKCSSCQA
QLGDIGTSSYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVYHLKCFTCSTCRNRLVPGDRFHYIN
GSLFCEHDRPTALINGHLNSGGSGGLEVLFQGPGGGSSTLYCQEWQTYRIPLMNGHDECVDAWPRRIFTDVVMC
LGKKGSTWLTEFGGGGSGGGGSNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDP
NEKRDHMVLLEFVTAAGITHGMDELYK

M3-LMO4

GSSLQNNQDVSFENIQWSIDPGADLSQYKAAATVIDTKDGSQSKLGGGGSGGHMGSGGLSWKRCAGCGGKIAD
RFLLYAMDSYWHSRCLKCSCCQAQLGDIGTSCYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGNVY
HLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHLNS

M668A-LMO4

GSSLQNNQDVSFENIQWSIDPGADLSQYKMDVAAADTKDGSQSKLGGGGSGGHMGSGGLSWKRCAGCGGKIA
DRFLLYAMDSYWHSRCLKCSCCQAQLGDIGTSCYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRAQGN
VYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHLNS

D669A-LMO4

GSSLQNNQDVSFENIQWSIDPGADLSQYKAAATVIDTKDGSQSKLGGGGSGGLEVLFQGPGGSGGLSWKRCAGC
GGKIADRFLLYAMDSYWHSRCLKCSCCQAQLGDIGTSCYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMR
AQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHLNS

V670A-LMO4

GSSLQNNQDVSFENIQWSIDPGADLSQYKMDVAAADTKDGSQSKLGGGGSGGLEVLFQGPGGSGGLSWKRCA
GCGGKIADRFLLYAMDSYWHSRCLKCSCCQAQLGDIGTSCYTKSGMILCRNDYIRLNSGACSACGQSIPASELVMR
AQGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHLNS

Appendix C – Primers

Oligonucleotides used as PCR primers are listed below

Name	Sequence (5'-3')
LMO4pgbt9REV	TCGCCCGGAATTAGCTTGGCTGCAGAATTCTCAGGAGTTCAAATGGCC
M668Afor	CTTTCTCAGTATAAAGCGGATGTTACTGTAATA
M668Arev	TATTACAGTAACATCCGCTTTATACTGAGAAAG
D669Afor	CTCAGTATAAAATGGCTGTTACTGTAATAGA
D669Arev	TCTATTACAGTAACAGCCATTTTATACTGAG
V670Afor	AGTATAAAATGGATGCTACTGTAATAGATAC
V670Arev	GTATCTATTACAGTAGCATCCATTTTATACT
LMO4 153 GEX R	CGTCTAGCAGTCAGTCAGTCGCTACTTAAG
GEX CtIP641 F	CCAAAATCGGATCTGGTTCCGCGTGGATCCTCTCTACAAAACAACCAAG
pET15gibson_For	AGCGGCCTGGTGCCGCGCGGCAGCCATATGTCTAAGGGCGAAGAACTGTTT
pET15gibson_Rev	TCCTTTCGGGCTTTGTTAGCAGCCGTCGACTCATTACTTGTACAGCTCATCCATAC
LMO4GFPForGib	GGTGGCGGCGGTTCTGGCGGTGGTGGATCCCTCTCCTGGAAGCG
330LDB1GFPRevGib	GCCACCACCGGAACCACCGCCACCGAATTCAAACTGGGTGTTCTCC
36nonsenseGFP	GCCACCACCGGAACCACCGCCACCGAATTCAAGGCACATTACAACAT
LDB1(295-318)Nonsense Rev	CCATGTTACAATGGCCATTCC
Duet1EGFPfor	AGCAGCCATCACCATCATCACCACAGCCAGGGCCTGGTGCCG
Duet1EGFPrev	AAGCTTGTCGACCTGCAGGCGCGCCGAGCTTCATTACTTGTACAGCTC
Duet2mchefor	ATTAGTTAAGTATAAGAAGGAGATATACATATGGTGAGCAAGGG
Duetmchery2rev	AAATTTCGCAGCAGCGGTTTCTTTACCAGACTTGTACAGCTCGTCC

Appendix D – SD Media

Media	Additional Components
SD-LW	200 mg/L adenine, 200 mg/L histidine
SD-LWH	200 mg/L adenine
SD-LWHA	None
X-α-gal	40 µg/mL where used