

Structural analysis of LMO2 for the development of a small molecule inhibitor

Helen Sewell (MBiolSci)

Submitted in accordance with the requirements for
the degree of Doctor of Philosophy

The University of Leeds

Leeds Institute of Molecular Medicine

April, 2010

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

I am very grateful to the Leukaemia and Lymphoma Charitable Trust for The Gordon Pillar studentship award. I would like to thank my supervisor Professor Terence Rabbitts for giving me the opportunity to carry out my PhD studies within the Section of Experimental Therapeutics and for his continued support and guidance over the last three years. I would also like to thank Dr. Tomoyuki Tanaka, Dr. Chi Trinh and Dr. Arnout Kalverda for their invaluable advice throughout my research. I would also like to thank Dr. Erika Mancini and Dr. Kamel El Omari for their collaborative support, which resulted in the crystal structure described in this thesis.

I would like to acknowledge Dr. James R. Ault and Dr. Jeff N Keen for the Mass spectrometry analysis; Andrew Baron for the Analytical ultracentrifugation data; Dr Narcis Fernandez-Fuentes for the *in silico* structural modeling and Dr Tomoyuki Tanaka for the LMO2 mutagenesis data.

Abstract

LMO2 is a nuclear LIM-only protein encoded by a gene located on chromosome 11p13 and was originally discovered through its activation due to specific chromosomal translocations in patients with T-cell acute lymphoblastic leukaemia (T-ALL). The specific chromosomal translocations occur with either the T-cell δ receptor gene (14q11) or T-cell β receptor gene (7q35) and result in aberrant LMO2 expression in T-cells. Transgenic mouse models of LMO2 induced T cell neoplasias showed that enforced LMO2 expression caused accumulation of immature thymic T cells, followed by clonal T cell tumours with long latency. LMO2 is therefore a specific therapeutic target as not only is it associated with chromosomal translocations but is also expressed in approximately 50% of T-ALL. The aims of this project were to structurally determine LMO2 for structure based drug development of small molecules that will target LMO2 protein-protein interactions.

LMO2 could not be purified alone as removal of the fusion tag resulted in severe precipitation of the free LMO2. Consequently, LMO2 was co-expressed with an antibody single domain termed VH#576, and purified to a high yield and purity. A final construct of LMO2, spanning residues 9 to 147, bound to VH#576 has been crystallised and the structure solved, to a medium resolution of 3.3Å, using phase information from single anomalous dispersion (SAD) data in combination with molecular replacement. Using a mammalian two-hybrid mutagenesis screen, key VH#576 binding residues have been identified. This data can be used, in combination with the crystal structure of VH#576/LMO2 Δ N7 Δ C11, to produce a Pharmacophore model for *in silico* screening and lead drug discovery. In addition to the crystallography approach, NMR was also investigated as a means to collect structural data on VH#576, in solution. A protocol has been developed to isotopically label and purify VH#576 along with unlabeled LMO2, in order to increase the stability of the antibody single domain for NMR data acquisition. Solving the structure of VH#576 by NMR requires further data collection. Advantageously, NMR solution structures represent more physiological environment and comparison of a VH#576 NMR structure and the crystal structure would enable the detection of any crystallisation artifacts.

Table of contents

Acknowledgements.....	i
Abstract.....	ii
Table of contents.....	iii
List of figures.....	ix
List of tables	xii
List of abbreviations	xiii
List of amino acids.....	xvi
1 Introduction	1
1.1 Human leukaemias	2
1.1.1 Chromosomal translocations and transcription factors.....	2
1.1.2 Factors associated with chromosomal translocations.....	3
1.1.3 Consistent features of leukaemic cells.....	3
1.1.4 Acute myeloid leukaemias.....	4
1.1.5 Acute lymphoid leukaemias	5
1.1.6 T-cell acute lymphoblastic leukaemia	5
1.2 LMO2 forms a DNA binding complex	7
1.3 The role of LMO2 in development.....	8
1.3.1 Control of LMO2 transcription.....	8
1.3.2 Role of LMO2 in haematopoiesis.....	8
1.3.3 Role of LMO2 in haematopoietic and endothelial development.....	11
1.3.4 Role of LMO2 in the activation of GATA1	12
1.3.5 Role of LMO2 in vascular development	13
1.4 LIM-only protein family	13
1.5 LMO2 expression in T-cells results in Leukaemia	16
1.5.1 T-cell acute lymphoblastic leukaemia	16
1.5.2 LMO2 DNA binding complex.....	18
1.5.3 LMO2 is not involved in normal T-cell development.....	18
1.5.4 Effect of LMO2 on T-cell development	18
1.6 T-ALL and associated translocations	21
1.6.1 V(D)J recombination	21
1.6.2 Alternative genes involved in T-ALL.....	22
1.6.3 Secondary mutations are required for the onset of T-ALL.....	25
1.6.4 X-linked severe combined immunodeficiency gene therapy trial	25

1.7	LMO2 expression in other forms of cancer.....	28
1.7.1	Diffuse large B-cell lymphomas.....	28
1.7.2	Prostate cancer.....	28
1.7.3	Pancreatic cancer.....	29
1.7.4	Role of LMO2 in tumour angiogenesis.....	29
1.8	T-ALL treatment.....	29
1.8.1	Alternative treatment for T-ALL.....	30
1.8.2	Anti-LMO2 intracellular antibody domain as the basis for small molecule drug design	31
1.9	Study aims.....	33
2	Methods and theory of structural techniques.....	35
2.1	DNA analysis.....	36
2.1.1	Agarose gel electrophoresis.....	36
2.1.2	Restriction endonuclease digestion.....	36
2.1.3	Quantification of DNA.....	36
2.2	Cloning methods.....	36
2.2.1	Polymerase chain reaction.....	36
2.2.2	Vector preparation and ligation reaction.....	38
2.2.3	Preparation of Electro-competent Cells.....	38
2.2.4	Electroporation.....	39
2.2.5	Plasmid miniprep purification.....	39
2.2.6	DNA sequencing service.....	39
2.2.7	Large scale plasmid purification.....	40
2.2.8	Plasmid glycerol stock.....	40
2.3	Protein purification screen at Oxford protein production facility.....	40
2.3.1	Preparation of vectors for protein expression screen at Oxford protein production facility.....	40
2.3.2	Protein expression screen at Oxford protein production facility.....	41
2.4	Protein purification methods.....	41
2.4.1	Expression of GST-LMO2.....	41
2.4.2	Expression of NusA-LMO2.....	42
2.4.3	Expression of MBP-LMO2.....	43
2.4.4	TEV protease purification.....	43
2.4.5	Expression of VH#576-LMO2.....	43
2.4.6	Methylation of surface lysine residues.....	44
2.4.7	Gel Filtration Chromatography.....	45
2.4.8	Ion exchange chromatography.....	45

2.5	VH#576 purification and stability trials	45
2.5.1	VH#576 expression vector preparation	45
2.5.2	Purification of VH#576	45
2.5.3	Removal of His tag from VH#576	46
2.5.4	VH#576 solubility trials	46
2.5.5	Addition of CHAPS to VH#576 sample buffer	46
2.6	Purification of isotopically labeled VH#576 with unlabeled LMO2	47
2.6.1	Purification of ¹⁵ N labeled VH#576/LMO2ΔN7 protein complex	47
2.6.2	Cloning of LMO2 expression vector	47
2.6.3	Cloning of NusA-Tev-LMO2 expression vector	47
2.6.4	Co-lysis of LMO2 and VH#576	47
2.6.5	Production of deuterated VH#576	48
2.6.6	Increasing the yield of deuterated VH#576 protein	48
2.6.7	VH#576/LMO2 stability trials	49
2.6.8	Purification of ¹⁵ N/ ¹³ C/ ² H labeled VH#576 with LMO2	49
2.6.9	Purification of ¹⁵ N labeled VH#576 with LMO2	49
2.7	Protein analysis	49
2.7.1	Protein quantification	49
2.7.2	SDS-PAGE	50
2.7.3	Western blot Analysis	50
2.7.4	Circular Dichroism	51
2.7.5	Analytical centrifugation	51
2.7.6	Mass spectrometry	52
2.7.6.1	MALDI-MS fingerprinting	52
2.7.6.2	Electrospray ionisation mass spectrometry	53
2.8	X-ray crystallography methods	53
2.8.1	Crystallography screen	53
2.8.2	Diffraction data collection	53
2.9	NMR methods	54
2.9.1	Protein sample preparation	54
2.9.2	Experiment preparation	54
2.9.3	¹ H- ¹⁵ N HSQC	54
2.9.4	¹ H- ¹³ C HSQC	55
2.9.5	HNCA and HN(CO)CA	55
2.9.6	NOESY and TOCSY	55
2.9.7	Data processing	55
2.10	Cell culture methods	55

2.10.1	Chinese hamster ovary cell transfection	55
2.10.2	Luciferase reporter assay	56
2.10.3	Western blot analysis	56
2.11	Commonly used reagents	57
2.11.1	Buffers and solutions	57
2.11.1.1	Buffers for DNA analysis	57
2.11.1.2	Buffers for protein analysis	57
2.11.2	Nutrients broths	58
2.12	Theory behind methods used for structural analysis	59
2.12.1	NMR theory	59
2.12.2	X-ray crystallography theory	62
3	Expression and purification of LMO2 for X-ray crystallography trials	66
3.1	Introduction	67
3.2	Protein expression and purification methods	68
3.3	Purification of GST-LMO2 fusion protein	68
3.3.1	Expression and purification of GST-LMO2	68
3.3.2	Strategy to reduce GST-LMO2 degradation	73
3.3.3	Identification of the 60 KDa contaminant protein and removal strategy	75
3.3.4	Size exclusion chromatography to remove GST only from GST-LMO2	77
3.3.5	Purification of GST-LMO2 via ion exchange chromatography	80
3.4	Expression and purification of NusA-LMO2	84
3.5	Protein expression and purification screen	87
3.5.1	OPPF Expression and purification screen	87
3.5.2	Comparison and further analysis of NusA-LMO2 and MBP-LMO2	89
3.5.3	Analysis of NusA-LMO2 and MBP-LMO2 by Circular Dichroism	91
3.5.4	MBP-LMO2 for crystallography trials	92
3.5.5	MBP-LMO2 rigid linker strategy	94
3.6	Expression and purification of LMO2 with antibody single domain, VH#576	99
3.6.1	X-ray crystallography screen with VH#576/LMO2 protein complex	103
3.6.2	Removal of N-terminal residues from LMO2 and expression with antibody single domain, VH#576	104
3.6.3	Lysine methylation of VH#576/LMO2 Δ N7 protein complex	110
3.6.4	Removal of C-terminal residues from LMO2 and expression with antibody domain VH#576	112
3.6.5	VH#576/LMO2 Δ N7 Δ C11 X-ray crystallography screens	116
3.6.6	Crystallisation optimisation	121
3.7	Discussion	124

4	Structural analysis of an anti-LMO2 antibody single domain by NMR	126
4.1	Introduction	127
4.2	Protein expression and purification methods	129
4.3	VH#576 protein purification and NMR results	129
4.3.1	Initial VH#576 purification	129
4.3.2	VH#576 size exclusion chromatography	131
4.3.3	VH#576 stability trials	133
4.3.4	Effect of pH on VH#576 NMR experiments	133
4.4	VH#576/LMO2 protein purification and NMR data acquisition	140
4.4.1	¹⁵ N labeled VH#576/LMO2ΔN7	140
4.4.2	Co lysis of VH#576 and LMO2	142
4.4.3	Deuterated VH#576	144
4.4.4	Increasing the yield of deuterated VH#576	146
4.4.5	Scale up of the purification of deuterated VH#576 with LMO2	148
4.4.6	VH#576 stability trial	150
4.4.7	Final VH#576/LMO2 sample for analysis	152
4.4.8	Comparison of free and bound VH#576 at 20°C	155
4.5	Discussion	157
4.6	Conclusion	158
5	Structural determination of LMO2 bound to anti-LMO2 antibody single domain and characterisation of critical interaction residues	160
5.1	Introduction	161
5.2	VH#576 and LMO2 mutagenesis	163
5.2.1	Analysis of VH#576 point mutations using a mammalian two-hybrid luciferase reporter assay	163
5.2.2	Mammalian two-hybrid method	165
5.2.3	Effect of VH#576 point mutations on the interaction with LMO2	165
5.2.4	LMO2 interaction region	169
5.3	VH#576/LMO2ΔN25 <i>in silico</i> structural model	171
5.3.1	The use of docking software	171
5.3.2	Interactions illustrated by the VH#576/LMO2ΔN25 <i>in silico</i> model	174
5.4	Solving the crystal structure of VH#576/LMO2ΔN7ΔC11 protein complex ..	175
5.4.1	Methods used to solve the structure	175
5.4.1.1	Diffraction data collection	175
5.4.1.2	Diffraction data processing	175
5.4.2	Structure Solution	177

5.4.2.1	Structure refinement.....	178
5.4.3	Crystal structure of VH#576/LMO2 Δ N7 Δ C11	181
5.4.4	Analysis of the interaction between VH#576 and LMO2 Δ N7 Δ C11	186
5.4.5	Interface area	191
5.5	Discussion	191
5.5.1	A comparison between VH#576/LMO2 Δ N7 Δ C11 crystal structure and <i>in silico</i> structural model.....	192
5.5.2	Comparison with other LIM structures	192
5.5.3	Effect of VH#576 upon LMO2 activity within the cell.....	193
5.6	Conclusion.....	194
6	General discussion and future work.....	196
6.1	General discussion.....	197
6.1.1	Medium resolution X-ray structure of VH#576/LMO2 Δ N7 Δ C11	197
6.1.2	The LMO2 DNA binding protein complex	198
6.1.3	VH#576 has the potential to inhibit the activity of LMO2.....	200
6.1.4	VH#576 and small molecule mimetics.....	201
6.1.5	LMO2 as a drug target.....	205
6.2	Conclusion.....	205
	References.....	207
	Appendix A.....	228
	Appendix B.....	238
	Appendix C.....	244
	Appendix D.....	257

List of figures

Figure 1.1: The distribution of translocation generated oncogenes among the acute lymphoblastic leukaemias.....	6
Figure 1.2: The haematopoietic tree	10
Figure 1.3: LMO2 protein and sequence	14
Figure 1.4: LMO2 gene translocations	17
Figure 1.5: A distinct LMO2 DNA binding protein complex found in leukaemic T cells....	17
Figure 1.6: Effect of LMO2 on T-cell development.....	20
Figure 1.7: Retroviral insertion results in ectopic LMO2 expression	26
Figure 2.1: NMR experiment setup	60
Figure 3.1: Western blot analysis of GST-LMO2 whole cell lysate	70
Figure 3.2: Purification of GST-LMO2 via glutathione sepharose affinity chromatography	72
Figure 3.3: Addition of chaperone expression vectors to reduce GST-LMO2 degradation..	74
Figure 3.4: Addition of ATP-MgCl ₂ for removal of bound chaperone	76
Figure 3.5: Western blot analysis of GST-LMO2 co-expressed with chaperone and trigger factor proteins	78
Figure 3.6: GST-LMO2 size exclusion chromatography	79
Figure 3.7: GST-LMO2 anion exchange chromatography	81
Figure 3.8: Mono Q anion exchange chromatography	83
Figure 3.9: Comparison of LMO2 fusion tags.....	85
Figure 3.10 NusA-LMO2 size exclusion chromatography.....	86
Figure 3.11: OPPF protein expression screen.....	88
Figure 3.12: Comparison of NusA-LMO2 and MBP-LMO2 purification	90
Figure 3.13: His-MBP-LMO2 3C protease digest.....	93
Figure 3.14: Expression and purification of MBP-LMO2 rigid linker constructs	95
Figure 3.15: Size exclusion chromatography of MBP-AAA-LMO2	96
Figure 3.16: Beckman XL-I Analytical ultracentrifugation	98
Figure 3.17: His-VH#576/LMO2 purification by Ni-NTA chromatography.....	100
Figure 3.18: VH#576/LMO2 size exclusion chromatography chromatogram	101
Figure 3.19: VH#576/LMO2 size exclusion chromatography load and pooled fractions...	102
Figure 3.20: Probability of disorder of LMO2 as predicted by RONN.....	105
Figure 3.21: Effect of various LMO2 N-terminal truncations on the purification with VH#576.....	107
Figure 3.22: Purification of VH#576/LMO2ΔN7 complex.....	108

Figure 3.23: VH#576/LMO2ΔN7 purification by size exclusion	109
Figure 3.24: The purification of lysine methylated VH#576/LMO2ΔN7 by size exclusion chromatography	111
Figure 3.25: Purification of LMO2, with various C-terminal truncations, with VH#576...	113
Figure 3.26: Size exclusion purification of VH#576/LMO2ΔN7ΔC11 expressed from C41	114
Figure 3.27: VH#576/LMO2ΔN7ΔC11 expressed from C41 and purified by size exclusion	115
Figure 3.28: Microscope image of crystal from reagent N5 of the Matrix screen	118
Figure 3.29: Time course of crystal growth.....	118
Figure 3.30: Silver stain analysis of VH#576/LMO2ΔN7ΔC11 crystals	120
Figure 3.31: Optimisation experiments	121
Figure 3.32: Crystal from three row optimisation dilution screen.....	123
Figure 4.1: Purification of VH#576 via N-terminal His tag	130
Figure 4.2: ¹ H- ¹⁵ N HSQC spectrum for antibody single domain VH#576.....	130
Figure 4.3: Purification of VH#576 via Ni-NTA affinity and size exclusion chromatography	132
Figure 4.4: ¹ H- ¹⁵ N HSQC spectrum for antibody single domain VH#576.....	134
Figure 4.5: ¹ H- ¹⁵ N HSQC spectra comparison for VH#576 at pH 7 and pH 8	136
Figure 4.6: ¹ H- ¹³ C HSQC spectrum for antibody single domain VH#576.....	138
Figure 4.7: ¹ H- ¹⁵ N HSQC spectra comparison for VH#576 and VH#576/LMO2ΔN7 complex.....	141
Figure 4.8: Co-lysis and Ni-NTA affinity chromatography purification of VH#576 and LMO2	143
Figure 4.9: Purification of 50% deuterated VH#576 bound to LMO2	145
Figure 4.10: Revised strategy for expression of 50% deuterated VH#576	147
Figure 4.11: Purification of 50% deuterated VH#576 bound to LMO2.....	149
Figure 4.12: NMR protein sample stability trial.....	151
Figure 4.13: VH#576 labeled with ¹⁵ N/ ¹³ C/ ² H bound to unlabeled LMO2.....	153
Figure 4.14: ¹ H- ¹⁵ N HSQC spectra comparison for VH#576 in uncomplexed and complexed states	154
Figure 4.15: ¹ H- ¹⁵ N HSQC spectra comparison for VH#576 in uncomplexed and complexed states	156
Figure 5.1 Mammalian two-hybrid luciferase reporter assay	164
Figure 5.2: VH#576 CDR mutagenesis to determine residues critical for the interaction with LMO2	166

Figure 5.3: VH#576 residues involved in the interaction with LMO2	168
Figure 5.4: Regions of LMO2 involved in the interaction with VH#576	170
Figure 5.5: VH#576/LMO2 Δ N25 <i>in silico</i> structural model.....	172
Figure 5.6: VH#576/LMO2 Δ N25 <i>in silico</i> structural model: The protein-protein interface	173
Figure 5.7: Diffraction pattern collected for a VH#576/LMO2 Δ N7 Δ C11 crystal	176
Figure 5.8: 2F _o -F _c electron density map of VH#576/LMO2 Δ N7 Δ C11	179
Figure 5.9: Crystal structure of VH#576/LMO2 Δ N7 Δ C11	182
Figure 5.10: LMO2 forms a dimer in the asymmetric unit.....	183
Figure 5.11: Superimposition of the FLINC2 and VH#576/LMO2 Δ N7 Δ C11 structures ...	185
Figure 5.12: Important VH#576 residues for the interaction with LMO2	188
Figure 5.13: Key interactions between VH#576 and LMO2 Δ N7 Δ C11	189
Figure 5.14: Important hydrophobic interactions between VH#576 and LMO2 Δ N7 Δ C11	190
Figure 6.1: Key interaction region between LMO2 and VH#576.....	203

List of tables

Table 1.1: The <i>LMO</i> gene family encode oncogenic proteins	15
Table 1.2: Translocation involving TCR genes in T-All	24
Table 2.1: PCR master mix for 4 X 50µl reactions	37
Table 2.2: General PCR protocol.....	37
Table 2.3: Ligation reaction.....	38
Table 2.4: CD parameters	51
Table 3.1: Fusion proteins and respective solubility	88
Table 3.2: Results of the pre crystallisation test	103
Table 3.3: Further crystallisation reagents identified	119
Table 5.1: Data Collection and refinement statistics for VH#576/LMO2ΔN7ΔC11 structure	180
Table 5.2: Atomic contacts between VH#576 and LMO2ΔN7ΔC11	187
Table 5.3: Further atomic contacts between VH#576 and LMO2ΔN7ΔC11	188
Table 5.4: Surface area of VH#576 required for the interaction with LMO2ΔN7ΔC11	191

List of abbreviations

AD:	Activation domain
ADME:	Adsorption, distribution, metabolism and excretion
Ag:	Antigen
ALL:	Acute lymphoblastic leukaemia
AML:	Acute myeloid leukaemia
ATP:	Adenosine triphosphate
β -Gal:	β -Galactosidase
bHLH:	basic helix-loop-helix
bp:	Base pairs
CCD:	Charge-coupled devices
CD:	Circular Dichroism
CDR:	Complementarity determining region
cDNA:	complementary DNA sequence
CHO:	Chinese hamster ovary cell line
COSY:	Correlation spectroscopy
CV:	Column volume
DBD:	DNA binding domain
DNA:	Deoxyribonucleic acid
dNTP:	2'-deoxynucleoside 5'-triphosphates (dATP, dCTP, dGTP, dTTP)
DMSO:	Dimethyl sulfoxide
DTT:	Dithiothreitol
E. coli:	Escherichia coli
ES:	Embryonic stem cells
FCS:	Foetal calf serum
g:	Grams
g:	Units of gravity
GDP:	Guanosine diphosphate
GFP:	Green fluorescent protein
GST:	Glutathione S-transferase
GTP:	Guanosine triphosphate
GuHCl:	Guanidine hydrochloride

HCl:	Hydrochloric acid
HEPES:	N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HSQC:	Heteronuclear Single Quantum Coherence
IDab:	Single domain intracellular antibody
Ig:	Immunoglobulin
IPTG:	Isopropyl-β-D-thiogalactopyranoside
Kb:	Kilo base
KDa:	Kilo dalton
L:	Liter
LMO	LIM-only protein
LID	LIM interaction domain
M:	Molar
MEL	Murine erythroleukaemia cell line
MgCl ₂ :	Magnesium chloride
mRNA:	messenger Ribonucleic acid
MALDI-MS:	Matrix assisted laser desorption mass spectrometry
MW:	Molecular weight
NaCl:	Sodium chloride
NaOH:	Sodium hydroxide
Ni:	Nickel
NOESY:	Nuclear Overhauser effect spectroscopy
nm:	Nano meter
NMR:	Nuclear magnetic resonance
OD:	Optimum density
O/N:	Over night
PBS:	Phosphate buffered saline
PCR:	Polymerase chain reaction
PEG:	Polyethylene glycol
PMSF:	Phenylmethylsulphonyl fluoride
PVDF:	Polyvinylidene fluoride
RNA:	Ribonucleic acid
RONN:	Regional order neural network
rpm:	Revolutions per minute
RT:	Room temperature

S:	Svedberg unit
SAR:	Structure activity relationship
SDS-PAGE:	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
scFv:	Single chain variable fragment
UV:	Ultra violet
V:	Volts
VH:	Variable heavy chain
VL:	Variable light chain
Wt:	Wild type

List of amino acids

Amino acid	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 Introduction

1.1 Human leukaemias

Some human cancers are associated with chromosomal mutations, either through a change in the number of chromosomes through nondisjunction or through a change in chromosome structure. Chromosomal aberrations can be described as a variation from the wild type chromosome structure. There are four common types of chromosomal mutations: deletions, duplications, inversions and translocations (Russell, P.J., 2002). All four are initiated by breaks in the chromosome. Reciprocal chromosomal translocations are a hallmark of leukaemias and lymphomas and invariably convert proto-oncogenes to oncogenes. This can result in the transformation from a differentiating cell to a tumour cell with abnormal proliferation.

Leukaemia generally arises from haematopoietic stem cells, which involve the peripheral blood and bone marrow. The leukaemias may be myeloid or lymphoid in origin and this can be determined by their cellular morphology (microscopy and cell surface markers), molecular cytogenetics (e.g. Fluorescent *in situ* hybridisation) and genomics (e.g. reverse transcription PCR). Leukaemia is further subdivided into acute or chronic forms (Harris et al., 1997). Chronic leukaemia is characterised by a variable level of terminal differentiation, as well as the presence of mature blood cells. In contrast, acute leukaemia is distinguished by low levels of differentiation, to mature haematopoietic cells, such as lymphocytes or neutrophils. Somaticly acquired chromosomal translocations have been found in 65% of the acute leukaemias (Look, 1997). These structural chromosomal rearrangements effect gene expression and act to disrupt normal cell proliferation and differentiation pathways.

1.1.1 Chromosomal translocations and transcription factors

The most frequent targets of chromosomal translocations are genes that encode transcription factors emphasising the critical role of these ‘master’ regulatory proteins in the control of blood cell development (Rabbitts, 1991). As such, nuclear transcription factors are often involved in oncogenic signaling pathways, activating the gene expression patterns that lead to malignancy. These balanced chromosome translocations result in the formation of gene fusions, which may lead to neoplastic tumour transformation in one of two ways. Translocations may activate proto-oncogenes by being juxtaposed with a strong promoter or tumour specific fusion proteins may be created. Translocations that inappropriately activate transcription factor genes in acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML) frequently result in a blockage of haematopoietic cells in defined stages of differentiation (Stegmaier et al., 1995).

1.1.2 Factors associated with chromosomal translocations

It has been suggested that DNA double strand breaks are required for chromosomal aberrations that result in gene fusions (Mitelman et al., 2007). The risk of developing a translocation has been linked to inducing agents such as DNA topoisomerase II poisons. DNA topoisomerase II, acts during replication and regulates DNA topology through transient DNA cleavage of both strands followed by strand passing and religation (Martin-Cordero et al., 2003). Chemotherapeutic agents, which poison DNA topoisomerase II, are thought to act by aborting the DNA rejoining step and thus inducing apoptosis. However it has been found, in for example breast cancer patients, that treatment can predispose patients to secondary leukaemias with balanced chromosomal rearrangements (Mistry et al., 2005). This suggests a role, for DNA topoisomerase II mediated cleavage of DNA, in forming translocations. Radiation has also been implicated in DNA double strand breaks and hence chromosome translocations.

Interphase (G1, S, G2) is a stage during the cell cycle in which DNA replicates. The chromosomes are not visible and the DNA appears as uncoiled chromatin. Chromosome structure during this stage may play a role in the formation of gene fusions (Russell, P.J., 2002). An example of this is the BCR-ABL1 fusion gene, the product of a translocation, which occurs in chronic myelocytic leukaemia, and accounts for 92% of cases. The loci involved in the BCR-ABL1 fusion have been found to be close to each other in the corresponding normal cell (Neves et al., 1999). Another factor, which enables illegitimate recombination, is shared sequence motifs at the chromosome breakpoints (Aplan, 2006). However shared sequence motifs account for a minority of gene rearrangements hence, the mechanism of gene fusions at a DNA level is a subject of ongoing research.

The point in time when translocations such as gene fusions occur, in different cell types, is also the focus of much research. In haematological malignancies there is evidence, from twin studies and PCR analysis, of specific gene fusions in Guthrie spots (sample of neonatal blood) (Greaves and Wiemels, 2003). This suggests the gene fusions are formed *in utero*, several years before overt leukaemia. This single genetic event is not sufficient for overt leukaemia and this is supported by a concordance rate, for leukaemia in twins, of 10%. Second genetic abnormalities may occur post-natally and this results in the onset of leukaemia (Riggi et al., 2005).

1.1.3 Consistent features of leukaemic cells

The leukaemic cells of any one patient all derive from a single mutant progenitor cell (Bonnet and Dick, 1997) whose transformation initiates clonal expansion and subsequent

overt, clinical leukaemia. The development of leukaemia is a stepwise process with an early initiating event such as a chromosome translocation and an increasing number of somatic mutations resulting in an increasingly transformed clonal population of cells. Leukaemic cells or blasts have an increased ability to survive conditions such as growth factor deprivation or stresses that would cause normal cells to undergo apoptosis. Thus, there are general mechanisms underlying leukaemic transformation including increased cell survival, increased proliferation capacity, increased self-renewal capacity, genomic instability and prevention of differentiation (Passegue et al., 2003).

1.1.4 Acute myeloid leukaemias

Myeloid cells are defined as any leukocyte that is not a lymphocyte. AML is characterised by a large number of abnormal cells that fail to differentiate into functional granulocytes or monocytes (Fialkow et al., 1987). The translocations associated with AML seem to aberrantly activate genes in primitive cells that have retained both multilineage and self-renewal capacity (Bonnet and Dick, 1997). The most common target of associated translocation is the acute myeloid leukaemia 1 core binding factor β (AML1-CBF β), a transcription factor which directly binds the enhancer core DNA sequence TGT/cGGT (Licht, 2001). AML1 has been found to fuse with ETO (eight-twenty-one); the fusion protein retains the ability to bind the enhancer core DNA sequence but not the transcriptional activation domain, and contains almost the entire length of ETO. The chimeric gene generated greatly enhances self-renewal of the haematopoietic stem cell population and blocks differentiation of committed progenitor cells (Okuda et al., 1998, Mulloy et al., 2002). Transformation of AML1-ETO may alter the transcriptional regulation of normal AML1 target genes.

MLL (mixed lineage leukaemia) is another example of a protein that forms a fusion protein and has been found to cause AML. Translocations t(4;11) (q23;p13.3), t(9;11) (q23;p13.3), t(11;19)(q23;p13.3), fuse MLL with AF-4 (ALL-1 fused gene from chromosome four), AF-9 (ALL-1 fused gene from chromosome nine) and ENL (eleven-nineteen-leukaemia) respectively (Rubnitz et al., 1999). These fusion partners are all serine/proline rich proteins with nuclear localisation signals. The C terminal of ENL has transcriptional regulation activity in vitro. MLL-ENL induces myeloid leukaemia when introduced into mice, however this was not seen with truncated MLL suggesting MLL-ENL mediates tumorigenic activity through a gain of function mechanism (Lavau et al., 1997). This mouse model also demonstrates the consistency of the association between these genetic aberrations and subtypes of leukaemia.

1.1.5 Acute lymphoid leukaemias

Acute lymphoblastic leukaemia (ALL) is the most common malignant disease of childhood and one of the leading causes of childhood death from diseases in developed countries (Rubnitz and Crist, 1997). ALL can also affect adults and involves either the T or B-cell lineage. T and B-cell genes are constantly rearranging their T-cell receptor and immunoglobulin genes respectively and therefore are susceptible to translocations (Zhang and Rowley, 2006). These rearrangements are thought to arise from errors in the recombination process. Furthermore, such translocations can result in the activation of an oncogene in a new transcriptional environment (Klein, 2000). In contrast to AML, the initiating translocation often generates an oncoprotein, which transforms committed progenitors at a stage of differentiation, which matches the majority of cells in the leukaemic clone. The distribution, of translocation generated oncogenes, among the acute lymphoblastic leukaemias is shown in Figure 1.1.

1.1.6 T-cell acute lymphoblastic leukaemia

Despite an increase in the knowledge of molecular pathways relating to leukaemias, 60% of adult patients diagnosed do not survive beyond five years (based on Cancer Research UK statistics). This has led to a novel approach to improve the efficacy of therapeutics available to treat leukaemia by targeting specific proteins in the disease pathway. LMO2 is potentially, one such therapeutic target as it is ectopically expressed in a subset of patients with T-cell acute lymphoblastic leukaemia (T-ALL). The *LMO2* gene is a consistent target of t(11;14)(p13;q11) translocations in T-ALL, as well as t(7;11)(q35;p13). *LMO2* over expression can also be demonstrated in T-ALL where these translocations are lacking (Ferrando et al., 2002). The translocations detailed lead to the juxtaposition of *LMO2* downstream of TCR promoter genes; *LMO2* is expressed as a normal protein product but with abnormal control of transcription (Larson et al., 1996). Initially the focus of the remainder of this chapter will be the normal role of *LMO2*, in haematopoiesis (Warren et al., 1994). This will be followed by an in depth evaluation of the involvement of *LMO2* in the development of T-ALL. The mechanism of tumorigenesis, initiated by aberrant *LMO2* expression, is not clear however theories will be discussed. The current treatment available for patients with T-ALL and options for development will also be reviewed.

Introduction

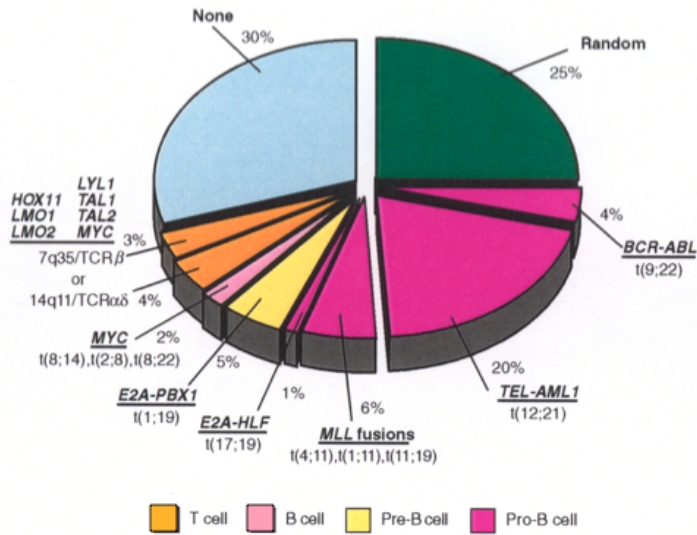


Figure 1.1: The distribution of translocation generated oncogenes among the acute lymphoblastic leukaemias

In acute lymphoblastic leukaemia (ALL), the gene fusions tend to be specifically associated with a specific lymphocyte subset: T-cell, B-cell, pre B-cell and pro B-cell. The T-cell leukaemias, account for 7% of ALL. Figure taken from (Look, 1997).

1.2 LMO2 forms a DNA binding complex

Lmo2 has been identified as an essential protein for erythropoiesis since null mutations of this gene result in a blockage in erythroid development (Warren et al., 1994). Mouse models with null mutations introduced into the genes Gata1, Gata2 and Tal1 (also known as SCL) were also found to prevent erythropoiesis, which implies a close functional relationship. Yeast two-hybrid experiments indicate a complex of Lmo2, TAL1, E47 and GATA1 forms in haematopoietic cells (Osada et al., 1995). It was shown TAL1 and E47 can bind the E-box motif as a heterodimer and GATA proteins bind to the GATA site however there is no evidence to suggest Lmo2 can bind DNA. Murine erythroleukaemia (MEL) cells have been used to analyse the LMO2 protein complex using CASTing (cyclic amplification and selection of targets) experiments (Wadman et al., 1997); the cellular extracts were mixed with a random oligonucleotide pool and specific complexes immunoprecipitated with anti-Lmo2 antiserum. After several rounds of binding and immunoprecipitation, the oligonucleotides were cloned and sequenced. This experiment identified bipartite DNA motif of an E-box followed, 9bp downstream, by a GATA site. In erythroid cells the complex has been characterised as E2A (E47/E12), Gata1, Ldb1, Lmo2 and Tal1 (Wadman et al., 1997). The LMO2 complex may vary throughout haematopoietic differentiation controlling various stage specific functions by regulating particular sets of genes.

A recent proteomics study also using MEL cells expanded the number of proteins present in the complex (Ldb1/Lmo2/Tal1/E2A/Gata1) to include HEB, Lmo4, and Lyl1 (closely related to Tal1) and a number of single stranded DNA binding proteins (Ssbp-1-4) important for the stability of the Ldb1 protein. In the proerythroblast-proliferating state, this complex was found to interact with another complex consisting of Gata1/Tal1/E47/HEB/Mtgr1/Eto2. This complex was also found to interact with the cell cycle regulator cdk9 and E2-2, with the equilibrium favouring the interaction of all these proteins to form a large multiprotein complex (Meier et al., 2006). Upon induction of differentiation of MEL cells, the level of Eto-2 drops whereas the level of Lmo4 rises and this may favour dissociation of the two smaller complexes. The repressive role of Eto-2 may be a mechanism for the activation of late erythroid genes. These findings highlight the dynamics of the LMO2 protein complex and reflect a mechanism for transcriptional control according to the differentiation stages of the erythroid cell; this will be discussed further in the next section.

1.3 The role of LMO2 in development

1.3.1 Control of LMO2 transcription

LMO2 is part of a transcriptional network that controls blood and endothelial development, it is composed of six exons and transcribed from two promoters, a distal promoter upstream of exon one and a proximal promoter upstream of exon three (Royerpokora et al., 1995). Eight enhancer elements have been identified (Landry et al., 2009) in mice, which appear to control the full pattern of endogenous *Lmo2* expression in embryos (midgestation). *In silico* comparative genomics and *in vivo* CHIP-chip studies were used to analyse the control of *Lmo2* expression. The extended proximal promoter was found to direct *Lmo2* expression in endothelial cells when bound by ETS transcription factors (SFPI1, FLI-1). *Lmo2*/*Tal1* and Gata-factors bind upstream (distal enhancer) haematopoietic elements and when both the distal and proximal promoters are activated, *Lmo2* is expressed in erythroid and fetal liver cells. Therefore, it is the combination of distal and proximal promoters, which confer the expression of *Lmo2* in haematopoietic tissues.

1.3.2 Role of LMO2 in haematopoiesis

Gastrulation is the period, after the blastocyst stage (inner cell mass from which the embryo arises) of embryonic development, in which cell masses become organised into three distinct germ layers. The mesoderm is one of three primary germ cell layers along with the ectoderm and endoderm. The mesoderm gives rise to blood cells in a process known as haematopoiesis (Figure 1.2). Prenatally, haematopoiesis occurs in the yolk sac, then in the liver and eventually in the bone marrow. In mouse, the yolk sac is the initial site of blood development where large nucleated erythroblasts first arise between embryonic day seven and eight. This is defined as primitive haematopoiesis and is followed by definitive haematopoiesis at embryonic day seven (Landry et al., 2008). All blood cells are derived from haematopoietic stem cells that are thought to originate from the Aorta-gonad-mesonephros region and then colonise in the fetal liver where definitive haematopoiesis is established. As will be discussed, LMO2 is essential for both definitive and primitive haematopoiesis (Yamada et al., 1998, Warren et al., 1994).

Lmo2 null mutation in mice leads to failure of yolk sac haematopoiesis and embryonic lethality around E10.5 (Warren et al., 1994). In fact, the knock out of transcription factors *Gata2/Gata1*, *Lmo2* and *Tal1* each result in the absence of primitive haematopoiesis (Porcher et al., 1996, Shivdasani et al., 1995, Robb et al., 1995, Fujiwara et al., 2004, Warren et al., 1994).

Introduction

The role of Lmo2 in adult mice was studied using chimeric mice made by injecting blastocysts with embryonic stem cells carrying the homozygous null mutations ($Lmo2^{-/-}$) of *Lmo2*. Analysis of these adult mice showed that the $Lmo2^{-/-}$ ES cells made no contribution to any haematopoietic lineage, indicating a critical role for Lmo2 in definitive haematopoiesis. Furthermore, introduction of an *Lmo2* expression vector rescues the $Lmo2^{-/-}$ embryonic stem cells and enables them to contribute to all haematopoietic lineages tested (Yamada et al., 1998). The early phenotype of these knockouts is explained by the observation that Ldb1, Cdk9, E2A, Lmo2, Gata1 and Eto2 are expressed around E9.5 in the Aorta-gonad-mesonephros region of the mouse embryo that gives rise to haematopoietic stem cells (Meier et al., 2006). Figure 1.2 shows the expression of LMO2 during haematopoietic stem cell development.

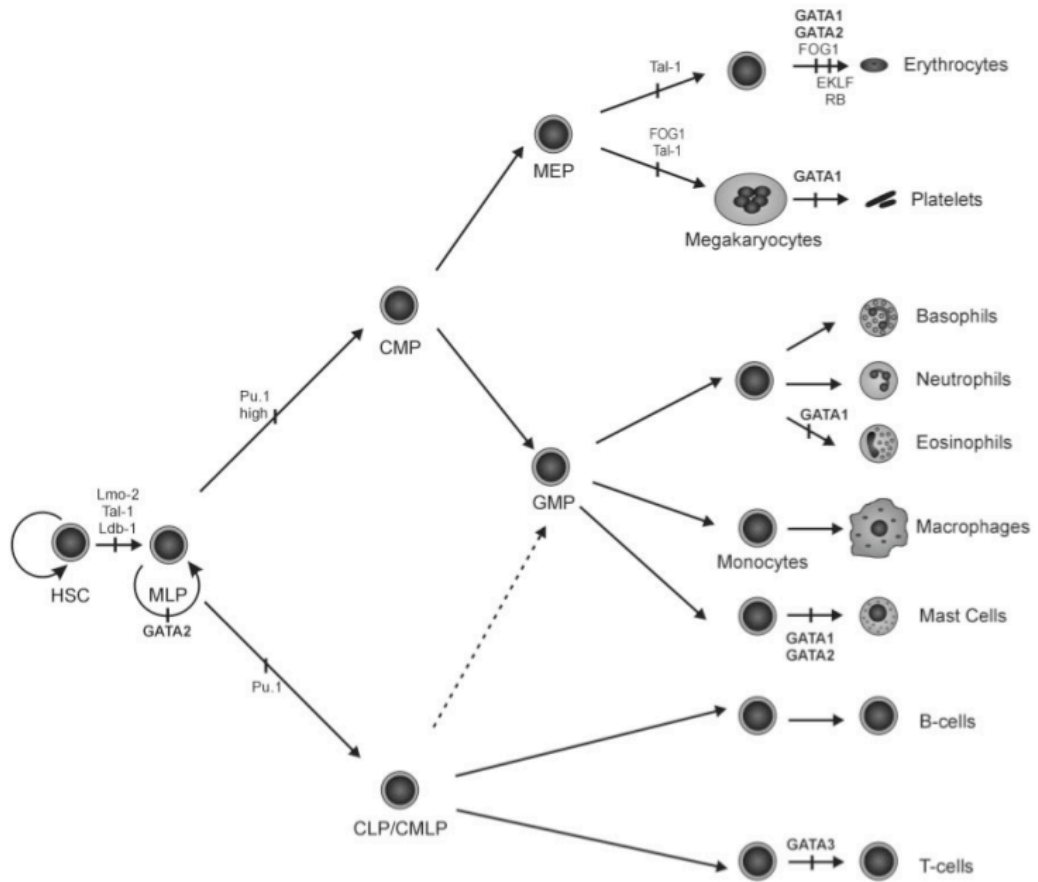


Figure 1.2: The haematopoietic tree

This diagram represents the main lineage commitment steps in haematopoiesis. The haematopoietic stem cell (HSC) gives rise to the multi-lineage progenitors (MLP), which can differentiate, into the haematopoietic lineages. MLPs differentiate into lymphoid and myeloid lineages in the common lymphoid progenitor (CLP) and common myeloid progenitor (CMP). CLPs can give rise to T and B cells only while CMPs give rise to megakaryocyte-erythrocyte progenitors (MEP) and granulocyte-monocyte progenitors (GMP). It has also been suggested that the first lineage commitment separates myeloid and erythroid potential, in the CMP, from myeloid lymphoid potential in the common myeloid lymphoid progenitor (CMLP). CMLPs can then further differentiate into B cells, T cells, and GMPs (as shown by the dashed line). Transcription factors relevant for the development of specific haematopoietic lineages are indicated. LMO2 has also been found to be expressed in erythroid and megakaryocytic cells (Warren et al., 1994). Figure taken from (Ferreira et al., 2005).

1.3.3 Role of LMO2 in haematopoietic and endothelial development

Tal1 and Lmo2 have been shown to drive haematopoietic and endothelial development through conversion of progenitors from early mesoderm to haemangioblasts, the putative progenitors contributing to both haematopoietic stem cells and vascular endothelial cells. It has also been shown that Tal1 and Lmo2 can synergistically act, during primitive haematopoiesis, to differentiate non-axial mesoderm into haemangioblasts in Zebrafish embryos (Gering et al., 2003).

Haematopoietic gene expression was analysed in Lmo2 null Zebrafish. *Lmo2* morpholinos (25-mer antisense oligonucleotide with modified bases containing a morpholine ring, making them very stable) were used to knock down gene expression. The effects were loss of haematopoiesis and disruption of endothelial development leading to loss of the dorsal aorta. In addition, these phenotypes are the same as those found for *Tal1* morphants (organism that has been genetically modified with a morpholino). This again suggests that Tal1 and Lmo2 work together in a multiprotein complex during mesoderm development and differentiation to haemangioblast. The only exception to this was a difference in the control of *Runx1* in the absence of Tal1 or Lmo2. *Runx1* is a heterodimeric transcription factor also involved in definitive haematopoiesis. The expression of *Runx1* was found to differ slightly in the posterior lateral plate mesoderm. In the absence of Tal1, *Runx1* began to recover at an earlier time point than in the absence of Lmo2 (Patterson et al., 2007). It may be that Lmo2 interacts with unknown genes that rescue expression of *Runx1* in the posterior lateral plate mesoderm in the absence of *Tal1*. The expression of glycoprotein *draculin* (anti-coagulant) and haematopoietically expressed homeobox protein *Hhex* were also analysed in *Lmo2* morphants and were found to be initially independent but after seven somites (mesodermal segments that form sequentially), increasingly dependent on Lmo2 expression.

Haematopoiesis rescue studies using *Tal1*^{-/-} ES cells demonstrated that the phenylalanine within the second helix of Tal1 is necessary for interaction with Lmo2 (Patterson et al., 2007). Lmo2 interaction with Tal1 was shown to be required for erythroid (*Gata1* and *Pu.1*), myeloid (*Pu.1*) and endothelial (*Flt4*) gene expression. Erythroid differentiation only occurs in the pronephric mesoderm when Lmo2 and Tal1 induce Gata1 expression. It was found that in the absence of Gata1, Tal1-Lmo2 induced haemangioblasts differentiate into endothelial cells (Gering et al., 2003).

Further proteins have been found to interact with the erythroid DNA binding complex (Tal1, Gata1, Ldb1 bound to E-box-GATA DNA sequence motif) including single stranded DNA-

binding proteins, Ssbp2 and Ssbp3. Specifically, Ssbp2 was found to augment the transcription of the Protein 4.2 (P4.2) in an Ldb1 dependent manner through interaction with the P4.2 proximal promoter (Xu et al., 2007). P4.2 is a major component of the erythrocyte cell membrane skeleton. Quantitative reverse transcription PCR experiments showed that over expression of Ssbp2, increased β -globin and P4.2 levels in murine erythroleukaemia (MEL) cells suggesting Ssbp2 has a role as a positive regulator in erythroid progenitors. Finally, Ssbp2 was found to inhibit the interaction of both Ldb1 and Lmo2 with E3 ubiquitin ligase, Rlim, and prevent Rlim-mediated Ldb1 ubiquitination and thus protect Lmo2 and Ldb1 from proteasomal degradation (Xu et al., 2007). Thus, single stranded DNA-binding proteins may play a role in regulating the abundance of LMO2.

The LMO2 complex has been found on the promoter regions of several other essential haematopoietic specific genes, such as retinaldehyde dehydrogenase 2, erythroid Kruppel-like factor, c-kit, α -globin and glycophorin A (GPA) (Wang et al., 2007). The ability of Lmo2 to mediate the expression of such a large number of genes is thought to be due to its ability to participate in multi-protein complexes. LMO2 is down regulated during the final stages of erythroid differentiation to release TAL1 and GATA1. Ectopic expression of LMO2 at this stage interferes with this release and inhibits erythropoiesis (Terano et al., 2005).

1.3.4 Role of LMO2 in the activation of GATA1

GATA1 is a haematopoietic transcription factor known to be essential for normal erythropoiesis (Pevny et al., 1995). LMO2 has been shown to be involved in the activation but not the repression of GATA1 target genes. To explore this further, a recent study examined the composition of GATA1 associated protein complexes at sites where GATA1 acts as an activator and at sites where it acts as a repressor using an erythroid cell line (Tripic et al., 2009). The TAL1, LMO2, LDB1, E2A complex was found at all sites where GATA1 activates transcription. In contrast, GATA1 fails to recruit this complex at sites where GATA1 acts as a repressor. In the same study, an *LMO2* shRNA (short hairpin RNA) was introduced in order to knock down LMO2 expression. The mRNA level of genes repressed or activated by GATA1 were measured using quantitative reverse transcription PCR (Tripic et al., 2009). This showed a decrease in the level of GATA1 activated genes (*Hbb-b1* and *Eraf3*), whereas the expression of genes usually repressed by GATA1 were unaffected (*Kit*, *Gata2*, *Lyl1*). This research indicates, the TAL1, LMO2, LDB1, E2A protein complex represents a tissue specific GATA1 co-activation complex but not repression. Alternatively the complex may be removed at repression sites by a proteasome mediated turnover of LMO2 or LDB1.

1.3.5 Role of LMO2 in vascular development

Development of a vascular system is essential for embryos to grow after they reach a certain size to allow blood flow to the rapidly developing embryonic tissues. The vascular system is constructed from two distinct processes, vasculogenesis and angiogenesis. Vasculogenesis forms the primary capillary network from haemangioblasts (primitive precursors specified from mesoderm) while angiogenesis is the process by which mature vascular structures are formed by the remodeling of endothelial cells from the existing capillary network. *Lmo2*-null ES cells were followed in mouse chimeras and were found to contribute to the capillary network normally until embryonic day nine (Yamada et al., 2000). Moreover, these ES cells do not contribute to endothelial cells of large vessel walls after day ten. These results suggest that *Lmo2* is required for angiogenesis but not for capillary formation from the mesoderm (vasculogenesis). Hence, the *Lmo2* transcription factor mediates specific phases of angiogenesis as well as haematopoiesis.

1.4 LIM-only protein family

LMO2 is one of four proteins in the LIM-only (LMO) family, a group of transcription factors that contain two tandem LIM domains. The LIM only protein family consists of LMO1, LMO2, LMO3 and LMO4 and all members have the ability to mediate specific protein-protein interactions (Foroni et al., 1992). These proteins act as adaptors to mediate the assembly of large protein complexes and have vital roles in both normal development and tumorigenesis (Table 1.1).

In general, LMO proteins are located in the nucleus but lack a nuclear localisation sequence; it is thought that binding of LDB1 maintains these proteins in the nucleus (Kenny et al., 1998). LMO proteins are made up of a type A LIM domain followed by a type B LIM domain. The domains are composed of two zinc binding LIM fingers (Figure 1.3) and are generally made up of four β strands followed by a short α helix. The only significant difference between type A and type B topologies is that type B contains a 3_{10} -helix, rather than a short α helix, after the final β hairpin (Dawid et al., 1998).

Introduction

A:

```
          10          20          30          40          50
MSSAIERKSL DPSEEPVDEV LQIPPSLLTC GGCQQNIGDR YFLKAIDQYW

          60          70          80          90          100
HEDCLSCDLC GCRLGEVGRR LYYKLGRKLC RRDYLRLFGQ DGLCASC DKR

          110         120         130         140         150
IRAYEMTMRV KDKVYHLECF KCAACQKHFC VGDYLLINS DIVCEQC IYE
```

WTKINGMI

B:

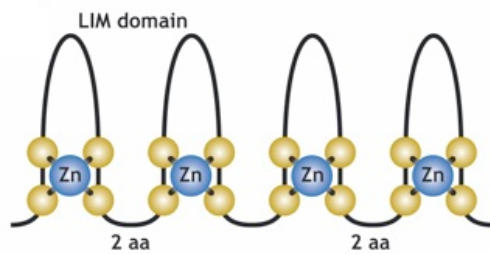


Figure 1.3: LMO2 protein and sequence

A: The amino acid sequence (in single letter code) of LMO2 with LIM1 and LIM2 represented in blue and purple respectively. The 158 amino acid protein has a molecular weight of 18.36 KDa and isoelectric point of 6.79.

B: LIM domains are composed of two zinc-binding LIM fingers. The residues that coordinate with each zinc ion are highlighted in yellow (cysteine, histidine, aspartate) in the sequence (A). Figure adapted from (McCormack and Rabbitts, 2004).

Introduction

LMO1 and *LMO2* (formerly known as *RBTN1* and *RBTN2*) were originally identified through their association with specific chromosomal translocations involved in T-ALL (Rabbitts et al., 1999), and have a sequence identity of 58.3%. Recently, *LMO3* has been identified as a potential translocation partner of T-cell receptor β locus through the use of chromatin conformation capture on chip (4C) technology (Simonis et al., 2009). The translocation t(7;12)(q35;p12.3) was identified in cell lines and samples from patients with T-ALL. Furthermore, *LMO3* has recently been implicated in neuroblastomas (Aoyama et al., 2005) and associates with HEN2, a basic helix-loop-helix protein. It has been proposed that *LMO3* may differentially regulate the expression of downstream target genes involved in neuronal differentiation or tumour formation. *LMO2* and *LMO3* have a sequence identity of 59.5%.

LMO4 was originally identified as a breast cancer auto antigen and is aberrantly expressed in 56% of breast tumours, of which 65% have an amplification of the *ERBB2* gene (Visvader et al., 2001). *LMO4* associates with BRCA1 and represses its transcriptional activity (Sum et al., 2002). Loss of *LMO4* results in G2/M arrest (Montanez-Wiscovich et al., 2009). *LMO2* and *LMO4* have 46.6% sequence identity. The oncogenic properties of the *LMO* gene family are summarised in Table 1.1.

Table 1.1: The *LMO* gene family encode oncogenic proteins

Gene	Chromosome	Translocation	Leukaemia	Others
<i>LMO1</i>	11p15	t(11;14)(p15;q11)	T-ALL	
<i>LMO2</i>	11p13	t(11;14)(p13;q11) t(7;11)(q35;q11)	T-ALL	Diffuse large B-cell lymphoma Pancreatic Prostate
<i>LMO3</i>	12p12.13	t(7;12)(q35;p12)	T-ALL	Neuroblastoma
<i>LMO4</i>	1p22.3			Breast cancer

1.5 LMO2 expression in T-cells results in Leukaemia

1.5.1 T-cell acute lymphoblastic leukaemia

T-cell acute lymphoblastic leukaemia (T-ALL) is a neoplastic disorder of the lymphoblasts (immature lymphocyte cells) committed to the T-cell lineage (Graux et al., 2006). Acute lymphoblastic leukaemia is the most common malignancy diagnosed in children (approximately 2 to 5 years of age) and occurs slightly more frequently in boys than girls. T-ALL is associated with an enlarged thymus and early spread to the cerebrospinal fluid. “Paediatric T-ALL is an aggressive malignancy of thymocytes that accounts for about 15% of ALL cases and for which treatment outcome remains inferior compared to B-lineage acute leukaemias” (van Vlierberghe et al., 2008). This neoplastic disorder originates in the thymus, due to a block in T-cell development and results in a deficiency in the number of mature lymphocytes, which can fight infection. Leukaemic transformation of immature thymocytes, is the result of a number of genetic abnormalities that permit uncontrolled cell growth (Mansour et al., 2007).

In many cases T-ALL is initialised by a disruption in the rearrangement of T-cell receptor (TCR) genes, which results in a chromosomal translocation, and thus activation of an oncogene such as HOX11, TAL1 or LMO2. LMO2 is found at the breakpoints of two translocation which in total account for 7% of T-ALL cases; t(11;14)(p13;q11) and t(7;11)(q35;p13) (Garcia et al., 1991). LMO2 is located on the short arm (p) of human chromosome 11 at band 13 (11p13). The chromosome 11p13 breakpoint has been found to be paired with either the TCR δ gene on chromosome 14 (band q11) or, the TCR β gene on chromosome 7 (band q35). LMO2 has two transcriptional promoters and the majority of known LMO2 translocations occur upstream of one or both of the LMO2 promoters (Boehm et al., 1991). Enforced LMO2 expression results from these translocations and causes a block in T-cell development.

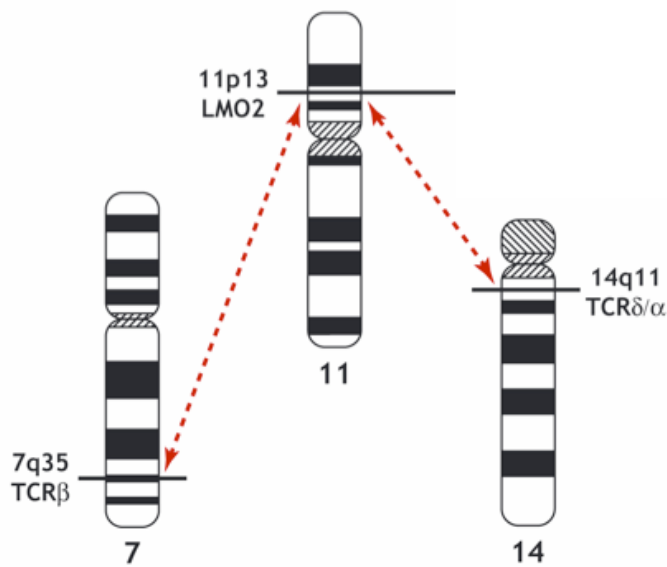


Figure 1.4: LMO2 gene translocations

Chromosomal translocation between LMO2, chromosome 11, and either TCR δ/α on chromosome 14 and TCR β gene on chromosome 7.

Adapted from (Nam and Rabbitts, 2006)

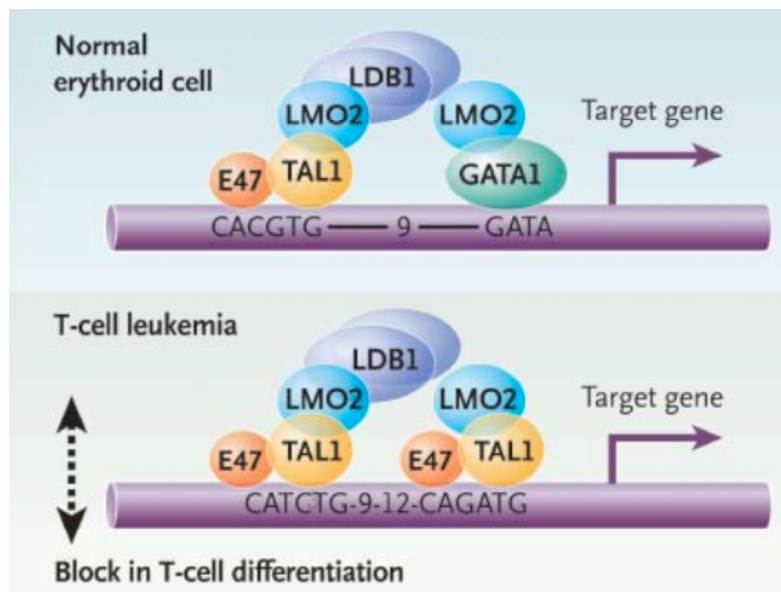


Figure 1.5: A distinct LMO2 DNA binding protein complex found in leukaemic T cells

A distinct complex of LMO2, TAL1, E47 and LDB1 was found in leukaemic T cells from Lmo2 transgenic mice. The complex found in normal erythroid cells (top) also contains GATA1. Adapted from (McCormack and Rabbitts, 2004).

1.5.2 LMO2 DNA binding complex

The LMO2 complex present in the T-cells of *Lmo2* transgenic mice and transgenic tumours was studied via electrophoretic mobility shift assay and antibody mediated super shifts (Grutz et al., 1998b) and was found to be distinct from the complex found in normal erythroid cells. The complex was found to bind two E-boxes with the most common spacer being 10bp long. In comparison with the erythroid-specific Lmo2 DNA binding complex, the complex was shown to recruit a second basic helix-loop-helix (bHLH) heterodimer in place of a GATA factor. The complex was limited to the immature double negative thymocyte subpopulation. A putative model (Figure 1.5) of the complex binding to a dual E box motif was produced in which LMO2 is represented as a bridging molecule, binding two bHLH heterodimers (TAL1, E47).

1.5.3 LMO2 is not involved in normal T-cell development

Functional LMO2 is not involved in normal T-cell development. An *Lmo2* conditional knockout mouse model, was utilised to study the effect of *Lmo2* knockout on lymphocyte development. In this model a mouse with *Lmo2* flanked by *loxP* sites (Cre recombinase recognition sites) was crossed with a mouse expressing Cre recombinase, under the control of lymphoid specific promoter (*Rag1*, *CD19* or *Lck*), thus leading to the lymphoid specific deletion of *Lmo2*. No effect on lymphoid development was observed suggesting LMO2 has no role in normal T-cell development (McCormack et al., 2003).

1.5.4 Effect of LMO2 on T-cell development

T-cells mature from common lymphocyte progenitors in the thymic cortex, their stage of development is classified according to the combination of clusters of differentiation (CD) markers expressed. The most immature thymic T-cells do not express CD4 or CD8 and are referred to as double negative (DN) cells. Upon differentiation the double negative population pass through four stages, from DN1 to DN4 and almost all follow the $\alpha:\beta$ pathway (Janeway et al., 2005). Phase one double negative (DN1) thymocytes express Kit and CD44 (Figure 1.6). Genes encoding both chains of the T-cell receptor are in the germline configuration however upon transient expression of CD25, expression of CD44 and Kit are reduced and the T-cell receptor β chain locus undergoes the process of V(D)J rearrangement (McCormack and Rabbitts, 2004). Recombinase activating genes (*RAG1* and *RAG2*) are switched on in order to rearrange the T-cell receptor genes. The β chain expressed is able to pair with pre T-cell receptor α chain to form the pre T-cell receptor which is expressed, along with CD3 molecules, on the cell surface (Janeway et al., 2005). This prevents further β chain rearrangements and expression of both CD4 and CD8 (double

Introduction

positive thymocytes) is followed by cell proliferation (Figure 1.6). Rearrangement of the α chain locus occurs once proliferation stops and cells become small double positive cells. Positive and negative selection occurs in the thymic cortex and medulla of the thymus, respectively. Double positive cells that can recognise self peptides, present on major histocompatibility complex molecules, can undergo positive selection and go on to mature and express high levels of T-cell receptor. These cells become either CD4 or CD8 single positive thymocytes (Janeway et al., 2005).

The effect of *Lmo2* expression in T-cells was analysed by the generation of a mouse model. *Lmo2* transgenic mice express *Lmo2* in thymocytes under the CD2 promoter (a lymphocyte specific promoter that is usually active at the common lymphoid progenitor stage). FACS (fluorescence activated cell sorting) analysis of thymocytes from CD2-*Lmo2* transgenic mice, with enforced expression of *Lmo2*, show an increase in the CD4⁻ CD8⁻ double negative population at the DN3 stage of development (Larson et al., 1995, McCormack et al., 2003) preceding the appearance of clonal T-cell tumours. To allow such clonal selection, the cell of origin must have the property of self-renewal, which is normally reserved for stem cells (McCormack, M.P., et al., 2010). *Lmo2* is normally down regulated at the DN3 stage and it is this ectopic expression of *Lmo2* that results in a block in T-cell development. Transgenic mice with enforced *Lmo2* expression develop thymic tumours with leukaemic blasts at approximately 6 months (Larson et al., 1994).

Tal1 ectopic expression in the thymus results in no tumour growth. However, *Lmo2-Tal1* double transgenic mice display an enhanced onset of T-cell leukaemia (Larson et al., 1996, McCormack et al., 2003), with tumour growth occurring at a faster rate of approximately three months. This may be due to an abolition of E2A protein function in T-cell development. The E2A gene encodes either E47 or E12, which form homodimers in order to regulate lymphoid development. *Tal1* may form heterodimers with E2A, preventing the formation of E2A homodimers, which leads to a block in transcriptional activity. In support of this, development of T-ALL in *Tal1* transgenic mice was enhanced when *E2A* was also knocked out (O'neil et al., 2004). Furthermore, E2a null mice develop leukaemia of a similar phenotype to CD2-*Lmo2* transgenic mice (Bain et al., 1997).

Lmo2 transgenic thymocytes have been found to have long term engraftment potential; this is restricted to double negative 3 thymocytes (CD44⁻ CD25⁺). This suggests LMO2 may promote the self-renewal of pre-leukaemic thymocytes. Thymocyte self-renewal provides a mechanism for committed T-cells to generate secondary mutations in other genes for leukaemic transformation. This may occur by reactivation of a haematopoietic stem cell

Introduction

(HSC) transcriptional program as microarray analysis showed up regulation of several HSC associated genes and down regulation of T-cell developmental genes (McCormack, M.P., et al., 2010).

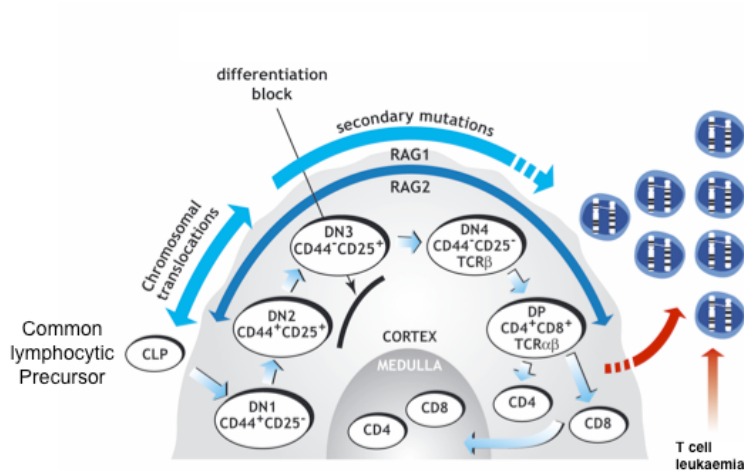


Figure 1.6: Effect of LMO2 on T-cell development

Progenitors from the bone marrow migrate to the outer edge of the cortex, in the sub capsular region of the thymus. Activation of LMO2 by chromosomal translocation results in a block in T-cell development at the CD44⁻CD25⁺ stage (DN3) during a period where RAG recombinase is actively expressed. Secondary mutations in other genes result in the onset of T-cell leukaemia. When no block in development occurs and positive selection is successful, newly mature cells leave the thymus through the medulla.

1.6 T-ALL and associated translocations

The human immune system is made up of adaptive and innate responses. The adaptive immune system has evolved to provide a more extensive means of defense and is made up of mainly T and B lymphocytes. The system is highly adaptive due to somatic hypermutation and V(D)J recombination (Janeway et al., 2005). T cells develop in the thymus and have T cell receptors (TCRs), specially adapted to detect antigens from foreign proteins or pathogens that have entered into the host cell and function to activate a T cell response. TCRs consist of a disulphide heterodimer of highly variable α and β chains expressed at the cell membrane as a complex with CD3. An alternative receptor, expressed in a subset of T cells, is made up of γ and δ chains expressed with CD3. Both of these receptors are expressed with a disulphide-linked homodimer of ζ chains, which has an intracellular signaling function (Janeway et al., 2005). Gene loci encoding α and δ chains are located on chromosome 14q11, while those encoding the β and γ chain are located on 7q35 and 7p15 respectively.

Cytogenetic analysis of lymphoblasts (leukaemic blasts) has found translocations activating oncogenes in 25 to 50% of T-ALL cases. Translocations involving TCR loci are found in approximately 35% of T-ALL (Graux et al., 2006). V(D)J recombination is a process of genetic rearrangement that occurs during the development of all T-cells and results in the specificity of antigen receptors (TCR). During V(D)J recombination many other genes are in an open chromatin conformation and as such are susceptible to the action of recombination enzymes. Illegitimate recombination can result in a juxtaposition of a transcription factor gene and a strong promoter of the TCR genes. In the case of LMO2, the breakpoint occurs upstream of the natural LMO2 promoter, resulting in ectopic expression of LMO2.

1.6.1 V(D)J recombination

Variable chains (α , β , γ and δ) are encoded by several gene segments, which are assembled, in developing lymphocytes, by somatic DNA recombination. Recombination activating genes, RAG-1 and RAG-2, are part of a complex of enzymes that carry out somatic V(D)J recombination (Mcblane et al., 1995). This pair of genes are expressed in developing lymphocytes only while they are engaged in assembling their antigen receptors. The RAG protein complexes specifically recognise recombination signal sequences (Finger et al., 1986). Recombination signal sequences are found adjacent to the coding sequences of V, D or J gene segments and consist of a heptamer-spacer-nonamer motif. The heptamer is a conserved block of seven nucleotides, the spacer is a non-conserved region of twelve or

Introduction

twenty-three nucleotides followed by another conserved region of nine nucleotides, the nonamer. In general, a gene segment flanked by a recombination signal sequence with a twelve base pair spacer can only be joined to one flanked by a twenty-three base pair spacer recombination signal sequence (Janeway et al., 2005).

Theoretically, translocations may occur due to the RAG proteins making a cut at one true recombination sequence which is at the V(D)J region, and another cryptic recombination sequence at the other chromosome involved, resulting in the aberrant rejoining of ends (Agarwal et al., 2006). Thus, the chromosomal loci containing such sequences might be more likely to participate in translocations. Sequences matching the V(D)J recombinase signal were found at the SIL-TAL1 translocation junction, reflecting an illegitimate recombinase activity in the TAL1 gene rearrangement in T-ALL.

Chromosomal translocations involving LMO2 t(7;11)(q35;p13) lead to sequences from chromosome 11 becoming inadvertently joined to TCR β (Garcia et al., 1991). An alternative translocation, t(11;14)(p13;q11) results in a breakpoint that pairs the LMO2 gene with TCR δ (Cheng et al., 1990). About 65% of these translocations are caused by illegitimate V(D)J recombination due to RAG mistargeting of many different LMO2 cryptic recombination signal sequences (sequences which resemble the TCR/Ig recombination signal sequences located next to the LMO2 breakpoint) (Dik et al., 2007). In other cases the breakage in TCR/Ig is induced by RAG while the break in the LMO2 locus is initiated by an unknown mechanism. Thymocytes from eight T-ALL patients were analysed for LMO2 expression levels using real-time quantitative PCR. This experiment provides evidence to support the theory that LMO2 activation in most t(11;14)(p13;q11) translocations, is due to decoupling of the negative regulatory element and not, as currently thought, due to juxtaposition of LMO2 to the TCR δ enhancer. However, the TCR δ enhancer was also shown, in the same experiment, to have a potent role in high LMO2 expression but this was in a less common inversional recombination at V δ 3-D δ 2 signal joints (join between two recombination signal sequences) (Dik et al., 2007).

1.6.2 Alternative genes involved in T-ALL

Another gene involved in T-ALL translocations is the HOX11 gene located on chromosome 10. It encodes a homeodomain protein that can bind DNA and activate transcription (Dear et al., 1993). Reciprocal translocations occur between HOX11 and either TCR δ or TCR β in 31% of adult T-ALL cases.

Introduction

Several DNA binding proteins, activated by translocations in T-ALL, have a bHLH domain i.e. TAL1. This motif has basic amino acids at the amino terminal region which mediate specific DNA recognition of the E box motif (Hsu et al., 1991). LYL1 (lymphoblastic leukaemia derived sequence 1) shares 90% sequence identity with TAL1 in the bHLH motif and have a largely overlapping pattern of expression (Giroux et al., 2007). Studies indicate that LYL1 may be involved in the development of a wide range of blood tumours (Meng et al., 2005). In a Lyl1 transgenic mouse model, driven by human elongation factor 1 α promoter, 30% of mice developed malignant lymphoma after 12 months and were associated with infiltration in multiple organs. Analysis of tumour cells showed they were mainly CD4 CD8 double positive or mature B cells (Zhong et al., 2007).

NOTCH1 mutations have also been found to be involved in T-ALL translocations (Lee et al., 2005). NOTCH1 encodes a heterodimeric receptor that regulates normal T-cell development. The mechanism of T-ALL driven by aberrant NOTCH1 signaling is unclear. Proteolytic processing of NOTCH receptors 1-4, mediated by γ -secretase leads to the cytoplasmic release of the ICN1 intracellular domain which translocates to the nucleus and forms a ternary complex (Pui, 2009). This complex is then able to recruit basal transcription machinery and thus activate NOTCH dependent target genes. Therefore NOTCH may activate downstream oncogenes such as MYC or the PI3K-AKT signaling pathway, which results in the development of T-ALL. NOTCH1 activating mutations have been discovered in more than 50% of patients with T-ALL. Interestingly, a recent study revealed that NOTCH1 mutations were detected in most molecular subtypes of human T-ALL including samples where TAL1 and LMO2 were also activated (Zhu et al., 2006).

Introduction

Table 1.2: Translocation involving TCR genes in T-All

(adapted from (Graux et al., 2006) and (van Vlierberghe et al., 2008))

Translocation involving TCR genes	Gene involved	Function of fusion gene/expressed oncogene
t(7;10)(q34;q24) t(10;14)(q24;q11)	<i>HOX11</i>	Transcription factor
t(5;14)(q34;q32)	<i>HOX11L2</i>	Transcription factor
inv(7)(p15q34), t(7;7)(p15;q34)	<i>HOXA cluster</i>	Transcription factor
t(1;14)(p32;q11) t(1;7)(p32;q34)	<i>TAL1</i>	Transcription factor
t(7;9)(q34;q32)	<i>TAL2</i>	Transcription factor
t(7;19)(q34;p13)	<i>LYL1</i>	Transcription factor
t(14;21)(q11.2;q22)	<i>BHLHB1</i>	Transcription factor
t(11;14)(p15;q11) t(7;11)(q34;p15)	<i>LMO1</i>	Protein-protein interaction
t(11;14)(p13;q11) t(7;11)(q35;p13)	<i>LMO2</i>	Protein-protein interaction
t(1;7)(p34;q34)	<i>LCK</i>	Signal transduction
t(7;9)(q34;q34.3)	<i>NOTCH1</i>	Fate determination, differentiation
t(7;12)(q34;p13) t(12;14)(p13;q11)	<i>CCND2</i>	Cell cycle activator

1.6.3 Secondary mutations are required for the onset of T-ALL

It is thought that the development of T-ALL requires two cytogenetic events such as a translocation and secondary mutation. The mechanism, which results in secondary mutations and onset of leukaemia, remains unclear. A possible explanation, which has already been discussed, is the mechanism of self-renewal to generate secondary mutations. Another possibility is that mutations are mediated by RAG V(D)J recombinase. This hypothesis has been tested by comparing tumorigenesis, in CD2-Lmo2 transgenic mice, in the presence or absence of the Rag1 gene. The proportion of mice, which developed T-ALL within 14 months, was the same for both groups. In addition, there was no significant difference between the tumour incidence curves for the two groups (Drynan et al., 2001). Therefore it can be postulated that Rag1 has no effect on the rate of T-cell tumour development and thus does not mediate secondary mutations.

1.6.4 X-linked severe combined immunodeficiency gene therapy trial

Boys with X-linked severe combined immunodeficiency (SCID) are deficient in the common γ chain of the interleukin-2 receptor, which causes failure of normal lymphocyte development. In a French gene trial CD34+ haematopoietic stem cells were transduced with a defective retroviral vector (moloney murine leukemia) that carried the human *IL2RC* gene and were then transplanted back into the patients. Five out of the twenty boys treated developed T-ALL, three to six years after treatment due to integration of the vector upstream of *LMO2* (Figure 1.7). As integration was in proximity of the *LMO2* promoter, aberrant transcription and expression of LMO2 resulted (Hacein-Bey-Abina et al., 2008). Gamma retroviral vectors preferentially integrate near the 5' end of actively transcribed genes and as a consequence LMO2 is a good target for insertional mutagenesis due to its expression during the early stages of haematopoiesis. A high incidence region of vector integration has been located near exon one of the *LMO2* locus (Yamada et al., 2009).

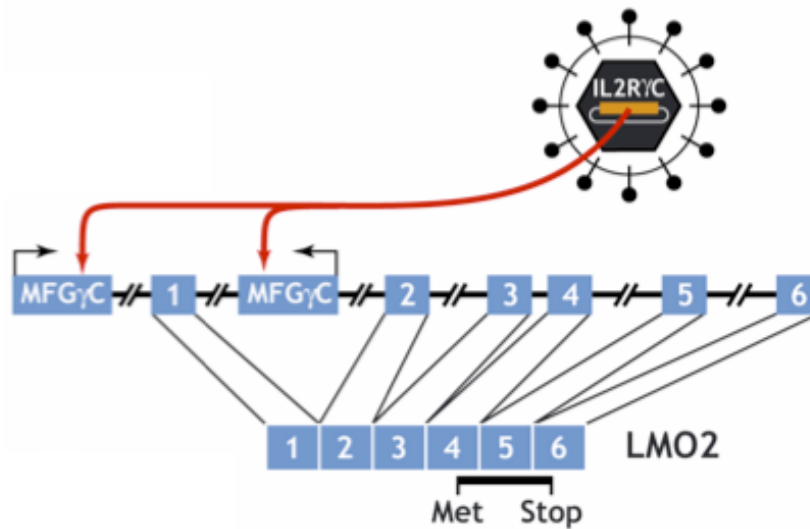


Figure 1.7: Retroviral insertion results in ectopic LMO2 expression

Ex vivo autologous bone marrow derived CD34⁺ cells transduced with a long terminal repeat driven MFG vector resulted in the development of a functional adaptive immune system in sixteen out of twenty-one patients treated. However five of the boys developed leukaemia; four of these cases were due to viral insertion close to the distal promoter of *LMO2*. Insertions were found either just ahead of LMO2 exon one (patient P5) or between exons one and two (patient P4). In both circumstances the LMO2 protein product is identical before and after the genetic abnormality has occurred, and it is the enforced expression of LMO2 that influences T-cell development and the onset of leukaemia.

Met = methionine protein translation-initiation codon,

Stop = protein-translation stop codon.

Figure adapted from (McCormack and Rabbitts, 2004).

Introduction

Two children (patients P4 and P5) that received retrovirus-mediated *IL2RC* gene transfer, as part of the trial, developed clonal proliferation of T lymphocytes thirty to thirty-four months after gene therapy. The loci of retroviral insertion in patient P4 and P5 clones were characterised by linear amplification mediated PCR sequencing of the 5' insertion-site fusion sequence and showed a 100% match to the 5' *LMO2* genomic DNA locus on chromosome 11 (Hacein-Bey-Abina et al., 2003). Clone P4 was mapped to the short arm of chromosome 11, close to the distal promoter of *LMO2* in the reverse orientation. The integration site of P5 was mapped 3Kbp upstream of the first exon of *LMO2*, in the forward orientation. Single nucleotide polymorphism (SNP) analysis supported findings that retroviral cis-activation results in monoallelic *LMO2* expression. The data suggests that viral long terminal repeat (LTR) has an enhancing effect on the activity of the *LMO2* distal (haematopoietic) promoter (Hacein-Bey-Abina et al., 2003). However, the possibility of disruption of a negative regulator of *LMO2* has not been excluded experimentally.

A recent study cloned and sequenced the retroviral integrations from five murine leukaemias containing insertional mutations (Dave et al., 2009). All the *Lmo2* insertions were located 5' upstream of exons four to six. To identify genes that may act together with *Lmo2* in tumorigenesis insertionally mutated genes were identified by ligation-mediated PCR. Microarray data from human leukaemias with up regulated *LMO2* expression was analysed with respect to the insertional mutations identified in murine leukaemias. Transcriptional profiles of these T-ALLs showed high expression of *TALI* and *LYLI* oncogenes as well as *GATA1* and *GATA2*. Down regulation of E2A activated genes and up regulation of E2A repressed genes were also found (Dave et al., 2009). This again suggests *Lmo2* redirects E2A activity by binding through its partners *Tall* or *Lyll*. *Lmo2* clonal tumours show low expression of *Notch1* target genes and higher expression of genes that are repressed by *Notch1* reflecting the insertion in this gene. Despite the sample size of this study being very small, the findings are consistent with the theory that leukaemia requires many "hits" in addition to activation of *LMO2*. In the gene therapy trial, leukaemias developed three to six years after engraftment suggesting a similar course of events (Dave et al., 2009). *LMO2* ectopic expression in T-cells results in an accumulation of cells with blocked differentiation which are susceptible to further mutations; secondary mutations result in leukaemia.

1.7 LMO2 expression in other forms of cancer

1.7.1 Diffuse large B-cell lymphomas

LMO2 is expressed in normal germinal centre B cells and a subset of germinal centre derived B cell lymphomas (Natkunam et al., 2007). The germinal cell provides an environment in which naïve B-cells can diversify their antigen receptors. Deregulation of this step contributes to germinal centre derived B cell lymphomas. It has been shown that the overall survival rate of patients with diffuse large B-cell lymphomas, the most common adult non-Hodgkin lymphoma, is longer if the patient has a gene expression profile similar to that of germinal centre derived B cell lymphomas (Alizadeh et al., 2000). As a result of these findings, a recent study (Natkunam et al., 2008) explored if LMO2 expression is similar. The study provided evidence that LMO2 has prognostic significance in germinal centre B-cell-like diffuse large B-cell lymphomas, when treated with anthracycline based regimens with or without rituximab. No chromosomal translocations or genetic aberrations are known to account for the over expression of LMO2 in this cancer. Gene expression studies have also shown *LMO2* mRNA expression to be the strongest predictor for a superior outcome in diffuse large B-cell lymphoma patients (Lossos et al., 2004). It may be that, in germinal centre B-cells, LMO2 protein interacts with transcription factors, which exert a specific effect on the cell.

1.7.2 Prostate cancer

LMO2 expression has been studied in human prostate tissue samples, cell lines and xenografts. Prostate cancer is the most common malignancy in men in the USA and UK (Rizzo et al., 2005). Analysis of prostate cancer cells revealed LMO2 was predominantly localised in the cytoplasm but also found in the nucleus suggesting LMO2 expression is tissue-type-specific. LMO2 expression was also shown to be associated with the advanced, tumour stage, of prostate cancer as well as the development of distant metastasis (Ma et al., 2007). LMO2 mRNA and protein were found in the more aggressive, androgen independent, cell lines and xenografts but not the androgen dependent. A possible mechanism, which promotes prostate invasion, maybe that LMO2 is able to repress E-cadherin. E-cadherin down regulation decreases the strength of cellular adhesion within a tissue, resulting in an increase in cellular motility thus promoting invasion to the surrounding tissues (Shapiro et al., 1995).

However, a second study (Gratzinger et al., 2009), using a monoclonal anti-LMO2 instead of the polyclonal (as for the previous study) antibody, showed no expression of LMO2 in the nucleus of prostate cancer cells and only weak cytoplasmic staining for a small number

of cases with uncertain significance. For that reason, the role of LMO2 in prostate cancer is unclear and requires further research.

1.7.3 Pancreatic cancer

The expression of LMO2 has been analysed in a large cohort of patients with pancreatic cancer (Nakata et al., 2009). mRNA levels were analysed in eleven different pancreatic cell lines and in cultures of normal pancreatic ductal epithelial cells. All eleven pancreatic cancer cell lines expressed *LMO2* mRNA, the normal pancreatic ductal epithelial cells did not express *LMO2* mRNA. In support of this, *LMO2* mRNA levels of pancreatic cancer tissues were higher than in normal pancreatic tissue. Tissue samples from one-hundred and thirteen pancreatic cancer patients were analysed for *LMO2* mRNA levels and survival curves constructed. Multivariate analysis (statistical analysis method) showed *LMO2* expression was associated with a better prognosis in pancreatic cancer.

1.7.4 Role of LMO2 in tumour angiogenesis

A blood supply is vital for tumour growth and metastasis and is created by remodeling existing blood vessel endothelium, a process known as angiogenesis. Lmo2 expression, has been found to be raised, in tumour endothelium such as mouse thymomas and human lung tumours. To investigate this, a mouse model was created to study the possible role of Lmo2 in tumour angiogenesis (Yamada et al., 2002). Tetracarcinomas (a tumour or group of tumours composed of tissue foreign to the site of growth) were induced in nude mice, by injecting *Lmo2-lacZ* targeted embryonic stem cells. β -galactosidase staining revealed 84% of tumours had an intrinsic Lmo2^{+/-} embryonic stem cell derived visible vascular system. This suggests, Lmo2 is an essential regulator of neo-vascularisation in tumours.

1.8 T-ALL treatment

The current strategy to treat patients with T-ALL is multi-agent chemotherapy using drugs of different classes and modes of action to maximise leukaemic toxicity while keeping resistance and side effects to a minimum (Bernard et al., 1998). Treatment for adults remains poor with only 30 to 40% of patients being long term survivors. Current treatments cure over 70% of children with acute lymphoblastic leukaemia, although many patients develop serious acute and late complications due to the side effects (van Vlierberghe et al., 2008). Some drugs used to treat T-ALL are nonspecific, acting by either blocking protein synthesis through hydrolysis of an amino acid essential for leukaemic growth or by interfering with the mitotic spindle apparatus. Such drugs often produce adverse cytotoxic effects in various normal tissues (Pui and Jeha, 2007, Crazzolara and Bendall, 2009). Drug

resistance can also develop through various cellular mechanisms such as an increased ability of the cell to repair DNA or decreased permeability of the cell (Aroui et al., 2010).

1.8.1 Alternative treatment for T-ALL

The aim now and for the future, is to reduce the toxicity of T-ALL treatment by designing new drugs that specifically interfere with leukaemic pathways and overcome chemo-resistance induced by common treatment regimes (Crazzolaro and Bendall, 2009). According to microarray analysis, LMO2 is expressed in 45% of T-ALL, even in the absence of chromosomal aberrations (Ferrando et al., 2002). Therefore, LMO2 is a potential drug target to treat patients with T-ALL (Nam and Rabbitts, 2006).

One possibility to target translocations, such as those that result in aberrant LMO2 expression, is to design zinc finger proteins to recognise unique chromosomal sites. The idea is to create binding elements which bind a region of the LMO2 gene and block its transcription (Pabo et al., 2001). Zinc fingers have a $\beta\beta\alpha$ structure that fold around a central ion. Tandem sets of fingers can contact neighbouring sites along the major groove of the DNA with the helix fitting into the groove. Still, a delivery strategy would also have to be developed such as retroviral expression vectors.

On an mRNA level, it may be possible to use short interfering double stranded RNA to down regulate *LMO2* by mRNA degradation. Short interfering RNA is an intermediate in the RNA interference pathway (Hannon, 2002) and has been used extensively in genome research to target and inactivate certain genes in order to study their function. The double stranded regions that are formed through binding of siRNAs are targets for the RNA induced silencing complex, which destroys the mRNAs by activation of the enzyme slicer. Cationic liposomes have been developed that can be administered safely *in vivo*. An example of an application of this liposome technology is the delivery of anti-bcl2 siRNA which was shown to have a strong anti-tumour activity, when administered intravenously, in the mouse model of liver metastasis (Yano et al., 2004). One downside to siRNA as a therapy is that the effects are transient and there is the possibility of off-target translational repression. Delivery of the siRNA is also still in the developmental stages with other possibilities including a delivery plasmid (e.g. viral vector) or hydrodynamic injections (McCaffrey et al., 2002).

On a protein level, an ideal way to inhibit activity is through the development of a small molecule, which specifically interacts with the target protein. This requires pockets on the surface of the protein and this may prove problematic, as the interface of protein-protein

interactions are often large and flat. Many successful small molecule therapeutics, designed to inhibit proteins, have targeted an enzyme. For example, the anti-kinase drug ST1571 (Gleevec) which targets the ABL portion of the BCR-ABL fusion protein (Druker et al., 2001). ST1571 causes arrest of growth or apoptosis of haematopoietic cells that express BCR-ABL but does not affect normal cells. The drug produced a cytogenetic response in 54% of patients treated in a phase one clinical trial including seven patients with complete cytogenetic remission. Myelosuppression occurred in 25% of patients, which may have been due to c-kit inhibition or haematopoiesis may have been compromised in patients with leukaemia. The drug binds to the ATP binding site situated within a deep cleft of the fusion protein. ST1571 only binds when the activation loop of the kinase is closed, and the drug stabilises the protein in this inactive conformation thus contributing to the drugs specificity. Resistance to ST1571 is primarily mediated by mutations within the kinase domain of BCR-ABL (Shah et al., 2004) and to a lesser extent due to the amplification of the BCR-ABL genomic locus. This has led to the development of further tyrosine kinase inhibitors.

One major problem posed by LMO2 inhibition is the probable development of anaemia. Therefore, it would be desirable to develop a delivery mechanism to specifically target an anti-LMO2 small molecule to T-cells.

1.8.2 Anti-LMO2 intracellular antibody domain as the basis for small molecule drug design

In response to the problem of developing an anti-LMO2 small molecule inhibitor, anti-LMO2 antibody single domains have been isolated in order to determine sites on LMO2 that have potential for small molecule drug design. A whole antibody is comprised of two heavy chains and two light chains held together by inter-chain disulphide bonds. Each chain has a variable and constant region: The variable regions form the antigen binding site and as such are highly variable between different antibodies, they display three external antigenic loops termed the complementarity determining regions (CDRs). Variable heavy (VH) and variable light (VL) antibody domains can be expressed without the constant region of the antibody; this yields the Fv fragment or single chain variable fragment (scFv) when linked together with a glycine linker. Initially scFv were thought to be the smallest unit capable of specific antigen binding however subsequent studies showed that VH domains can retain a large percentage of the initial binding affinity (Jaton et al., 1968). Furthermore, antibodies lacking VL domains have been found to exist naturally in Camelids (Camels, dromedaries, llamas etc.) (Hamerscasterman et al., 1993).

Introduction

Anti-LMO2 scFv (ALR3) has been isolated and shown to reduce the block in T-cell development in *CD2-Lmo2* transgenic mice (Nam et al., 2008). An anti-LMO2 peptide aptamer has also been identified which binds specifically to Lmo2 and prevents its function as a T-cell oncogenic protein in a mouse transplantation assay (Appert et al., 2009). Antibody fragments or peptide aptamers can be used as a guide to functionally relevant parts of a target protein. Furthermore, the macrodrugs have the capacity to validate LMO2 as a potential drug target. The remainder of this section will discuss the isolation of an anti-LMO2 antibody single domain, VH#576.

Isolation of an anti-LMO2 VH antibody domain, termed VH#576, has been successful using the third generation intracellular antibody capture method (Tanaka and Rabbitts, 2010). Briefly; the antigen LMO2 (amino acids 26-158) was fused to a DNA binding domain (DBD) and a library of single domain antibody fragments fused to an activation domain (AD). The human, intracellular antibody domain (iDab) libraries screened were generated by randomising codons in the CDR three regions. Interaction between bait (DBD) and prey (AD) results in a hybrid transcription factor complex that can activate yeast reporter genes. Both LexA and GAL4 yeast two-hybrid systems were used; interaction results in the reporters activating histidine expression allowing growth in media lacking the amino acid or β -galactosidase expression detected using X-gal. Yeast carrying cDNA encoding the antibody fragment were isolated and the plasmid extracted from yeast for sub library construction. Sub libraries were constructed using assembly PCR with PCR primers designed to randomise the amino acids whose side chains are most exposed in CDR one and CDR two (in direct contact with the antigen based on single domain crystal structure PDB 2UZI). The first round positives were used to randomise CDR two and, usually, the second round positives used to randomise CDR one however, in this case, CDR one affinity maturation was not required. True iDab binders were identified from those that could grow in the absence of histidine by extracting the yeast single domain expression plasmids and re-testing them in Gal4 and LexA baits. The binding affinity of VH#576 for LMO2 was shown, by Surface plasmon resonance, to have an equilibrium dissociation constant (Kd) of 94nM (Tanaka and Rabbitts, 2009).

These antibody single domains can be termed macrodrugs as they have the potential as research tools to validate relevant protein interactomes of disease cells using mouse models. Macrodrugs also have potential as therapeutics however currently there are no successful delivery systems to deliver macrodrugs, such as VH#576, into the cell. An alternative strategy currently under development in the THR laboratory, is to solve the crystal structure of the macrodrug bound to the target protein. This not only identifies key regions of the

target protein to focus on but also provides the information required to screen small molecule compound libraries *in silico* (Goodford, 1985). Information from the crystal structure can be utilised for structure-based lead discovery. Drugs on the market which have originated from this structure-based design approach include those used to treat Human immunodeficiency virus (Agenerase) and were developed using the crystal structure of HIV protease (Goodford, 1985, Tie et al., 2004).

1.9 Study aims

The possibility of eventually specifically targeting the cell-specific LMO2 complex is a plausible therapeutic strategy but as yet, no specific drugs or other reagents are available for treatment. The aim of this project was to develop small molecules that will target LMO2 protein-protein interactions. This work required structural information of LMO2 and of LMO2 bound to the macrodrug, VH#576.

Previous structural work carried out on LMO2 (Matthews et al., 2001) resulted in an NMR solution structure (PDB 1J20) of the N terminal LIM domain (residues 26-87) bound, through a C-terminal flexible linker of eleven residues, to the LIM interaction domain (LID) of Ldb1 (residues 300-339). The crystal structure of LMO4, a protein that shares 46.6% sequence identity with LMO2, has been solved to a high resolution of 1.3Å (Deane et al., 2004). The construct comprised of near full length LMO4 (residues 16-152 of the 166 amino acids) linked to LID via a flexible linker (PDB 1RUT). Upon embarking on this project no other LMO2 structural information was available and as such the main purpose of this project was to crystallise LMO2 and collect X-ray diffraction data in order to solve the structure. Initially, attempts were made to purify LMO2 recombinant protein with and without the anti-LMO2 macrodrugs. However, as will be discussed, recombinant LMO2 could not be purified alone except as a fusion protein because removal of the fusion tag resulted in precipitation of the free LMO2.

Consequently, a bipartite approach was initiated to obtain structural information on VH#576 for the development of a pharmacophore model. Both NMR and X-ray crystallography techniques were utilised in parallel. NMR analysis was carried out on VH#576 both as a free protein and in complex with LMO2. As will be discussed, further NMR data would enable the structure of VH#576 to be solved in solution. In this case, X-ray crystallography proved more successful resulting in the crystal structure of LMO2 (residues 9-147) bound to VH#576. Completion of a VH#576 NMR structure (in the complexed state) would provide a

Introduction

comparison for this crystal structure, which would either verify the structure or identify any crystal contacts, which disturb the local structure.

The second aim of this project was to identify a VH#576 hotspot region; amino acid residues which make a significant energetic contribution to the interaction with LMO2. This aim was fulfilled through the collection of VH#576 mutagenesis data. Vital binding residues and important interacting regions of both VH#576 and LMO2 were identified. This information may also be used to facilitate drug design as the target area has been significantly reduced. Furthermore, these results provide information required for a third structural approach; one of homology modeling followed by docking of the VH#576 and LMO2 models. This resulted in a structural prediction of how the two proteins might interact.

This investigation has led to a crystal structure of LMO2 that provides an insight into the mechanism of interaction of LMO2 and reflects the intrinsic flexibility of the protein. The structure defines the topography of the VH#576 binding interface. Results of a recent erythroid differentiation assay (T. Tanaka, personal communication) has provided evidence that VH#576 can interfere with LMO2 within the cell, presumably by blocking its interaction with other proteins in the DNA binding complex. The VH#576/LMO2 structure appears to represent a conformation of LMO2, which is unable to interact with other proteins in the complex, e.g. LDB1. Therefore if a small molecule can be designed to lock LMO2 in this conformation its interacting ability will be blocked, preventing its downstream tumorigenicity.

Work carried out within this PhD has contributed to a publication (Appendix D):

Appert, A., Nam, C. H., Lobato, N., Priego, E., Miguel, R. N., Blundell, T., Drynan, L., Sewell, H., Tanaka, T. & Rabbitts, T. (2009) Targeting LMO2 with a Peptide Aptamer Establishes a Necessary Function in Overt T-Cell Neoplasia. *Cancer Research*, 69, 4784-4790.

The contribution of H. Sewell towards this paper is detailed as follows:

Design and development of an effective strategy for the expression and purification of an anti-LMO2 peptide displayed on a thioredoxin scaffold (PA-207). H. Sewell purified the recombinant protein, which was then passed onto T. Tanaka for analysis by Surface plasmon resonance.

2 Methods and theory of structural techniques

2.1 DNA analysis

2.1.1 Agarose gel electrophoresis

In order to separate and determine the size of DNA, agarose gel electrophoresis was performed. 1 to 2 % w/v of electrophoresis grade agarose (Invitrogen, Paisley, UK) was dissolved in 100ml final volume of 1 X TBE (Tris/borate/EDTA) by heating the solution in the microwave for 2 minutes. 7 µg of ethidium bromide (10 mg/ml solution) was then added to the agarose. The solution was then poured into a gel casting tray with an appropriate well forming comb and left to set. The DNA was mixed with 6 X DNA loading dye and loaded into the wells along side a 1 Kb DNA marker (Invitrogen, Paisley, UK). The gels were run at 65 V for 1 hour and photographed using a UV transilluminator (Gel documentation system, Bio-Rad).

2.1.2 Restriction endonuclease digestion

Typically 1 µg of DNA was digested using 1 unit of the appropriate restriction enzyme (New England Biolabs, MA, USA) and buffer in a 20µl reaction. When specified, 0.2 mg/ml BSA was added. The digestion was incubated at the temperature recommended, usually 37°C, for 1 hour. Preparative digests for ligation of DNA fragments contained 30 to 50 µg of DNA in a 50 µl reaction and incubated at the appropriate temperature for 3 to 4 hours.

2.1.3 Quantification of DNA

Accurate quantification of DNA was obtained by reading the absorbance of purified DNA at 260nm using the nucleic acid module of the NanoDrop spectrophotometer (Thermo scientific, UK). The spectrophotometer was blanked using 2 µl of TE (10mM Tris-HCl pH 8.0, 1mM EDTA) buffer or water as appropriate. 2 µl of DNA sample was then placed on the optical surface and the absorbance at 260 nm measured. A DNA concentration is given in ng/µl (as calculated by Beer-Lambert law). The 260/280 ratio of the sample absorbance was also noted as this should be 1.8 for “pure” DNA.

2.2 Cloning methods

2.2.1 Polymerase chain reaction

The Peltier thermal cycler (Biorad, Hertfordshire, UK) and Platinum *Pfx* (Invitrogen, Paisley, UK), DNA polymerase (proof reading exonuclease activity for higher fidelity DNA amplification) were used for PCR amplifications. The following reaction mixture and thermo cycling conditions were used (Table 2.1 and Table 2.2). Oligonucleotides were ordered from Invitrogen and 100 µM stock solutions were prepared by adding the

appropriate volume of sterile water and stored at -20°C. Annealing temperatures were calculated by subtracting 5°C from the melting temperature (T_m) of the smallest primer. The stock solution was diluted 1 in 10 and the appropriate volume added to the PCR master mix.

Table 2.1: PCR master mix for 4 X 50µl reactions

Component	Volume (µl)
10 X <i>Pfx</i> buffer	20
10mM dNTP	6
50mM MgSO ₄	4
Upstream primer	6
Downstream primer	6
<i>Pfx</i> polymerase	1.6
Sterile water	152.4
Template DNA or water for the negative control (both added after mix had been divided into aliquots)	1

The master mix was divided into 49 µl reactions (0.2 ml eppendorf tubes used) and either DNA template or water (for the negative control) added as appropriate. A DNA engine Tetrad 2 peltier thermal cycler (Biorad, Hertfordshire, UK) was programmed with the protocol shown in Table 2.2 below.

Table 2.2: General PCR protocol

Number of cycles	Temperature (°C)	Time (Seconds)
1	95	300
35	95	30
	58 (annealing temperature may vary)	30
	68	30
1	68	600
	4	Forever

The PCR reaction was purified using a QIAquick nucleotide removal kit (Qiagen, Crawley, UK), according to the manufacturer's instructions, and digested with the appropriate

restriction enzymes. The PCR product was then loaded onto an agarose gel (up to 2 % w/v agarose Tris/Borate/EDTA gel) and the DNA visualised using ethidium bromide and a UV trans illuminator (long wave to prevent DNA damage). The appropriate band was extracted, using a clean scalpel, and the DNA purified using a QIAquick gel extraction kit (Qiagen, Crawley, UK) according to the manufacturer's instructions. A negative control, with no template DNA, was also analysed, by gel electrophoresis, to check for contamination.

2.2.2 Vector preparation and ligation reaction

Plasmid DNA was digested, sequentially, with the appropriate enzymes using the QIAquick nucleotide removal kit (Qiagen, Crawley, UK) between digests. The final digest reaction was run on a 1 % w/v agarose Tris/Borate/EDTA gel (Sigma-Aldrich, Dorset, UK) and the appropriate band excised using a scalpel. The plasmid DNA was purified using QIAquick gel extraction kit (Qiagen, Crawley, UK). The DNA concentrations of plasmid and insert were quantified using a NanoDrop spectrophotometer. If required the DNA concentration of the cut vector was increased using a Concentrator 5301 (Eppendorf, UK). The standard ligation reaction (3:1 insert:vector molar ratio) used is detailed in Table 2.3. T4 DNA ligase and T4 DNA ligase buffer (250 mM Tris-HCl pH 7.6, 50 mM MgCl₂, 5 mM ATP, 5 mM DTT and 25% PEG-8000) were obtained from Invitrogen. The ligation reactions was set up according to Table 2.3 and incubated overnight at 16°C.

Table 2.3: Ligation reaction

Component	Amount
5 X ligase buffer	4µl
Plasmid	200ng
Insert	Calculated from 3:1 molar ratio
T4 DNA Ligase	1µl
Sterile water	Make up to total reaction volume of 20 µl

2.2.3 Preparation of Electro-competent Cells

5 ml of SOB was inoculated with a single DH5α colony and grown up overnight at 37°C, 255 rpm. The starter culture was diluted (2ml in 198ml SOB) and grown up until the OD at 600 nm read 0.4 (6305 Spectrophotometer, Jenway). The culture was kept on ice for 30 minutes and then centrifuged at 4000 X g for 15 minutes. The cells were resuspended in 200 ml ice cold milli Q H₂O and centrifuged using the same conditions. The cells were then

resuspended in 100 ml ice cold milli Q H₂O. The cells were again centrifuged and resuspended in 10 ml ice cold milli Q H₂O with 10 % glycerol (sterile). After a final centrifuge step, 1 ml ice cold milli Q H₂O with 10 % glycerol was added to the cells. The resuspended cells were divided into aliquots (45 µl). Cells were used directly for electroporation and those not used were frozen by immersion in liquid nitrogen and stored at -80°C.

2.2.4 Electroporation

5 µl of ligation reaction was added to 45 µl competent cells, transferred into an electroporation cuvette (0.2 mm electrode, Biorad) and kept on ice for 30 minutes. The same was done with the cut vector only (amount added to ligation reaction divided by 4) and uncut starting vector to act as negative and positive controls respectively. The cells were electroporated using a Biorad Gene Pulser on the following settings: 2.5 KV field strength, 200 Ω resistance and 25 µF capacitance. The cells were removed from the cuvette and placed into an eppendorf by addition of 150 µl LB and incubated at 37°C, 220 rpm for 1 hour. The cells were plated onto LB (plus appropriate antibiotic) agar plates and incubated overnight. Vector DNA was purified from any resulting colonies by miniprep.

2.2.5 Plasmid miniprep purification

5 to 24 colonies were picked from the ligation plate (depending on the number of colonies present on the negative control plate) and grown up in 5 mls of LB (plus appropriate antibiotic) by overnight incubation at 37°C, 255 rpm. The cells were harvested by centrifugation at 4000 X g for 15 minutes at 4°C. The plasmid was purified using Qiagen Miniprep plasmid purification kit, according to the manufacturer's instructions, and eluted in 30 µl TE (10 mM Tris pH 8.0, 1 mM EDTA) buffer. A diagnostic restriction enzyme digest was performed using the enzyme sites used to ligate the insert. If positive, the digest should release a DNA fragment the size of the insert when analysed by 1 % w/v agarose Tris/Borate/EDTA gel. 2 of the positive clones were sent for sequencing (The sequencing service, The University of Dundee).

2.2.6 DNA sequencing service

DNA sequencing was performed by DNA Sequencing & Services (MRCPPU, College of Life Sciences, University of Dundee, Scotland) using Applied Biosystems Big-Dye chemistry (version 3.1) on an Applied Biosystems model 3730 automated capillary DNA sequencer.

2.2.7 Large scale plasmid purification

A single colony of DH5 α (transformed with the positive clone) was used to inoculate 1 ml 2TY (plus appropriate antibiotic) and incubated for approximately 8 hours at 37°C, 255 rpm. The starter culture was diluted 100 μ l in 50 ml of selective 2TY medium and incubated overnight at 37°C, 255 rpm. The cells were harvested by centrifugation at 4000 X g for 15 minutes at 4°C. The plasmid was purified using Qiagen Hispeed plasmid Midi kit, according to the manufacturer's instructions, and eluted in 500 μ l TE buffer.

2.2.8 Plasmid glycerol stock

5ml of 2TY broth (and appropriate antibiotic) were inoculated with a single colony of DH5 α transformed with the correct clone (as checked by DNA sequencing) and incubated overnight at 37°C, 255 rpm. The culture was centrifuged for 5 minutes at 4000 X g. The cells were resuspended in 2 ml 50% glycerol (sterilised by autoclaving) and incubated on ice. The cells were divided into aliquots and snap frozen by immersing in liquid nitrogen. The tubes were stored at -80°C. To obtain single colonies of bacteria containing the plasmid DNA of interest, a stab of the glycerol stock was streaked, using aseptic techniques, onto an agar plate containing the appropriate selection antibiotics.

2.3 Protein purification screen at Oxford protein production facility

2.3.1 Preparation of vectors for protein expression screen at Oxford protein production facility

The appropriate primer extensions (see Appendix C) were used to enable In-Fusion cloning (Clontech, France) into the prepared pOPIN vectors to derive the desired His-tagged protein constructs. PCR reactions were performed using KOD Hi-Fi polymerase (Novagen, Nottingham, UK) with 30 pmol of each forward and reverse primers, and 1 μ l of plasmid DNA (100ng/ μ l) per 50 μ l reaction. PCR products were purified using AMPure magnetic beads (Beckman-Coulter, UK) according to the manufacturer's instructions and eluted with 50 μ l of TE buffer. Approximately 5 μ l of purified PCR product (10 to 200 ng in total) and 100ng of the appropriately linearised pOPIN vector were mixed in the wells of an In-Fusion Dry-Down 96-well plate (Clontech, France) and incubated at 42°C for 30 min. All reactions were diluted 1:5 with TE buffer and 5 μ l used to transform OmniMaxII T1-phage resistant cells (Invitrogen, Paisley, UK), in 96 tube format. Transformants were selected by plating on 24 well culture plates containing 1ml of LB agar per well, supplemented with the appropriate antibiotic, 0.02% w/v X-Gal and 1mM IPTG. Plates were incubated overnight at

37°C. For each ligation reaction 4 white colonies were used to inoculate separate aliquots of 1.5 ml LB supplemented with the appropriate antibiotic in deep well, 96-well plates. The cultures were grown overnight (37°C, 200 rpm) before harvesting by centrifugation at 5000 X g for 10 min at 4°C. Plasmids were prepared from the cell pellets using a QIAgen BioRobot 8000 and QIAgen Turboprep kits (Qiagen, Crawley, UK), according to the manufacturer's instructions. The resulting plasmids were screened by PCR with a standard T7 forward primer and the reverse primer used for the initial insert amplification. PCR products were analysed, for the correct size, by agarose gel electrophoresis.

2.3.2 Protein expression screen at Oxford protein production facility

Expression constructs were checked by PCR and transformed into Rosetta LysS *E. coli* in a 96 tube format. 35 µg/ml Chloramphenicol was added to maintain the pRareLysS plasmid. Plates were incubated for 18 hours at 37°C before individual colonies were used to inoculate, 500 µl GS96 broth (QBioGene, Cambridge, UK) supplemented with 0.05% v/v glycerol, 1% w/v glucose and 50 µg/ml carbenicillin in 96 well deep well plates. The plates were sealed with gas permeable adhesive seals and shaken at 225 rpm at 37°C for 18 hours. Cells were grown up in auto induction media by inoculating, in 24-well deep-well plates, 2.5 ml of overnight express instant TB media (Novagen, Nottingham, UK) supplemented with 50 µg/ml carbenicillin and with 50 µl of each overnight culture. The diluted cultures were grown at 37°C with shaking at 225 rpm, for 3 hours before reducing the temperature to 25°C and shaking for a further 24 hours at 25°C. A 1.5ml aliquot of culture from each well was then transferred to a 2ml 96 well deep well plate and harvested by centrifugation at 6000 X g for 10 min at 4°C. Cell pellets were frozen at -80°C for at least 30 minutes prior to screening for soluble protein expression using a QIAgen BioRobot 8000 robotic platform and a standard Qiagen Ni-NTA magnetic bead protocol (as per manufacturer's instructions). Soluble protein eluted from the Ni-NTA beads was analysed by SDS-PAGE (Invitrogen NuPAGE Novex 10% Bis-Tris Midi gels with MES buffer system) and visualised with SafeStain (Invitrogen, Paisley, UK).

2.4 Protein purification methods

2.4.1 Expression of GST-LMO2

pGEX-His-Tev-LMO2 (Ampicillin resistance) was used to transform Rosetta pLysS (Chloramphenicol resistance), C41 or B834, by electroporation. A single *E. coli* colony from the appropriate expression host strain was used to inoculate 5 ml LB containing appropriate antibiotics and grown overnight at 37°C, 225 rpm. This culture was used as a starter culture to inoculate 1L LB medium containing appropriate antibiotics and allowed to

grow, under the same conditions, to OD_{600nm} 0.6 before induction of recombinant protein expression by the addition of IPTG to a final concentration of 0.5mM at 16°C. The cells were harvested after 12 to 16 hours of cultivation at 16°C. Please note, for the expression of LMO2, ZnSO₄ was added prior to induction at a final concentration of 0.1mM. Cells were collected by centrifugation (Beckman coulter Avanti J-26 XP with rotor JS 4) at 4000 X g for 20 minutes and resuspended in 30ml of lysis buffer (10mM Na₂HPO₄, 1.8mM KH₂PO₄, pH 7.3, 2.7mM KCl 140mM NaCl) containing EDTA-free protease inhibitor cocktail tablets (Roche Diagnostics, Mannheim, Germany) and 0.1mM ZnSO₄. Cells were lysed using a cell disrupter (Constant cell disruption systems, UK) at 27Kpsi and the lysate centrifuged at 50,000 X g for 50 minutes using a centrifuge (Beckman coulter Avanti J-26 XP), with rotor JA 25.50, to remove the cell debris.

Glutathione-sepharose 4B resin was diluted to a 50% slurry and equilibrated in wash buffer. The cell lysate was incubated with 4ml of glutathione sepharose (per liter prep) for 2 hours at 4°C. The slurry was transferred into a disposable chromatography column (Biorad, Hertfordshire, UK) and the unbound fraction collected for analysis. The column was washed with 50 ml of wash buffer (10 mM Na₂HPO₄, 1.8 mM KH₂PO₄ pH 7.3, 2.7 mM KCl, 140 mM NaCl, 0.1 mM ZnSO₄) and soluble GST-LMO2 eluted with 10 mM glutathione elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM reduced glutathione, 0.1 mM ZnSO₄).

2.4.2 Expression of NusA-LMO2

pET43a-NusA-LMO2 plasmid was used to transform *E. coli* C41. Bacterial cells were cultured in 1L of LB plus Ampicillin (100µg/ml) at 30°C, to an OD_{600nm} of 0.4 to 0.5. Again, ZnSO₄ was added to a final concentration of 0.1 mM and protein expression induced with IPTG (final 0.5mM) at 16°C for 14 hours. Cells were harvested, lysed and cell debris removed as in 2.4.1. In this case, the lysis buffer was made up of 50 mM Na₂HPO₄ pH 8, 300 mM NaCl, 10 mM imidazole, 0.1 mM ZnSO₄, 10 mM β-mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany).

Protein purification was via nickel-agarose chromatography (Novagen, Nottingham, UK) and a disposable column. The bound protein was washed with buffer (50 mM Na₂HPO₄ pH 8, 300 mM NaCl, 20 mM imidazole, 0.1 mM ZnSO₄, 10% glycerol, 10 mM β-mercaptoethanol) and the NusA-LMO2 protein eluted with 50 mM Na₂HPO₄ pH 8, 300 mM NaCl, 250 mM imidazole, 0.1 mM ZnSO₄, 10% glycerol, 10 mM β-mercaptoethanol. The eluted protein was dialysed overnight against 50 mM Na₂HPO₄ pH 7, 150 mM NaCl.

2.4.3 Expression of MBP-LMO2

pOPIN_M-LMO2 plasmid was used to transform *E. coli* Rosetta pLysS (Chloramphenicol resistant). Bacterial cells were cultured in 1L of LB plus carbenicillin (50 µg/ml) and Chloramphenicol (34µg/ml), at 37°C, to an OD_{600nm} of 0.4 to 0.5. Prior to induction, ZnSO₄ was added at a final concentration of 0.1 mM. Protein expression was induced with IPTG (0.5 mM final concentration) at 16°C for 14 hours. Cells were harvested, lysed and cell debris removed as in 2.4.1. MBP-LMO2 protein was extracted from the cell pellet by cell disruption (Constant cell disruption systems, UK) in 20 mM Tris-HCl, pH 8, 250 mM NaCl, 0.1 mM ZnSO₄, 10 mM β-mercaptoethanol and complete, EDTA free, protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany). Protein purification was via amylose resin (New England Biolabs, MA, USA) and a disposable column. The bound protein was washed with lysis buffer without protease inhibitor. MBP-LMO2 protein was eluted in 10 mM Maltose, 20 mM Tris-HCl, pH 8, 250 mM NaCl, 0.1 mM ZnSO₄, 10 mM β-mercaptoethanol. The sample was concentrated using an Amicon Ultra-15 centrifugal filter device (Millipore, MA, USA).

2.4.4 TEV protease purification

pRET3a (Ampicillin resistance) plasmid was used to transform *E. coli* B834. Bacterial cells were cultured in 1L of LB plus Ampicillin (100 µg/ml) at 37°C to an OD_{600nm} of 0.5 and protein expression induced with IPTG (final 0.5 mM) at 22°C for 4 to 6 hours. Tev protease was extracted from the cell pellet by cell disruption (Constant cell disruption systems, UK) in 50 mM Tris-HCl pH 8.0, 500 mM KCl, 1% Triton-X 100, 1 mM PMSF, 10 mM β-mercaptoethanol. The cell lysate was spun down at 35,000 X g for 50 minutes. Protein purification was via nickel-agarose chromatography (Novagen, Nottingham, UK). The bound protein was washed with wash buffer: 50 mM Tris-HCl pH 8.0, 1 M KCl, 10% glycerol, 20 mM imidazole, 1 mM PMSF, 10 mM β-mercaptoethanol. Tev protease was eluted with 50 mM Tris-HCl, pH 8.0, 1 M KCl, 10% glycerol, 250 mM imidazole, 1 mM PMSF, 10 mM β-mercaptoethanol. The sample was dialysed against 50 mM Tris-HCl pH 8, 150 mM NaCl, 20% glycerol, 1 mM DTT and concentrated using an Amicon Ultra-15 centrifugal filter device (Millipore, MA, USA). Tev protease was divided into 500 µl aliquots before storage at -20°C.

2.4.5 Expression of VH#576-LMO2

VH#576-LMO2 was expressed from pRK-His-Tev-VH#576-LMO2 however various N and C-terminal truncated versions of LMO2 were also expressed. To remove these residues the expression plasmid (pRK-His-Tev-VH#576-LMO2) was engineered. This was done using PCR mutagenesis and assembly PCR (see Appendix C for primer design). Primers were

designed to maintain the ribosome binding site in-between VH#576 and LMO2. This was followed by ligation of the fragment back into the expression vector using EcoRI and KpnI restriction enzyme sites.

pRK-His-Tev-VH#576-LMO2 plasmid was used to transform *E. coli* C41. Bacterial cells were cultured in 1 L of LB plus Ampicillin (100µg/ml) at 37°C, 225 rpm and allowed to grow until OD_{600nm} 0.6 at which stage protein expression was induced by adding IPTG to a final concentration of 0.5 mM, at 16°C and cultivated for 14 hours. ZnSO₄ was added prior to induction at a final concentration of 0.1 mM. VH#576-LMO2 protein was extracted from the cell pellet by cell disruption (Constant cell disruption system) in 20 mM Tris-HCl pH 8, 250 mM NaCl, 20 mM imidazole, 1 µM ZnSO₄, 10 mM β-mercaptoethanol and EDTA free, protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany). Protein purification was via nickel-agarose chromatography (Novagen, Nottingham, UK). The bound protein was washed with lysis buffer without protease inhibitor. VH#576-LMO2 protein complex was eluted with 20 mM Tris-HCl, pH 8, 250 mM NaCl, 250 mM imidazole, 0.1 mM ZnSO₄ and 10 mM β-mercaptoethanol. Tev protease (see previous section for preparation) was added to the eluted protein at a ratio of 1:10 and dialysed (14.3 mm inflated tubing dialysis, Scientific laboratory services) overnight against elution buffer without imidazole. After passing the protein through a second Ni-NTA chromatography column (to remove the Tev protease), the sample was concentrated using an Amicon Ultra-15 centrifugal filter device, 10 KDa cut-off (Millipore, MA, USA) for gel filtration chromatography. A Hi-Load superdex 75 column (GE Healthcare, Uppsala, Sweden) was equilibrated with 2 column volumes of 20 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM DTT at a flow rate of 1 ml/min. The sample was loaded onto the column and eluted in 2 column volumes.

2.4.6 Methylation of surface lysine residues

VH#576-LMO2 protein was purified by Ni-NTA chromatography as described in the previous section, and then dialysed into 50 mM HEPES, pH 7.5, 250 mM NaCl during Tev protease digest and then diluted to a concentration of less than 1 mg/ml. After passing the protein through a second Ni-NTA chromatography column (to remove Tev protease), 20 µl 1 M dimethyl-amine-borane complex and 40 µl 1M formaldehyde were added per ml of protein solution and incubated at 4°C, for 2 hours, with gentle shaking. A further 20 µl 1M dimethyl-amine-borane complex and 40 µl 1M formaldehyde were added, per ml of protein solution, and incubated at 4°C, again for 2 hours. Following a final addition of 10 µl 1M dimethyl-amine-borane complex, per ml of protein solution, the reaction was incubated overnight at 4°C. Size exclusion using 20 mM Tris-HCl, pH 7.5, 200 mM NaCl quenched the reaction and the purified protein sample was analysed, for modification of surface lysine

residues, by positive electrospray ionisation mass spectrometry.

2.4.7 Gel Filtration Chromatography

Gel filtration chromatography was carried out using superdex 200 10/300 GL column (GE Healthcare, Uppsala, Sweden) or Hi-Load superdex 75 column (GE Healthcare, Uppsala, Sweden) and an ÄKTA Xpress Instrument (GE Healthcare, Uppsala, Sweden). The column was equilibrated with 2 column volumes (CV) of distilled H₂O at a flow rate of 0.3 ml/min (Superdex 200) or 1 ml/min (Superdex 75) followed by 2CV of equilibration buffer. The sample was prepared by concentration to a volume between 500 µl and 250 µl using a 10 kDa Amicon filter (Millipore, MA, USA). All buffers were filtered using 0.2 µm filters (Whatman, UK) and the sample was centrifuged at 16,000 X g for 10 minutes prior to loading. The protein was loaded onto the column using a super loop and the chromatography run at 0.3 ml/min (Superdex 200) or 1 ml/min (Superdex 75). Protein was eluted from the column using 2 CV of equilibration buffer. The eluate was monitored, by measuring the UV absorption at 280nm (A_{280nm}) and 0.5 ml or 1 ml fractions were collected using a fraction collector.

2.4.8 Ion exchange chromatography

Ion exchange chromatography was carried out using a Q sepharose column (GE Healthcare, Uppsala, Sweden) and ÄKTA Xpress Instrument (GE Healthcare, Uppsala, Sweden). All buffers and samples were filtered using 0.2 µm filters (Whatman, UK). For GST-LMO2 purification, the column was equilibrated using 40 mM Tris-HCl, pH 8.5. A liner gradient of 0 to 1 M NaCl over a volume of 400 ml was used to elute the protein from the column. The eluate was monitored, by measuring the UV absorption at 280nm.

2.5 VH#576 purification and stability trials

2.5.1 VH#576 expression vector preparation

A VH#576 expression vector was prepared by removal of the LMO2 cDNA sequence from vector pRK-His-Tev-VH#576-LMO2 using restriction enzymes HindIII and EcoRI. The purified vector was re-ligated using a standard, T7 ligase, ligation reaction. Positive clones, identified by the release of no fragment upon Hind III and Eco RI double digest, were sequenced (DNA sequencing and services, Dundee).

2.5.2 Purification of VH#576

The expression vector pRK-His-Tev-VH#576 (Ampicillin resistance) was used to transform *E. coli* C41. Cell cultures were grown in labeled rich media (Silantes, München, Germany)

or labeled minimal media, at 37°C, 250 rpm, until the OD_{600nm} measured 0.6. The temperature was then reduced to 20°C and after approximately 1 hour, IPTG was added to a final concentration of 0.5mM and the culture left overnight (12 to 16 hours). Cells were centrifuged at 4000 X g for 20 minutes (Beckman coulter, Avanti J-26 XP, rotor; JS 4.0) and resuspended in 30ml lysis buffer (20 mM Tris-HCl, pH 8, 250 mM NaCl, 20 mM Imidazole, 10 mM β-mercaptoethanol), per liter prep. EDTA-free protease inhibitor cocktail tablets (Roche Diagnostics, Mannheim, Germany) were also added to the lysis buffer (1 per 50ml). The cells were lysed by cell disruption (Constant cell disruption systems, UK) at 27Kpsi and centrifuged at 50,000 X g for 50 minutes to remove cell debris (Beckman coulter, Avanti J-26 XP, rotor; JA 25.50). The whole cell lysate was incubated, for 1 hour, with 2ml Ni-NTA agarose (Qiagen, Crawley, UK) per liter prep. The Ni-NTA resin was washed (lysis buffer without protease inhibitor cocktail) and eluted with elution buffer (20 mM Tris-HCl, pH 8, 250 mM NaCl, 250 mM Imidazole, 10 mM β-mercaptoethanol).

2.5.3 Removal of His tag from VH#576

The His tag was removed from VH#576 by Tev digest (1:10 Tev protease:protein) during dialysis against 20 mM Tris-HCl, pH 8, 150 mM NaCl, 10 mM β-mercaptoethanol overnight. This was followed by a size exclusion chromatography step using a superdex 200 10/300 GL column. Chromatography was monitored, by measuring the UV absorption at 280nm, and the protein fractions analysed by SDS-PAGE.

2.5.4 VH#576 solubility trials

VH#576 was prepared as described in the previous two sections and concentrated to 2 mg/ml using Amicon 15 filtration columns, 10 KDa cut off (Millipore, MA, USA). A 24 well plate was prepared with grease around each well and 1 ml of the appropriate solution pipetted into each well. 2 µl of protein was pipetted on a glass slide and 1 µl reservoir solution. The slide was inverted and sealed onto the relevant well. Plates were stored for one week at room temperature and observed under the microscope. A score of 0 to 4 was given to each well

(0 = no precipitate 4 = heavy precipitate).

2.5.5 Addition of CHAPS to VH#576 sample buffer

3-[(Cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS) was obtained from Fisher scientific and added to the purified protein sample at a final concentration of 2 mM prior to concentration.

2.6 Purification of isotopically labeled VH#576 with unlabeled LMO2

2.6.1 Purification of ¹⁵N labeled VH#576/LMO2ΔN7 protein complex

C41 was transformed with bi-cistronic expression vector pRK-His-Tev-VH#576-LMO2ΔN7 (Ampicillin resistance). Cell cultures were grown up using ¹⁵N labeled minimal media. Protein expression was induced and VH#576/LMO2ΔN7 was purified as described in section 2.4.5.

2.6.2 Cloning of LMO2 expression vector

The LMO2 expression vector was produced by removal of His-Tev-VH#576 from pRK-His-Tev-VH#576-LMO2 using two XbaI restriction enzyme sites. The digest was run on a 1% agarose gel and the linear vector extracted and purified using a kit (Qiagen, Crawley, UK). The cut vector was re-ligated in a standard T7 ligase reaction. Positive clones were identified by the release of no fragment upon XbaI digest and sequenced (DNA sequencing and services, Dundee).

2.6.3 Cloning of NusA-Tev-LMO2 expression vector

The NusA-Tev-LMO2 expression vector was produced by PCR amplification of Tev-LMO2, from vector pGEX-His-Tev-LMO2, with inclusion of SpeI and AvrII restriction enzyme sites at either end (see Appendix C for primer sequences). The PCR product was cloned into SpeI-AvrII sites of pET43a. A SpeI, AvrII restriction enzyme double digest was used to identify positive clones and these clones were sequenced (DNA sequencing and services, Dundee).

2.6.4 Co-lysis of LMO2 and VH#576

C41 *E. coli* host was transformed with either pRK-LMO2ΔN7 or pET43a-Tev-LMO2 (NusA-Tev-LMO2 expression vector). 500 mls of each were grown up in LB media at 37°C, 255 rpm until the OD_{600nm} measured 0.6. ZnSO₄ was added to a final concentration of 0.1 mM and the temperature was then reduced to 16°C. After 1 hour, expression was induced with 0.5 mM IPTG. Concurrently, C41 *E. coli* transformed with pRK-His-Tev-VH#576 were grown up at 37°C, 255 rpm, until the OD_{600nm} measured 0.6, at which point the temperature was reduced to 20°C and expression induced with 0.5 mM IPTG (final concentration). All cultures were left overnight (12 to 16 hours) and harvested by centrifugation at low speed (20 minutes, 4000 X g, 4°C).

The following cell pellets were resuspended together in lysis buffer (20 mM Tris-HCl, pH 8, 250 mM NaCl, 20 mM Imidazole, 10 mM β -mercaptoethanol), with protease inhibitor cocktail (EDTA free, Roche):

- pRK-LMO2 Δ N7 and pRK-His-Tev-VH#576
- pET43a-Tev-LMO2 and pRK-His-Tev-VH#576

Cells were lysed and proteins co-purified via Ni-NTA affinity chromatography (see section 2.4.5). The His tag was removed from VH#576 and the NusA tag from NusA-Tev-LMO2 by Tev protease digest (1:10 protease:protein) during dialysis against imidazole free buffer. The dialysed samples were run through a disposable column with 2 ml Ni-NTA resin.

2.6.5 Production of deuterated VH#576

75% D₂O minimal media was prepared by substituting 750 ml of H₂O for 750 ml D₂O (Silantes, Germany) per liter of media. C41 *E. coli* strain transformed with pRK-His-Tev-VH#576 was adapted to growth on deuterated medium by repeat sub streaking of colonies, gradually increasing the deuterium content of the agar medium. LB agar (plus Ampicillin) plates were made with increasing percentages of D₂O: 20%, 40%, 60%, 80%. A single colony was selected from the 80% D₂O, LB agar plate and a starter culture grown up in 10 mls 75% D₂O minimal media (37°C, 255 rpm) overnight. The starter culture was used to inoculate 90mls 75% D₂O minimal media and this culture was incubated (37°C, 255 rpm) until the OD_{600nm} measured 0.7. The temperature was reduced to 20°C and after 1 hour, IPTG added to a final concentration of 0.5mM. The culture was left over night (12 to 16 hours) and spun down in a centrifuge at low speed (20 minutes, 4000 X g, 4°C). 100ml of pET43a-Tev-LMO2 culture was also grown up in LB, induced and spun down as described in 2.6.4. The two cell pellets were resuspended together, lysed and purified as previously described in 2.4.5.

2.6.6 Increasing the yield of deuterated VH#576 protein

Protocol one: C41 were grown up in LB until OD_{600nm} of 0.7 and then the cell pellet re-suspended in 4 times less 75% D₂O minimal media (Marley et al., 2001). The cells were left to recover for 1 to 2 hours followed by induction of recombinant protein expression with 0.5 mM IPTG at the appropriate temperature (in this case 20°C). Protocol two: C41 were grown up in LB until OD_{600nm} of 3 to 5, then the cell pellet was resuspended in the same volume of 75% D₂O minimal media (Sivashanmugam et al., 2009). Again, this was followed by a 1 to 2 hour recovery period and induction initiated by the addition of 0.5 mM IPTG. Both cultures were left overnight and purified as previously described (section 2.4.5).

2.6.7 VH#576/LMO2 stability trials

Aliquots of 50% deuterated VH#576 bound to LMO2 (expressed and purified according to protocol two of the previous section) were exchanged into the appropriate buffer (20 mM Tris-HCl, 1 mM DTT and either 50 mM, 100 mM or 150 mM NaCl) using Amicon concentration filter units (Y3) with molecular weight cut off 10 KDa and concentrated to 10µl at 10mg/ml. Samples were left for one week at the appropriate temperature. After one week, all samples were centrifuged at 16 000 X g for 15 minutes and the protein transferred to new eppendorfs being careful not to disturb any precipitate. The concentration was measured using a Nanodrop spectrophotometer (samples denatured). Samples were also analysed by SDS-PAGE and visualised using Coomassie brilliant blue.

2.6.8 Purification of ¹⁵N/¹³C/²H labeled VH#576 with LMO2

For VH#576 expression, protocol two (section 2.6.7) was used to generate cell mass in unlabeled LB media and then transferred into the same volume of minimal media (75% D₂O) labeled with ¹⁵N labeled NH₄Cl (Silantes, Germany) and ¹³C labeled (Sigma-Aldrich, Dorset, UK) glucose. 4.5 liters of minimal media was used to express triple labeled VH#576. 4.5L of LB was used to express NusA-LMO2 and the proteins co-purified as previously described (section 2.4.5).

2.6.9 Purification of ¹⁵N labeled VH#576 with LMO2

VH#576 was expressed in 5 L minimal media with ¹⁵N labeled NH₄Cl (Silantes, Germany). 5 L of LB was used to express LMO2ΔN7ΔC11. The cell pellets of each culture were re-suspended together, in lysis buffer, and then co-lysed by cell disruption (Constant cell disruption systems, UK) at 27 KPsi. The VH#576/LMO2ΔN7ΔC11 complex was purified as described in 2.4.5.

2.7 Protein analysis

2.7.1 Protein quantification

Protein was denatured using 6 M guanidium HCl as this allows improved estimation of protein concentration from the sum of Trp, Tyr, and Phe components. A blank of 6 M guanidium HCl was used to calibrate the NanoDrop 8000 (Thermo Scientific). The protein sample was diluted (1 µl protein to 9 µl 6M guanidium HCl) and 2 µl loaded onto the optical surface of the NanoDrop. The “protein A280nm” analysis module was used to measure protein absorbance at 280nm. The sample type was set to “Other protein (E+MW)” and the protein’s molar extinction coefficient (M⁻¹cm⁻¹) and molecular weight (MW) in kilodaltons entered. The protein concentration is automatically calculated in mg/ml and this

was multiplied by 10 to get the final concentration of the protein sample. The A260/280 ratio was also noted, as this should be approximately 0.57, anything above this indicates DNA contamination.

2.7.2 SDS-PAGE

0.75 mm glass plates were assembled and 5 ml of 10%, 12% or 15% resolving gel prepared and pipetted into the plates. Isopropanol was overlaid to promote polymerisation by preventing oxygen diffusing into the gel. After 30 minutes, the overlay was poured away and the gel top washed with milli Q H₂O. The stacking gel solution was pipetted directly onto the resolving gel and a clean Teflon comb inserted. Samples for analysis were prepared by adding 15 µl to 5 µl of 4X loading buffer and were heated at 100°C for 5 minutes along with 20 µl broad range, protein marker (Invitrogen, Paisley, UK). 1X sample buffer was prepared as a blank for any spare lanes. The gel was mounted in the electrophoresis equipment and 1X Tris-glycine buffer added to the top of the reservoirs. All protein markers and samples were loaded onto the gel and a voltage of 180V applied until the dye front reached the bottom of the gel. The glass plates were removed from the electrophoresis apparatus and carefully pulled apart. The gel was placed in Coomassie brilliant blue stain for 1 hour and then into destain for 30 minutes.

2.7.3 Western blot Analysis

SDS-PAGE was run as detailed previously. Polyvinylidene fluoride (PVDF) membrane was soaked in methanol for 5 minutes, followed by a further 5 minute incubation in ice cold transfer buffer. The gel was also incubated for 5 minutes in ice cold transfer buffer. The PVDF membrane was placed on the gel in the blotting cassette with 2 pieces of filter paper either side and one sponge either side. Electrophoresis was used to drive the negatively charged protein bands onto the positively charged membrane using 300 mA for 1 hour. The membrane was then rinsed with TBST and placed in blocking buffer for 1 hour. Antibodies were diluted in blocking buffer as follows:

- Anti-LMO2 – 1:3000 (Obtained from T. Rabbitts. Polyclonal)
- Anti-HIS – 1:3000 (Sigma. Monoclonal, anti-polyhistidine peroxide conjugated therefore secondary antibody was not required.)
- Anti-GST – 1:2000 (Cell signaling technology)

After rinsing with TBST, 10ml of the appropriate diluted antibody was added to the membrane and incubated overnight, at 4°C, with gentle shaking. The membrane was then incubated with Tris buffered saline with tween 20 (TBST); 4 fast washes, followed by 4 washes of 15 minutes each. The appropriate secondary antibody (for both anti-LMO2 and anti-GST primary antibodies this was anti-rabbit horseradish peroxidase conjugated

antibody, derived from Sigma) was diluted 1:10,000 in blocking buffer and incubated with the membrane for 1 hour at room temperature. The membrane was then washed with TBST (4 fast washes, followed by 4 washes of 15 minutes each). The Western blot was analysed using ECL kit (Thermo fisher scientific, Loughborough, UK). Standard X-ray film was exposed to the membrane at ambient temperature behind a Hi Speed intensifying screen for an appropriate length of time, depending on the strength of the signal. Films were developed in a SRX 101A developer (Photon Imaging, Swindon, UK).

2.7.4 Circular Dichroism

The protein samples were prepared at a concentration of approximately 0.22 mg/ml in 25 mM NaH₂PO₄, pH 8. 350 µl of protein was pipetted into the cell (1mm path length). CD spectra were recorded on a Jasco J-715 spectropolarimeter equipped with peltier temperature control. CD data were collected over a wavelength range of 190 to 260 nm and with a step resolution of 1 nm, and a response time of 8 seconds (see Table 2.4). Final spectra were the sum of three scans accumulated at a speed of 50 nm/min.

Table 2.4: CD parameters

Wavelength Scan	
Start	260
End	190
Step res	1nm
Speed	50nm/min
Scans accumulated	1
Response	8 seconds
Sensitivity	50mdeg

2.7.5 Analytical centrifugation

2 sector cells were used for this experiment. MBP-LMO2 was analysed in 20 mM Tris-HCl, 250 mM NaCl, 0.1 mM ZnSO₄, 10 mM β-mercaptoethanol and at three different concentrations, 0.1 mg/ml, 1 mg/ml and 7 mg/ml. A blank of buffer without 10mM β-mercaptoethanol was used. 420 µl of sample was loaded into each cell. Sedimentation velocity was performed over 2 hours, at 50,000 X g, using a Beckman XL-I Analytical Ultracentrifuge. The protein was then resuspended and the experiment repeated at 60,000 X g over 2 hours. The Rayleigh interference system detects solute concentration

differences by changes in the refractive index. A graph of S-value distributions normalised to loading concentration was plotted.

2.7.6 Mass spectrometry

Further identification was performed using Mass spectrometry (MS). MALDI-MS fingerprinting was used for protein identification by partially matching the sequence of the protein to the data generated. MALDI-MS fingerprinting does not give a native molecular weight as it is carried out under denaturing conditions. Positive ionisation electrospray mass spectrometry was used to ascertain the precise molecular weight of the protein both under native and denaturing conditions.

2.7.6.1 MALDI-MS fingerprinting

MALDI-MS (Matrix assisted laser desorption ionisation) fingerprinting was carried out by Dr. Jeff N. Keen. This was done through protein separation using SDS-PAGE, the protein band of interest was then excised using a scalpel. The protein slice was cut into 4 pieces, 2 of which were transferred to duplicate wells of a microtitre plate for automated trypsin digestion using a MassPREP workstation (Waters UK Ltd). The Coomassie brilliant blue-stained gel pieces were first subjected to automated destaining using 50 mM ammonium bicarbonate/50% (v/v) acetonitrile. The proteins were reduced using 10 mM dithiothreitol (in 100 mM ammonium bicarbonate, 30 minutes) and alkylated using 55 mM iodoacetamide (in 100 mM ammonium bicarbonate, 20 minutes), then the gel pieces were washed with 100 mM ammonium bicarbonate and dehydrated using acetonitrile prior to the addition of 25 μ l trypsin (Promega, Southampton, UK) solution (6ng/ μ l in 50mM ammonium bicarbonate). Digestion was allowed to proceed for 5 hours at 37 °C. Peptides were then extracted using 30 μ l 1% (v/v) formic acid/2% (v/v) acetonitrile and an aliquot (1 μ l) applied to a stainless steel MALDI plate together with 1 μ l matrix solution (2mg/ml α -cyano-4-hydroxycinnamic acid in 60% (v/v) acetonitrile/0.08% aqueous TFA). This mixture of peptide fragments were analysed by MALDI. The dried plate was transferred to a M@LDI L/R mass spectrometer (Waters UK Ltd) and the digest was analysed in reflectron mode using standard operating parameters. Briefly, the instrument used a N₂ laser at 337 nm, source voltage was set at 15000 V, microchannel plate detector voltage was set at 1950 V, pulse voltage was set at 2450 V, reflectron voltage was set at 2000 V, coarse laser energy was set to low, with fine adjustment used for each sample to \square optimise the signal. At least 100 laser shots were accumulated and combined to produce a raw spectrum. Spectra were processed (background subtraction, smoothing and peak centroiding) and calibrated externally using a tryptic digest of alcohol dehydrogenase (mass error typically less than 100 ppm). The data is mapped to the theoretical sequence to determine identity.

2.7.6.2 Electrospray ionisation mass spectrometry

Protein samples were analysed on the Synapt HDMS (Waters UK Ltd) mass spectrometer by Dr. James R. Ault. The samples were analysed both in 150mM ammonium acetate buffer (Sigma-Aldrich, Dorset, UK) and under denaturing conditions where the sample was diluted to a buffer composition of 50% methanol/0.1% aqueous formic acid. A minimum protein concentration of 40 μ M was used. The instrument was calibrated using a separate injection of sodium iodide. The mass spectrum shows multiply charged ions and these were then transformed on to a molecular mass scale using the maximum entropy processing technique.

2.8 X-ray crystallography methods

2.8.1 Crystallography screen

Protein was concentrated to approximately 10 mg/ml using Amicon Ultra-15 centrifugal unit, 10 KDa cut-off (Millipore, MA, USA) and subjected to a pre crystallisation test (PCT, Hampton Research, CA, USA) to check the concentration was suitable. Reagents from commercially available crystallisation screening kits were reformatted into sets of 96 in deep well blocks. Precipitant solutions were then dispensed into reservoirs of a Greiner 96 well plate. Using a robot (Cartesian) 100nl from each reservoir was dispensed onto each platform along with 100 nl of protein to form a single droplet. Each plate was sealed manually, using self adhesive transparent film (Viewseal, Greiner). Crystallisation plates were stored at 21 °C and imaged regularly.

2.8.2 Diffraction data collection

Diffraction data was collected at Diamond light source, Oxfordshire, using macromolecular crystallography beamline IO2 and detector type ADSC Q315 CCD. 30% glycerol was added to the mother liquor to act as a cryoprotectant. The crystal was placed in a loop and mounted onto the goniometer head, in a stream of liquid nitrogen. The three point centering method was used to align the crystal accurately in the X-ray beam. An adsorption edge scan was collected to ascertain the wavelengths required for multiple wavelength anomalous diffraction (MAD) data collection. MAD data was collected at the Zn K-edge. The crystal was rotated through 360° and images collected at each degree of rotation. This was done for all three wavelengths.

2.9 NMR methods

2.9.1 Protein sample preparation

Proteins were uniformly labeled with stable isotopes ^{15}N and/or ^{13}C by expressing proteins in *E. coli* in a minimal media that contains ^{15}N -labeled NH_4Cl (Silantes, München, Germany) and/or ^{13}C -labeled glucose (Silantes, München, Germany) as the sole nitrogen and/or carbon sources. Protein was purified and concentrated, using an Amicon concentration filter, 10 KDa cut off (Millipore, MA, USA), to a minimum volume of 320 μl and concentration of at least 0.3 mM.

2.9.2 Experiment preparation

10% Deuterium oxide (DTO) and 0.5 mM Azide, to prevent bacterial growth, were added to the sample. The sample was loaded into an outer 5mm glass Shigemi Advanced NMR microtube (Sigma-Aldrich, Dorset, UK) using a long-tip pipette (Sigma-Aldrich, Dorset, UK). The inner Shigemi tube was inserted so that there were no bubbles between the inner and outer tubes and sealed using parafilm. The length of the sample portion was set to 17mm. In most cases the instrument temperature was set to 10°C and then the tube was loaded into the magnet. The probe was tuned as variations in the polarity and dielectric constant of the lock solvent affect tuning. Tuning was carried out, until the power versus frequency display read zero, for the resonance frequency of the nucleus being measured. Pulse length was altered to give a 90° rotation. Field shimming was done to make the magnetic field more homogeneous and to obtain better spectral resolution. Data was acquired at 10°C using a Varian Unity Inova spectrophotometer unless otherwise stated.

2.9.3 ^1H - ^{15}N HSQC

^1H - ^{15}N HSQC experiments were recorded using a 600 MHz NMR instrument (Varian Unity Inova spectrometer). A ^1H - ^{15}N HSQC provides a 2D heteronuclear chemical shift correlation map between directly bonded ^1H and ^{15}N . Due to the neutral pH that is required to maintain VH#576 in solution, a pulse sequence that contains water flip-back pulses was applied to maintain maximum signal for the exchanging amides. The following parameters were used to run a HSQC experiment. Pulse widths for ^{15}N and ^1H were 45 μs and 10 μs respectively. Decoupling ^{15}N pulse width was 220 μs . A total of 1024 complex points for the acquisition dimension with a spectral width of 8012 Hz and 128 complex points for the indirect dimension with a spectral width of 2083 Hz, were recorded with 160 scans per increment for an overnight experiment. Experiments were recorded at various temperatures from 10 to 25 °C

2.9.4 ^1H - ^{13}C HSQC

^1H - ^{13}C HSQC experiments were recorded using a 600 MHz NMR instrument (Varian Unity Inova spectrometer). The experiment was recorded with 120 (^{13}C) x 1024 (^1H) complex points and spectral windows of 12, 000 (^{13}C) and 10, 000 (^1H) Hz. 64 scans per increment were used at a temperature of 15°C.

2.9.5 HNCA and HN(CO)CA

HNCA and HN(CO)CA were run sequentially using a 750 MHz and 500 MHz spectrometer respectively. For the HNCA, 32 (^{13}C) x 24 (^{15}N) x 1024 (^1H) complex points and spectral windows of 5500 (^{13}C), 2750 (^{15}N) and 10, 473 (^1H) Hz were used, with 64 scans per increment, at 10°C. For the HN(CO)CA, 28 (^{13}C) x 24 (^{15}N) x 1024 (^1H) complex points and spectral windows of 3300 (^{13}C), 1800 (^{15}N) and 7509 (^1H) Hz were used, with 80 scans per increment, at 10°C.

2.9.6 NOESY and TOCSY

A 3D ^{15}N NOESY-HSQC was recorded on a 900 MHz NMR system at Birmingham (HWB NMR large scale facility), with a mixing time of 150ms. The pulse sequence used was provided through the Biopack library (Varian). The experiment was recorded with 116 (NOE) x 26 (^{15}N) x 1024 (^1H) complex points and spectral windows of 10, 000 (NOE), 3050 (^{15}N) and 12, 608 (^1H) Hz. 8 scans per increment were used.

A 3D ^{15}N TOCSY-HSQC was also recorded at 900 MHz with a mixing time of 40ms. This was recorded with 86 (TOCSY) x 24 (^{15}N) x 1024 (^1H) complex points and spectral windows of 10, 000 (TOCSY), 3050 (^{15}N) and 12, 608 (^1H) Hz, with 16 scans per increment. Both spectra were acquired at 10°C.

2.9.7 Data processing

Spectra were processed using nmrPipe (Delaglio et al., 1995) and analysed in CCPN analysis (Vranken et al., 2005).

2.10 Cell culture methods

2.10.1 Chinese hamster ovary cell transfection

Chinese hamster ovary (CHO) cells were harvested from growing cultures and seeded into 12 well plates, at a density of 10^5 cells per well, in 2mls of DMEM glutamax (Invitrogen, Paisley, UK) and 10% FCS. The plate was incubated over night at 37°C, 5% CO_2 . After 16 hours the cells were 50 to 80% confluent and lipofection was performed. In a sterile tube

(A) 24 μ l of Lipofectamine (Invitrogen, Paisley, UK) was added to 576 μ l Optimem (Invitrogen, Paisley, UK). In a sterile tubes (B) 300 ng pM3-LMO2, 300 ng pEF-VP16-VH#576 (or appropriate mutant), 60 ng pRL-CMV, and 600 ng pG5 luciferase reporter vector were added to 50 μ l Optimem. For the negative control pEF-VP16-VH#576 was replaced with pEF-VP16-Y#6VH. 50 μ l of (A) was added to each reaction (B) and left to incubate at ambient temperature for 45 minutes. 400 μ l of Optimem was added to each complexed sample. The cells were aspirated, washed with Hanks media (Invitrogen, Paisley, UK) and the 500 μ l of Optimem complexed sample added. The cells were incubated for 3 to 5 hours at 37°C, 5% CO₂. After incubation the media was replaced, with 2 mls of DMEM glutamax and 10% FCS per well, and the cells left for 48 hours to express (37°C, 5% CO₂).

2.10.2 Luciferase reporter assay

Cells were aspirated and washed with Hanks media (Invitrogen, Paisley, UK), 150 μ l of Trypsin/EDTA was added to each well and once cells were trypsinised, 150 μ l of DMEM glutamax and 10% FCS were added to quench the reaction. Cells were transferred to labeled eppendorfs and 75 μ l of each added to a white 96 well, flat bottomed plate. The remainder was centrifuged at 13,000 X g for 3 minutes and the cell pellet frozen at -80°C for analysis by Western blot. To the appropriate wells of the 96 well plate, 75 μ l of dual glo reagent (Promega, Southampton, UK) was added and left for 10 minutes and a luminescence reading was taken. 75 μ l of stop glo combined with 1/100 activator (Promega, Southampton, UK) was then added to the appropriate wells and left for 10 minutes before a second reading was taken (this gives the renilla levels for use as a transfection efficiency control). Luminescence readings were performed using a Mithras LB940 plate reader (Berthold Technologies, Harpenden, UK), for 2 seconds per well. A graph of Firefly/Renilla was plotted for VH#576 mutants and controls.

2.10.3 Western blot analysis

The cell pellet was resuspended in 37.5 μ l of PBS/EDTA and 12.5 μ l of 4X SDS loading buffer added before the sample was heated at 100°C for 5 minutes. 20 μ l of each sample was loaded onto a 15% acrylamide gel, the proteins separated by SDS-PAGE and transferred to a PVDF membrane. After an incubation of 1 hour with 4% blocking buffer, anti-VP16 (14-5 Santa Cruz biotechnology) was added at 1:1000 and left overnight at 4°C. After four 15 minute washes with TBST, anti-mouse conjugated HRP antibody was added at a concentration of 1:1000 in 4% blocking buffer. After four 15 minute washes with TBST, the Western blot was analysed using ECL kit (Thermo fisher scientific, Loughborough, UK). Standard X-ray film was exposed to the membrane at ambient temperature behind a Hi Speed intensifying screen for an appropriate length of time, depending on the strength of the

signal. Films were developed in a SRX 101A developer (Photon Imaging, Swindon, UK).

Commonly used reagents

2.10.4 Buffers and solutions

All buffers were sterilised by filtration through a 0.2µm nalgene filter.

2.10.4.1 Buffers for DNA analysis

- **1L of 5 X TBE buffer (Tris/borate/EDTA):** 54 g Tris-HCl base, 27.5 g boric acid, 20 ml 0.5 M EDTA, pH 8.
- **1% agarose:** 1g agarose (Sigma-Aldrich, Dorset, UK), 100ml 1 X TBE buffer
- **6 X DNA loading buffer:** 30% (v/v) glycerol, 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF

2.10.4.2 Buffers for protein analysis

- **12ml SDS gel loading buffer (4X):** 3ml 1M Tris-HCl pH 6.8, 4.8 ml 10% SDS, 2.4 ml 100% glycerol, 0.024 g bromophenol blue, 0.58 ml milli Q H₂O, 10% β-mercaptoethanol (add to stock solution just prior to use).
- **5 X Tris glycine:** 15.1 g Tris Base, 94 g glycine, 50 ml 10% SDS, milli Q H₂O to 1 L.
- **12% SDS-PAGE gel:** 3.3 ml milli Q H₂O, 4 ml 30% acrylamide, 3.8 ml Tris-HCl (1.5 M, pH 8.8), 0.15 ml 10% SDS, 0.15 ml 10% ammonium persulfate, 6 µl TEMED.
- **5% stacking gel for Tris-glycine SDS-PAGE:** 2.7 ml milli Q H₂O, 0.67 ml 30% acrylamide, 0.5 ml Tris-HCl (1M, pH 6.8), 0.04 ml 10% SDS, 0.04 ml 10% ammonium persulfate, 4 µl TEMED.
- **Coomassie brilliant blue stain:** 50 ml milli Q H₂O, 40 ml methanol, 10 ml acetic acid, Coomassie brilliant blue (0.2%).
- **Coomassie brilliant blue de-stain:** 50 ml milli Q H₂O, 40 ml methanol, 10 ml acetic acid.
- **Transfer Buffer:** 5.8 g Tris base, 2.9 g glycine, 200 ml methanol, milli Q H₂O to 1 L.
- **10X TBS:** 24.23 g Tris-HCl, 80.06 g NaCl, 800 ml milli Q H₂O. pH to 7.6 with HCl and add milli Q H₂O to 1 L.
- **TBS-T: 100ml TBS (10X):** 900 ml milli Q H₂O, 1 ml Tween 20.
- **Blocking Buffer:** 4 g fat-free milk powder, 100 mls TBST

2.10.5 Nutrients broths

Luria Bertani (LB) broth: 10 g bactotryptone, 10 g bacto yeast, 5 g NaCl, milli Q H₂O to 1 L.

LB agar plates: LB plus 15 g bactoagar per liter.

2TY: 16 g bactotryptone, 10 g bacto yeast, 5 g NaCl, milli Q H₂O to 1 L.

SOC: 20 g bactotryptone, 5 g bacto yeast, 0.5 g NaCl, 10 ml 250 mM solution of KCl, pH to 7.0 with NaOH. Add milli Q H₂O to 975 ml. Sterilised and just prior to use, 5 ml of sterile 2 M MgCl₂ and 20 ml of sterile (by 0.2 µm filter) 1 M glucose solution were added.

M9 minimal media (components autoclaved separately): 200 ml 5 X M9 salts (see below), 2 ml 1 M MgSO₄, 20 ml 20% glucose (sterilised by filtration), 1 M CaCl₂, 10 ml 100 X MEM vitamins (Invitrogen, Paisley, UK) and made up to 1 L with sterile H₂O.

5 X M9 salts (1L): 64 g Na₂HPO₄•7H₂O, 15 g KH₂PO₄, 2.5 g NaCl, 5 g NH₄Cl and made up to 1 L with sterile H₂O.

LB media and agar, SOC medium and 2TY medium were all sterilised by autoclaving for 20 minutes at 15 lb/sq on a liquid cycle. Unless otherwise stated the antibiotic concentrations of the plates and media were:

Ampicillin (Amp): 100 µg/ml

Chloramphenicol (Cam): 34 µg/ml

Tetracycline (Tet): 20 µg/ml

LB agar was made up as stated. If required, antibiotics were added to cooled agar (40 to 50°C) and 25 ml poured immediately into petri dishes, using aseptic techniques, under a fume hood. Plates were allowed to set at room temperature and stored at 4°C for a maximum of two weeks.

2.11 Theory behind methods used for structural analysis

2.11.1 NMR theory

Biomolecular NMR requires proteins enriched with ^{15}N or both ^{15}N and ^{13}C . Atomic nuclei ^1H , ^{15}N and ^{13}C have net nuclear spins as a result of the magnetic moments of their component protons and neutrons. For spin $\frac{1}{2}$ nuclei the energy levels split into parallel and anti parallel orientations, upon application of a magnetic field. Spins aligned parallel with external magnetic fields are of a slightly lower energy than those aligned in an antiparallel orientation. The orientation of nuclear spins can be altered by pulses of electromagnetic radiation, in the radio frequency range (Rhodes, 2006). Nuclei in different chemical environments absorb different frequencies of energy and this allows specific nuclei to be detected by their specific absorption energy. The result is called the chemical shift and is expressed in parts per million (ppm). Chemical shifts signify the position of a resonance along a frequency axis and are defined relative to a standard such as 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Nuclear spins also interact magnetically, exchanging energy with each other, in a process called spin-spin coupling. As nuclei must be within a few bonds of each other to couple, this effect can be used to determine which nuclei are neighbours within the molecule (Whitford, 2005).

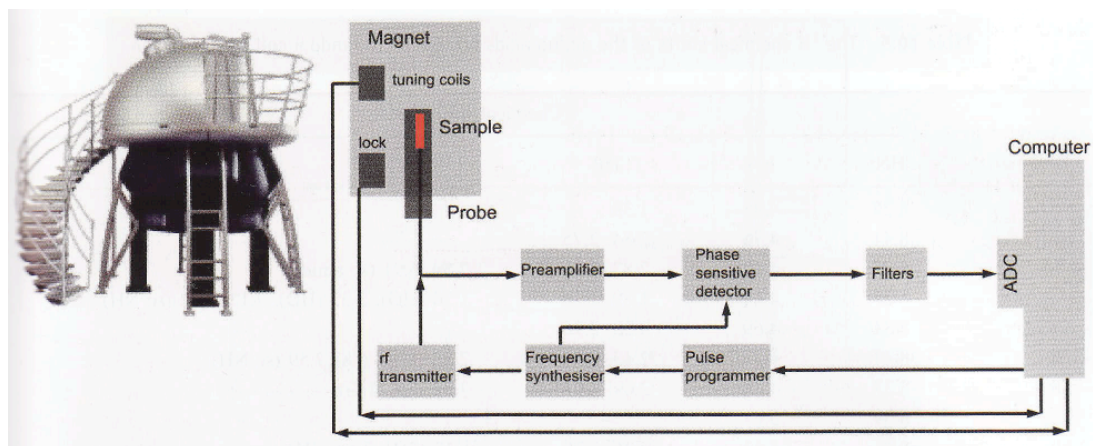


Figure 2.1: NMR experiment setup

In this diagram of an NMR instrument, the major components are the magnet containing the probe, the 'shim' to maintain homogeneity and electronics to 'lock' the field at a given frequency. The radio frequency transmitters provide pulses designed by a pulse programmer and a frequency synthesiser together with a detection system of numerous amplifiers, filters and analogue-digital converters (ADC).

Figure taken from (Whitford, 2005).

Nuclear spins in a sample can be viewed as precessing about an axis, z , aligned with the magnetic field H (Figure 2.1). The excess of spins aligned with the field means there is a net magnetisation vector pointing in the positive direction along z at equilibrium. The receiver coils, encircle the x -axis and detect radio frequency radiation in the xy plane. Detection of a radio frequency requires that the net magnetisation vector be moved into the xy plane and this is achieved by the application of a 90° pulse of radio frequency at the correct intensity, applied along the x axis. The resulting radio frequency signal contains the characteristic absorption frequencies of all the nuclei in the sample from which chemical shifts can be derived.

Nuclear spins lose energy to their surroundings, a process known as spin lattice relaxation. The longitudinal relaxation time T_1 reflects the rate at which magnetisation returns to the longitudinal axis after a pulse and is measured in seconds. In solution, T_1 is correlated with the overall rate of tumbling of a macromolecule but is also affected by internal molecular motion arising from conformational flexibility. The transverse relaxation time T_2 describes the decay rate of transverse magnetisation in the xy plane. T_2 is always shorter than T_1 and is correlated with dynamic processes occurring within a protein. The most important factors, which influence T_2 are molecular mass, temperature, solvent viscosity and exchange processes (Whitford, 2005).

Two dimensional (2D) NMR spectroscopy involves the application of successive pulses that lead to magnetisation transfer between nuclei (Aue et al., 1976). The mechanism of transfer proceeds either through bond (scalar mechanism) or through space (dipolar) interaction. A 2D experiment consists of 4 time periods. The preparation period is a single pulse or series of pulses and delays, which create the magnetisation terms that, will make up the indirect dimension. The terms evolve during the evolution period. The evolution of magnetisation is followed by a second pulse that initiates the mixing period which results in magnetisation transfer to other spins via through bond or through space interactions. The decay of the free induction decay signal is detected during the acquisition period. The whole pulse scheme is repeated to allow for signal averaging and other instrumental factors including relaxation. The experiment is then repeated by incrementing the evolution period to build up a series of free induction decays recorded at different intervals. A 2D experiment contains two time variables and Fourier transformation of this dataset yields a 2D contour plot, where the precision frequencies occurring during the evolution and detection periods determine peak positions in the 2D plot. This can be extended to three dimensional (3D) NMR by including a second indirect evolution period and a second mixing period.

As amide nitrogens are sequentially connected through alpha and carbonyl carbons a ^{15}N and ^{13}C labeled protein sample is used to carry out three dimensional (3D) experiments in order to assign backbone amide signals. HNCA and HN(CO)CA, are generally run sequentially (Kay et al., 1990). The experiment name denotes the observed signal correlations in the spectra, for example, HNCA provides correlations of signals from the amide (HN) in the i th residue to the attached C_α in the i th and $i-1$ residue. HN(CO)CA provides correlation between HN in the i th residue to the C_α in the $i-1$ residue. In this case the CO is a relay nucleus and, therefore, its frequency is not detected.

Assignment is a critical stage of determining protein structure by NMR and can prove problematic. The known protein sequence is used to connect nuclei of amino acid residues that are neighbours in the sequence using information from the spectra (Dubs et al., 1979). Upon assignment of a HSQC spectrum, the crucial information for structure determination, comes from Nuclear Overhauser effect (NOE) measurements that provide distance information between pairs of protons. Supplementary constraints can be derived from through bond correlations in the form of dihedral angles (Guntert, 1998). Furthermore, chemical shift data provides information on the type of secondary structure and hydrogen bonds can be detected via through-bond interactions (Spera and Bax, 1991) or inferred from slow hydrogen exchange.

2.11.2 X-ray crystallography theory

X-ray diffraction results from the interaction between incident X-rays and the electrons of atoms within the crystal. As electrons are localised around the nucleus of atoms, the electron density is a good indication of atomic position within the molecule. Bragg's law (Equation 2.1) is the relationship between the crystal structure and its diffraction pattern (Bragg, 1913). Each spot on the diffraction pattern is a Bragg reflection or reflection.

Equation 2.1

$$\text{Bragg's law} = n\lambda = 2d \sin \theta$$

(where λ = wavelength d = spacing between the planes and θ = the glancing angle)

Schematically a crystal can be described as a lattice (Blow, 2005). A series of equally spaced parallel lines will cover all the points in a lattice. All the possible sets of planes, which cover all the lattice points, can be described by different values of h , k and l (Miller indices). The interplanar distance associated with any reflection can be determined directly by observation of its θ angle and use of Bragg's law. The diffraction pattern of a crystal lattice is also a lattice but with dimensions which are inversely proportional. A set of planes

(h , k and l) produce the reflection h , k and l in the diffraction pattern. Diffracted rays reflect from hkl planes of spacing d_{hkl} only at angles (θ) for which $2d_{hkl} \sin \theta = n\lambda$ (an integral number of wavelengths). The crystal is rotated through a small angle ($1-2^\circ$), such that the new oscillating range overlaps the previous one slightly. From this new position, oscillation produces additional reflections in the second frame. This process is continued until all unique reflections have been recorded. When screw rotational symmetry is present, there are systematic absences among reflections in lattice planes perpendicular to the screw axis.

The result of X-ray data collection is a list of intensities, each assigned an index hkl corresponding to its position in the reciprocal lattice. Indices are counted from the origin, which lies in the direct path of the X-ray beam (indices 000). As indices increase, there is a corresponding decrease in the spacing d_{hkl} of the real space planes represented by the indices and thus these reflections carry far more information about the fine details of the structure. Software is used to index the reflections, determine unit cell parameters and spacegroup. Indexing takes into account the mosaic spread of the reflections collected. Mosaicity is an angular measure of the degree of long range order of the unit cells within a crystal. Lower mosaicity indicates more ordered crystals and hence better X-ray diffraction. Indexing is followed by integration, a process that converts all images into a single file. It provides the miller index of each reflection and intensities. Scaling is used to scale the relative images so that they have a consistent intensity scale. Data analysis yields a list of reflections (positions) and their intensities (amplitudes) but the phases of the scattered X-rays are unknown.

All reciprocal lattices possess a symmetry element called a centre of symmetry or point of inversion at the origin (Rhodes, 2006). The computed sum for the reflection hkl is called the structure factor F_{hkl} . The amplitude of F_{hkl} is proportional to the square root of reflection intensity I_{hkl} . The frequency of the structure factor is equal to $1/d_{hkl}$, making the wavelength the same as the spacing of the planes producing the reflection. In three dimensions, Friedel's law requires that $|F_{hkl}| = |F_{-h,-k,-l}|$. Reflections are known as Friedel's pairs. The structure factors from Friedel's pairs have the same amplitude but opposite phase angles. If the amplitude and the phase of the structure factor are known for all reflections, the electron density map can be calculated and the crystal structure solved. However, no phase information is directly available from the intensities of the diffracted X-rays. Therefore various experimental methods are in use to solve the phase problem.

One method employed to solve the phase problem is anomalous scattering. The zinc present in LMO₂, for example, have the ability to scatter X-rays anomalously. The use of multiple

wavelengths near the absorption edge of a heavy atom can be used to identify heavy atom positions, providing powerful phase information. Multiple wavelength anomalous dispersion (MAD) data is collected at three different wavelengths; peak, edge and remote. At the peak wavelength, anomalous scattering should have its largest value. At the edge the normal scattering is reduced to its lowest possible value and at the remote normal scattering is close to its normal value. Atoms, which scatter anomalously, do not obey Friedel's law as the reflections hkl and $-hkl$ are not equal in intensity. The difference of scattered intensities between I_{hkl} and $I_{-h,-k,-l}$ is known as Bijvoet difference and the difference between $|F_{hkl}|$ and $|F_{-h,-k,-l}|$ Bijvoet amplitude difference (Bijvoet, 1954). Changing the wavelength changes the intensities of the individual reflections, which can be used to directly calculate the phase angle for each reflection. The magnitude, of anomalous differences, may be indicated by an R factor (Equation 2.2).

Equation 2.2

$$R_{\text{anom}} = \text{mean Bijvoet amplitude difference} / \text{mean amplitude}$$

Phase information may be obtained from measurement of $|F_{\text{PH}}|$ (reflection of amplitude for protein plus heavy atom) for each member of the Friedel pair. In this case only data collected at the peak wavelength was processed (single anomalous diffraction) and used to identify the position of zinc atoms.

Another method for phase determination is molecular replacement. Molecular replacement can be used to solve a structure when you have a good model for a reasonably large fraction of the structure in the crystal. The model should have at least 30% sequence identity. Both maximum likelihood and Patterson based molecular replacement were used for the work described in this chapter. The Patterson function can be computed without phases. A preliminary model of the crystal structure is obtained, by first orienting (rotating) and then positioning (translating) the model molecule in the crystal lattice (Rossmann and Blow, 1962). The phases are then calculated from the model and combined with the observed structure factor amplitudes. The electron density map produced using this method contains a strong bias towards the starting model however values are enough for refinement.

“The accuracy of a 3D structure depends on the refinement, the resolution and the restraints that are introduced in the structure analysis (Blundell et al., 2002).” Before refinement begins a series of test reflections are picked at random (5%) and are not available for the refinement procedure; the R_{free} . This is done to make sure that any improvements are an

accurate representation of the model. The difference between the R_{factor} and the R_{free} should be no more than 5%. The R_{factor} is used to judge the crystal model, it is a measure of how well the modified electron density fits the observed structural amplitudes and can be calculated, as shown in Equation 2.3. R factors for good structures should be approximately $1/10^{\text{th}}$ of the resolution of the data.

Equation 2.3

$$R = \sum |F_{\text{obs}} - F_{\text{cal}}| / \sum F_{\text{obs}}$$

(Where F_{obs} = observed structural factors. F_{cal} = calculated structural factors)

3 Expression and purification of LMO2 for X-ray crystallography trials

3.1 Introduction

Non random chromosomal translocations are common in the malignant cells of patients with acute and chronic leukaemia. LMO2 (LIM domain only 2) is encoded by a gene located on chromosome 11p13 and is associated with specific chromosomal translocations $t(11;14)(p13;q11)$ and $t(7;11)(q35;p13)$ in patients with T-cell acute lymphoblastic leukaemia (T-ALL). T-ALL is an aggressive malignancy of thymocytes in which multiple genetic defects combine during pathogenesis resulting in uncontrolled cell growth (Van Vlierberghe et al., 2008). At present, treatment involves multi-agent chemotherapy (Bernard et al., 1998) however only 30 to 40% of adult patients are long term survivors. Current treatments cure over 70% of children with acute lymphoblastic leukaemia, although many patients develop serious acute and late complications due to the side effects (van Vlierberghe et al., 2008).

LMO2 has been shown to have an important role in T-ALL (Grutz et al., 1998b). Transgenic mouse models of LMO2-induced T cell neoplasias showed that enforced LMO2 expression caused accumulation of immature thymic T cells, followed by clonal T cell tumours with long latency (Larson et al., 1995). LMO2 is therefore a specific therapeutic target as not only is it associated with chromosomal translocations and retroviral insertions (Hacein-Bey-Abina et al., 2003) but also is expressed in approximately 50% of T-ALL patients (Aplan, 2004). In addition, LMO2 has been implicated in diffuse large B cell lymphomas (Natkunam et al., 2007) and its aberrant expression has been detected in prostate tumours (Ma et al., 2007) and pancreatic cancer (Nakata et al., 2008, Nakata et al., 2009).

LMO2 is a nuclear LIM only protein made up of two tandem LIM domains. Each LIM domain is composed of two zinc binding finger like motifs (LIM fingers). LMO proteins have been identified as important motifs that mediate specific protein-protein interactions. As such LMO2 is an important protein to understand and a potential drug target. The structure based drug design approach, described in this thesis, requires a three-dimensional structure of the drug target, LMO2. The two routinely used methods, to study macromolecular structure, are X-ray crystallography and Nuclear magnetic resonance (NMR). The focus of this chapter is the crystallisation of LMO2 for the purpose of X-ray crystallography.

The first stage of X-ray crystallography is purification of the target protein to both a high concentration and purity. A commonly used strategy is to express different versions of the

target protein until a well expressing, soluble, and correctly folded construct is identified (Braun and LaBaer, 2004). Further variations of the protein, such as point mutations or amino or carboxy-terminal truncations may be required in order for crystallisation to be successful (Williams et al., 2000).

The first half of this chapter explores the use of three different N-terminal fusion tags, GST, NusA and MBP, to aid the solubility and purification of LMO2. All three constructs proved problematic and as a result an alternative strategy was investigated. An anti-LMO2, heavy chain variable immunoglobulin domain, VH#576, had been previously isolated (Tanaka and Rabbitts, 2010). VH#576 was co-expressed with LMO2 and used to co-purify the heterodimeric complex. This construct failed to crystallise even after methylation of the surface lysine residues. Deletion analysis of LMO2 amino (N)-terminus and carboxy (C)-terminus revealed seven N-terminal residues could be removed and eleven C-terminal residues, while still retaining protein solubility and complex stability. Finally, a complex of human LMO2 spanning amino acid residues 9 to 147, bound to an antibody single domain, was crystallised. X-ray diffraction data was collected for structural determination as will be discussed in chapter five.

3.2 Protein expression and purification methods

The general methods employed for recombinant protein expression and purification are outlined in chapter two, section 2.4. The crystallography screens were set up as described in section 2.8. Any modifications are stated in the appropriate section of this results chapter.

3.3 Purification of GST-LMO2 fusion protein

Glutathione S-transferase (GST) is a 26 KDa protein that can be fused to a protein of interest and purified via glutathione affinity chromatography. The method is based on the affinity of GST to the glutathione ligand coupled to a matrix. GST fusion proteins can be competitively eluted using reduced glutathione, which does not affect the proteins native structure and function. It is important to know, when considering the stoichiometry of a GST fusion protein, that GST can undergo dimerisation (Nishida et al., 1998).

3.3.1 Expression and purification of GST-LMO2

Expression of GST-LMO2 was induced in the *E. coli* Rosetta strain harbouring the expression plasmid pGEX-His-Tev-LMO2 (Ampicillin resistant). Protein expression was induced and the whole cell lysate was analysed by Western blot (Figure 3.1).

Two negative controls were also included as a means of confirming the expression of GST-LMO2:

- I. Rosetta cells transformed with pGEX-His-Tev-LMO2 and not induced with IPTG (transformed only)
- II. Rosetta cells which have not been transformed with an expression vector but induced with IPTG (induced only)

The GST-LMO2 construct also has a His tag, as shown in the expression vector map (Appendix C). Figure 3.1 shows anti-LMO2 and anti-His Western blots, both of which show a band corresponding to the size of GST-LMO2 (46.03kDa) in lane two (induced and transformed). This band also appears in lane three of both Western blots, the transformed only control, and this suggests the T7 promoter is slightly “leaky”. Analysis of untransformed Rosetta *E. coli* cells (lane four) confirms that the Western blot signal is specific to transformation with the appropriate construct. Both Western blots show break down of GST-LMO2 represented by the bands below that of GST-LMO2. This is likely due to proteolytic degradation via host cell protease.

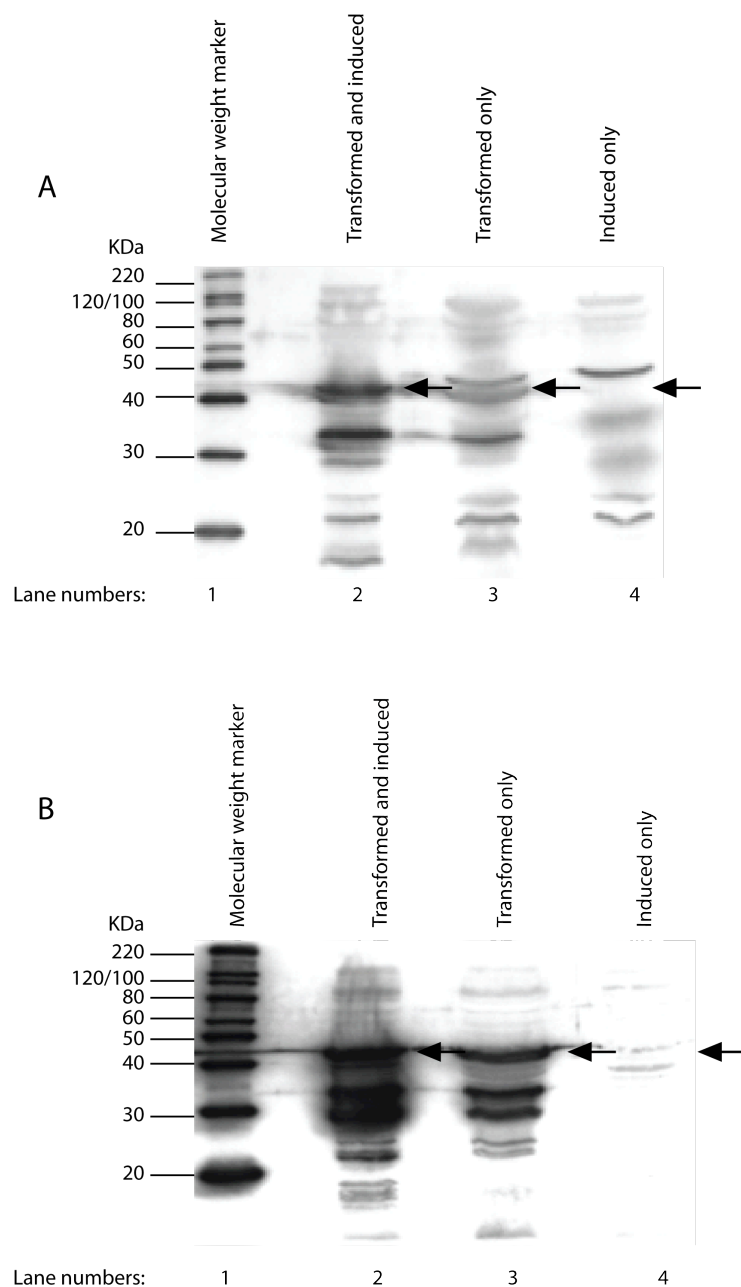


Figure 3.1: Western blot analysis of GST-LMO2 whole cell lysate

A: Western blot analysis of cell lysate (soluble fraction) using anti-LMO2 polyclonal primary antibody. A crude extract of cellular proteins was separated by 12% SDS-PAGE, transferred onto PVDF membrane and incubated with affinity purified anti-LMO2 rabbit antiserum followed by horseradish peroxidase conjugated anti-rabbit antibody. Arrows indicate the presence or absence of GST-LMO2.

B: Further analysis of the cell lysate was performed using anti-polyhistidine peroxidase conjugate antibody. PVDF membrane was incubated with monoclonal anti-polyhistidine peroxidase conjugate, mouse IgG2a isotype (Sigma-Aldrich, Dorset, UK). Both Western blots show GST-LMO2 is expressed in the transformed and induced cells but also in the transformed only cells. Again, arrows indicate the presence or absence of GST-LMO2.

GST-LMO2 was then purified via glutathione sepharose chromatography and eluted using 10mM reduced glutathione. SDS-PAGE analysis of the eluate (Figure 3.2) shows a band running at the correct size for GST-LMO2 (46.03 KDa), along with an additional intense band representing a protein of approximately 60 KDa. Two further bands are also visible which run below GST-LMO2. The bands of lower molecular weight were shown to be degradation products by Western blot. The 60 KDa contaminant protein does not contain GST, His or LMO2, as it is not present in any of the Western blots. Further experiments revealed this 60 KDa contaminant protein could not be removed by extensive column washing with 0.5% triton X-100 (data not shown).

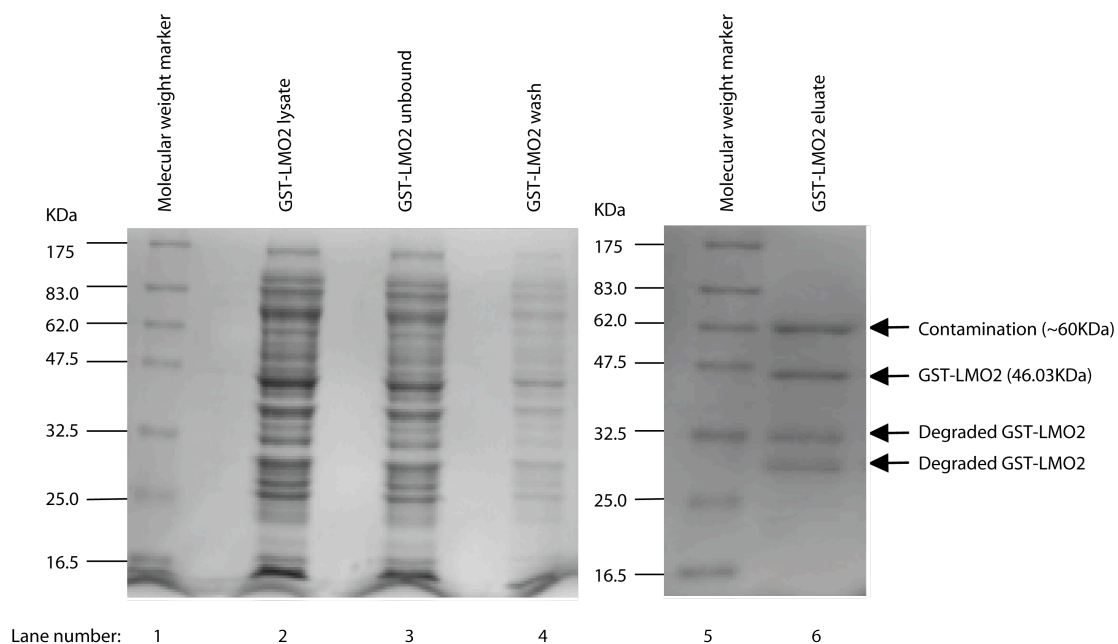


Figure 3.2: Purification of GST-LMO2 via glutathione sepharose affinity chromatography
Purification of GST-LMO2, after recombinant expression in *E. coli* (Rosetta). Samples were taken at each stage of the chromatography (load, unbound fraction, wash fraction and eluate) and analysed by 12% SDS-PAGE. The gel was stained with Coomassie brilliant blue. Arrows indicate the proteins present in the eluate. GST-LMO2 (46.03 KDa) is shown to undergo a degree of degradation and there is a strong contaminating protein of approximately 60 KDa.

3.3.2 Strategy to reduce GST-LMO2 degradation

As protease inhibitors (Roche Diagnostics, Mannheim, Germany) did not prevent the degradation of GST-LMO2, a strategy was investigated to decrease the rate of protein degradation. An alternative explanation for the rapid degradation observed during GST-LMO2 purification (Figure 3.2) is the inability of the protein to form a correct tertiary structure due to anomalies in protein folding. In some cases, over expression of the *E. coli* chaperone complex GroEL-ES has facilitated correct protein folding and enhanced the production of active proteins (Wall and Pluckthun, 1995). It was anticipated that the controlled over expression of GroEL-ES would enhance GST-LMO2 stability and thus reduce degradation.

Many newly synthesised proteins reach their folded state spontaneously, without assistance, however folding efficiency may be limited by side-reactions such as aggregation promoted by transiently expressed hydrophobic surfaces. The GroEL-ES chaperone complex is essential for *E. coli* growth and designed to ensure protein folding occurs efficiently (Whitford, 2005). GroEL and GroES are composed of seven and fourteen identical subunits, respectively and are organised to form a large isolation chamber within which proteins can fold.

Three different chaperone vectors were obtained; pA6, pGroES-EL and pGTF2 (Madrid, Spain). pA6 and pGroES-EL express GroEL-GroES chaperones. pGTF2 expresses GroEL-GroES chaperones and trigger factor, a protein thought to facilitate protein folding by interacting with GroEL and strengthening substrate binding (Nishihara, K. et al, 1999). Each plasmid was co-transformed into *E. coli* (C41) along with pGEX-His-Tev-LMO2 and protein expressed as previously described. C41 was used, in place of Rosetta pLysS, as all three chaperone expression plasmids have a Chloramphenicol selection marker, as does Rosetta pLysS. C41 is defective in OmpT and Lon protease production and therefore may aid in the expression of fusion proteins by minimising the effects of proteolytic degradation by the host (Miroux and Walker, 1996). The cell lysates were purified by glutathione sepharose chromatography and eluates analysed by SDS-PAGE (Figure 3.3). The gel shows less degradation occurred when GST-LMO2 was co-expressed with GroEL-GroES and trigger factor (pGTF2).

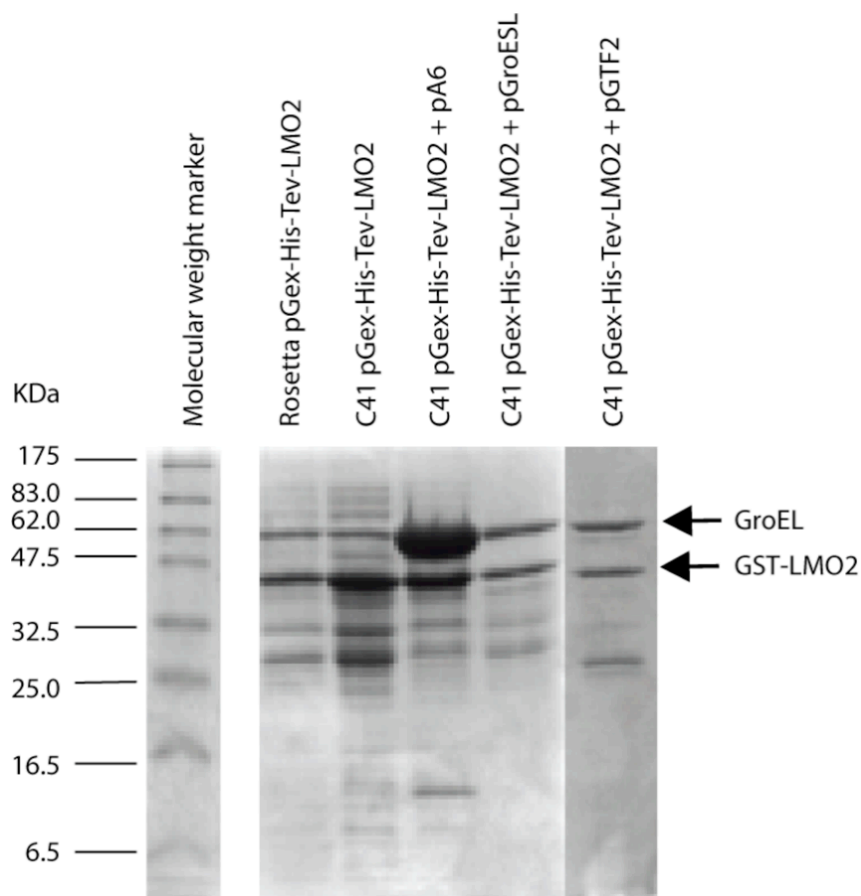


Figure 3.3: Addition of chaperone expression vectors to reduce GST-LMO2 degradation
12% SDS-PAGE stained with Coomassie brilliant blue. The gel suggests degradation of GST-LMO2 is reduced by the over expression of GroEL-GroES and that the pGTF2 vector, expressing GroEL-GroES chaperone complex and trigger factor (TF), was the most effective.

3.3.3 Identification of the 60 KDa contaminant protein and removal strategy

Figure 3.3 shows the 60 KDa protein band increases in intensity for cells over expressing chaperone GroEL, this suggests that the original 60 KDa contaminant was in fact endogenous GroEL bound to GST-LMO2. A similar problem has been described during the purification of a Papillomavirus capsid protein, HPV16 L1 (Chen, X.S. et al., 2001). In the case of Chen *et al.* the 60 kDa contaminant protein, which could not be removed by extensive washing, size exclusion, or ion exchange chromatography, was identified as GroEL by N-terminal sequencing. The addition of ATP-MgCl₂ (2mM, 5mM respectively) has been demonstrated to releases GroEL from its substrate (Chen, X.S. 2001). pGTF2 (Chloramphenicol resistance) and pGEX-His-Tev-LMO2 (Ampicillin resistance) were co-transformed into C41 and expression induced in the usual way. The cell lysate (soluble fraction) was split into three aliquots and ATP-MgCl₂ added to a final concentration of:

- 1) 2mM ATP and 5mM MgCl₂
- 2) 4mM ATP and 10mM MgCl₂
- 3) 8mM ATP and 20mM MgCl₂

Glutathione sepharose resin (50% slurry) was added to each lysate and incubated for 2hours, at 4°C, with gentle rotation. Each column was washed with 100ml of washing buffer containing ATP and MgCl₂ at the appropriate final concentration (listed above). This was followed by a 50ml wash without ATP-MgCl₂. The eluate of each column was analysed by SDS-PAGE (Figure 3.4). The addition of ATP-MgCl₂ virtually eliminated GroEL from each eluate with 2mM ATP and 5mM MgCl₂ being most effective.

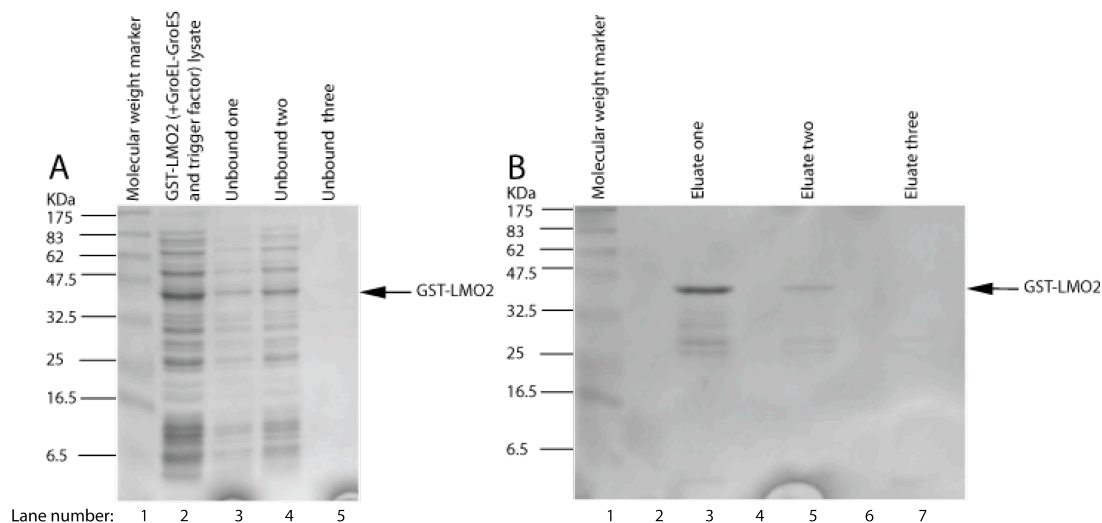


Figure 3.4: Addition of ATP-MgCl₂ for removal of bound chaperone
12% SDS-PAGE stained with Coomassie brilliant blue.

A: The cell lysate (soluble fraction) was analysed along with the three unbound fractions. Lanes three, four and five represent unbound fractions from lysates which have mixed with: 1) 2mM ATP and 5mM MgCl₂ 2) 4mM ATP and 10mM MgCl₂ 3) 8mM ATP and 20mM MgCl₂ respectively. No unbound fraction can be seen in lane five and this may be due to experimental error.

B: Each column eluate was concentrated and analysed. The gel shows 2mM ATP and 5mM MgCl₂ are the most effective concentrations for removal of the GroEL contamination while still maintaining a practical yield of 1.25mg per liter culture.

3.3.4 Size exclusion chromatography to remove GST only from GST-LMO2

Figure 3.5 shows more than 60% of the protein eluted from the glutathione column to be GST only; this is a common problem when purifying proteins fused with GST. Therefore, in order for the protein to be pure enough for crystallography studies, an additional chromatography step was required. The first method tested was size exclusion chromatography. Previous size exclusion chromatography had little success (data not shown) so a larger column (25mm diameter by 600mm in length) was used to increase the resolution of separation. The column resin was superdex 75, which separates proteins in the range of 3 to 600 KDa. A total of 4mg (28.89 μ M) GST-LMO2 in 2ml phosphate buffer was loaded onto the column and a flow rate of 2.8ml/min was used. 4ml fractions were collected, over an elution volume of 300mls, and analysed by SDS-PAGE (Figure 3.6).

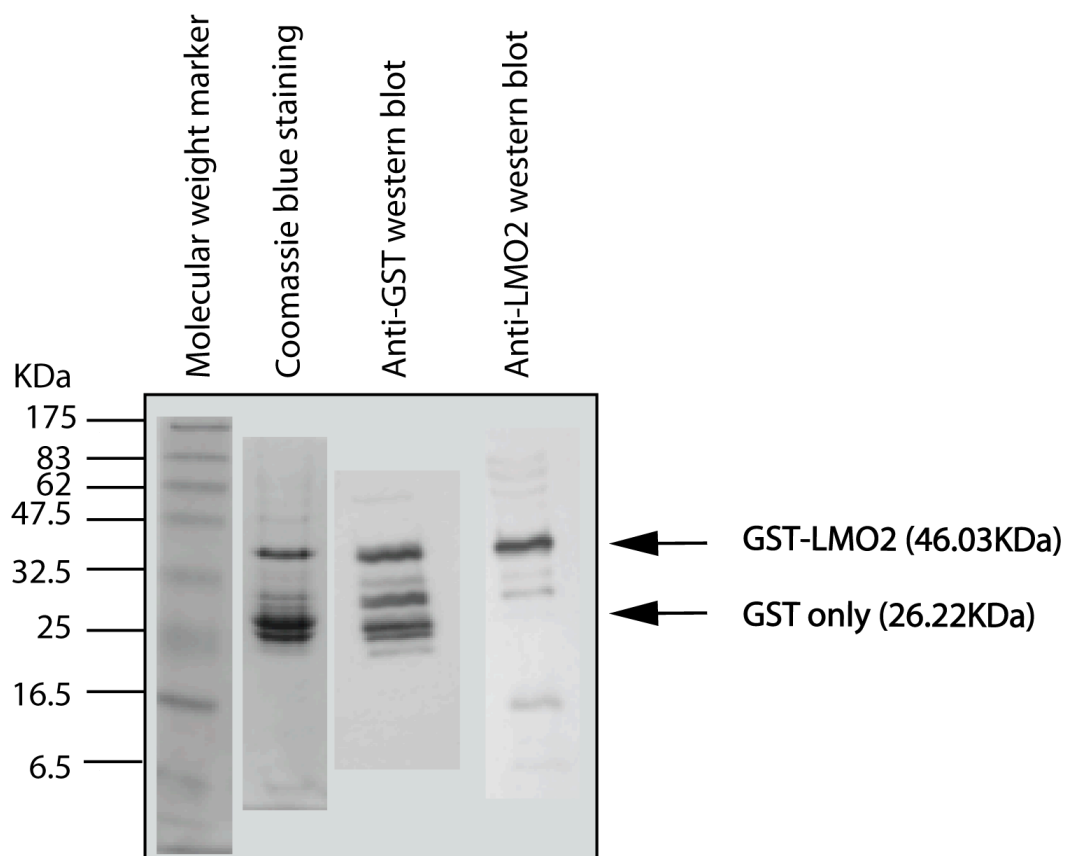
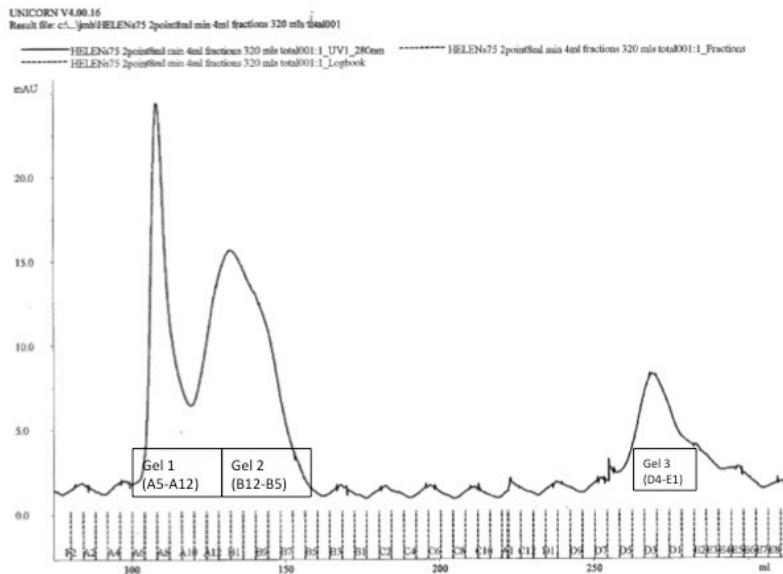


Figure 3.5: Western blot analysis of GST-LMO2 co-expressed with chaperone and trigger factor proteins

The eluate was separated by size using 12% SDS-PAGE and then transferred onto PVDF membrane. One membrane was incubated with anti-GST antibody solution and the other anti-LMO2 antibody solution. Both membranes were then incubated with horseradish peroxidase conjugated anti-rabbit antibody. The figure compares the Western blots, to the Coomassie brilliant blue stained gel, and reveals the majority of GST-LMO2 degradation is GST only.

A:



B:

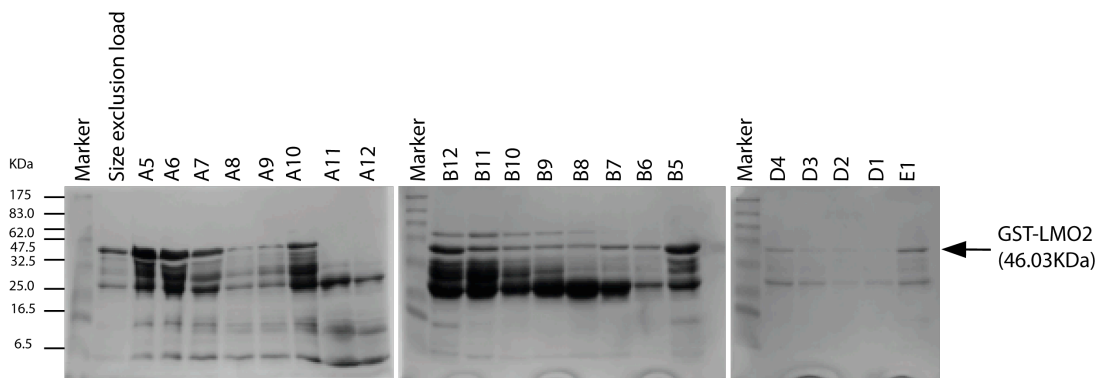


Figure 3.6: GST-LMO2 size exclusion chromatography

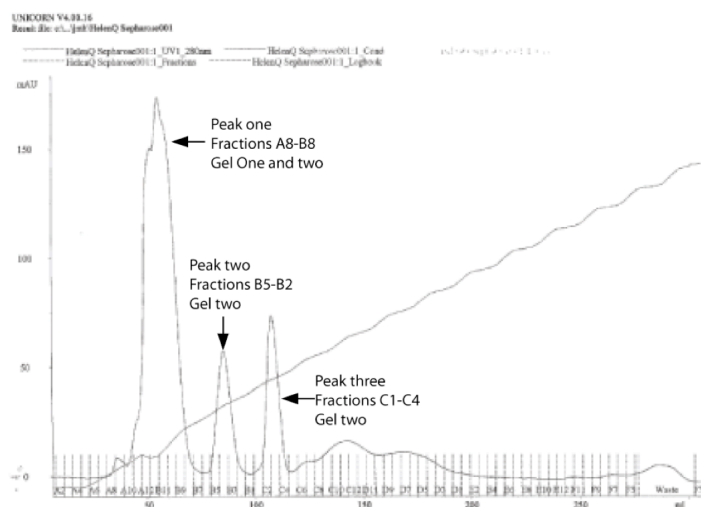
A: A chromatogram to show the UV absorbance at 280nm over the course of the size exclusion chromatography. The fractions relating to each gel lane are highlighted. 4ml fractions were collected using a fraction collector in serpentine mode.

B: 12% SDS-PAGE stained with Coomassie brilliant blue. Fractions B12 to B7 show a higher molecular weight protein (60 KDa), which is likely to be GroEL. All fractions show GST-LMO2 degradation.

3.3.5 Purification of GST-LMO2 via ion exchange chromatography

SDS-PAGE analysis of the size exclusion fractions B12 to B8 (Figure 3.6) shows evidence of GroEL chaperone protein. GroEL is made up of 14 subunits each of which are 60 kDa. Thus, the interaction of GST-LMO2 with GroEL would greatly increase the overall molecular weight and cause GST-LMO2 to be eluted early in the size exclusion run. This suggests a fraction of chaperone GroEL remains tightly bound to GST-LMO2, even after incubation with 2mM ATP and 5mM MgCl₂. The next step was to investigate the option of ion exchange chromatography; a high resolution technique, which separates proteins according to differences in their net surface charge. The estimated pI of GST-LMO2 is 6.88 therefore Q sepharose anion exchange resin was selected, with a pH 8.5 buffer. At pH 8.5 GST-LMO2 would have a net negative charge and bind the positive cationic groups of the resin in the process of anion exchange. 7mg of GST-LMO2 was loaded onto the column and a gradient of 0 to 1M NaCl was used to elute the protein over a volume of 400ml. Approximately 90% of the protein loaded was eluted in the first 80ml (Figure 3.7) with GST only being the major component of all of these fractions.

A:



B:

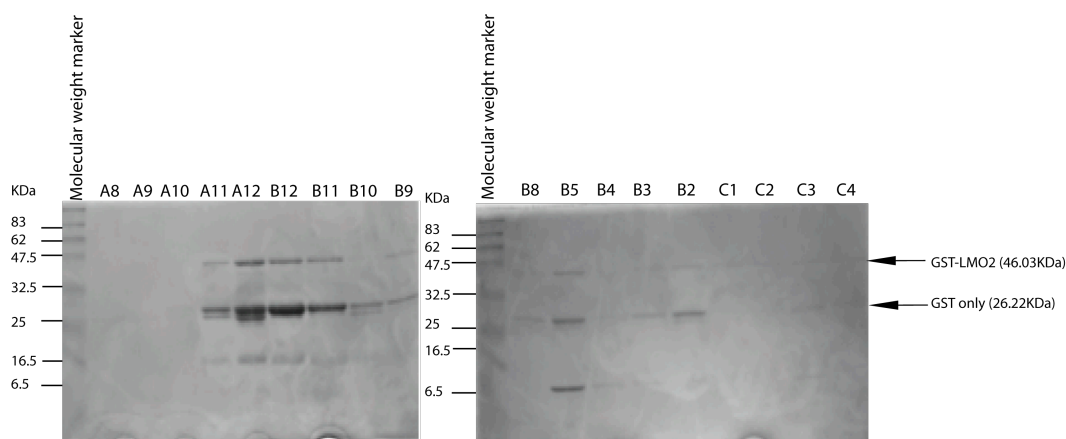


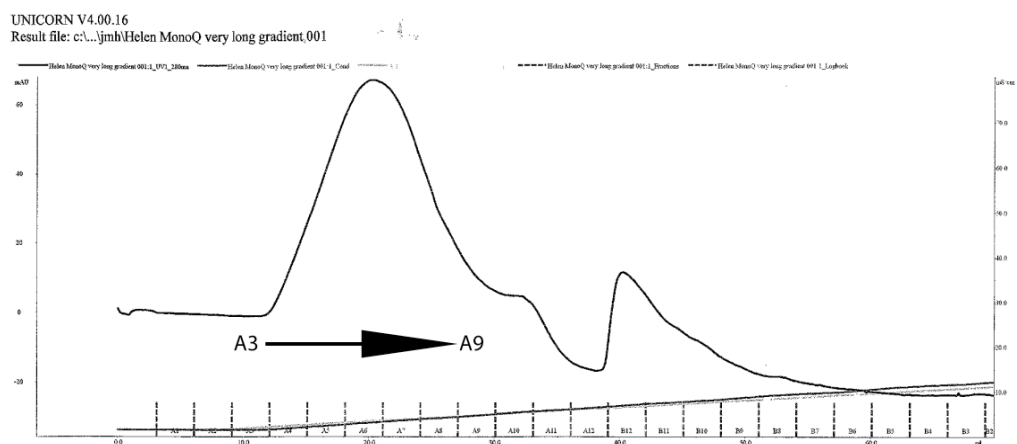
Figure 3.7: GST-LMO2 anion exchange chromatography

A: A chromatogram to show the UV absorbance at 280nm and the NaCl elution gradient (0-1M over 400mls). The fractions collected are also indicated.

B: 4ml fractions were concentrated and analysed by 12% SDS-PAGE stained with Coomassie brilliant blue. The fractions analysed are indicated on the chromatogram (A). The majority of GST-LMO2 (46.03 KDa) was eluted in the first peak along with, over 60%, GST only (26.22 KDa).

Anion exchange appears to have removed GroEL from the protein sample completely. However, the majority of GST-LMO2 was eluted in the main peak along with a large percentage of GST only contaminant (Figure 3.7). A Mono Q anion exchange column (5/50GL) was the next option as it has a higher resolution capacity due to the efficiency of small, perfectly spherical monodispersed particles. Approximately 11.7mg of GST-LMO2, in 60ml of 40mM Tris-HCl, pH 8.5 was loaded onto the column. A flow rate of 2.0 ml/min was used and a gradient of 0 to 1M NaCl, over 500ml, was used to elute the protein. However the column failed to separate the proteins further (Figure 3.8). The GST only protein remained a major contaminant; this could potentially be removed by a TEV protease cleavage of GST, followed by further purification of LMO2 such as size exclusion or ion exchange chromatography. However, due to the large number of problems described, a new strategy was employed and this will be discussed in the next section.

A:



B:

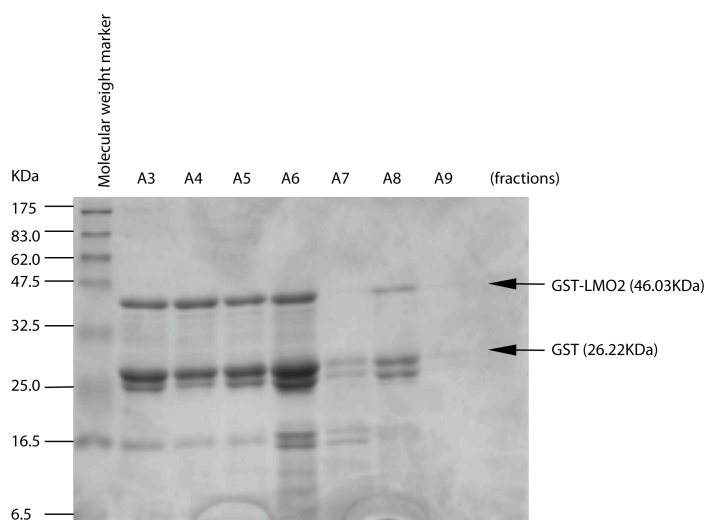


Figure 3.8: Mono Q anion exchange chromatography

A: A chromatogram to show the UV absorbance at 280nm and the NaCl elution gradient. The fractions collected are also indicated.

B: 3ml fractions were concentrated and analysed by 12% SDS-PAGE stained with Coomassie brilliant blue. The fractions analysed are from the peak indicated on the chromatogram (A). Again the peak contains both GST-LMO2 (46.03 KDa) and GST only (26.22 KDa).

3.4 Expression and purification of NusA-LMO2

The problems with GST-LMO2 purification, highlighted in the previous section, warranted experimentation of an alternative approach for the expression of LMO2. Different gene fusions have been shown to increase the solubility and stability of a target protein (Arnau et al., 2006). NusA (55 KDa) was chosen as an alternative fusion protein due to its high solubility when over expressed in *E. coli* (Harrison R.G., 2000). It cannot itself be used as an affinity tag therefore the fusion construct employs an N-terminal His tag for purification. His tags are widely used for purification via immobilized-metal affinity chromatography. In this case chelated Ni²⁺ ions were used as the affinity ligand. The Ni²⁺ is complexed with an immobilised chelating agent, Nitrilotriacetic acid (NTA). NTA occupies four of the six ligand binding sites in the coordination sphere of the nickel ion, leaving two sites free to interact with the His tag.

E. coli host strains may also affect recombinant protein expression therefore two were tested in this investigation. C41 and B834 *E. coli* strains were transformed with expression vector pET43a-NusA-LMO2 (Ampicillin resistant) and the expression levels from both strains compared to that of GST-LMO2 from C41. Cell cultures were grown up and protein purified as previously described. Figure 3.9 shows a comparison of the LMO2 fusion proteins purified. In both cases (C41 and B834 strains) the purification of NusA-LMO2 had a high background of contaminant proteins; this was later reduced by the addition of a size exclusion chromatography step. *E. coli* strain C41 provided a higher level of NusA-LMO2 protein expression and was therefore selected as the host strain for further purification experiments. NusA-LMO2 (78.5 KDa) size exclusion chromatography (superdex 200, 10/300 GL) gave an efficient separation of proteins (Figure 3.10) with a final NusA-LMO2 purity of around 95% as estimated by the assessment of SDS-PAGE (Figure 3.10).

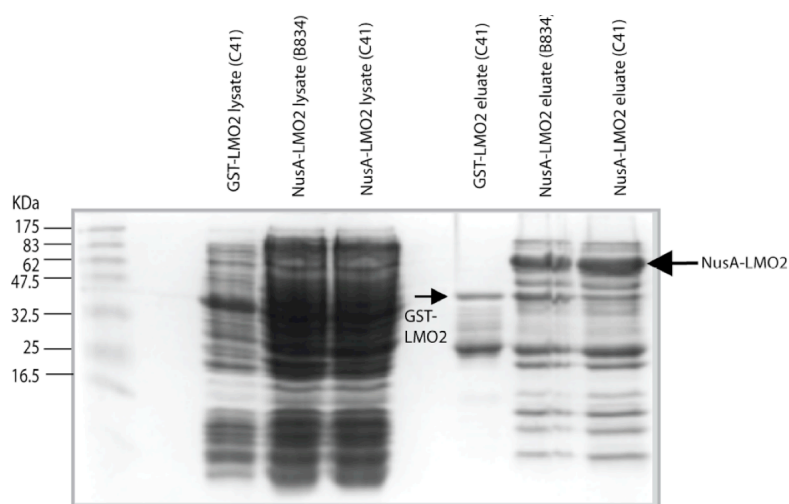


Figure 3.9: Comparison of LMO2 fusion tags

12% SDS-PAGE analysis shows NusA-LMO2 (78.5 KDa) is expressed at a higher yield than GST-LMO2 (46.03 KDa). *E. coli* strain C41 provided a higher level of NusA-LMO2 protein expression than B834. However, the purity of NusA-LMO2 was low due to a high level of background contaminant proteins, which were later removed by size exclusion chromatography.

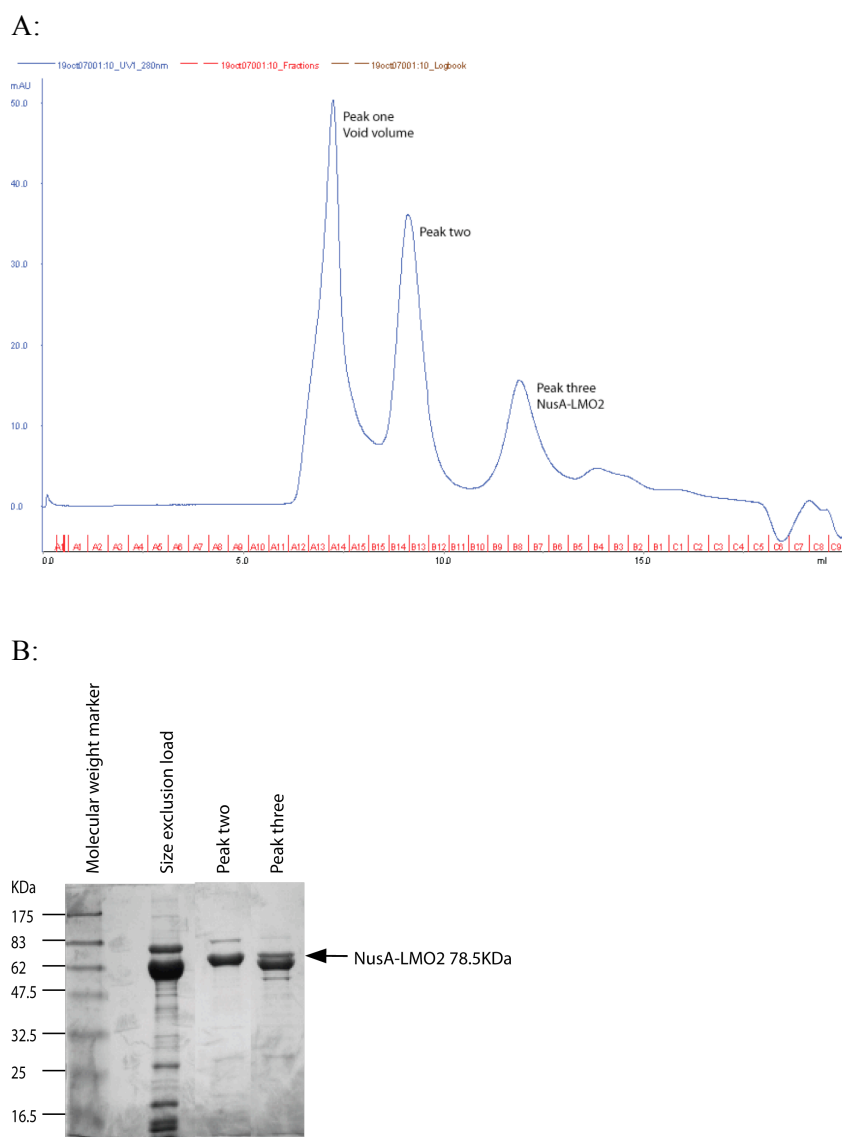


Figure 3.10 NusA-LMO2 size exclusion chromatography

A: Size exclusion chromatography analysed by measuring UV absorbance at 280nm. 0.5ml fractions were collected as indicated in red.

B: 12% SDS-PAGE stained with Coomassie brilliant blue. The fractions analysed are indicated on the chromatogram (A). Fractions from each peak were pooled and concentrated. From the calibration curve for this column (Appendix A, figure 1), peak three corresponds most accurately to the weight of NusA-LMO2 (78.5 KDa). There is a larger contaminant protein of approximately 80 KDa causing a percentage of NusA-LMO2 to elute in peak two.

The use of NusA as a fusion partner has greatly improved the purification of LMO2 as a fusion protein. NusA-LMO2 shows limited degradation in comparison with GST-LMO2 and can be further purified effectively by size exclusion chromatography. Nevertheless, the next section will describe the implementation of a protein expression screen to ascertain if further improvements could be made upon the purification of LMO2 as a fusion protein.

3.5 Protein expression and purification screen

Oxford protein production facility (OPPF) specialise in high throughput structural biology with particular expertise in vector construction and protein expression screening. This section will describe work, carried out by H. Sewell, at this research complex. For cloning and protein purification methods please refer to chapter two, section 2.3.

3.5.1 OP PF Expression and purification screen

To investigate whether LMO2 could be purified more successfully, a high throughput expression screen was carried out at Oxford protein production facility (OPPF). The OP PF vector suite was used to make expression vector constructs for LMO2, LDB1 and TAL1_{bh1h}, using the In-fusion (Clontech, France) ligation independent cloning method. LDB1 and TAL1 interact with LMO2 in a DNA binding complex therefore, it may be possible to co-express LMO2 with either TAL1_{bh1h} (minimal interaction region of the protein) or LDB1 to increase its stability and solubility. The expression vectors produced are listed in the Table 3.1.

The expression vectors were designed so all fusion proteins expressed are also fused to a His tag. Ni-NTA magnetic beads were used to purify all proteins expressed in the screen in a high throughput, semi automated manner. The results (Figure 3.11) showed the best candidates for scale up were:

- I. His-MBP-3C-LMO2
- II. His-MBP-3C- TAL1_{bh1h}
- III. His-MBP-3C-LDB1

The maltose binding protein (MBP) and LMO2 fusion protein expressed and purified to the highest yield of more than 5mg/ml. MBP (44 KDa) can be used for affinity purification with amylose resin followed by mild elution using maltose. The advantages of this method are that both binding capacity and specificity are high meaning high yields and purity can be achieved in one step.

Table 3.1: Fusion proteins and respective solubility

Lane	Protein	Expression Tag	Expected size	Soluble expression
M	Protein marker			
1	LMO2	His-SUMO	30.65	
2	TAL1 _{bhlh}	K His (kanomycin resistance)	10.09	
3	TAL1 _{bhlh}	MAHHHHHHSSGLEVLFGQP	10.09	
4	TAL1 _{bhlh}	His-GST-3C	33.8	<0.5mg/l
5	LMO2	K His (kanomycin resistance)	20.44	
6	LMO2	MAHHHHHHSSGLEVLFGQP	20.44	
7	LMO2	His-GST-3C	44.15	<0.5mg/l
8	LMO2	His-MBP-3C	58.9	>5mg/l
9	LDB1	His-GST-3C	72.3	
10	LDB1	His-MBP-3C	87.05	0.5-5mg/l
11	LDB1	His-SUMO	58.8	
12	TAL1 _{bhlh}	His-SUMO	20.3	
13	TAL1 _{bhlh}	His-MBP-3C	48.55	0.5-5mg/l
14	LDB1	K His (kanomycin resistance)	48.59	
15	LDB1	MAHHHHHHSSGLEVLFGQP	48.59	

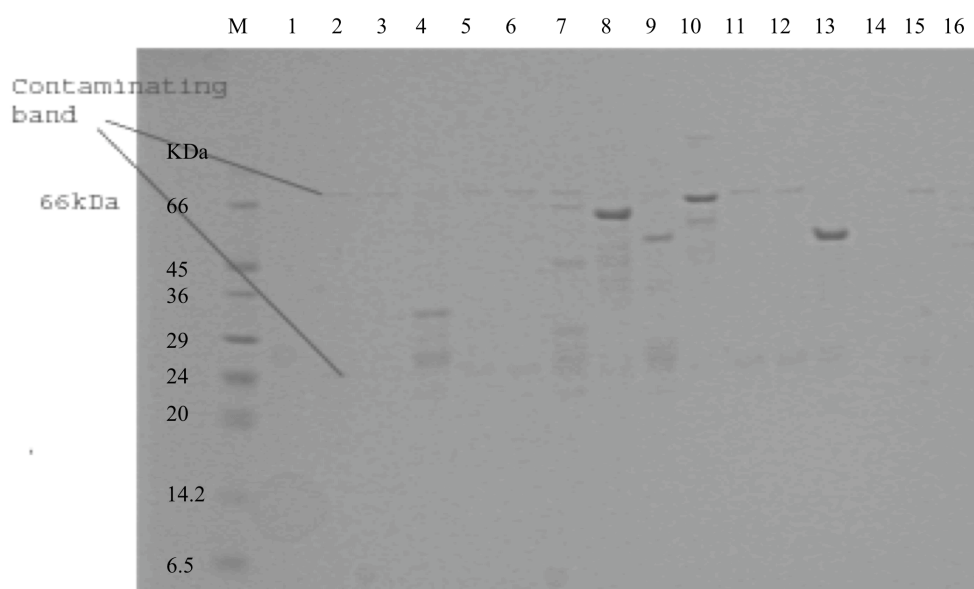


Figure 3.11: OPPF protein expression screen

Fusion protein expression was analysed by SDS-PAGE and stained using SafeStain. Table 3.1 indicates the protein and fusion partner purified along with its expected size, solubility and lane number. The results show that MBP protein is an effective fusion partner for LMO2, LDB1 and TAL1_{bhlh} as all constructs purify to a yield of more than 0.5mg per liter. Other fusion proteins expressed and purified successfully, although to a lower yield (less than 0.5mg/l), were His-GST-3C-LMO2 and His-GST-3C-Tal1_{bhlh}.

3.5.2 Comparison and further analysis of NusA-LMO2 and MBP-LMO2

SDS-PAGE analysis provided an initial assessment of protein sample quality and purity. Mass spectrometry was then employed to ensure the identity of the purified protein. Protein identification of NusA-LMO2 and MBP-LMO2 was confirmed by peptide mass fingerprinting using mass spectrometry. This work was carried out by Dr. Jeff N. Keen (Leeds Institute of Genetics, Health and Therapeutics, The University of Leeds). Peptides were generated by trypsin digest of the fusion protein and the mass of each peptide was accurately measured using a matrix assisted laser desorption time of flight (MALDI-TOF) mass spectrometry method. The mass/charge (m/z) ratios were recorded along with their relative abundance in the form of a m/z spectrum, which is shown in Appendix A, figure 2 and 3. Peptide peaks were matched between those predicted and the actual spectrum. Both proteins were confirmed to be correct with NusA-LMO2 and MBP-LMO2 giving 49% and 60% peptide sequence coverage respectively.

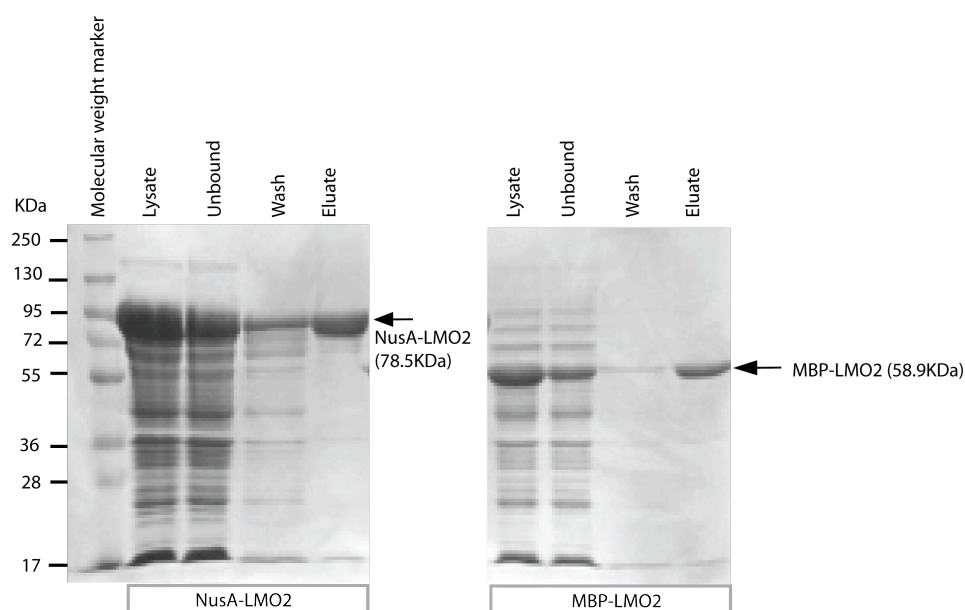


Figure 3.12: Comparison of NusA-LMO2 and MBP-LMO2 purification

10% SDS-PAGE stained with Coomassie brilliant blue. Samples were taken at each stage of both purifications (lysate, unbound, wash and eluate) and analysed as shown. Eluate bands from both NusA-LMO2 and MBP-LMO2 were excised and used to carry out MALDI-MS peptide mass fingerprinting to confirm the identity of the proteins.

3.5.3 Analysis of NusA-LMO2 and MBP-LMO2 by Circular Dichroism

Further analysis of the quality of the LMO2 fusion proteins purified took the form of Circular Dichroism. “Circular Dichroism spectroscopy plays an important role in the study of protein folding and it allows the characterisation of secondary and tertiary structures of proteins in native, unfolded and partially folded states” (van Mierlo and Steensma, 2000). Although Circular Dichroism spectra would mainly reflect the structure of the large LMO2 fusion tags, data was collected to determine if NusA-LMO2 and MBP-LMO2 were folded or not. Circular Dichroism spectroscopy measures differences in absorbance of right and left circularly polarised light. Peptide bonds dominate the far-UV region (190-250 nm) of the spectra and can be used to characterise the secondary structure.

The Circular Dichroism spectra of NusA-LMO2 and MBP-LMO2 can be found in Appendix A, figure 7 and both have the characteristic minimum at around 215nm; this is typical for a structure with considerable β -sheet content. From this data, we can confirm that both proteins are folded, as the spectra of both are negative at more than 208nm. CDSSTR (Johnson, 1999) obtained from DICHROWEB (Whitmore and Wallace, 2004) was used to analyse the spectra further. CDSSTR uses a set of 43 proteins as a reference for fitting the experimental spectrum. The content of various secondary structures obtained for NusA-LMO2 by this method were: 38% α -helix, 16% β -sheet, 22% β -turn and 24% unordered. For MBP-LMO2 the results, by this method were: 20% α -helix, 30% β -sheet, 26% β -turn and 24% unordered.

3.5.4 MBP-LMO2 for crystallography trials

For crystallography purposes protein tags are ideally removed usually by digestion with a protease, providing that an appropriate recognition site has been engineered into the expression vector. This is because tags may influence the structure of the target protein and may also be flexible, inhibiting crystal formation. As such, a favourable strategy for LMO2 crystallisation would be to remove the N-terminal His-MBP purification tag by 3C protease cleavage. HRV 3C protease is a recombinant form of the 3C protease from human rhinovirus type 14. In this case, the 3C protease (22 KDa) was fused to an N-terminal His tag providing a means to separate the protease, along with His-MBP, from the target protein post cleavage. This was done using a second Ni-NTA chromatography step.

An initial titration experiment was designed to estimate the appropriate ratio of enzyme to target protein. From this experiment it could be concluded that a 1:5 (protease:target protein) ratio was the most effective, with the digest going to completion after 1 hour. A large-scale digest was then carried out overnight at 4°C, during dialysis against imidazole free buffer. Conversely, during this protease reaction, a large amount of free LMO2 precipitated (Figure 3.13). The precipitate was removed by spinning down the dialysed solution (4000 X g) and retained for analysis. The dialysed solution was then passed through 2ml of Ni-NTA resin in a disposable column. Theoretically, LMO2 should pass through the column while His-MBP and His tagged 3C protease should remain bound to the column. However as shown by Figure 3.13, some His-MBP is present in the flowthrough. Nevertheless, the beads were eluted with 250mM imidazole and the eluate analysed showed some His-MBP was retained by the Ni-NTA resin (Figure 3.13).

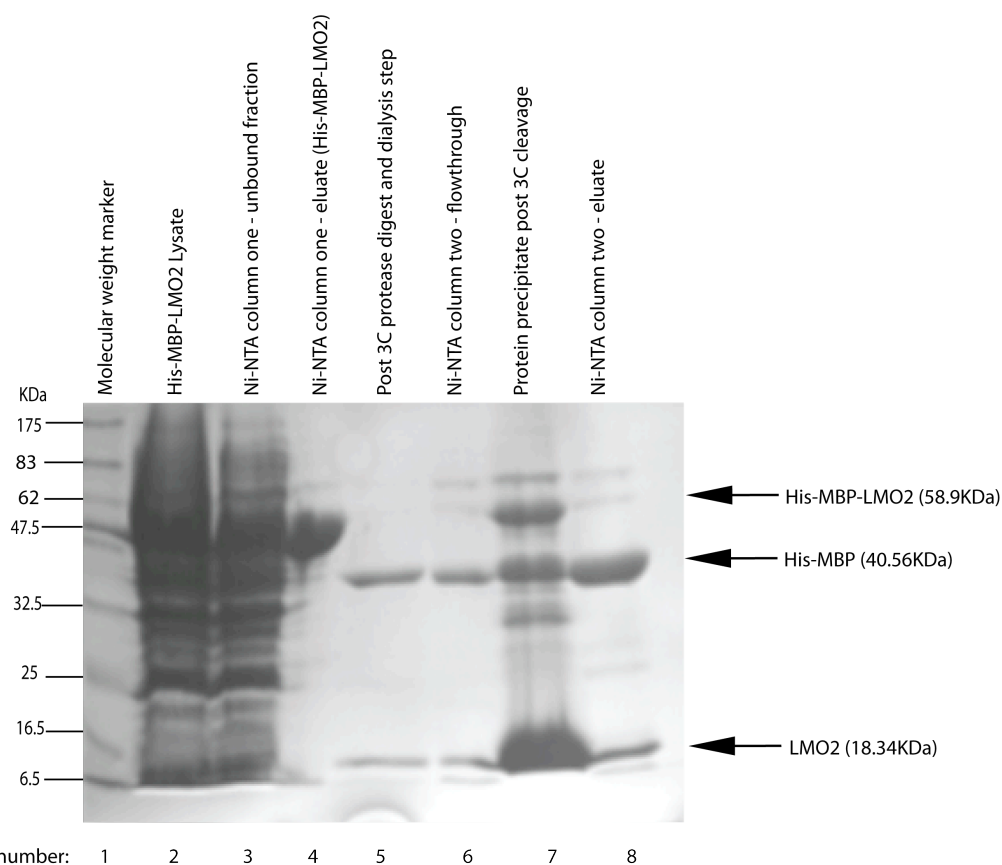


Figure 3.13: His-MBP-LMO2 3C protease digest

12% SDS-PAGE stained with Coomassie brilliant blue. Unfortunately most of the LMO2 protein precipitates upon removal of the His-MBP tag (lane 7). Separation of His-MBP from LMO2, via Ni-NTA chromatography is poor. This can be seen by the presence of His-MBP in the flowthrough fraction (lane six). However the Ni-NTA resin did retain some His-MBP as can be seen by the analysis of the Ni-NTA eluate (lane 8).

3.5.5 MBP-LMO2 rigid linker strategy

Cleavage of the MBP expression tag, from MBP-LMO2, resulted in precipitation of free LMO2, even at low concentrations (Figure 3.13, lane 7). One strategy investigated, to prevent the precipitation of LMO2 upon cleavage from MBP, was to leave the affinity tag in place for crystallisation trials. Leaving the MBP tag in place would have some advantages for crystallography, such as providing a known structure to facilitate phase determination by molecular replacement. The flexible linker region, between MBP and LMO2, was engineered to give a short rigid linker with the anticipation that this would reduce the potential for conformational heterogeneity (Smyth et al., 2003).

The MBP expression vector, pOPIN_M, was modified to include a NotI site. This provided a DNA sequence to code for two alanine amino acids with potential for addition of further residues to create the rigid linker. This was achieved using appropriately designed primers (see Appendix C for primer sequences) to amplify LMO2 and at the same time incorporate 5' additions (linker region). Three expression plasmids were constructed to express:

- I. MBP-AAA-LMO2
- II. MBP-AAAAA-LMO2
- III. MBP-AAAEF-LMO2

The expression vectors were used to transform *E. coli* Rosetta and protein expression induced. Amylose resin was used to purify the protein, which was then eluted in 10mM Maltose. All constructs were expressed and highly soluble with a yield of approximately 40mg per liter.

The appropriate protein bands from the gel shown in Figure 3.14 were excised and analysed by MALDI-MS fingerprinting (carried out by Dr. Jeff N. Keen) to confirm the identity of the purified products (see Appendix A, figure 4). Size exclusion chromatography was used to determine the quaternary structure of MBP-AAA-LMO2 i.e. if it is forming a dimer/trimer/tetramer. The protein eluted at a volume equivalent to a protein size of 120 KDa, which would make MBP-AAA-LMO2 (60.22 KDa) a dimer although, this was further investigated by Analytical ultracentrifugation as described in the next section.

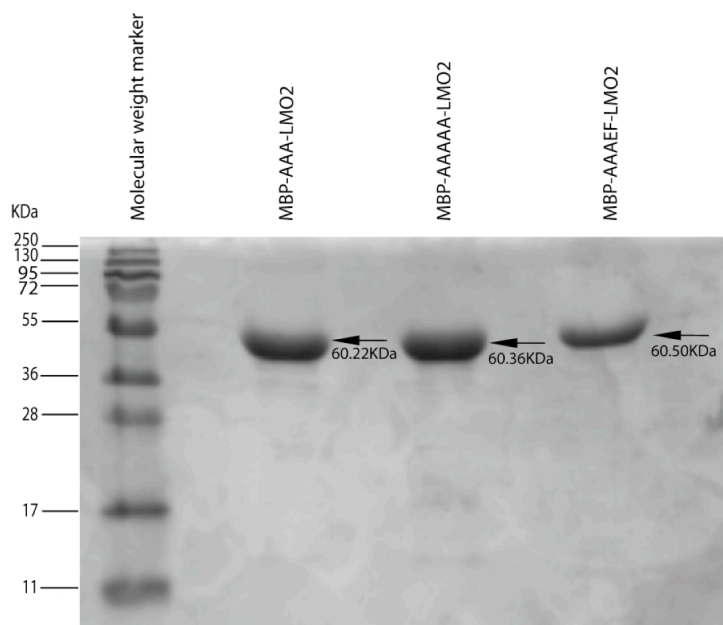


Figure 3.14: Expression and purification of MBP-LMO2 rigid linker constructs

12% SDS-PAGE stained with Coomassie brilliant blue. The gel shows all constructs can be successfully expressed and purified to a high degree of purity, estimated to be more than 90%.

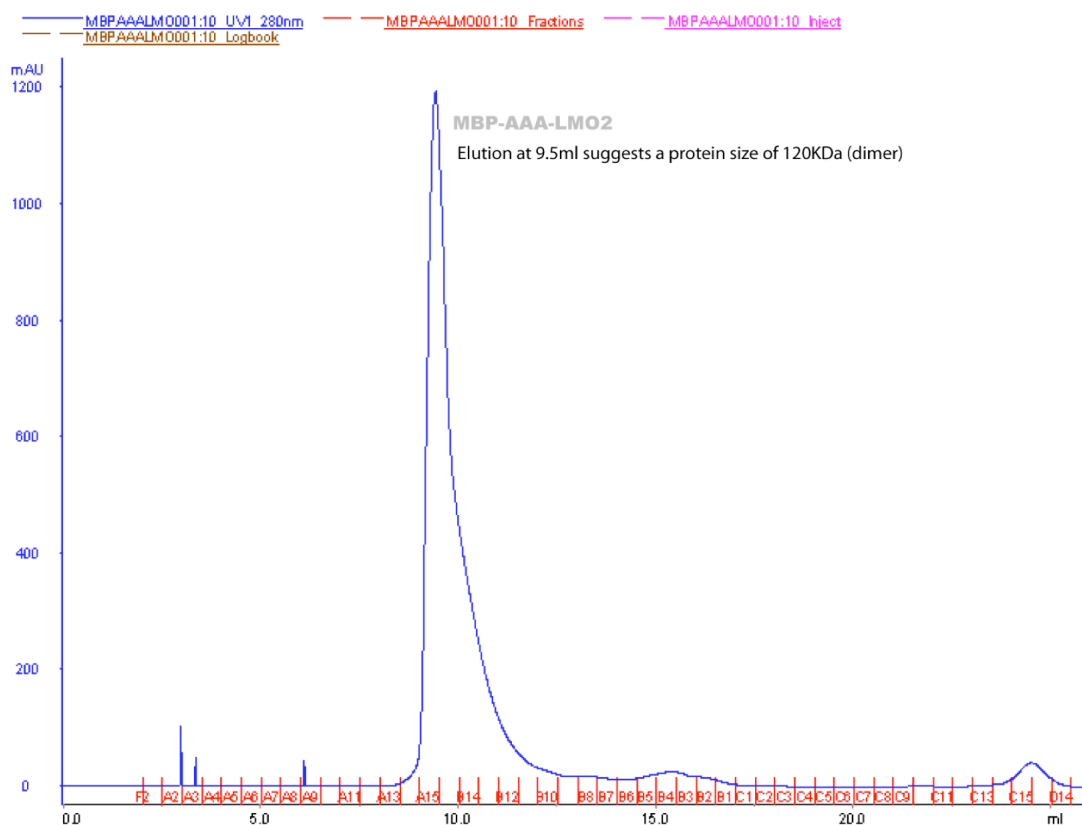


Figure 3.15: Size exclusion chromatography of MBP-AAA-LMO2

MBP-AAA-LMO2 protein was concentrated to 10mg/ml and loaded onto a Superdex 200 10/300 GL column (void volume of 8mls). The chromatography was analysed by measuring UV absorbance at 280nm. The protein sample eluted at a volume of 9.5mls, which equates to a protein size of 120 KDa; this suggests MBP-AAA-LMO2 is forming a dimer. However, further analysis confirmed MBP-AAA-LMO2 was actually a soluble aggregate so eluting in the void volume.

Size exclusion can be used to predict native protein molecular weight although limitations include poor resolution for larger proteins. Furthermore, dilution during the chromatography may reverse any weak interactions as the protein concentration decreases (Liu et al., 2006). MBP-AAA-LMO2 eluted at an elution volume close to the void volume of the size exclusion column and consequently, Analytical ultracentrifugation was utilised to determine a more precise molecular weight for the protein under native conditions. MBP-AAA-LMO2 was analysed at three different concentrations. A form of Analytical ultracentrifugation, termed Sedimentation velocity, was used to analyse the rate at which MBP-AAA-LMO2 formed a unique boundary at a characteristic speed. The velocity and shape of this moving boundary was used to estimate the sedimentation coefficient and molecular weight of the protein. The results showed that all samples contained aggregated material sedimenting in the range of 10 to 50 S (Svedbergs), representing molecular masses in the range of 200 KDa to 3 MDa. Larger particles have higher Svedberg values, which are measured using equation 3.1:

Equation 3.1

Sedimentation Coefficient (S) = Constant speed of sedimentation (ms^{-1}) / acceleration applied (ms^{-2})

In all samples, a small proportion of material sedimented in the range 3.5 to 4.8 S. The molecular mass of this material was calculated at between 35 and 100 KDa, however the data was not good enough to support a more definite figure. Figure 3.16 shows the results in the form of a graph of S-value distributions normalised to loading concentration, which indicates MBP-AAA-LMO2 soluble aggregate formation is dependent on concentration. Figure 3.16 gives the exact percentage of soluble aggregate for each concentration of protein analysed. Another observation that supported the conclusion MBP-AAA-LMO2 was a soluble aggregate was the ease at which the sedimented protein could be resuspended. This rendered MBP-AAA-LMO2 rigid linker protein a non-viable option for crystallography.

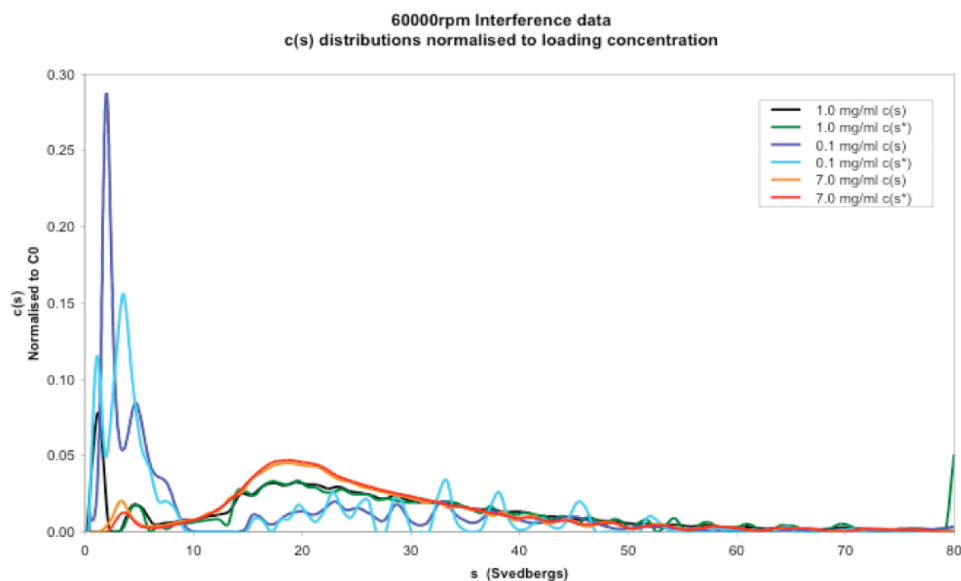


Figure 3.16: Beckman XL-I Analytical ultracentrifugation

This is a plot of S-value distributions normalised to loading concentration and shows that the proportion of aggregated MBP-AAA-LMO2 protein was dependent on concentration. For each protein concentration of 0.1mg/ml, 1mg/ml and 7mg/ml, the percentage of soluble aggregate was calculated as 39%, 85% and 93% respectively.

3.6 Expression and purification of LMO2 with antibody single domain, VH#576

The option of co-expressing LMO2 with an antibody domain was explored as a means to stabilise LMO2, preventing its aggregation in solution. In the past antibody fragments have also been used to aid crystallisation (Kovari et al., 1995). An anti-LMO2 antibody, variable heavy chain, single domain (VH#576) has been previously identified using yeast intracellular antibody capture technology (Tanaka and Rabbitts, 2010).

LMO2 and VH#576 were co-expressed from a single plasmid, pRK-His-Tev-VH#576-LMO2 (obtained from T. Tanaka, THR lab) made up of bicistronic mRNA with an internal ribosome entry site. The complex was purified using nickel-agarose affinity chromatography (Figure 3.17) followed by further purification step using a Hi load 16/60 superdex 75 (S75) prep grade size exclusion column. The UV absorbance at 280nm (Figure 3.18) was monitored and fractions from the appropriate peak pooled and concentrated. The protein eluted at a volume, which reflects the molecular weight of a 1:1 protein complex, 32.3 KDa. The final protein sample, in a buffer of 20mM Tris-HCl, pH 8, 200mM NaCl, and 1mM DTT, was concentrated to 11.3mg/ml and used to set up crystallography trials (Figure 3.19).

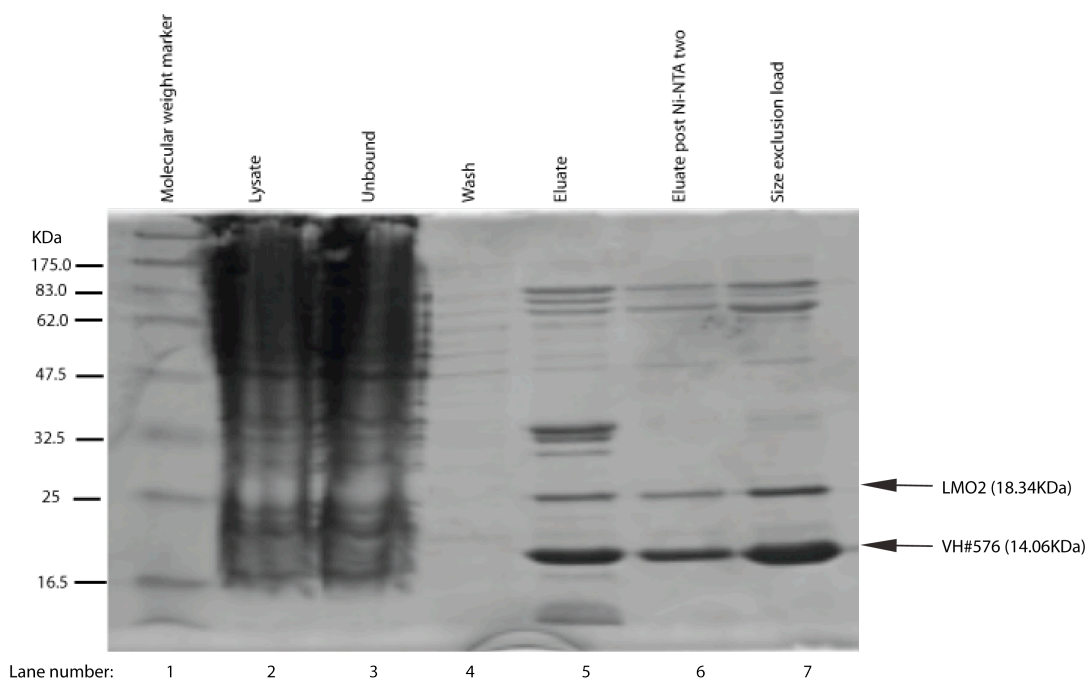


Figure 3.17: His-VH#576/LMO2 purification by Ni-NTA chromatography

15% SDS-PAGE stained with Coomassie brilliant blue. VH#576/LMO2 complex was purified via the N-terminal His tag of VH#576 and the eluate is shown in lane five. The eluate of the second Ni-NTA chromatography step is shown in lane six. The concentrated eluate (size exclusion load) shown in lane seven suggests the complex stoichiometry is not 1:1 with more VH#576 present than LMO2. There are also some high molecular weight contaminant proteins present, with molecular weights greater than 62 KDa.

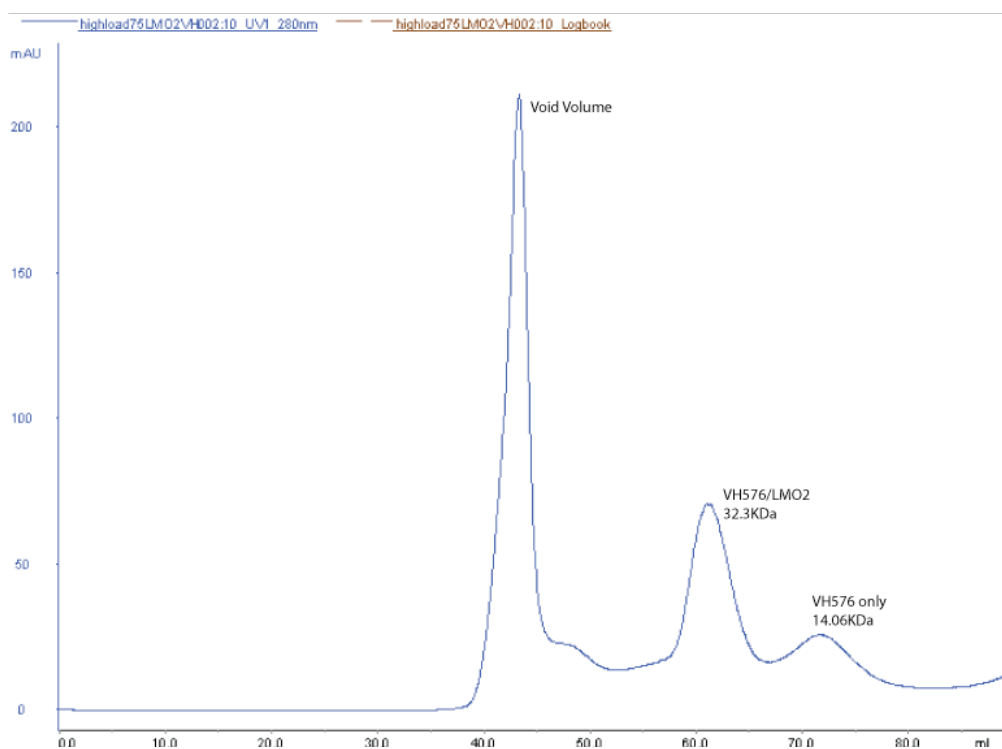


Figure 3.18: VH#576/LMO2 size exclusion chromatography chromatogram

VH#576/LMO2 complex was concentrated to 500 μ l at 10mg/ml and loaded on S75 16/60 size exclusion column. Large molecular weight proteins and aggregated VH#576/LMO2 were eluted in the void volume (first peak). The excess VH#576 protein was separated from the LMO2 bound VH#576 as shown by the final peak. The complex elutes at a volume, which represents an appropriate molecular weight of 32.3 KDa, and this reflects a protein complex stoichiometry of 1:1. Fractions from the appropriate peak (peak two) were pooled and concentrated to 11.3mg/ml for crystallography screens.

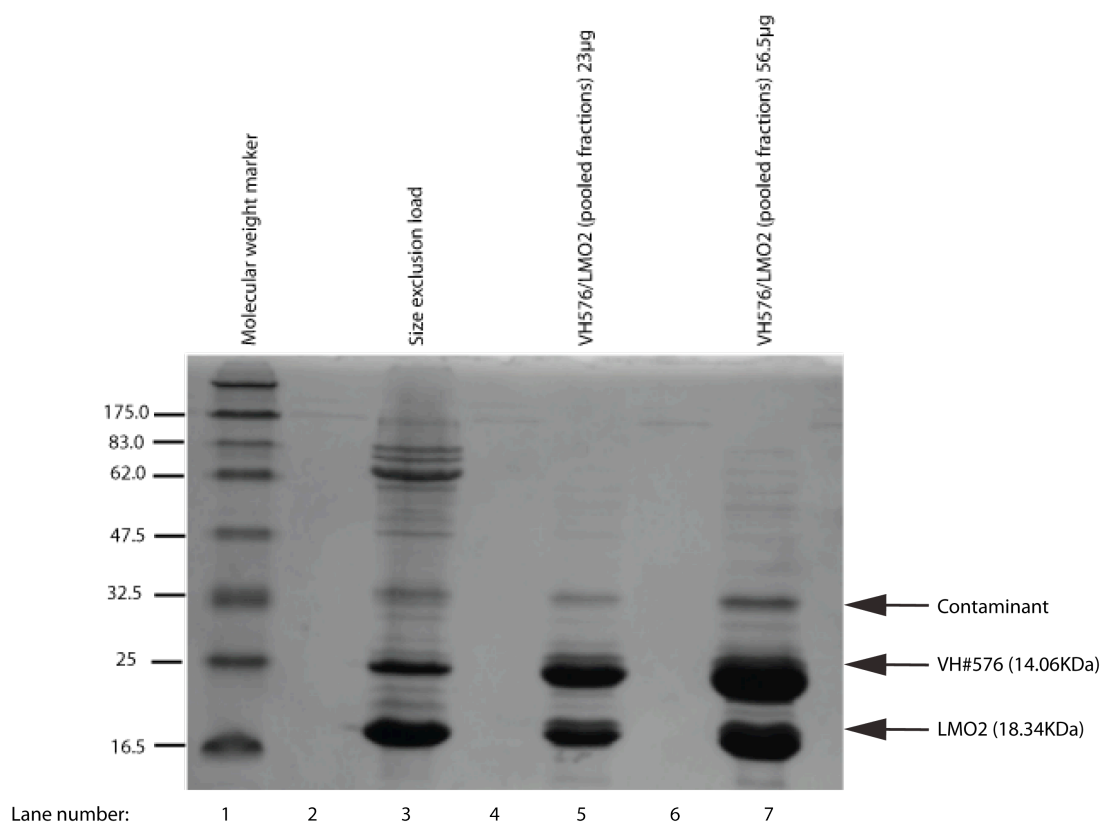


Figure 3.19: VH#576/LMO2 size exclusion chromatography load and pooled fractions

15% SDS-PAGE stained with Coomassie brilliant blue. The purified protein was overloaded on the gel so that any impurities could be visualised. The final protein product was estimated to be around 90% pure with the major contaminant being Tev protease. The protein sample (lane seven) was used immediately to set up crystallisation screens.

3.6.1 X-ray crystallography screen with VH#576/LMO2 protein complex

The first step in setting up initial crystallography screens was making sure the target protein was at an appropriate concentration, for crystallography, using a pre crystallisation test (PCT, Hampton Research, CA, USA). Setting up crystallisation screens with an optimised protein concentration should reduce the number of clear and precipitate results thus increasing the chances of crystallisation. The PCT was carried out using VH#576/LMO2 protein at a concentration of 11.3mg/ml. The results, which are shown in Table 3.2, indicated that lighter crystallisation screens (weaker precipitants) were more appropriate. The screens were performed with the protein at 11.3mg/ml.

Table 3.2: Results of the pre crystallisation test

	PCT reagent A1/B1 result	PCT reagent A2/B2	Suggested action
A1/A2	Heavy amorphous precipitate	Heavy amorphous precipitate	Dilute sample 1:1
B1/B2	Clear	Light granular precipitate	Perform screen

Initial crystallisation trials were performed using the facilities at the Astbury centre for structural molecular biology (The University of Leeds). The sitting drop vapour diffusion method for crystallisation was implemented along with the following commercial screening kits (Hampton Research, CA, USA):

- I. Hampton I and II
- II. Salt RX
- III. Index
- IV. Natrix
- V. MembFac

Ten 96 well plates (MRC2, Hampton Research, CA, USA) were used for the crystallisation trials. The plates were set up by dispensing 100µl (per well) of the screening solution from the master block to the reservoir of the crystallisation plate. An Impax automatic protein crystallisation system robot (Douglas Instrument, Hungerford, UK) was then used to dispense 1µl drops onto the platform of each well: 0.5µl of reservoir solution plus 0.5µl of protein (VH#576/LMO2 complex). The plates were then sealed with clear seal film (Hampton Research, CA, USA). Plates were prepared in duplicate so that one could be stored at 4°C and the other at 18°C.

Plates were left for one week and then inspected under a microscope (Leica MZ6 and Leica MZ12) for crystals, precipitate, crystalline precipitate, or clear drops. Any crystals were then tested for the presence of protein. This can be done in many ways:

- I. IZIT dye staining (Hampton Research, CA, USA); IZIT is a small molecule dye which will fill the solvent channels in protein crystals, colouring the crystals blue. Salt crystals do not possess these large solvent channels therefore, the dye cannot enter the crystals and they remain clear.
- II. Gel electrophoresis and silver staining.
- III. Mount the crystal and analyse the diffraction pattern.
- IV. Shine UV light on the crystal, if it absorbs and emits the UV light it is likely to be protein.

In this case option four was used, however no protein crystals resulted from any of the crystallisation screens.

3.6.2 Removal of N-terminal residues from LMO2 and expression with antibody single domain, VH#576

Disordered regions of a protein may affect the solubility and crystallisation of a protein. “Many proteins contain local regions of such disorder, and some appear to be totally unfolded in their native states” (Yang et al., 2005). Structure determination by X-ray crystallography relies on ensembles of identical structures to amplify the experimental signal and therefore it was crucial to identify any disordered regions of LMO2. To predict disorder an algorithm has been developed, termed Regional order neural network (RONN) (Yang et al., 2005), based on data sets of protein disorder from the protein data bank (PDB). Figure 3.20 shows the probability of disorder of LMO2 as predicted by RONN. The graph suggests the first twenty N-terminal residues of LMO2 have a high probability (greater than 0.5) of being disordered.

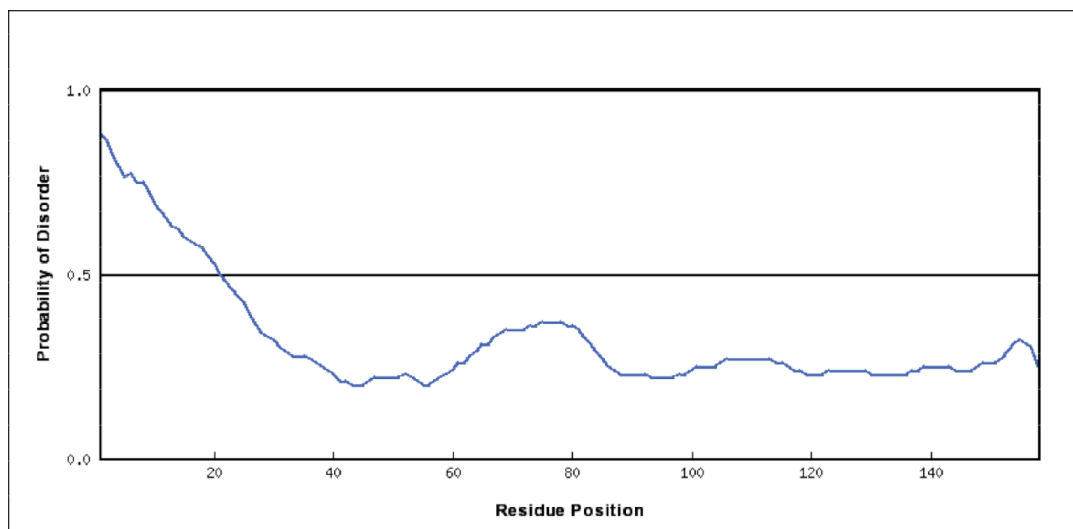


Figure 3.20: Probability of disorder of LMO2 as predicted by RONN

The graph suggests the first twenty residues of LMO2 have a high probability (greater than 0.5) of being natively disordered, i.e. do not form a well defined three-dimensional structure. The results of this algorithm suggest the N-terminal tail of LMO2 may hinder crystal formation.

To remove these residues the expression plasmid (pRK-His-Tev-VH#576-LMO2) was engineered, removing varying numbers of amino acids from the N-terminus of LMO2. A protein expression screen was carried out, purifying each of the constructs by Ni-NTA chromatography only. As shown by Figure 3.21, removing more than seven N-terminal residues, from LMO2, reduced solubility and stability of the complex. Consequently, VH#576/LMO2 Δ N7 was fully purified (Figure 3.22 and Figure 3.23), concentrated to 8mg/ml and used to set up further crystallisation trials. Unfortunately, all screens were once again unsuccessful.

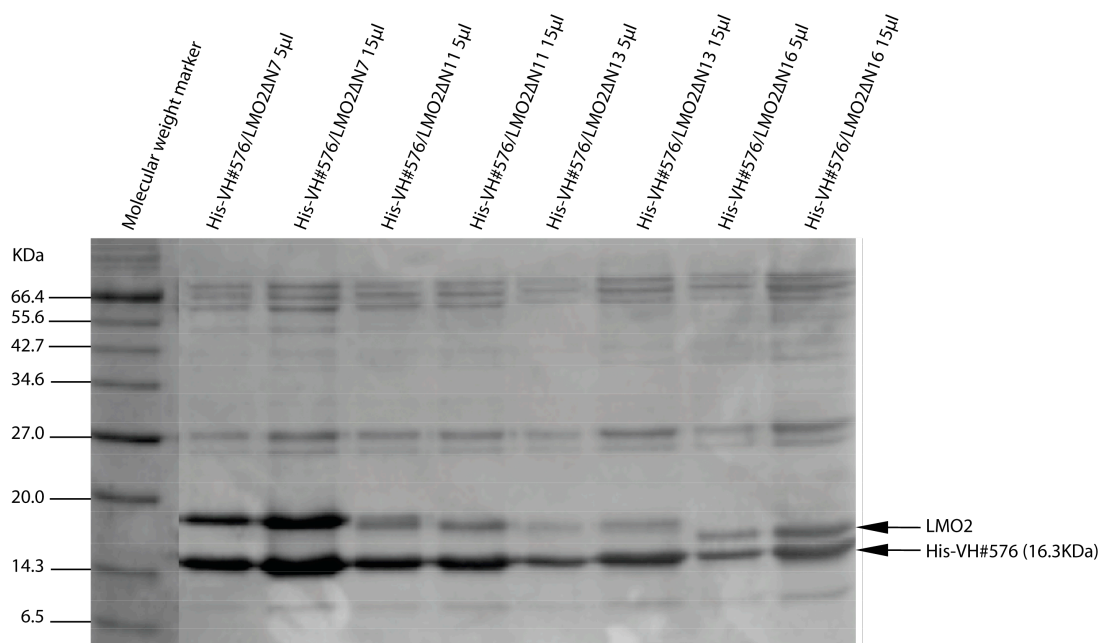


Figure 3.21: Effect of various LMO2 N-terminal truncations on the purification with VH#576

15% SDS PAGE stained with Coomassie brilliant blue. The gel shows that the construct expressing LMO2ΔN7 was co-purified most efficiently with VH#576. Removing more than seven residues from the N-terminal of LMO2 reduces the yield of the complex purified.

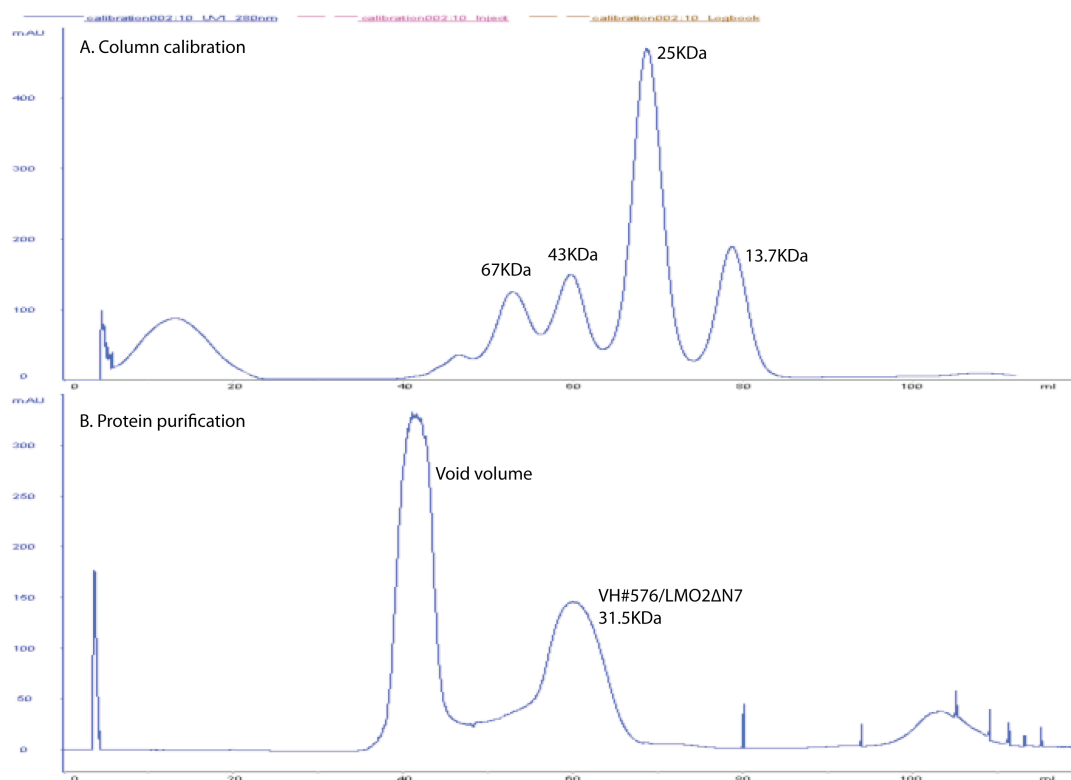


Figure 3.22: Purification of VH#576/LMO2 Δ N7 complex

A: Calibration of Hi load 16/60 superdex 75 prep grade. The chromatogram represents the UV absorbance at 280nm.

B: Purification of VH#576/LMO2 Δ N7 protein complex by size exclusion chromatography using a Hi load 16/60 superdex 75 (S75) prep grade column. The complex eluted at a volume, which reflects a protein size of 31.5 KDa

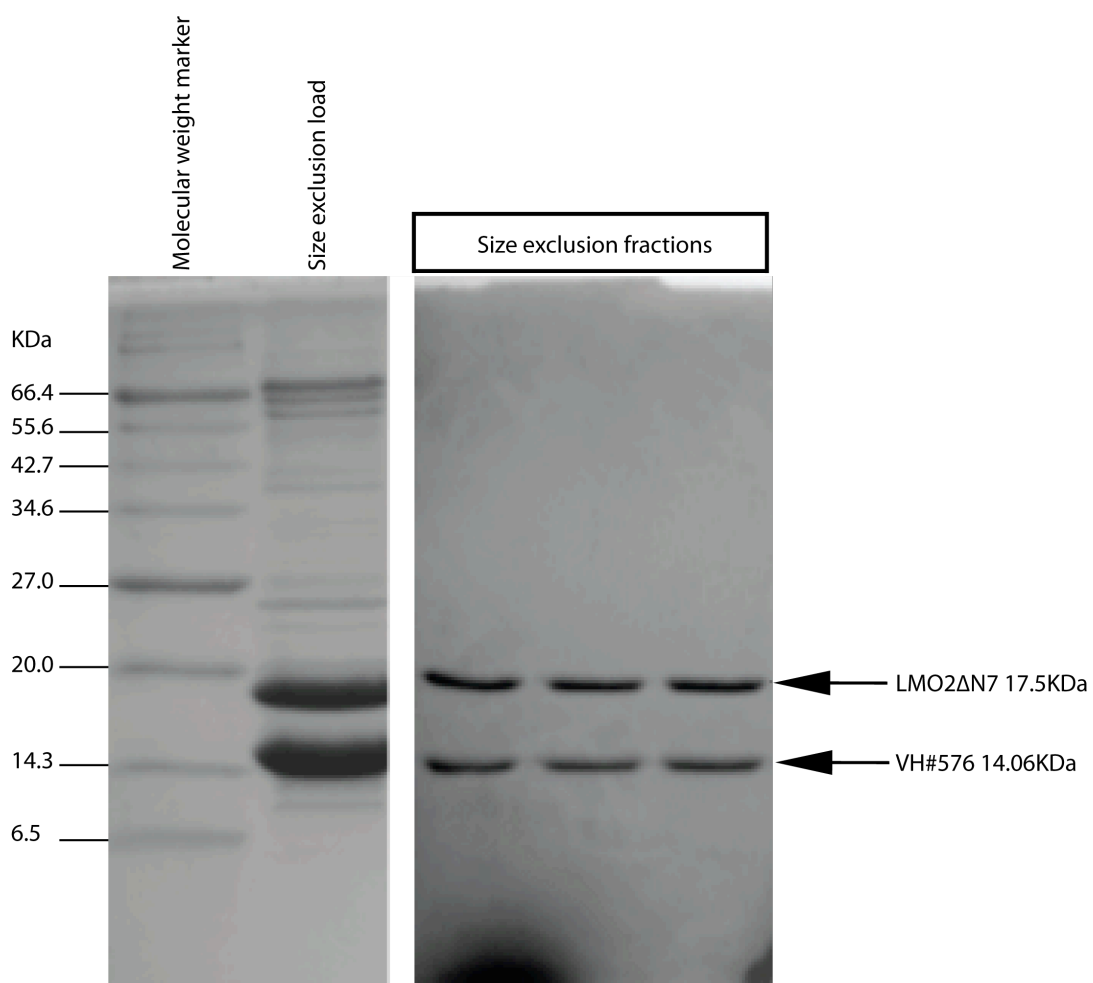


Figure 3.23: VH#576/LMO2ΔN7 purification by size exclusion

15% SDS-PAGE gel stained with Coomassie brilliant blue. The gel shows the VH#576/LMO2ΔN7 size exclusion chromatography protein load and the fractions from the relevant peak (Figure 3.22).

3.6.3 Lysine methylation of VH#576/LMO2 Δ N7 protein complex

Lysine methylation was suggested (Personal communication from T. Walter) as an alternative strategy to improve the chance of successful crystallisation. Recently, methylation of surface lysines has been used to chemically modify the surface of a protein, reducing surface entropy with a view to increasing the chance of crystal formation (Kim et al., 2008). Only lysine residues on the surface of the protein are methylated and not those at the interface of a strong protein-protein interaction. Primary amines of lysine residues and N-termini are modified to tertiary amines, by reductive methylation. Mass spectrometry was used to determine the extent of methylation. For structures solved using this method, methylated lysine residues were found at crystal contact points, therefore surface methylation may promote protein-protein interaction (Walter et al., 2006).

VH#576/LMO2 Δ N7 was expressed and purified by Ni-NTA chromatography as per the usual protocol. This was followed by removal of the His tag from VH#576 and subsequent methylation of the surface lysine residues. The reaction was quenched using size exclusion chromatography and this resulted in a well resolved peak, at the expected elution volume (Figure 3.24). The overall yield was 1mg per liter prep.

Positive electrospray mass spectrometry (Appendix A, figure 5) was performed by Dr. James Ault (The University of Leeds) and used to measure the molecular weight of the purified proteins. The mass spectrum shows multiply charged ions, and this data was transformed on to a molecular mass scale using the maximum entropy processing technique. The results indicated VH#576 had four lysines methylated and LMO2 had ten and therefore all lysine residues in the complex had been successfully methylated plus the N terminal amine of both proteins.

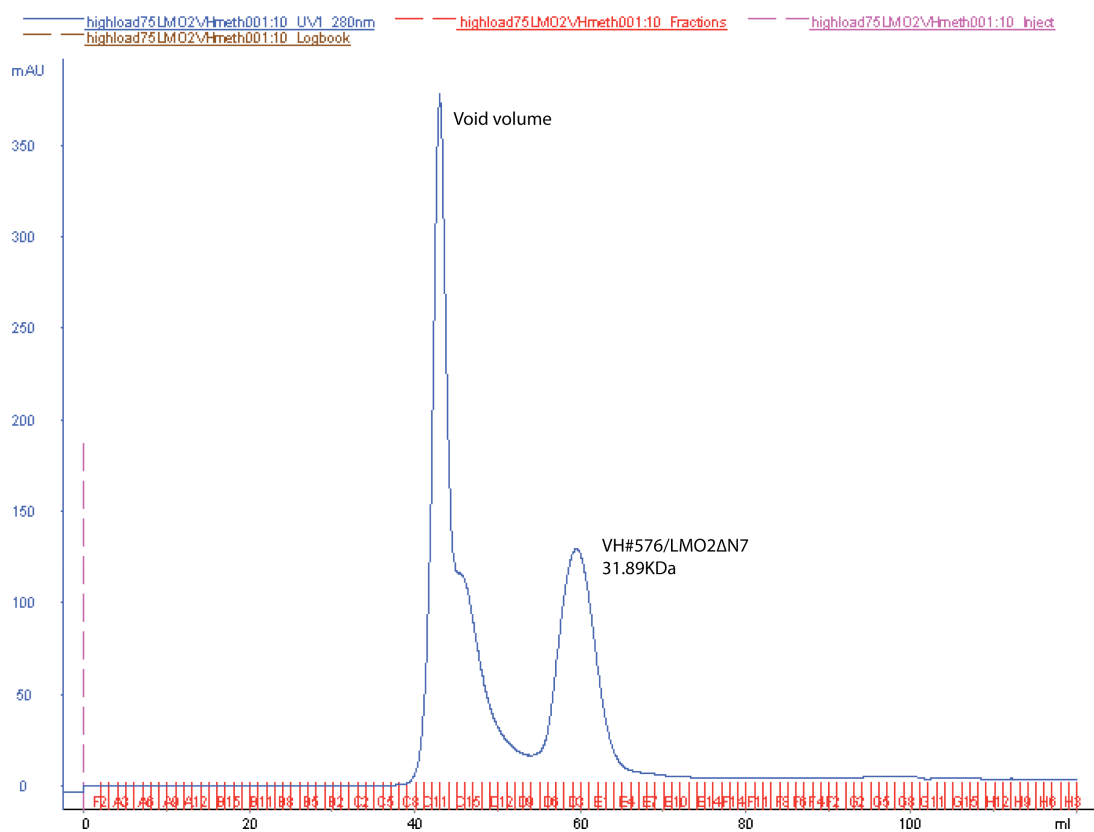


Figure 3.24: The purification of lysine methylated VH#576/LMO2ΔN7 by size exclusion chromatography

The chromatogram represents the UV absorbance at 280nm and shows a well resolved peak for the methylated VH#576/LMO2ΔN7 protein sample.

Crystallisation screens were set up using VH#576/LMO2 Δ N7 lysine methylated protein complex at a concentration of 8mg/ml (determined using pre crystallisation kit from Hampton Research, CA, USA). In this case lysine methylation did not aid crystallisation as, again, all screens were unsuccessful.

3.6.4 Removal of C-terminal residues from LMO2 and expression with antibody domain VH#576

The final approach was to remove any C-terminal residues from LMO2 not involved in the interaction with VH#576. The C-terminus of LMO2 may disrupt the formation of crystals if it is flexible and undergoes large-scale motions. Four versions of the expression plasmid (pRK-His-Tev-VH#576-LMO2 Δ N7) were engineered removing varying numbers of residues from the C-terminus of LMO2. The four constructs were expressed and purified by Ni-NTA chromatography (Figure 3.25). As well as C41 *E. coli* host strain, Shuffle T7 express lysY (New England Biolabs, MA, USA) was also tested for protein expression with a view to improving the conformation and thus stability of the protein expressed. Shuffle T7 express lysY is an engineered *E. coli* B strain designed to provide disulfide bond formation in the cytoplasm and constitutively express disulfide bond isomerase (DsbC), which assists in protein folding and correction of misoxidation.

The construct, with the largest number of C-terminal residues removed, which still retained its stability and solubility, was VH#576/LMO2 Δ N7 Δ C11. This complex was expressed from *E. coli* strain C41 on a large scale and purified using both Ni-NTA and size exclusion chromatography (Figure 3.26). VH#576/LMO2 Δ N7 Δ C11 protein was also purified via the lysine methylation strategy. These two samples (methylated and un-methylated) were analysed by positive electrospray mass spectroscopy (Appendix A, figure 6). This was both to ensure surface lysine methylation had been successful and also to ensure the un-methylated protein sample had the correct molecular weight and had not been truncated or cleaved in any way. Mass spectrometry analysis of methylated VH#576/LMO2 Δ N7 Δ C11 indicates all thirteen possible methylation sites have been successfully methylated. An additional species at a higher molecular mass of 58.014 KDa was also identified which may be a slight GroEL chaperone protein contamination. The VH#576 monomer was also detected under native conditions.

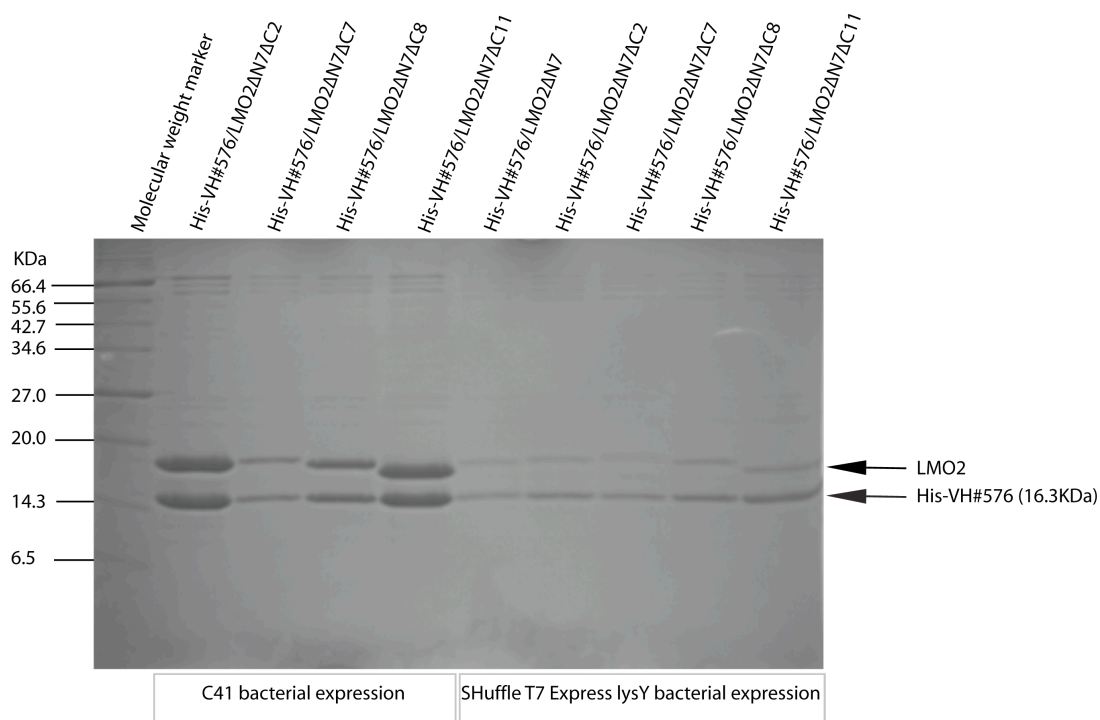
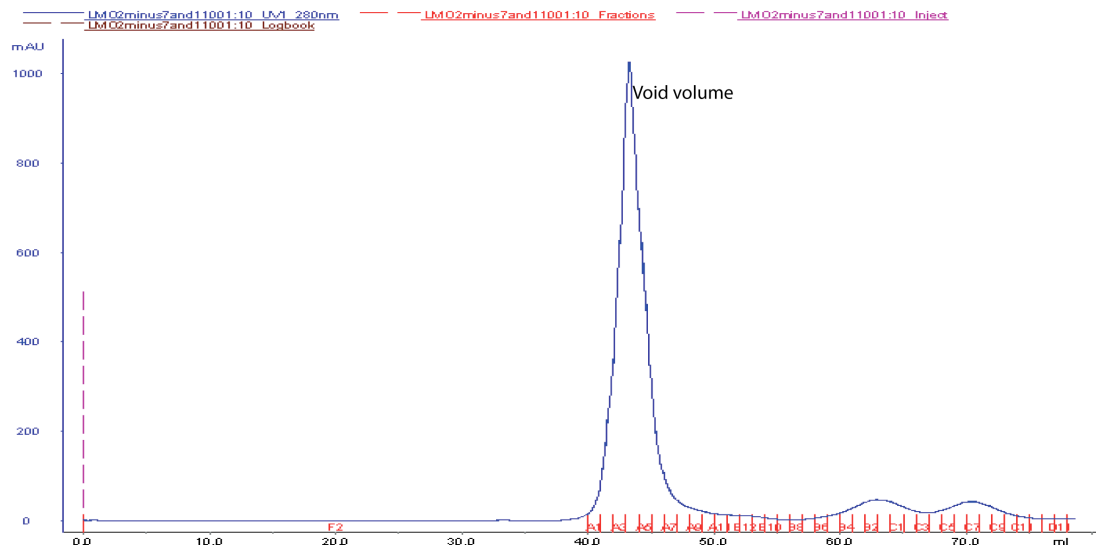


Figure 3.25: Purification of LMO2, with various C-terminal truncations, with VH#576
15% SDS PAGE stained with Coomassie brilliant blue. The gel shows that VH#576/LMO2ΔN7ΔC11 complex was expressed most efficiently. The gel also shows the protein yield is higher when expressed from C41 rather than Shuffle.

A:



B:

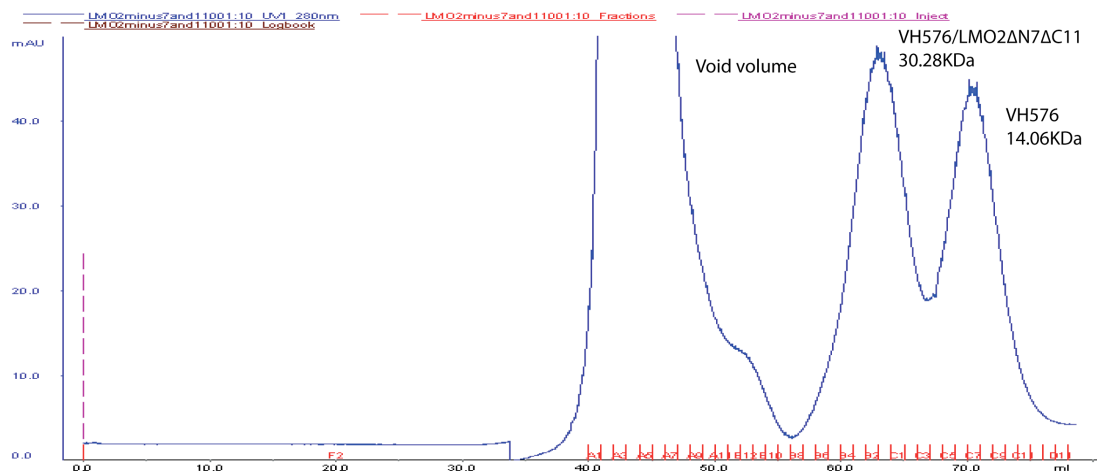


Figure 3.26: Size exclusion purification of VH#576/LMO2ΔN7ΔC11 expressed from C41

A: VH#576/LMO2ΔN7ΔC11 purified by size exclusion using Hi load 16/60 superdex S75 prep grade size exclusion column. The chromatogram is a measure of UV absorbance at 280nm and is shown in full.

B: UV_{280nm} chromatogram focusing on the peak corresponding to VH#576/LMO2ΔN7ΔC11.

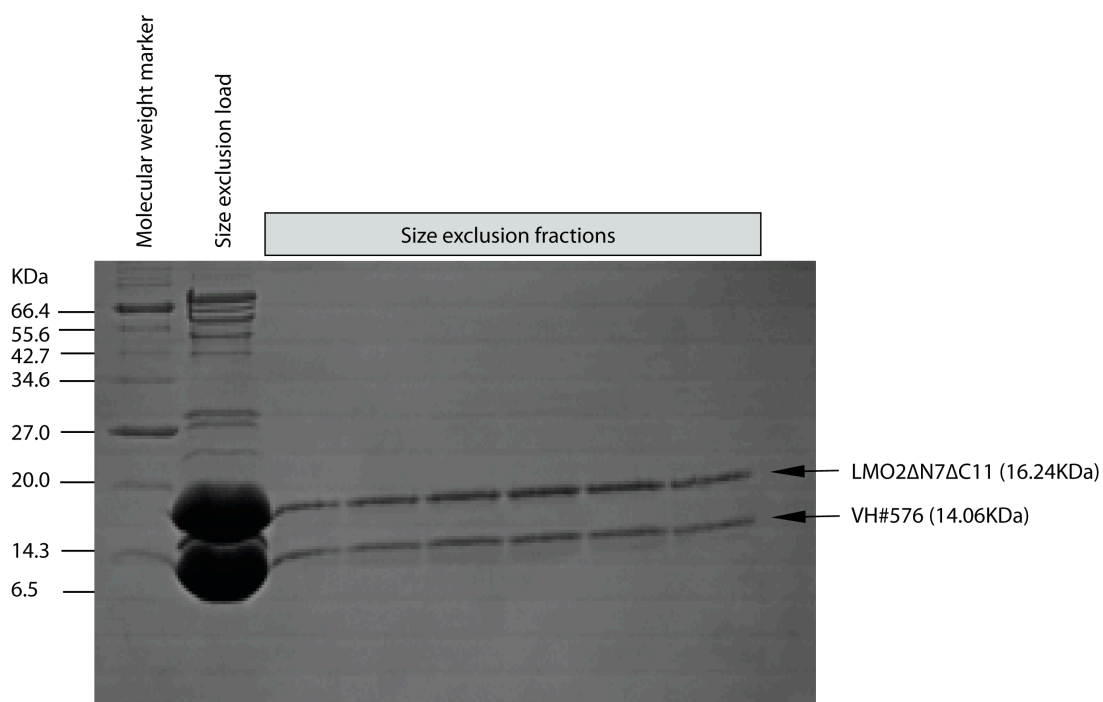


Figure 3.27: VH#576/LMO2ΔN7ΔC11 expressed from C41 and purified by size exclusion 15% SDS-PAGE stained with Coomassie brilliant blue. The gel shows the VH#576/LMO2ΔN7ΔC11 size exclusion load and the fractions from the relevant peak (Figure 3.26).

The VH#576/LMO2 Δ N7 Δ C11 complex was subjected to far-ultraviolet (UV) Circular Dichroism spectropolarimetry (see Appendix A, figure 8) in order to determine if the protein complex was folded and to estimate the proportions of different secondary structural elements present. The Circular Dichroism spectrum has a characteristic minimum at around 215nm; this is typical for a structure with considerable β -sheet content. From these data, we could confirm the protein complex was folded as the spectrum was negative at more than 208nm. The Circular Dichroism spectrum was analysed by CDSSTR (Johnson, 1999) obtained from DICHROWEB (Whitmore and Wallace, 2004). The content of various secondary structures obtained, for VH#576/LMO2 Δ N7 Δ C11, by this method were: 8% α helix, 38% β sheet, 24% β turn and 30% unordered. These values are a fair indication that both the VH antibody single domain and LMO2 were present, as a complex, in the purified protein sample.

3.6.5 VH#576/LMO2 Δ N7 Δ C11 X-ray crystallography screens

VH#576/LMO2 Δ N7 Δ C11 protein complex, expressed from *E. coli* strain C41, was taken forward for crystallography trials. For this trial a simple protein sample buffer of 20mM Tris-HCl, pH 8 was chosen with the aim of increasing the degree to which the reagent buffers, of the crystal screen, could manipulate the pH and overall environment of the protein sample. A sample of protein (final concentration 8mg/ml) was snap frozen by immersion in liquid nitrogen and sent on dry ice to Oxford protein production facility (OPPF). The OPFF have developed a high throughput crystallisation facility (Walter et al., 2003, Walter et al., 2005).

Initially six blocks (576 conditions) of crystallisation screening kits (Hampton Research, CA, USA) were used:

- I. Hampton screens I and II
- II. PEG/ Ion, Grid screens PEG 6000 and Ammonium sulphate
- III. Natrix and cryoscreens
- IV. Gris screens PEG/Li, NaCl, MPD and 'Quick screen' Phosphate
- V. Salt RX
- VI. Index

Crystallisation screens were performed (by Thomas S. Walter) using the Cartesian Technologies MicroSys MIS4000 pipetting instrument to execute nano-liter-scale sitting drop vapor diffusion crystallisation experiments. Drops dispensed consisted of 100nl protein and 100nl reagent. Crystallisation plates were stored at 21°C and imaged automatically thus maintaining a digital time course of crystal growth (Walter et al., 2005). Images of each well of the crystallisation screens were accessed remotely using a web browser allowing

rapid assessment of each drop. As shown by Figure 3.28 and Table 3.3, this crystallisation trial resulted in the identification of a number of reagents with the capacity to crystallise VH#576/LMO2ΔN7ΔC11 using the sitting drop vapour diffusion method.



Figure 3.28: Microscope image of crystal from reagent N5 of the Matrix screen

Crystal growth was observed for reagent N5 of the Matrix crystallisation screen (0.2 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate, 0.05 M MES monohydrate pH 5.6, 5% w/v Polyethylene glycol 8,000). The crystal was weakly birefringent (typical of biological crystal).

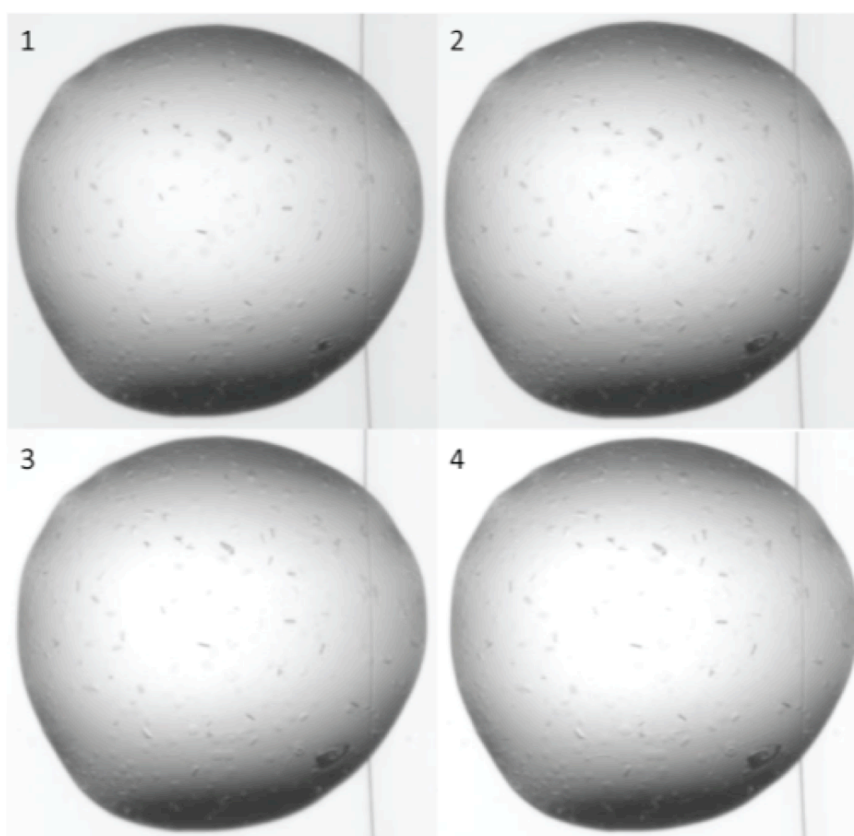

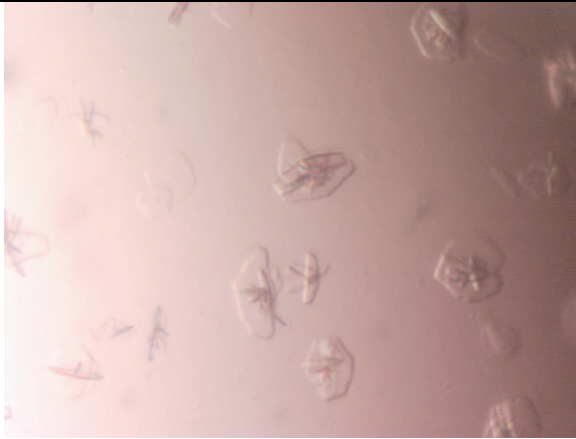
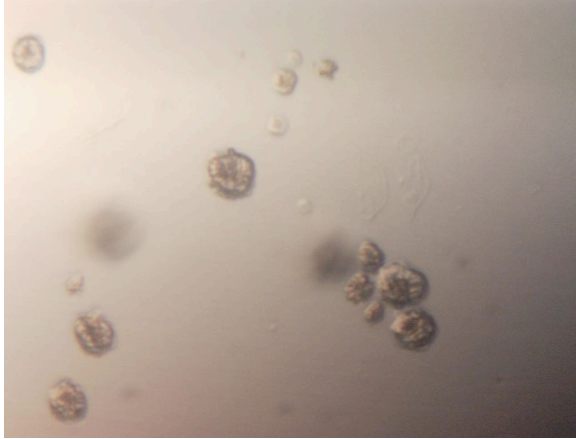


Figure 3.29: Time course of crystal growth

A time course of crystal growth observed for reagent N5 of the Matrix crystallisation screen. The crystal of interest (bottom right of drop) can be seen at 1 day 11 hours (image 1) and reaches full size by 4 days 19 hours (image 4).

Table 3.3: Further crystallisation reagents identified

Microscope image	Condition
	Natrix N3 100mM Magnesium acetate tetrahydrate, 50mM MES monohydrate, pH 5.6, 20% v/v 2-methyl-2,4-pentenediol
	Natrix N38 5% PEG 4000 200mM Ammonium acetate 150mM Magnesium acetate, 50mM Sodium HEPES, pH 7
	Ammonium sulphate A3 100mM MES monohydrate pH 6.0, 0.8 M Ammonium sulfate

As the crystals could potentially be formed of VH#576 alone, LMO2 Δ N7 Δ C11 alone or from the complex (VH#576/LMO2 Δ N7 Δ C11), four crystals grown from the initial screen were tested for protein content by SDS-PAGE and silver stain analysis (Figure 3.30). Crystals from the Natrix screen, condition N3, were shown to be formed from the complex.

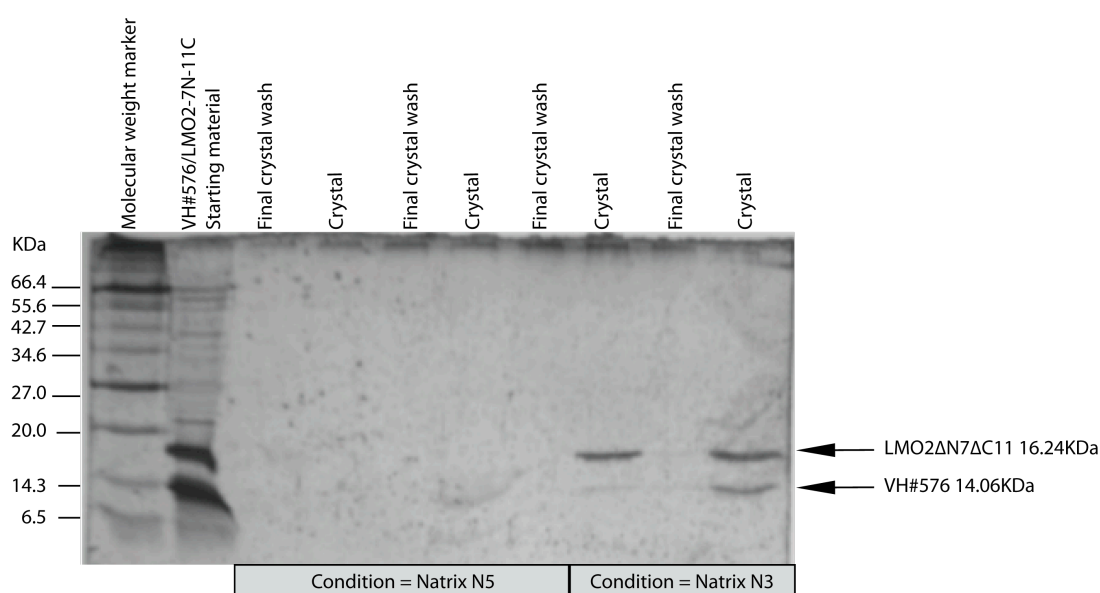


Figure 3.30: Silver stain analysis of VH#576/LMO2 Δ N7 Δ C11 crystals

Analysis of two crystals from the Natrix N5 reagent failed. Analysis of two crystals from the Natrix N3 reagent showed the VH#576/LMO2 Δ N7 Δ C11 complex to be present in both cases. No protein was present in the final crystal washes indicating that the crystals had been washed effectively with no protein residue present from the drop.

3.6.6 Crystallisation optimisation

The following optimisation work was carried out by H. Sewell with guidance from T. S. Walter, Dr. E. J. Mancini and Dr. K. El Omari. Six conditions were taken forward for crystallisation optimisation experiments. Standard sitting drop vapour diffusion experiments were again utilised varying pH, reservoir dilution and protein concentration as shown by Figure 3.31. Additive screens (Hampton Research, CA, USA) were also set up. For a full list of the optimisation screens carried out and the results see Appendix A, table 1.

A:

	100%	97%	94%	91%	88%	85%	82%	79%	76%	73%	70%	67%	Protein/reservoir volumes
F													100nl+100nl
G													200nl+100nl
H													100nl+200nl

B:

	100%	85%	70%	55%	100%	85%	70%	55%	100%	85%	70%	55%	Protein/reservoir volumes
F													100nl+100nl
G													200nl+100nl
H													100nl+200nl
pH	X -0.5			X				X +0.5					

Figure 3.31: Optimisation experiments

A: Three row dilution optimisation; the top row shows how the reservoir concentration was decreased from 100% to 67%. The final column shows how the protein to reservoir volume ratios were changed for each row, thus varying protein concentration. The bottom three rows (F, G and H) of a Greiner 96 well plate were used.

B: Three row pH optimisation; the pH was shifted from that of the original condition by +0.5 and -0.5. Within these pH changes, reservoir concentration was also varied from 100% to 55% (as indicated by the top row). Protein to reservoir drop volumes were also different for each row as shown by the final column.

These optimisation trials resulted in a number of crystals that were subsequently tested for diffraction. Diffraction data was collected by Dr. E. J. Mancini and Dr. K. El Omari at The European Synchrotron Research Facility (ESRF) in Grenoble and provided information on how well crystals from each condition were able to diffract. Crystals from the Ammonium sulphate A3 additive screen diffracted to the highest resolution (3.5Å to 3Å) and as such, this condition was optimised further by analysing the effect of different additives. This work was carried out by H. Sewell. Six optimisation plates were set up using the three row dilution technique (100 to 78%, 2% steps), one for each of the additives listed below. 100nl of the following additives were added to all wells of each optimisation plate:

- Jeffamine M-600
- 2, 5 Hexanediol
- 1, 3 Butanediol
- 1, 6 Hexanediol
- Pentaerythritol ethoxylate
- Glycerol

Crystals were visible after two days and continued to grow up until day four. Of these optimisation screens the additives 1, 3 Butanediol, 1, 6 Hexanediol, Jeffamine M-600 and Glycerol gave hexagonal plate shaped crystals with sharp edges (Figure 3.32). Optimisation plates were also set up by hand (by H. Sewell) using a similar method: 24 well Cryschem plates (Hampton Research, CA, USA) were used with reagent dilution varied horizontally (100 to 75%, 5% steps) and drop ratio vertically (1µl:1µl, 1µl:2µl, 2µl:1µl, 2µl:2µl). Three plates were set up, one for each additive; 1, 3 Butanediol, 1, 6 Hexanediol, and Glycerol. All three plates gave crystals.



Figure 3.32: Crystal from three row optimisation dilution screen

Reagent: Ammonium sulphate (A3): 100mM MES monohydrate pH 6.0, 0.8 M Ammonium sulfate

Additive: 1,3 Butanediol

Well G6: 200nl protein/100nl reagent/100nl additive, reagent at 90%

The crystal measured 100 μ m by 40 μ m.

As will be discussed in chapter five, the structure of VH#576/LMO2 Δ N7 Δ C11 was solved to 3.3Å using diffraction data collected at Diamond light source. The crystal diffracted was grown in 0.8M Ammonium sulphate, 0.1M MES monohydrate pH 6 (diluted with H₂O to 88%) and additive 1, 6 Hexanediol with a drop ratio of 200nl of protein to 100nl of reagent.

3.7 Discussion

This chapter describes the successful crystallisation of LMO2. The overall pipeline, from gene to structure included cloning, expression, purification, crystallisation screening and finally optimisation of conditions to produce structure yielding crystals. The use of VH#576 effectively transformed LMO2 from an aggregated protein to a soluble, monodispersed protein complex suitable for crystallisation.

The use of VH#576 to enhance the stability and hence solubility of LMO2 is analogous to the strategy utilised by Deane *et al.* The group stabilised the LIM1 domain of LMO2 through interaction with the LIM interacting domain (LID) of LDB1, a member of the LMO2 DNA binding complex (Deane et al., 2001). The authors found that upon incubation with LDB1(LID), micro-gram quantities of LMO2 could be recovered. However, the complex degraded over time and therefore the group engineered a flexible linker between the C-terminus of the LIM1 domain of LMO2, and LDB1(LID) (Deane et al., 2001). LDB1(LID) was shown to stabilise LIM1 of LMO2 and enable its purification to sufficient quantities for structural studies by NMR (Matthews et al., 2001).

In this case, it was necessary to make LMO2 more amenable to crystallisation by removal of amino and carboxy terminal residues, in order to reduce flexible regions and improve the crystal packing. Terminal residues of LMO2, not involved in the interaction with VH#576, were systematically deleted, thus eliminating any flexible regions that might interfere with crystallisation. A final construct of LMO2 spanning residues 9 to 147 (LMO2 Δ N7 Δ C11) bound to antibody single domain, VH#576, was crystallised under a number of conditions.

The crystallisation screen employed the sitting drop vapour diffusion method and nano liter scale drops (Walter et al., 2003) to provide favourable kinetics for the crystallisation reaction. Many factors influence crystallisation including temperature, rate of equilibration, precipitant type and concentration, pH, protein concentration, detergents and impurities. These are all factors that were optimised when trying to produce crystals suitable for X-ray diffraction and improve the resolution of diffraction. Ammonium sulphate, A3 (Hampton Research, CA, USA); 0.8 M Ammonium sulfate, 100mM MES monohydrate pH 6.0, with

additive 1, 6 Hexanediol gave hexagonal plate shaped structure yielding crystals. The addition of additive 1, 6 Hexanediol to this condition dramatically improved the quality of crystals grown and this may be due to manipulation of protein-protein or protein-solvent interactions or perturbation of water structure.

Co-expression of a target protein with an antibody fragment may increase its stability upon purification (Warke and Momany, 2007) and there have been instances of successful applications of this approach previously, an example of which is HIV capsid protein p24. When the protein was purified alone, interactions between the oligomerisation domains in the viral coat caused the viral capsid protein to aggregate (Prongay et al., 1990). Upon purification with a Fab fragment, the complex could be crystallised. Interestingly, the crystal contact points were primarily between Fab fragments. The advantage of using a VH single domain over a Fab fragment is the absence of the flexible hinge region that could potentially inhibit crystallisation.

Many research groups have identified the potential of generating antibody domains for use as co-crystallisation chaperones (Nettlehip et al., 2008, Tereshko et al., 2008), to enable the crystallisation of a protein or improve the resolution of its crystal structure. Crystallisation is often hindered by the inherent conformational heterogeneity of proteins and this limits the number of productive lattice contacts that are available for crystallisation. Crystallisation chaperones, such as VH or Fab antibody domains, bind to a specific conformation and thus minimise conformational heterogeneity and increase the chances of gaining well-ordered crystals. As such, the antibody single domain strategy described in this chapter may well be applicable as a generic tool to facilitate protein purification, crystal growth and improve crystal quality. Furthermore, VH#576 is a member of the VH₃ family of heavy chain antibody domains, which are particularly suitable for this application as the framework is stable and well expressed. As will be shown in chapter five, the antibody single domain can also provide valuable phasing information.

In conclusion, we have exposed the relative instability of free LMO2 and demonstrated its purification with an anti-LMO2 VH antibody single domain. Upon removal of the flexible N and C-terminal residues of LMO2, the complex was crystallised. A high throughput mode of crystallisation screening was utilised to identify a number of crystallisation conditions, a subset of which were optimised. Diffraction data has been collected and the structure solved to a resolution of 3.3Å, as will be described in chapter five.

4 Structural analysis of an anti-LMO2 antibody single domain by NMR

4.1 Introduction

LMO2 is an attractive drug target for the treatment of a subset of patients with T-cell acute lymphoblastic leukaemia (T-ALL). Despite this there are currently no anti-LMO2 therapies. Antibodies have the qualities of specificity and affinity required for such a therapy. However, LMO2, as with most oncogenic proteins, is located inside the cell and not available for antibody mediated targeting. Consequent developments in antibody engineering have enabled stable antibody domains to be expressed in the reducing environment of the cell with no requirement for disulphide bond formation for correct folding (Tanaka and Rabbitts, 2008). It has been shown that single variable region domains are highly efficient as intrabodies. Accordingly, intracellular antibody capture (Tanaka and Rabbitts, 2010), based on *in vitro* yeast two-hybrid screening has been used to derive a single domain VH intrabody which binds specifically to LMO2 with high affinity (equilibrium dissociation constant of 94nM). This antibody domain, termed VH#576, has the potential to inhibit LMO2 from interacting with its natural protein partners and thus may prevent its aberrant activity as a transcription factor within T-cells of patients with T-ALL. However, due to significant problems with the delivery of antibody domains it is necessary to evolve VH#576 into a small drug like molecule.

Structural data for small molecule drug design can be either explicit, for example the availability of a single X-ray structure of the target protein, or implicit, for example enough structural data on the ligand, or in this case the antibody domain, to develop a three dimensional pharmacophore map. Such a pharmacophore is then utilised in the usual way, for example, to search a three dimensional database *in silico* in order to identify mimetic compounds for focused screening. There are instances in which the rational structure based drug design of inhibitors can be performed without the target structure (McGregor, 2007).

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for determining protein structure. Various NMR experiments are implemented in order to acquire enough information to assign peaks of the spectral profile to individual atoms and eventually assemble a structure. However, as protein size increases the complexity of the task becomes greater as larger proteins tumble more slowly and their relaxation process becomes more efficient therefore the signal decays rapidly. Thus, VH#576/LMO2 was considered too large for structural analysis by NMR. VH#576, however, is far more suitable for detailed NMR investigations. Furthermore, the structure can be utilised to aid anti-LMO2 small molecule drug design. NMR is also an ideal tool for mapping the precise interaction sites of antibody domains (Wilkinson et al., 2009). Therefore, the aim of the work described in this chapter

was to gain structural data on the anti-LMO2 VH domain, VH#576, both in the uncomplexed and complexed state. This chapter focuses on the use of NMR to collect structural data; though the work was carried out in parallel with crystallisation and X-ray crystallography studies of the VH#576/LMO2 complex.

VH#576 was identified from a library that belongs to the VH₃ subgroup (Tanaka and Rabbitts, 2003). Previous studies on purified human VH fragments, from the same human VH₃ subgroup, have shown poor expression and aggregation probably due to the exposed VH-VL interface (Davies and Riechmann, 1994, Dudgeon et al., 2009). However this chapter shows VH#576 can be expressed and purified from *E. coli* to a level sufficient for NMR. This is most likely due to the long CDR three region (17 residues as opposed to the more usual 12 residues) folding back onto the former, hydrophobic, VH-VL interface (Riechmann and Muyldermans, 1999).

Previously a VH human antibody domain (VH-P8), from the human VH₃ gene subgroup, has been studied by solution NMR. This was enabled by three mutations to the VH/VL interface region (based on the Camelid VH domains) and the addition of CHAPS to the sample buffer. These modifications enabled 90% of the ¹H and ¹⁵N main chain signals to be assigned. Studies were carried out to analyse the interaction between VH-P8 and Protein A. In this case the binding site of Protein A lies in the VH not in the Fc portion, which is the more commonly known Ig isotype binding site. Upon binding, the chemical shift changes were identified for many residues (Riechmann and Davies, 1995). The ability to bind Protein A is a unique trait of VH domains originating from the human VH₃ subgroup of genes.

Isolated human VH domain HEL4 (based on human germline VH₃ framework) is highly soluble at concentration of more than 3mM and its structure has been solved by X-ray crystallography to 2Å resolution. The structure shows the standard fold of an immunoglobulin variable domain with nine anti-parallel β strands. The structure also reveals, interestingly, the Trp 47 is flipped into a cavity formed by Gly 35 of CDR1 increasing the hydrophilicity of the VH-VL interface (Jespers et al., 2004).

This chapter describes the NMR structural studies of VH#576 and the strategies employed in order to stabilise the antibody single domain in solution. A protocol was developed to purify a complex of ¹⁵N, ¹³C labeled VH#576 bound to unlabeled LMO2. Amide chemical shift changes after complex formation have been observed however they cannot be mapped onto the sequence due to limitations encountered during assignment. Further data collection

is required for assignment of VH#576 to be possible.

4.2 Protein expression and purification methods

The general methods employed for recombinant protein expression and purification are outlined in chapter two, section 2.5 and 2.6. Any modifications are stated in the appropriate section of this results chapter. NMR data acquisition was carried out as described in chapter two, section 2.9.

4.3 VH#576 protein purification and NMR results

4.3.1 Initial VH#576 purification

For the initial NMR experiments His-VH#576 was expressed from *E. coli* C41 cells grown up in 500ml ¹⁵N labeled rich growth media (E.coli-OD2 N Silantes, Germany) and purified by Ni-NTA affinity chromatography. His-VH#576 (16.3 KDa) was then dialysed into 20mM Tris-HCl, pH 8, 150mM NaCl followed by concentration of the protein sample to 320µl at 0.12mM. The results of this purification are shown in Figure 4.1.

HSQC experiments are often used to screen proteins, for structure determination potential by NMR, and as such this semi-purified VH#576 protein sample, was used immediately for this purpose. The spectrum of VH#576 (Figure 4.2) looked of sufficient quality to support a structure determination by NMR, with good chemical shift dispersion of a number of signals. The protein remained stable for approximately 48 hours and therefore, the purpose of subsequent experiments was to optimise the stability of the protein sample to increase the time period available for NMR data acquisition.

Natural antigen binding antibodies, lacking light chains, have been found in Camels. As such a series of papers, relating to antibody fragment stability, suggest mutating residues in the framework region of the VH domain, near CDR2 (former VH/VL interface) to the corresponding residue present in Camelid VH domains (G44E, L45R, W47I) (Barthelemy et al., 2008, Davies and Riechmann, 1994). However, when such mutations were introduced in the VH framework of the anti-RAS VH, antigen binding activity was destroyed *in vivo* as revealed by the results of a luciferase reporter assay (Tanaka et al., 2003). VH#576 and VH#6 (anti-RAS VH) are based on the same intracellular antibody capture consensus scaffold and this suggests that modifications may not be applicable to improving the stability of VH#576.

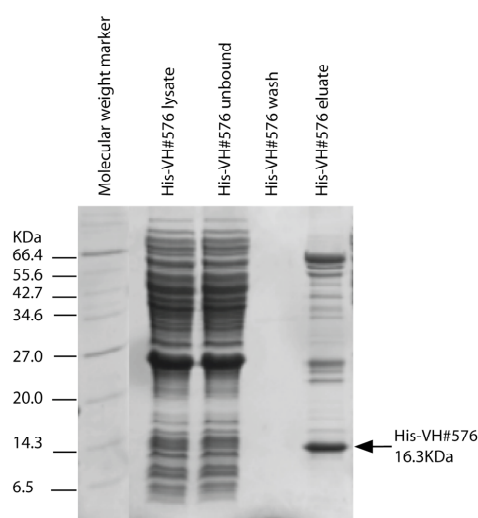


Figure 4.1: Purification of VH#576 via N-terminal His tag

15% SDS-PAGE stained with Coomassie brilliant blue. Samples were taken at each stage of the purification and analysed as shown above. The purification yield was 1.4mg per liter culture. The eluate was concentrated to 320 μ l at 0.12mM.

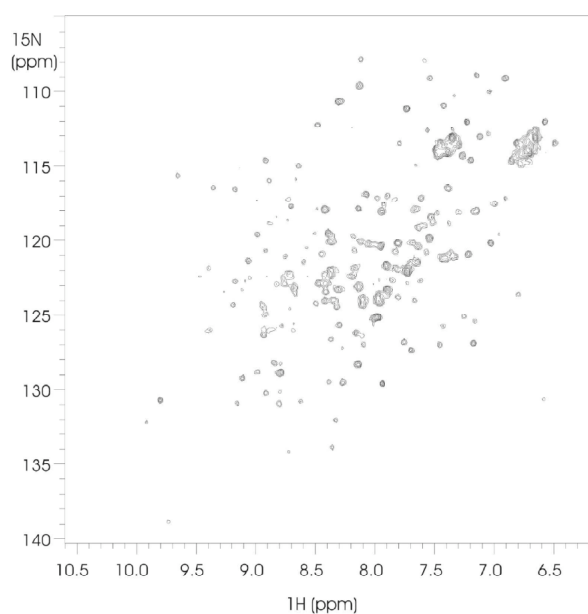


Figure 4.2: ^1H - ^{15}N HSQC spectrum for antibody single domain VH#576

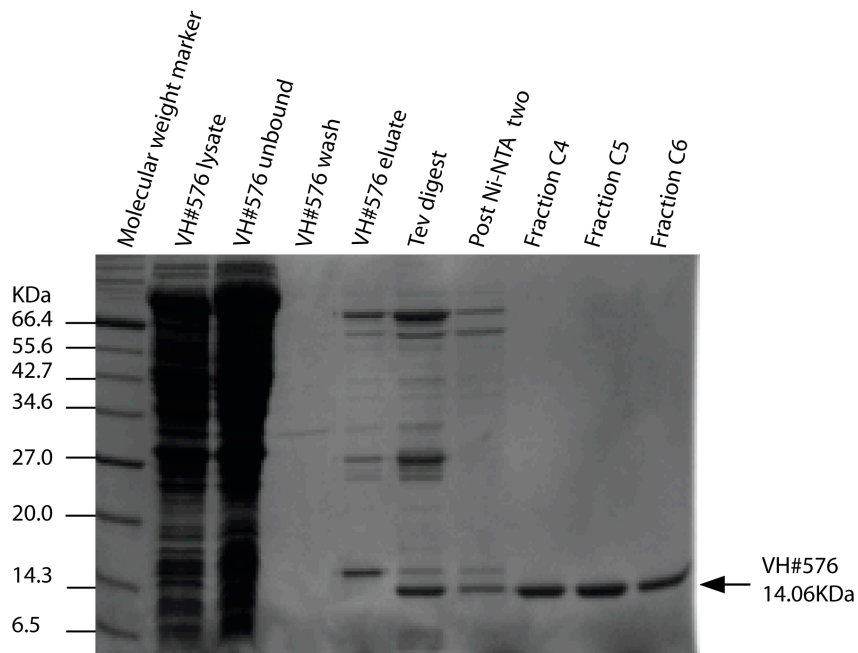
^1H - ^{15}N HSQC spectrum for VH#576 protein suspended in 20mM Tris-HCl, pH 8, 150mM NaCl. A HSQC spectrum provides a map of the protein backbone with approximately one cross peak for each amino acid (each H-N amide pair), except for proline. Peaks are well dispersed representing a well folded protein. The number of peaks should roughly equal the number of residues in the protein, in this case 147 residues, of which approximately 120 are seen on the spectrum. Peaks for the remaining residues may not have been detected due to the high pH resulting in amide exchange with water molecules.

4.3.2 VH#576 size exclusion chromatography

The initial ^1H - ^{15}N HSQC spectrum recorded for His-VH#576 was encouraging and in light of this, modifications were made to the expression conditions and purification protocol. This was to improve both the solubility and stability of VH#576.

His tags have been found to cause aggregation (personal communication from T. Tanaka) and therefore, to improve the stability of VH#576 in solution the His tag was removed via a Tev protease digest. In view of the large number of contaminant proteins, the Tev digest was followed by a size exclusion chromatography step. The results of this purification are shown in Figure 4.3. VH#576 fractions from the appropriate size exclusion peak were analysed by SDS-PAGE and were estimated to be around 90% pure. The final yield of 0.7mg per liter prep was adequate for the purpose of NMR. The VH#576 purification strategy was optimal at this point nevertheless, to further improve the long-term stability of the protein, a solubility trial was completed as will be described in the next section.

A:



B:

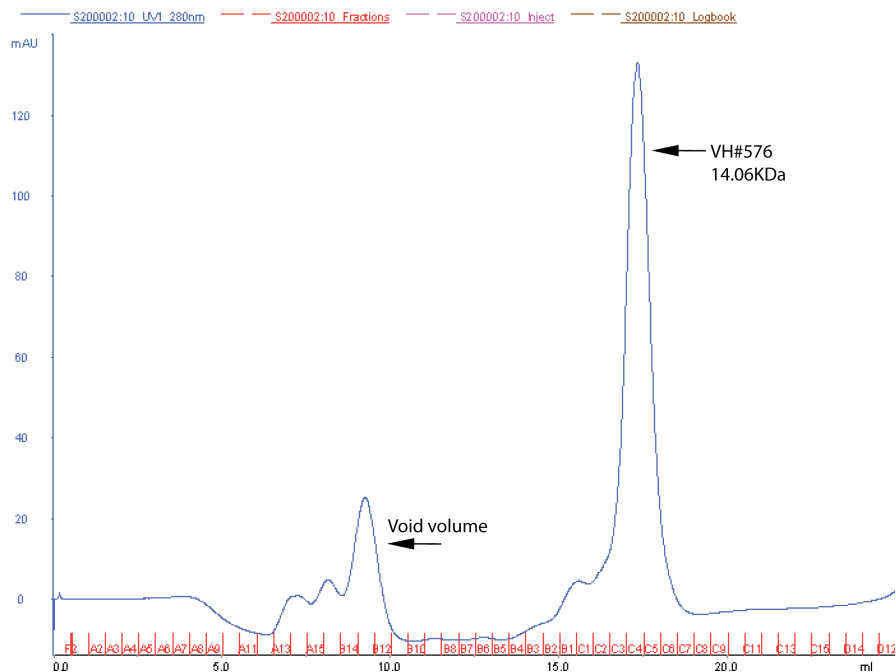


Figure 4.3: Purification of VH#576 via Ni-NTA affinity and size exclusion chromatography

VH#576 protein was expressed in 2L M9 minimal media with N^{15} labeled ammonium chloride. Samples were taken at each stage of the purification and analysed as shown above.

A: 15% SDS-PAGE stained with Coomassie brilliant blue.

B: VH#576 size exclusion UV_{280nm} chromatogram.

Purification yield was 1.4mg from 2L. The appropriate size exclusion fractions were pooled and concentrated to 320 μ l at 0.3mM for NMR data acquisition.

4.3.3 VH#576 stability trials

As mentioned previously, VH#576 precipitated over the course of two days in the NMR tube at ambient temperature. In order to identify a sample buffer which improves the long-term stability of the protein sample, a microdrop screening experiment was set up to examine different concentrations, pH, buffer type and stabilisers (Lepre and Moore, 1998). The most stable condition was identified as 100mM NaCl, 100mM HEPES, pH 7. CHAPS was also shown to improve the long term stability of VH#576. To establish the effectiveness of 100mM NaCl, 100mM HEPES, pH 7 as a sample buffer a HSQC spectrum was acquired for ¹⁵N labeled VH#576. Once again the protein precipitated out of solution over a period of 48 hours, despite the positive result during the screen. The screen did however highlight the value of CHAPS in order to stabilise VH#576 and therefore, it was decided to revert back to the original buffer of 20mM Tris-HCl pH 8, 150mM NaCl, 1mM DTT, but with the addition of 2mM CHAPS; this dramatically improved the stability of the protein during NMR data acquisition thus increasing the time frame for data collection.

4.3.4 Effect of pH on VH#576 NMR experiments

Over the course of the experiments carried out to increase stability of the VH#576 protein the optimal purification strategy identified was: VH#576 purified by Ni-NTA affinity chromatography followed by removal of the His tag and further purification by size exclusion chromatography. The most appropriate buffer identified was 20mM Tris-HCl, pH 8, 150mM NaCl, 1mM DTT with the addition of detergent, 2mM CHAPS. The maximum protein concentration obtained was 0.3mM. NMR experiments were carried out at a temperature of 10°C again to maintain the long-term stability of the protein in solution; this gave a window of approximately one week for data acquisition. An example of a HSQC run under these conditions is shown in Figure 4.4.

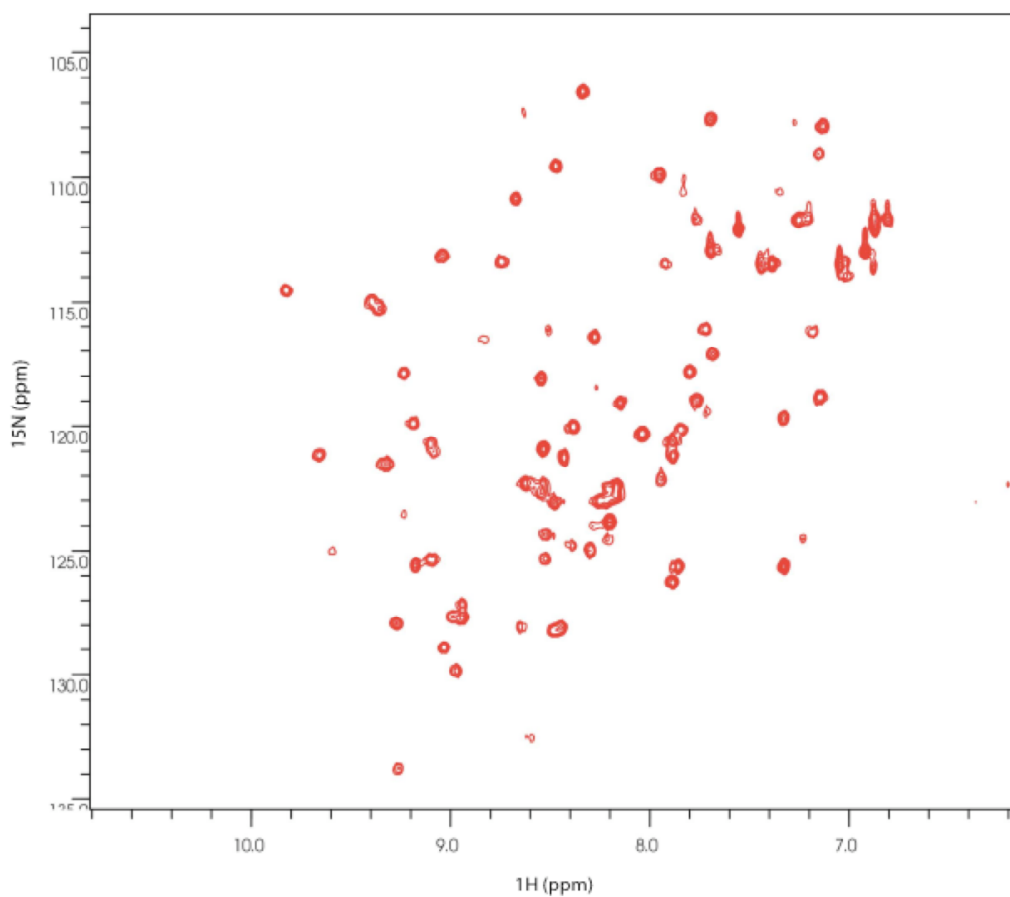


Figure 4.4: ^1H - ^{15}N HSQC spectrum for antibody single domain VH#576

^1H - ^{15}N HSQC spectrum recorded overnight. VH#576 protein was suspended in 20mM Tris-HCl pH 8, 150mM NaCl, 1mM DTT, 2mM CHAPS. The HSQC spectrum was recorded at 10°C. 123 residues should be present of which approximately 80 can be seen and this is potentially due to the high pH of the sample buffer.

NMR analysis was carried out at 10°C as the lower temperature also has a strong, positive influence on protein stability. The HSQC recorded had 80 of the 123 residues expected presumably due to the high pH of the protein sample buffer. The protein was most stable at pH 8 however at such a high pH, labile amide protons exposed to the solvent are likely to exchange rapidly with the solvent protons reducing the sensitivity of the experiment (Vu *et al.*, 2009).

An experiment was carried out to ascertain if the pH of the VH#576 sample buffer could be reduced while still maintaining the protein stability. Two samples were prepared, one at pH 7 and one at pH 8 and HSQC spectra were recorded for both samples as displayed in Figure 4.5. Approximately fifteen more peaks were observed at pH 7 due to the decreased amide exchange with the solvent protons. For the purpose of residue assignment, NOESY-HSQC and TOCSY-HSQC three dimensional spectra were also collected using the same sample at pH 7, which remained stable.

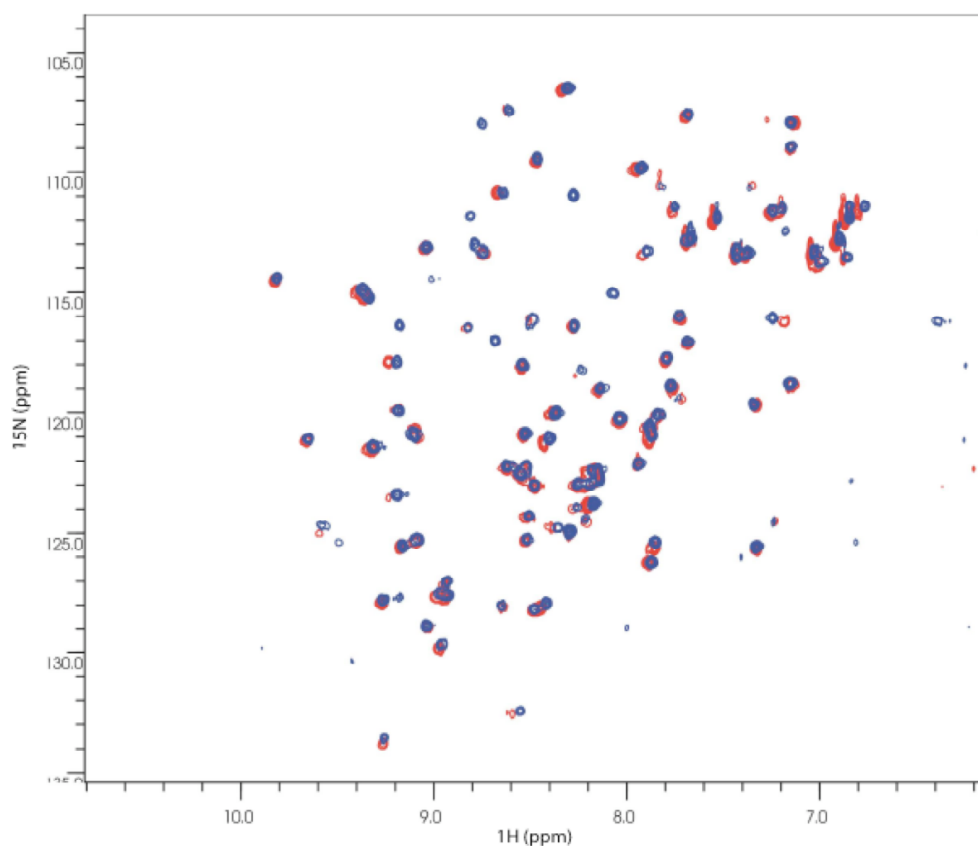


Figure 4.5: ^1H - ^{15}N HSQC spectra comparison for VH#576 at pH 7 and pH 8

The blue spectrum represents the protein at pH 7 and the red spectrum at pH 8. The exchange rates of backbone amide protons with water are minimal at pH 3 to 4 and increase linearly with the concentration of the hydroxy ions. Signal loss upon pre-saturation becomes significant at a pH above 5 to 6 and increases rapidly with the pH. However, usually antibody domains are only stable at above neutral pH therefore, a pH was chosen to be the minimum value at which the antibody sample remains stable during the length of the NMR procedure. 123 residues should be present of which approximately 95 can be seen at pH 7 and approximately 80 at pH 8.

In order to produce a $^{15}\text{N}/^{13}\text{C}$ -labeled VH#576 sample for further NMR experiments, C41 *E. coli* cells were grown up in double labeled ($^{15}\text{N}/^{13}\text{C}$) minimal media and purified as previously described. The pH of the final sample buffer was pH 7. To check the quality of the NMR sample prepared, a ^1H - ^{13}C HSQC spectrum was recorded overnight (Figure 4.6). For ^1H - ^{13}C HSQC spectrum, each peak represents a proton linked to a carbon. Unlike ^1H - ^{15}N HSQC, ^1H - ^{13}C HSQC are not affected by amide exchange, therefore most peaks should be seen for each residue.

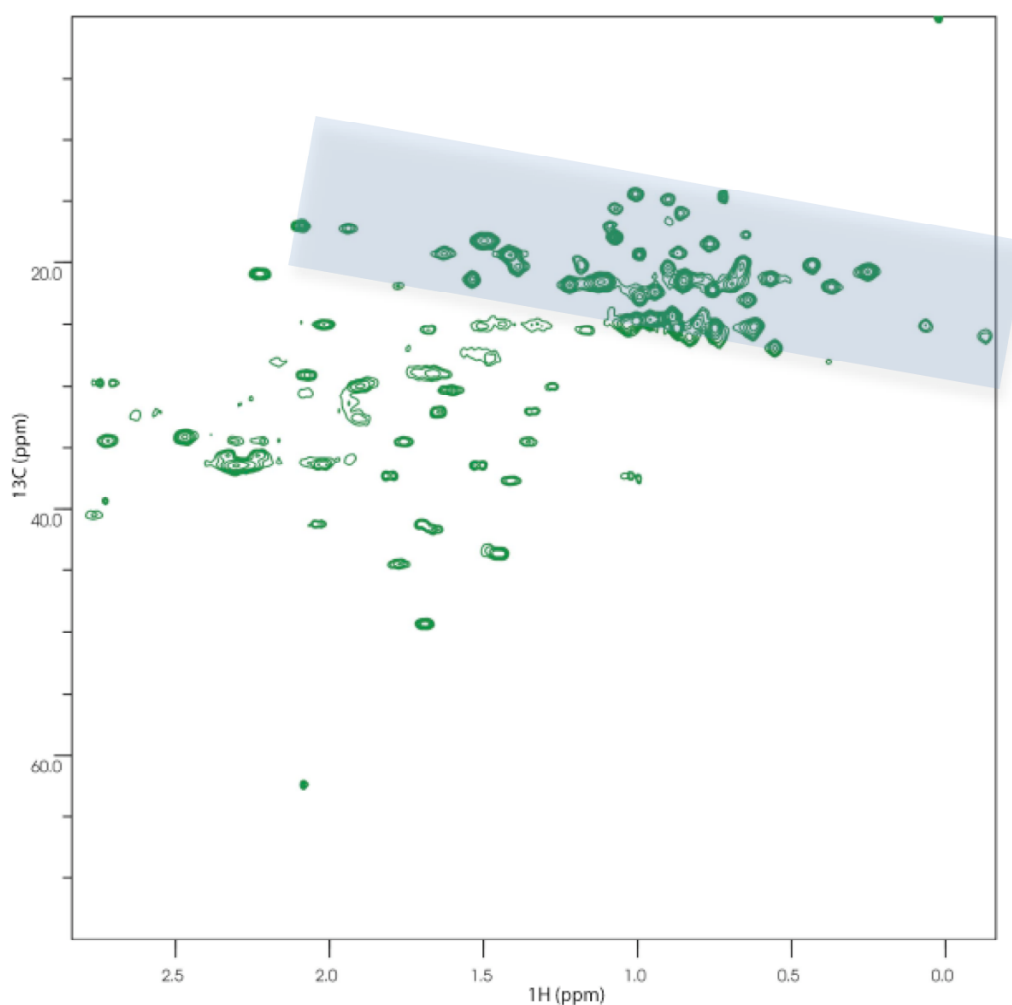


Figure 4.6: ^1H - ^{13}C HSQC spectrum for antibody single domain VH#576

^1H - ^{13}C HSQC spectrum recorded overnight at 10°C . This spectrum shows the proton to carbon bonds therefore each peak represents part of an amino acid side chain. VH#576 protein was suspended in 20mM Tris-HCl pH 7, 150mM NaCl, 1mM DTT, 2mM CHAPS. The section of the spectrum shaded in grey shows the methyl fingerprint region of methionine, alanine, threonine, valine, isoleucine and leucine (left to right). All 41 amino acids (with a methyl group in their side chain) out of the 41 present in the VH#576 sequence, are represented by peaks in the section. This demonstrates the well folded nature of VH#576, despite the reduced number of cross peaks that are visible in the ^1H - ^{15}N HSQC.

70% of the expected number of peaks, for VH#576, were observed in the ^1H - ^{13}C HSQC spectrum recorded (Figure 4.6). This result was very encouraging particularly given that the protein stability has also been significantly improved. Identification of optimal protein sample conditions enabled collection of three dimensional spectra HNCA, HN(CO)CA, NOESY-HSQC and TOCSY-HSQC. However the signal was too weak for recording less sensitive spectra such as HN(CA)CB, HN(COCA)CB and HN(CA)CO. Furthermore, the TOCSY-HSQC only showed cross peaks to the $\text{H}\alpha$ but not further into the side chain. As the amino acid type identification for the cross peaks is very reliant on $\text{C}\beta$ chemical shift information, a sequential assignment was impossible in the absence of spectra correlating the $\text{C}\beta$ shift with the amide chemical shift.

The low signal strength observed can be attributed to the low concentration of VH#576 (0.2mM to 0.3mM) purified and unfavourable NMR conditions (low temperature of 10°C and addition of CHAPS detergent). In response to this problem, a strategy was developed to purify VH#576/LMO2 complex, with only VH#576 isotopically labeled, on the basis that this complex was extremely stable when stored for crystallography studies. The aim was to increase the concentration and long-term stability of the purified protein sample and thus increase NMR signal strength.

4.4 VH#576/LMO2 protein purification and NMR data acquisition

It was reasoned that purification of VH#576 as a complex with LMO2 may well stabilise VH#576 and allow subsequent concentration of the complex for detailed NMR analysis with the purpose of structural determination. The average peak number detected for VH#576 was 86 out of a possible 123. This leaves 37 residues missing which roughly equates to the number of residues that make up the CDR regions (33). A well documented (Vranken et al., 2002) problem when analysing antibody domains by NMR, is the flexibility of the hypervariable loops which sometimes means the residues can not be detected. This flexibility results in a conformational exchange process on a microsecond-millisecond time scale and fast hydrogen exchange. VH#576 bound to LMO2 has a specific conformation and thus reduced flexibility. Subsequent NMR analysis was likely to result in an increase in both sensitivity and the number of residues observed. It would be difficult to assign the complex fully due to its large size (32 KDa) therefore, once it was demonstrated that the complex was stable during NMR data acquisition, a protocol was developed to purify isotopically labeled VH#576 bound to unlabeled LMO2. Furthermore, any chemical shift changes upon binding could also be detected by overlaying HSQC spectra of complexed VH#576 with uncomplexed VH#576.

4.4.1 ¹⁵N labeled VH#576/LMO2ΔN7

Prior to the development of a strategy to purify labeled VH#576 bound to unlabeled LMO2, the stability of the complex was tested using fully labeled (¹⁵N) VH#576/LMO2ΔN7 protein complex (Figure 4.7). The final sample buffer was 20mM Tris-HCl pH 7, 50mM NaCl, 1mM DTT, with a protein concentration of 0.2mM.

The HSQC spectra recorded for the VH#576/LMO2ΔN7 complex (Figure 4.7) was weak with a lot of peaks in the complex being too weak to detect. This was due, in part, to the low sample concentration (0.2mM). Furthermore, the relatively large size of the complex means NMR peak height was decreased due to a decrease in the rate of molecular diffusion. Therefore, a higher protein concentration and/or a deuterated protein sample would be required to study this complex further. However, in this case, the aim was to collect NMR data on the VH#576 half of the complex only and therefore, based on the long term stability of this complex (more than two weeks), a strategy was developed to isotopically label VH#576 and purify this protein with unlabeled LMO2.

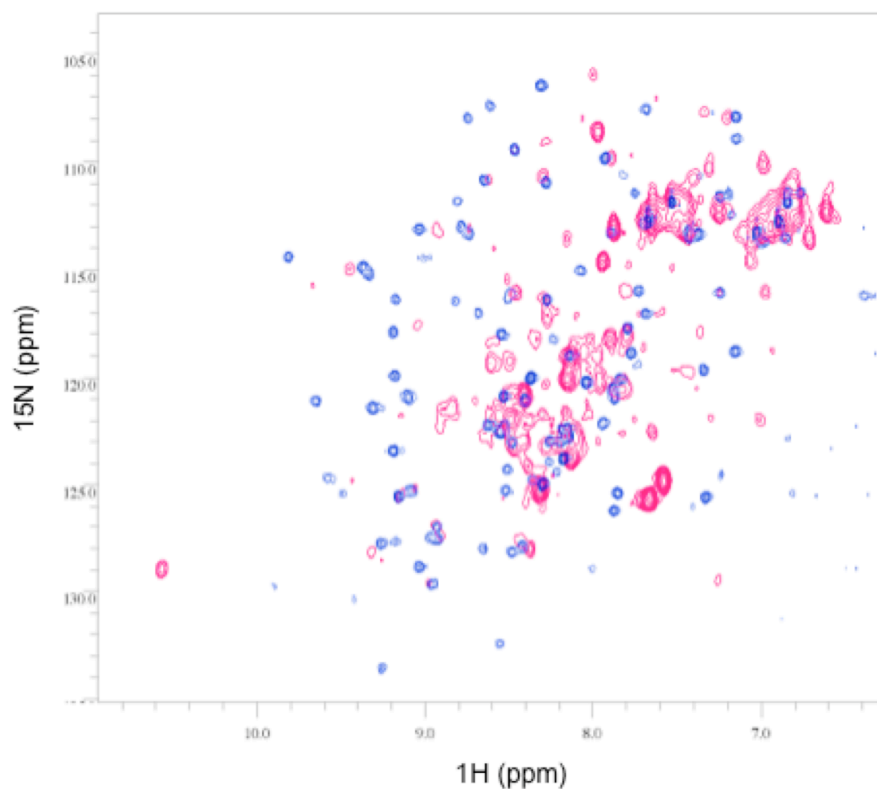


Figure 4.7: ^1H - ^{15}N HSQC spectra comparison for VH#576 and VH#576/LMO2 Δ N7 complex

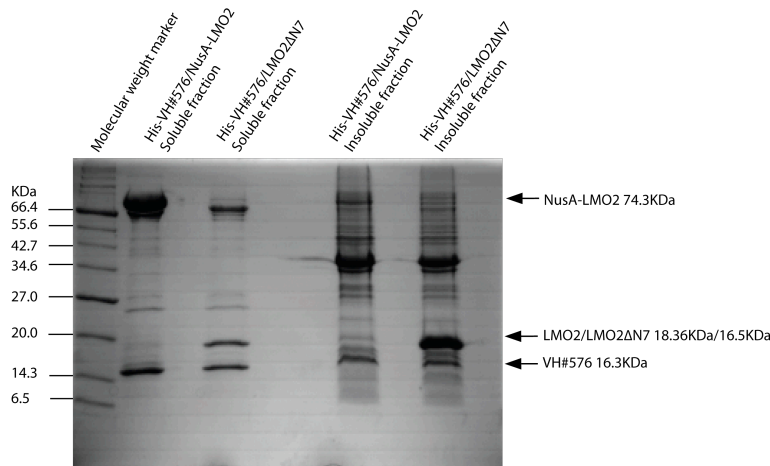
The blue spectra represents uncomplexed VH#576 and the red spectra VH#576/LMO2 Δ N7. The spectra for VH#576/LMO2 Δ N7 is weak compared to that of VH#576 due to the low concentration of the protein. However, importantly, the VH#576/LMO2 Δ N7 complex remained stable in the NMR tube during experiments and for more than two weeks afterwards.

4.4.2 Co lysis of VH#576 and LMO2

LMO2 is a protein of low solubility and stability in solution and as such cannot be purified alone. As a result of this, it was necessary to develop a strategy to effectively purify unlabeled LMO2 with uniformly labeled VH#576. In order to investigate this experimentally, two LMO2 expression vectors were produced; one expressing LMO2 Δ N7 alone and one expressing LMO2 as a fusion protein with NusA (NusA-LMO2). A Tev protease recognition site was engineered between NusA and LMO2 making cleavage of NusA possible. LMO2 (NusA tagged or untagged) and VH#576 were expressed from separate cultures, co-lysed and purified as a complex. The two purified protein complexes were analysed by SDS-PAGE (Figure 4.8). When LMO2 was expressed as a fusion protein (NusA-LMO2) and purified with VH#576 the yield was 6mg per liter (therefore approximately 2mg per liter of VH#576/LMO2 complex without NusA). In comparison the yield obtained when LMO2 Δ N7 was expressed alone and purified with VH#576 was 0.12mg per liter. The NusA-LMO2 co-lysed with His-VH#576 was taken forward as the best option due to its relatively high level of expression and solubility upon purification.

Structural analysis of an anti-LMO2 antibody single domain by NMR

A:



B:

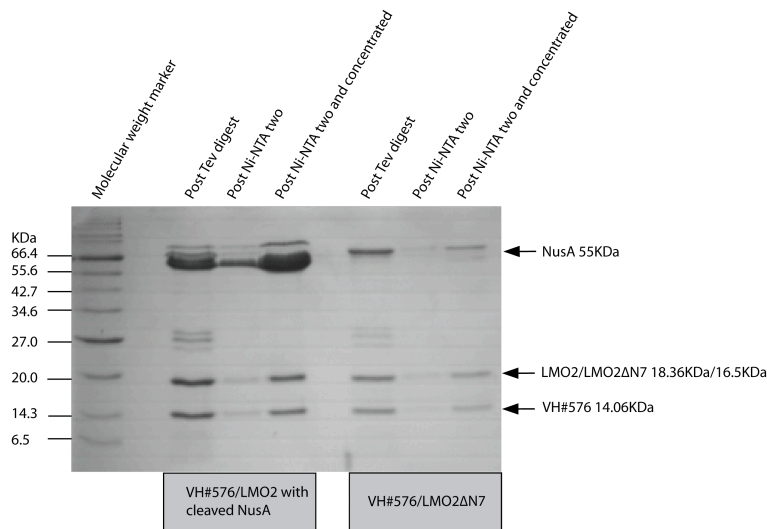


Figure 4.8: Co-lysis and Ni-NTA affinity chromatography purification of VH#576 and LMO2

15% SDS-PAGE stained with Coomassie brilliant blue

A: LMO2 Δ N7 or NusA-LMO2 were co-lysed with VH#576 by cell disruption. Purification of the complex was via the N-terminal His tag of VH#576 and Ni-NTA purification. The eluate (soluble fraction) was analysed by SDS-PAGE, and shows both complexes were purified successfully. The cell pellet, from each purification, was re-suspended and analysed (insoluble fraction); a large band of insoluble LMO2 Δ N7 can be seen.

B: The eluate from Ni-NTA chromatography was digested with Tev protease in order to cleave the His tag from VH#576 and NusA from NusA-LMO2. Fractions were analysed post Tev digest and post Ni-NTA affinity chromatography two (to remove Tev protease and His tag). The yield was highest for VH#576/LMO2 protein complex purified from the NusA-LMO2 fusion protein.

4.4.3 Deuterated VH#576

Large proteins (more than 30 KDa) have increased transverse relaxation rate constants that decrease the sensitivity of NMR experiments as less magnetisation persists, through pulse sequence delays, until detection. These resonances have larger line widths. Transverse relaxation rate constants can be reduced in larger proteins by eliminating unfavourable relaxation pathways such as relaxation of spins by the ^1H - ^{13}C dipolar interaction or the relaxation of ^1HN spins resulting from dipolar interactions with proximal aliphatic ^1H spins. An increase in both sensitivity and resolution can be achieved by reducing specific interaction pathways, thereby reducing the relaxation properties for both ^{13}C and ^{15}N spins (Venters et al., 1995). This is done by replacing a percentage (more than 50%) of carbon-bound hydrogen atoms in a protein with deuterium. ^1HN spins are reintroduced through amide proton solvent exchange by dissolving the protein in H_2O buffer. Reintroduction of ^1HN spins is slow for amide moieties that are highly protected from solvent exchange and a reduction in sensitivity for these residues may be observed. As the overall complex size will be 32.4 KDa, ideally, VH#576 needed to be perdeuterated (over 50%) and labeled with ^{15}N and ^{13}C . Non labile hydrogens in proteins are deuterated using D_2O instead of H_2O in the protein expression medium.

A protocol was developed to purify 50% deuterated (final) VH#576, labeled with ^{15}N and ^{13}C . A series of experiments were carried out using deuterated minimal media (no ^{15}N / ^{13}C) to check the viability of purification prior to the preparation of triple labeled VH#576 protein sample complexed with unlabeled LMO2.

Initially, VH#576 was expressed in 100mls of 75% D_2O minimal media and purified with NusA-LMO2 (Figure 4.9). C41 *E. coli* strain transformed with pRK-His-Tev-VH#576 was adapted for growth on deuterated medium by repeat sub streaking of colonies. The proteins were expressed and purified as described in chapter two, section 2.6.5. The yield was too low to give an informative chromatogram from the size exclusion chromatography step however this problem was solved by the development of an alternative cell culture method. The method was developed based on previous research (Sivashanmugam et al., 2009) (Cai et al., 1998, Marley et al., 2001) which described a protocol designed to generate cell mass in unlabeled rich media followed by exchange into labeled media at high cell densities.

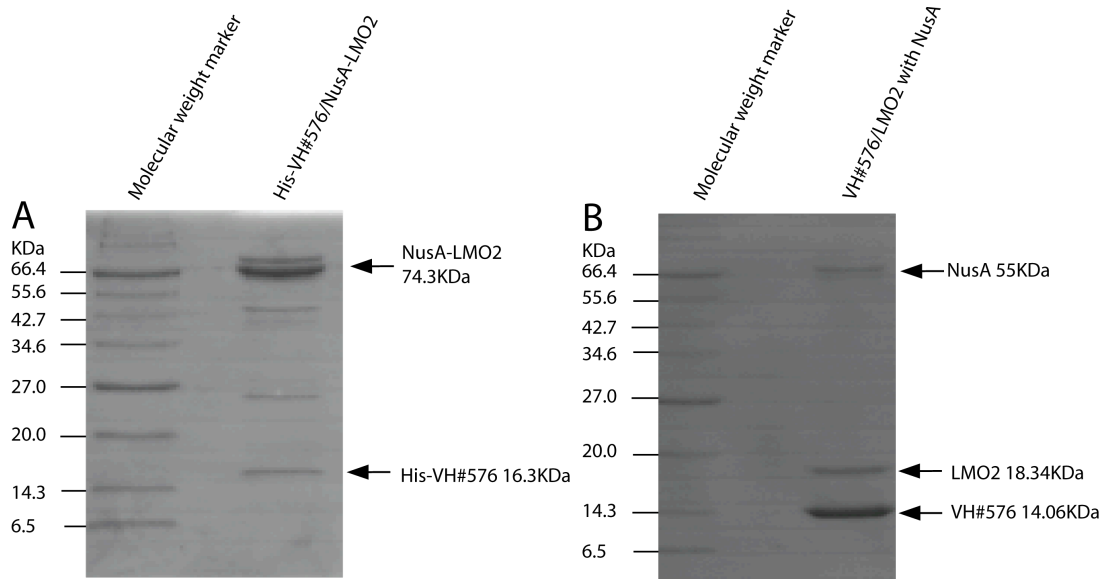


Figure 4.9: Purification of 50% deuterated VH#576 bound to LMO2

15% SDS-PAGE stained with Coomassie brilliant blue.

A: VH#576 was expressed in 75% D₂O minimal media while NusA-LMO2 was expressed in LB. The complex was purified by Ni-NTA affinity chromatography and the eluate analysed. The yield from a 100ml culture was 0.3mg.

B: Both the His tag and NusA tag were removed from VH#576 and LMO2 respectively, using Tev protease. The protein complex was then passed through a further Ni-NTA affinity chromatography column to separate both the His tag and Tev protease from the VH#576/LMO2 complex. The yield from a 100ml culture was 0.15mg (total protein).

4.4.4 Increasing the yield of deuterated VH#576

To improve the yield of VH/LMO2 purification further two protocols, based on increasing cell mass before induction of recombinant protein expression, were investigated. As the protocols varied in the two papers indicated, they were both tested and the final yields compared (see chapter two, section 2.6.6). Protocol one (Cai et al., 1998, Marley et al., 2001) entailed growing up *E. coli* cells in LB media until an OD_{600nm} of 0.7, the cells were then re-suspended in four times less labeled minimal media. Protocol two (Sivashanmugam et al., 2009) was similar however cells were grown up in LB and then re-suspended in the same volume of minimal media. Uniform isotope labeling is essential for NMR, thus it is vital the expression vector is not translated before media exchange. Western blot analysis, of a sample taken just prior to growth media exchange, confirmed there was no 'leaky' expression of the vector (Figure 4.10). Ni-NTA purification yields, per 50ml culture, were 0.2mg and 0.3mg for protocols one and two respectively. These results were very promising and therefore, based on a higher yield, protocol two (Sivashanmugam et al., 2009) was the chosen media exchange protocol, for the next experiment.

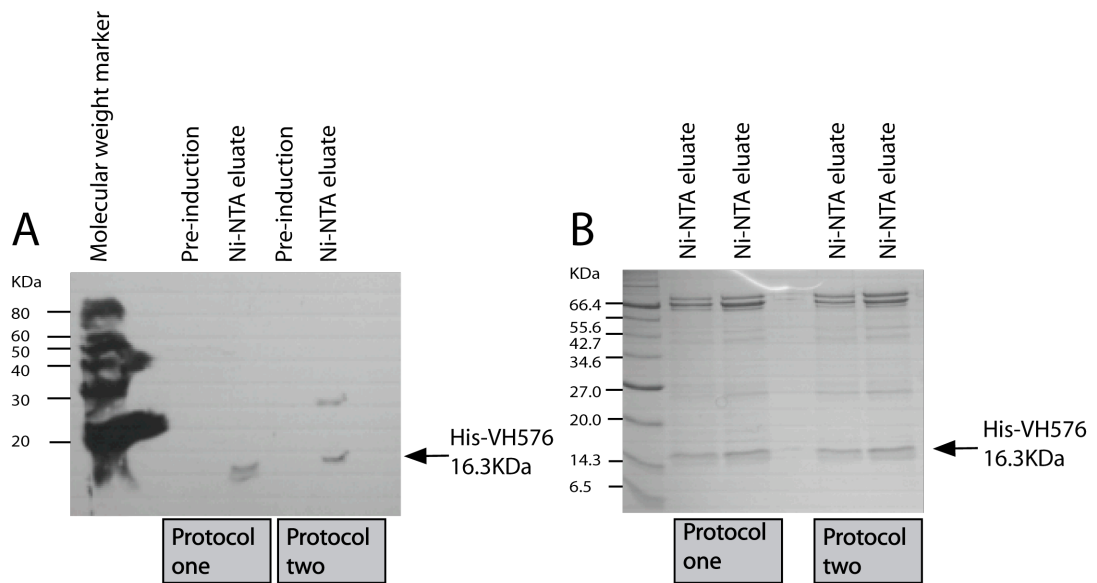


Figure 4.10: Revised strategy for expression of 50% deuterated VH#576

15% SDS-PAGE acrylamide gel

A: Anti-His Western blot analysis of samples taken just prior to media exchange. As a positive control, a relative amount of the Ni-NTA purified His-VH#576 sample was also analysed. The Western blot shows no protein induction prior to media exchange.

B: Analysis of Ni-NTA chromatography eluates by SDS-PAGE. Protocol one gave a His-VH#576 yield of 0.2mg (4mg per liter) and protocol two gave a yield of 0.3mg (6mg per liter).

4.4.5 Scale up of the purification of deuterated VH#576 with LMO2

The previous trial prep was not analysed by size exclusion chromatography. To achieve this, a scale up of the culture volumes was required; 500mls of 75% D2O minimal media (His-VH#576) and 500mls LB (NusA-LMO2) were used for bacterial growth and protein expression. This was to check the viability of a larger prep for the production of triple labeled VH#576 bound to LMO2. The total protein yield obtained from the first Ni-NTA chromatography step was 12mg. Protein was then subjected to a Tev digest during dialysis and the complex concentrated for size exclusion chromatography (superdex 200 10/300 GL column). Figure 4.11 shows that fractions from the appropriate peak of the size exclusion chromatogram contain both the VH#576/LMO2 complex and a NusA contaminant. This is due to the resolution capacity of the column. Consequently, a further chromatography step was investigated, ion exchange. A pH of 6 was chosen so that NusA would have a net negative charge and should bind the anion exchange resin whereas the majority of VH#576/LMO2 should remain unbound. Two anion exchange resins were tested; DEAE sepharose and Q sepharose. The relevant size exclusion fractions were pooled and diluted with sodium phosphate buffer, pH 6. Anion exchange flowthrough and eluate samples were concentrated and analysed by SDS-PAGE (Figure 4.11). Both resins retained NusA, separating it from the VH#576/LMO2 complex however, DEAE sepharose retained less of the VH#576/LMO2 complex so was utilised for the final purification experiment.

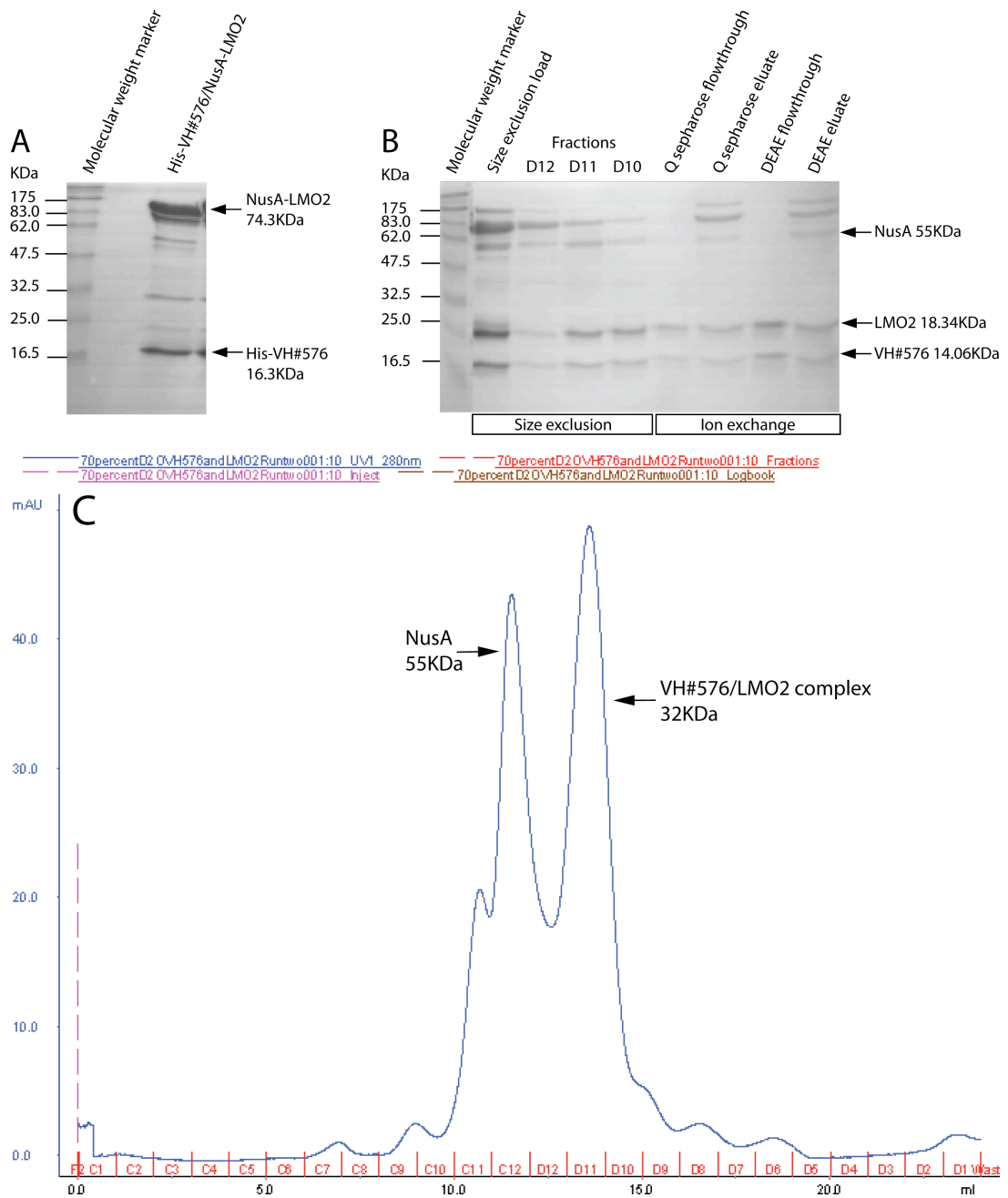


Figure 4.11: Purification of 50% deuterated VH#576 bound to LMO2

A: 15% SDS-PAGE stained with Coomassie brilliant blue. 50% deuterated VH#576 co-lysed with unlabeled NusA-LMO2 and purified by Ni-NTA affinity chromatography.

B: Purification of the VH/LMO2 complex by size exclusion and ion exchange. The size exclusion fractions analysed are indicated in C. Size exclusion chromatography was unable to resolve NusA and VH/LMO2 proteins and therefore two anion exchange resins were tested for their ability to retain NusA at pH 6 with DEAE being most suitable.

C: A280nm chromatogram for the size exclusion chromatography step.

4.4.6 VH#576 stability trial

As discussed previously the optimisation of NMR sample conditions is essential, as the protein must remain stable in solution for the duration of the NMR experimental procedure, which can be more than one month. From previous work with the complex it was known to be stable at pH 7 and therefore a stability trial was designed to test the effect of NaCl concentration and temperature on stability. Analysis by SDS-PAGE (Figure 4.12) shows degradation of LMO2 in all buffers at room temperature but not at 16°C. The most effective NMR sample conditions were identified as: 20mM Tris-HCl pH 7, 50mM NaCl, 1mM DTT at a temperature of less than 16°C. 10°C was chosen so that the spectra could be more directly compared to those of uncomplexed VH#576.

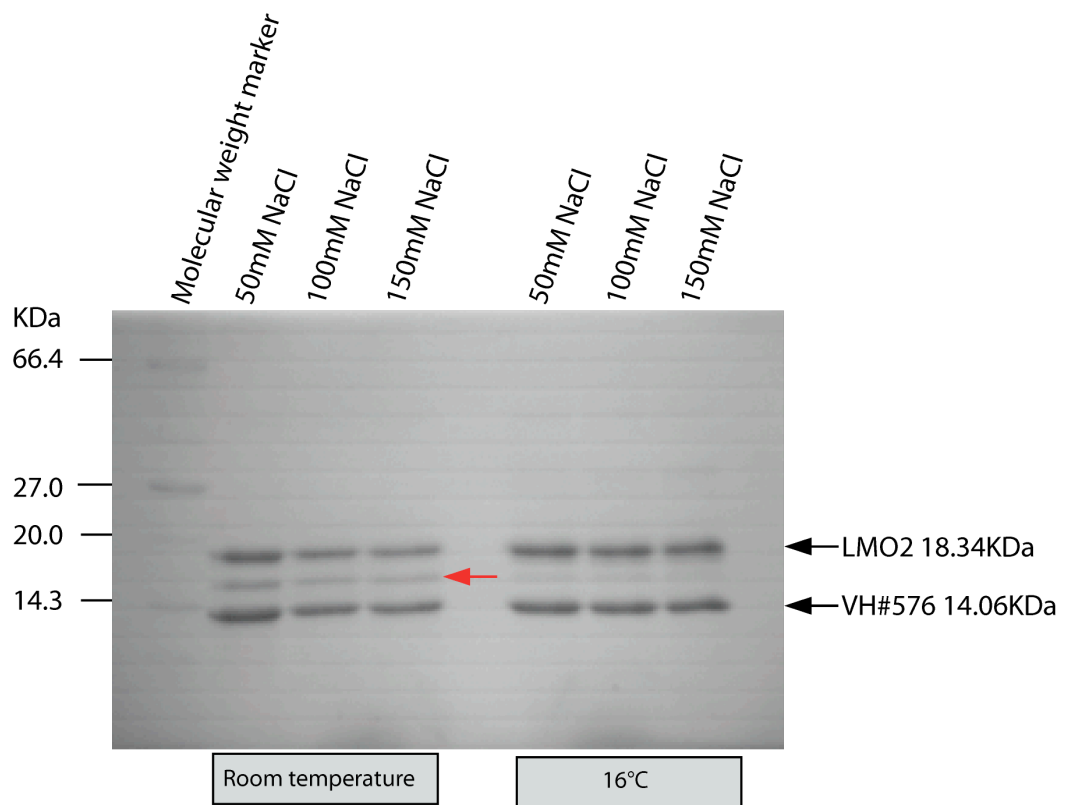


Figure 4.12: NMR protein sample stability trial

Samples of 50% deuterated VH#576 purified with unlabeled LMO2 were exchanged into 20mM Tris-HCl, 1mM DTT buffer with various NaCl concentrations. Protein samples were stored for one week at either 16°C or room temperature. Samples were then analysed by 15% SDS-PAGE stained with Coomassie brilliant blue. The red arrow indicates slight degradation of LMO2 when stored at room temperature; this did not occur at 16°C. The gel also shows the complex remains stable in 20mM Tris-HCl, pH 7, 50mM NaCl, 1mM DTT.

4.4.7 Final VH#576/LMO2 sample for analysis

50% deuterated VH#576 labeled with both ^{13}C and ^{15}N was purified successfully with LMO2 using the strategy developed (Figure 4.13). The sample was concentrated to 320 μl at 0.4mM. 900 MHz NMR Spectrometer (Oxford Instruments magnet) with 4-channel Varian INOVA console (Birmingham Biomolecular NMR Facility) was used to record a ^1H - ^{15}N HSQC spectrum for the protein in Tris-HCl buffer, pH 7.0, 50mM NaCl, 1mM DTT, at a temperature of 10°C. Despite thorough stability and temperature trials to define optimum conditions for the stability of the protein complex, $^{15}\text{N}/^{13}\text{C}/^2\text{H}$ labeled VH#576 bound to unlabeled LMO2 precipitated out of solution, consequently no spectra were collected. The protein sample was re-purified by size exclusion (superdex 200 10/300 GL column) and concentrated to 0.1mM. The size exclusion step showed evidence of aggregation as a large percentage of the complex eluted in the void volume. This aggregation may have occurred due to the higher concentration of the initial protein sample (13mg/ml) or it may have been due to the slight NusA protein contamination. Despite this loss of protein, a ^1H - ^{15}N HSQC spectrum was successfully recorded for the re-purified sample, at 10°C, using 750MHz Varian Unity Inova spectrometer and this was compared to a HSQC recorded for uncomplexed VH#576 at 10°C (Figure 4.14).

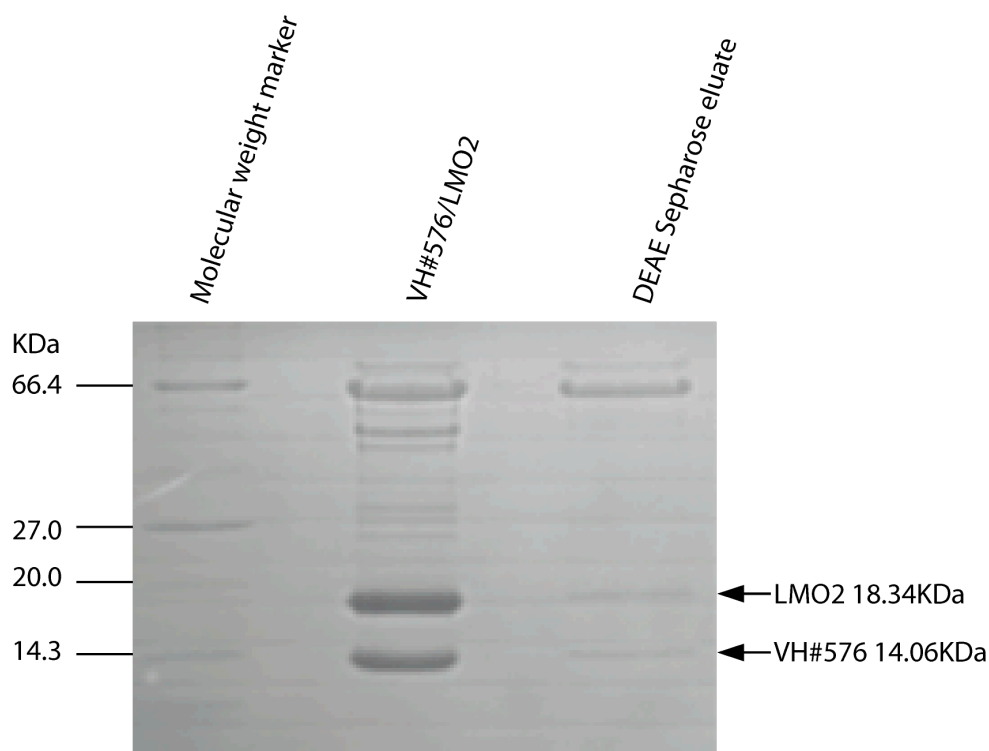


Figure 4.13: VH#576 labeled with $^{15}\text{N}/^{13}\text{C}/^2\text{H}$ bound to unlabeled LMO2

50% deuterated VH#576 was purified with unlabeled LMO2 by a process of Ni-NTA affinity, size exclusion and ion exchange chromatography. The purified protein complex was exchanged into 20mM Tris-HCl pH 7, 50mM NaCl, 1mM DTT and analysed by 15% SDS-PAGE stained with Coomassie brilliant blue as show. The DEAE sepharose anion exchange step successfully removed some of the NusA protein contaminant as shown in the eluate (final lane of the gel). However, the final sample still contains a slight NusA protein contamination therefore, can be estimated to be approximately 85% pure.

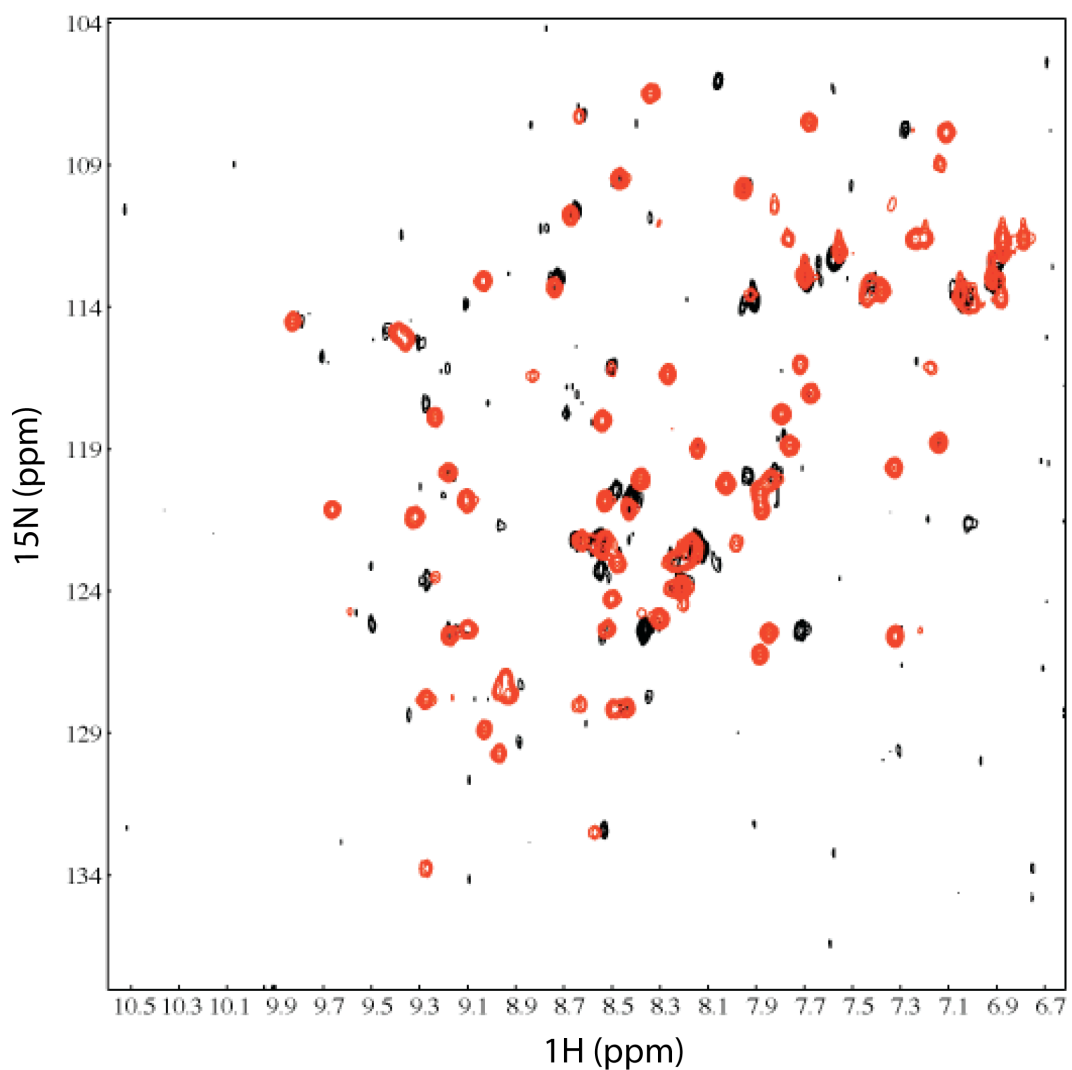


Figure 4.14: ^1H - ^{15}N HSQC spectra comparison for VH#576 in uncomplexed and complexed states

Data acquisition was carried out at 10°C. The black spectra represents complexed VH#576 and the red spectra uncomplexed VH#576. Due to the low concentration of the re-purified VH#576/LMO2 sample (0.05mM), the spectra for complexed VH#576 is weak (black) therefore, an accurate comparison with uncomplexed VH#576 cannot be made. However, chemical shift changes can be seen for approximately ten residues.

4.4.8 Comparison of free and bound VH#576 at 20°C

As a result of aggregation of the previous purification, ¹⁵N labeled VH#576 was purified with LMO2ΔN7ΔC11 as it may be that the flexible ends of LMO2 were responsible for the aggregation of the complex. C41 *E. coli* host was transformed with pRK-LMO2ΔN7ΔC11 and grown up in LB media as described previously (section 5.4.1). VH#576 was expressed in ¹⁵N labeled minimal media without deuterium; therefore no media exchange process was required. The purified complex was concentrated to 0.2mM in a buffer of 20mM Tris-HCl pH 7, 50mM NaCl, 1mM DTT. The HSQC recorded for the complex was of an improved quality compared to the previous spectrum (Figure 4.15). Moreover, the complex did not aggregate and was stable for more than one week. Thus, the strategy developed represents a mechanism to study VH#576 in complex with LMO2. Further data collection could potentially result in a solution structure for VH#576.

Chemical shift mapping is a highly sensitive tool for identification of protein binding sites and as such an overlay of uncomplexed and complexed VH#576 HSQC ¹H-¹⁵N correlation spectra was produced using nmrPipe (Figure 4.15). Perturbation of chemical shifts can be seen for a number of residues as a result of complex formation. This would suggest that the residues, missing from the free VH#576 HSQC spectrum, are not entirely from the CDR regions (LMO2 interacting sites). However, resonances in the spectrum may be shifted because the associated nuclei are proximal to the interaction interface or because conformational changes upon binding have resulted in altered local magnetic environments, even for nuclear spins distant from the interaction site (local structural effects) (Cavanagh et al., 2007). Spectral assignment will be necessary to identify the residues associated with each chemical shift change. Until then it can only be predicted that these chemical shifts are associated with the residues of the three CDR regions of VH#576.

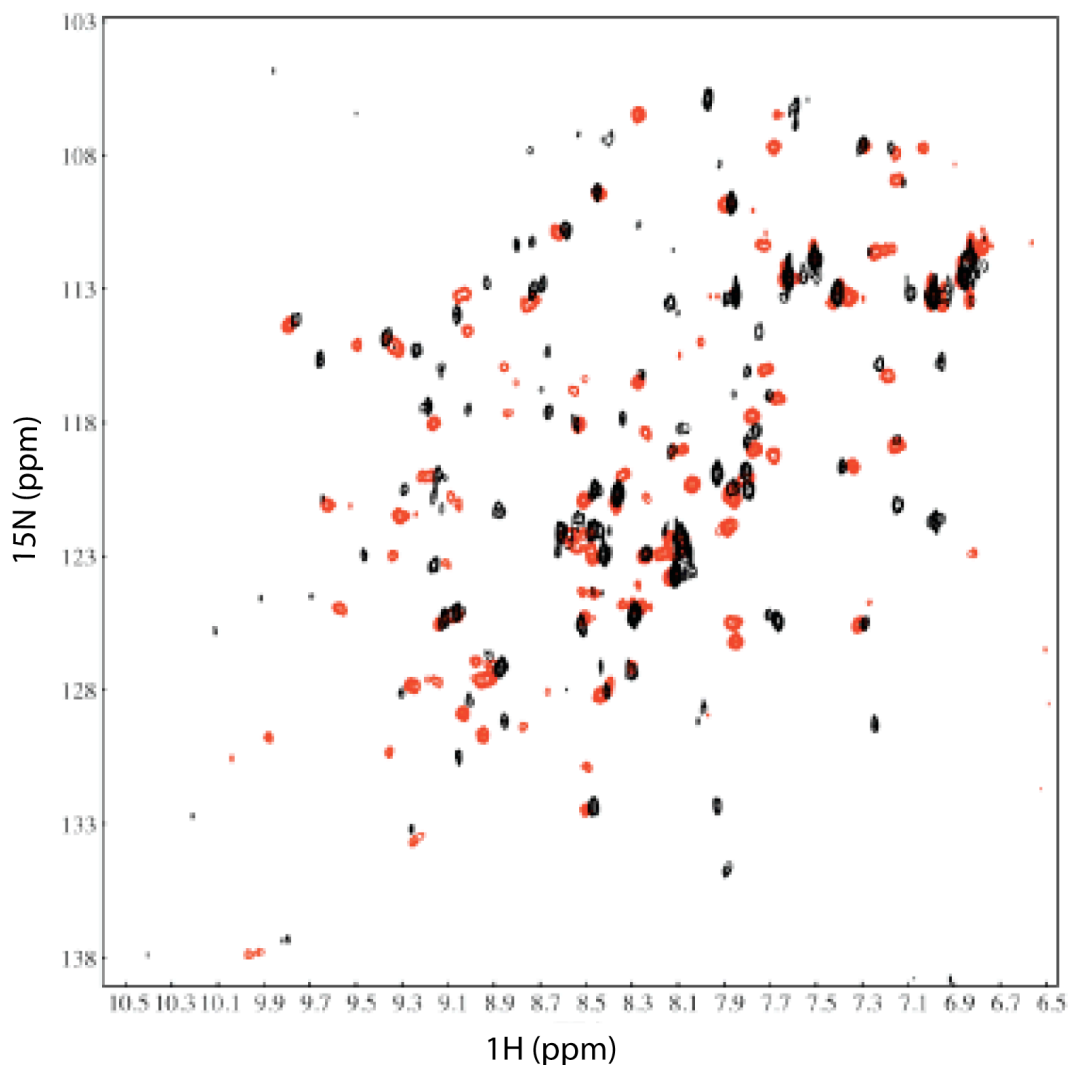


Figure 4.15: ^1H - ^{15}N HSQC spectra comparison for VH#576 in uncomplexed and complexed states

Data acquisition was carried out using ^{15}N labeled VH#576 bound to unlabeled LMO2 $\Delta\text{N7}\Delta\text{C11}$, at 20°C. The black spectra represents complexed VH#576 and the red spectra uncomplexed VH#576. Chemical shift changes can be seen for many residues however some are larger than others. Larger chemical shift changes are likely to represent residues which interact with LMO2.

4.5 Discussion

In this chapter VH#576 was taken forward for structural studies due to an initial well dispersed HSQC spectrum recorded for the antibody single domain reflecting a well folded protein. The long CDR three loop is thought to aid the solubility of VH#576 by folding over the side chain of hydrophobic amino acids at the former VH/VL interface, however this may be transient due to the loops conformational mobility (Bond et al., 2003). Strategies to stabilise VH#576 in solution were developed effectively and enabled NMR data acquisition.

As with VH-P8, CHAPS was used to increase the stability of VH#576 in solution making it possible to collect three dimensional spectra HNCA, HN(CO)CA, NOESY-HSQC and TOCSY-HSQC. These data enabled sequential linking of amino acids. However, signal strength was too low for less sensitive three dimensional experiments, making assignment impossible. Low signal strength was due to the low concentration of the VH#576 protein sample (0.2mM to 0.3mM) and unfavourable NMR conditions of low temperature (10°C) and addition of a detergent (CHAPS). Furthermore, spin systems were often incomplete as a result of the flexible CDR regions undergoing a chemical exchange process on a slow microsecond-to-millisecond time scale.

Evidence was collected to support the theory that VH#576 in complex with LMO2 would be more stable in solution and could be purified to a higher concentration without the need for the addition of CHAPS to the sample buffer. Structure determination of large proteins (greater than 30 KDa) by NMR is complex and therefore a protocol was developed to produce labeled VH#576 bound to unlabeled LMO2 thus reducing the complexity of the task by analysis of bound VH#576 only. The strategies described in this chapter represent methods that not only transform LMO2 from an insoluble protein but also allow separate analysis of the two halves of the VH#576/LMO2 complex via NMR. A HSQC spectrum was recorded for complexed VH#576 but the sample was not at a high enough concentration for further data collection. The strategy has scope for further improvements with the potential of data collection for structural determination of complexed VH#576.

Comparison of VH#576 HSQC two dimensional correlation spectra, in the complexed (with LMO2) and uncomplexed state, reveals chemical shift changes which may be associated to the residues interacting with LMO2, in the CDR regions, or may be associated to residues proximal to the interaction interface. Alternatively they may be attributed to conformational changes upon binding. Residue assignment upon further data collection would enable these residues to be mapped.

The purpose of this work was to gain enough structural information on VH#576 to produce a three dimensional pharmacophore model for screening small molecule libraries. However, the X-ray crystallography approach carried out in parallel to this research proved successful leading to a crystal structure of VH#576/LMO2 Δ N7 Δ C11 (as will be described in chapter five) therefore, it is debatable whether further collection of NMR data for structure determination is still a requirement in order to fulfill the overall aim of developing an anti-LMO2 small molecule inhibitor. From a structural point of view it would be interesting to compare NMR solution structures of complexed and uncomplexed VH#576 with further comparisons between NMR solution structures and the crystal structure. Comparison of the NMR structure and crystal structure would enable the detection of any crystallisation artifacts due to crystal packing as an NMR solution structure represents a more physiological environment. Crystal packing can reduce local protein dynamics, particularly in solvent exposed loop regions, or distort relative domain orientations and protein-protein interactions (Eyal et al., 2005). This is particularly significant in this case as VH#576 has many exposed loops and the rigid nature of the crystal structure may not portray accurately the range of possible orientations for these flexible loops. Furthermore, the crystal structure demonstrates interactions between LMO2 and the former VH/VL interface of VH#576 and this would be confirmed by the observation of chemical shift changes in this region.

In terms of feasibility, it may well be possible to solve the NMR solution structure of both complexed and uncomplexed VH#576 with the aid of high magnetic field strength (900MHz) and cryogenically cooled probes to improve experiment sensitivity.

4.6 Conclusion

The addition of CHAPS detergent has enabled the NMR analysis of an anti-LMO2 antibody single domain. HNCA, HN(CO)CA three dimensional spectra have been collected for free VH#576 and allow sequential linking of amino acid residues however residues could not be identified due to an extremely weak HNCACB spectra.

VH#576 chemical shift changes were observed upon binding LMO2 but until assignment of uncomplexed VH#576 has been completed it is not known whether they are associated with residues of the CDR regions. Future work would include recording HNCACB, CBCA(CO)NH for uncomplexed VH#576 to enable complete assignment, identify residues missing from the spectra and residues that undergo chemical shift changes upon binding LMO2.

VH#576 structure determination would involve collecting details of the local backbone geometry by measuring nuclear Overhauser effects (NOEs). Comparison of a complexed VH#576 NMR solution structure and complexed VH#576 crystal structure may show slight differences due to the influence of crystal packing.

5 Structural determination of LMO2 bound to anti-LMO2 antibody single domain and characterisation of critical interaction residues

5.1 Introduction

It has been shown that LMO2 has a vital role in T-cell acute lymphoblastic leukaemia (T-ALL) and has been implicated in diffuse large B-cell lymphoma, where it has prognostic significance (Natkunam et al., 2007). Current, the only available treatment for T-ALL involves intensive chemotherapy, which has significant short and long term side effects (Aifantis et al., 2008). Therefore, LMO2 is an important target for development of novel therapies, in particular new drugs to modulate the interaction of LMO2 with its natural partners. The protein complex in which LMO2 is involved is an attractive target for therapy development.

Preliminary work, in response to this problem included the identification of VH#576, an antibody single domain that binds to LMO2 with high affinity and specificity. VH#576 has the capacity to modulate the function of LMO2 inside the cell (T. Tanaka, personal communication). Using VH#576, a macrodrug, as a therapeutic agent presents a number of serious problems among them delivery, degradation, and immunogenicity. The long-term goal therefore, is to evolve VH#576 into a small, drug-like, chemical compound. Computational tools can be used to complement and guide the drug discovery process. Structure-based drug design requires the three-dimensional (3D) structure of the protein of interest and as such, the 3D structure of LMO2 bound to VH#576 was solved as described in this chapter. The information obtained from the crystal structure will be used to guide the design of small molecules able to mimic VH#576, i.e. able to bind to LMO2, preventing the interaction with its natural partners.

Historically, the targeting of protein-protein interactions has proved difficult because the surfaces of the proteins involved are often large and flat, with no potential binding pockets or grooves. In addition, protein-protein interactions typically have a large interface of 750-1500Å² (Petros et al., 2006). Therefore, it may not be possible to inhibit the whole interaction with a single small molecule. Nevertheless, there have been a number of successes in the discovery of small molecules able to modulate protein-protein interactions (Wells and McClendon, 2007); the majority of these were due to the realisation that most of the binding energy can be ascribed to a set of complementary residues, a hot spot, surrounded by weaker interactions. Mutation of residues at the interface of a protein-protein interaction, to alanine or glycine, can identify these critical regions of interaction. Upon mutation of a critical residue, binding is ablated or greatly impaired. Moreover, identification of a hotspot significantly aids the design of small molecules that alter protein-protein interactions (Zhang and Palzkill, 2004) as the target area is significantly reduced.

Analogous analysis of an anti-RAS VH provided biochemical data that matched the precise interaction of residues determined by X-ray crystal structure (Tanaka et al., 2007).

An example of the successful development of a small molecule to inhibit a protein-protein interaction is the identification of anti HDM2 (human protein double minute 2) small molecule inhibitor, termed Nutlin-3. HDM2 interacts with and blocks the activity of P53, an important transcription factor and tumour suppressor. P53 has many functions to maintain genetic stability such as cell cycle arrest, apoptosis and senescence (Gaidarenko and Xu, 2009). Therefore blocking the interaction between HDM2 and P53 could lead to activation of P53 and thus inhibition of tumour growth. The crystal structure of MDM2 (mouse homologue) bound to a P53 transactivation domain (an α helical region of fifteen amino acids), shows the interface is mainly hydrophobic. Mutational analysis of the P53 (Picksley et al., 1994) transactivation domain by alanine scanning identified a hotspot of three amino acids (Phe19, Trp23 and Leu26) that were critical for interaction with MDM2. A synthetic chemical library screen identified an MDM2 inhibitor, Nutlin-3, which binds in the same area as the hotspot residues of P53. The high binding compound displaces P53 from the MDM2 protein with a half maximal inhibitory concentration of 90nM, as determined by Surface plasmon resonance. The binding mode of Nutlin-3 has been illustrated by a high resolution X-ray structure (Vassilev et al., 2004), which shows the P53 hotspot residue Phe19 is mimicked by the isopropoxyloxy phenyl moiety of Nutlin-3.

For the purpose of structure determination, NMR and X-ray crystallography were investigated in parallel. A third approach, which was considered in the interim, when no crystallography data was available, was homology modeling. Proteins with similar amino acid sequences have similar conformations therefore proteins of known structure can be used as a basis for building structural models. Models of LMO2 and VH#576 were produced in this way and docked *in silico*. Mutagenesis of residues in the CDR regions of VH#576 enabled the identification of crucial VH#576 residues for the interaction with LMO2 and this information was utilised by the docking program (HADDOCK) to produce a structural model of the complex (Dominguez et al., 2003). Finally, once VH#576/LMO2 Δ N7 Δ C11 crystallisation was successful; diffraction data was collected and processed. The structure has been solved to a medium resolution of 3.3Å. The structure reveals hydrophobic interactions between LMO2 and the former VH/VL interface and demonstrates the plasticity of LMO2. Information derived from the VH#576/LMO2 Δ N7 Δ C11 crystal structure will be essential to guide the process of structure based drug design.

5.2 VH#576 and LMO2 mutagenesis

5.2.1 Analysis of VH#576 point mutations using a mammalian two-hybrid luciferase reporter assay

The critical, or hotspot, VH#576 residues that mediate the interaction with LMO2 were identified using a mutagenesis screen. Residues were mutated to glycine and alanine as they are non-polar, aliphatic amino acids with non-reactive side chains of a single hydrogen atom and methyl group respectively. The interaction between VH#576-VP16 and GAL4-LMO2 Δ N24 (LMO2 residues 25 to 147) was measured through a mammalian two-hybrid luciferase reporter assay in which VP16 activation domain can interact with GAL4 DNA binding domain (Figure 5.1). This interaction induces functional transcriptional activation from RNA polymerase II basal promoters with upstream GAL4 binding sites and thus cis-activation of firefly luciferase reporter genes. The mammalian GAL4 dependent reporter plasmid (pG5), contains five GAL4 responsive binding sites in front of the luciferase cassette. pRL-CMV plasmid for the expression of *Renilla* luciferase acts as an internal control to which firefly expression can be normalised; this minimises experimental variability caused by, for example, differences in cell viability and transfection efficiency. The expression of luciferase is quantified sequentially, using a luminometer.

An amino acid sequence alignment of VH#576 with anti-Ras antibody single domain, VH#6 (see Appendix B, figure 1) was used to aid the choice of residues to mutate. Residues were selected based on a combination of differences in the sequence alignment and residues, which were shown to be important for the interaction between RAS and anti-RAS (VH#6).

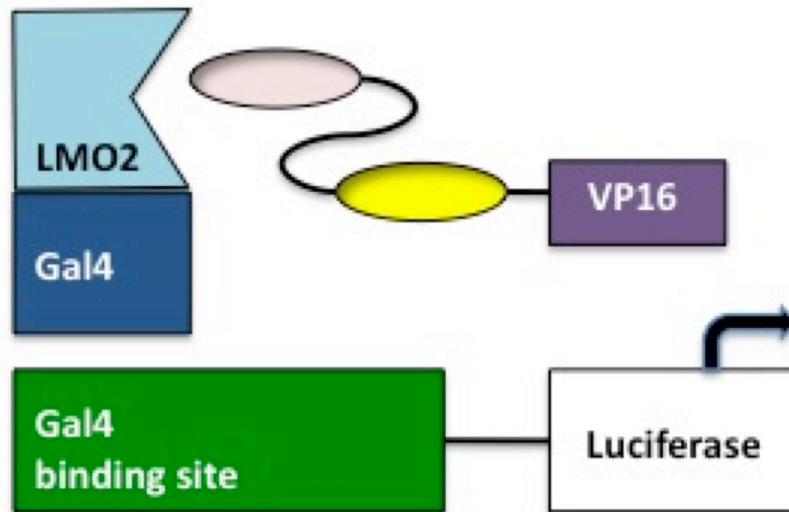


Figure 5.1 Mammalian two-hybrid luciferase reporter assay

Co-transfection of mammalian Chinese hamster ovary (CHO) cells with vectors expressing GAL4-LMO2 Δ N24 and VP16-VH#576. LMO2 and VH#576 binding results in a close interaction of GAL4 and VP16 domains. The GAL4 promoters are then functionally activated with associated increases in the luciferase reporter expression.

5.2.2 Mammalian two-hybrid method

Mammalian expression vectors pEF-VP16-VH#576 and pM3-GAL4-LMO2 Δ N24 were obtained from T. Tanaka (THR laboratory). Selected residues in the CDR regions of VH#576 were mutated, from the base vector template, to either glycine or alanine (see Appendix C for primer sequence) by point mutation followed by assembly PCR. The PCR product was cloned back into pEF-VP16. All expression vectors were sequenced using ‘The sequencing service’ (The University of Dundee). The effect of each point mutation was analysed using mammalian two-hybrid luciferase reporter assay, each clone was tested in duplicate. If the first round mutation resulted in no interaction then the reciprocal glycine/alanine mutation was also tested. Five rounds of mutagenesis experiments were completed with a positive and negative control for each. The positive control used was original VH#576 and the negative control, anti-Ras VH#6. The expression of the mutated VH#576-VP16 fusion protein was checked by anti-VP16 Western blot (for full blots see Appendix B).

5.2.3 Effect of VH#576 point mutations on the interaction with LMO2

Figure 5.2 brings together results of all five rounds of mutagenesis experiments as analysed by a mammalian two-hybrid luciferase reporter assay. It is important to note that a given point mutation of VH#576 may destroy the interaction with LMO2 in one of three ways. It may be that the residue mutated was critical for the interaction. Secondly, the mutation may result in a complete change in conformation or unfolding of the VH#576 antibody fragment. Thirdly, the mutation may prevent VH#576 expression however, this was checked for by Western blot analysis. That said, the residues selected for mutation were all located in the CDR loops, which are highly flexible, exposed regions of the protein, therefore it is likely that if the mutation suppresses the interaction it has a critical role.

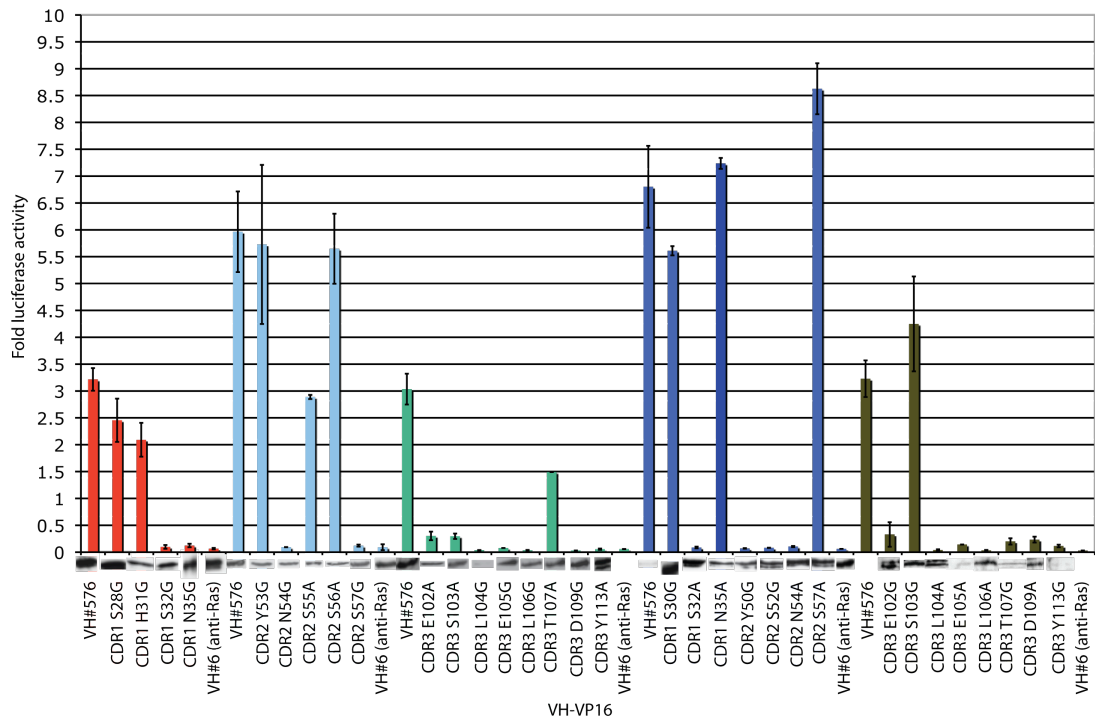


Figure 5.2: VH#576 CDR mutagenesis to determine residues critical for the interaction with LMO2

The five colours used for the bars represent five separate experiments each with a positive (VH#576 original protein) and negative (VH#6 anti-Ras) control. Each bar represents an average of luciferase activity measured for two wells (replicates) and the bar extensions indicate standard deviation. If interaction was inhibited by a glycine/alanine mutation in the first round then the reciprocal mutation was also analysed. The results suggest the majority of the interaction occurs through the CDR3 region with a total of seven residues identified as critical for the interaction. One residue in CDR1 was found to be involved in the LMO2 interaction and two for CDR2. Anti-VP16 Western blot analysis was employed to analyse the prey (VH#576-VP16) expression levels and results are shown underneath each bar in the chart. All mutated versions of VH#576 were expressed, except for the Y113G mutated form. This suggests that Tyr 113 is important structurally for the stability of VH#576.

Many of the VH#576 residues found to be critical for the interaction, such as Ser 32, Tyr 50, Ser 52, Asn 54, Thr 107, have polar side chains with hydroxyl groups that are able to form hydrogen bonds. Therefore, this is a likely mechanism of interaction with LMO2. The CDR3 region contains two glutamic acids (Glu 102, Glu 105) and one aspartic acid (Asp 109). These residues have side chains with a carboxyl group of negative charge, under physiological conditions, and thus the potential to form electrostatic interactions with positively charged groups. The side chain of leucine (Leu 104, Leu 106) is hydrophobic and likely to interact with other non-polar side chains such as tryptophan and phenylalanine. Figure 5.3 shows the side chains of each residue of VH#576 found to be critical for the LMO2 interaction.

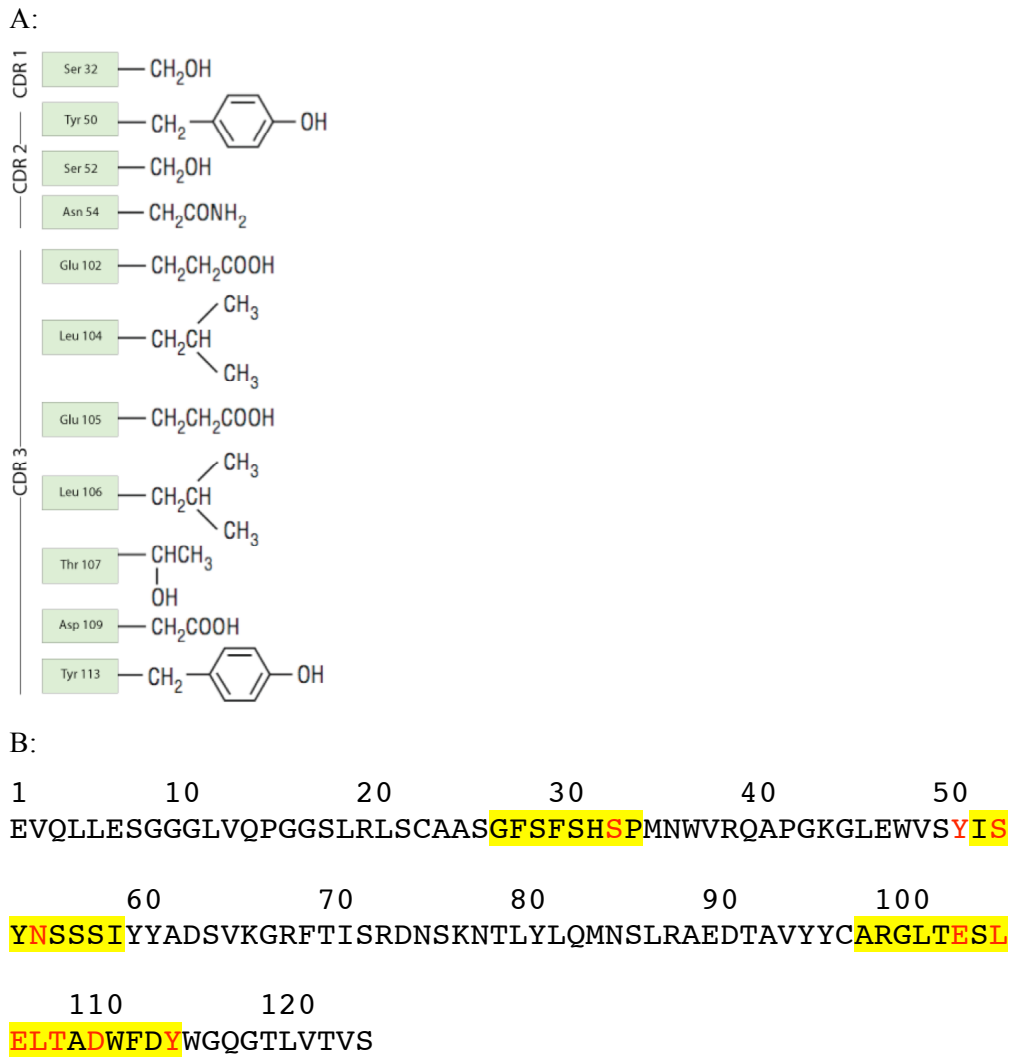


Figure 5.3: VH#576 residues involved in the interaction with LMO2

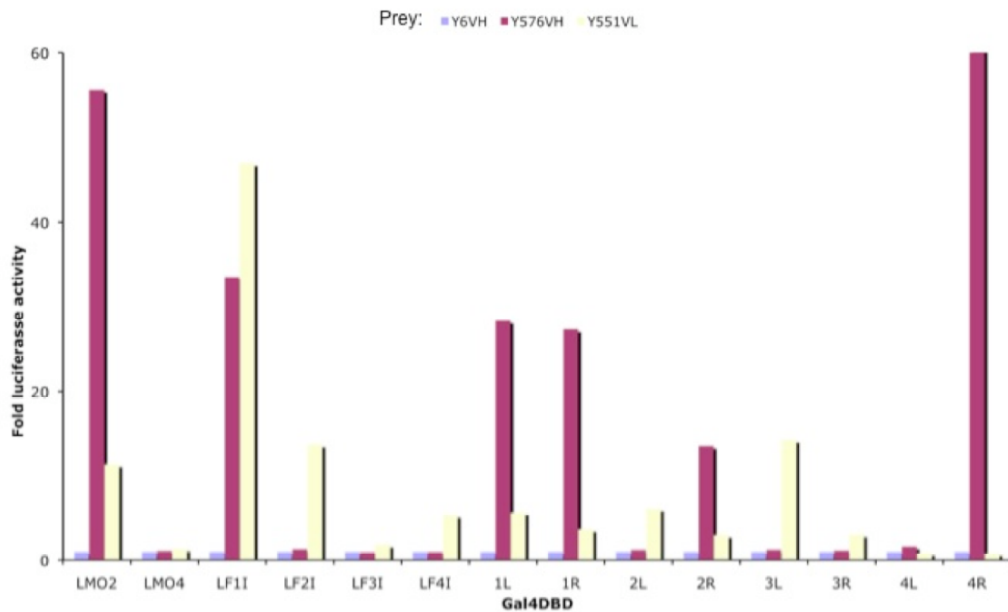
A: A schematic representation of the interacting residues of VH#576. Hydroxyl (OH) groups are likely to form hydrogen bonds with residues of LMO2 and negative acidic side chains are likely to form electrostatic interactions.

B: VH#576 amino acid sequence (single letter code) with CDR regions highlighted in yellow and vital LMO2 interacting residues highlighted in red. Note that the nomenclature used for describing the heavy chain domain structure is the IMGT (Lefranc, 2001) numbering system.

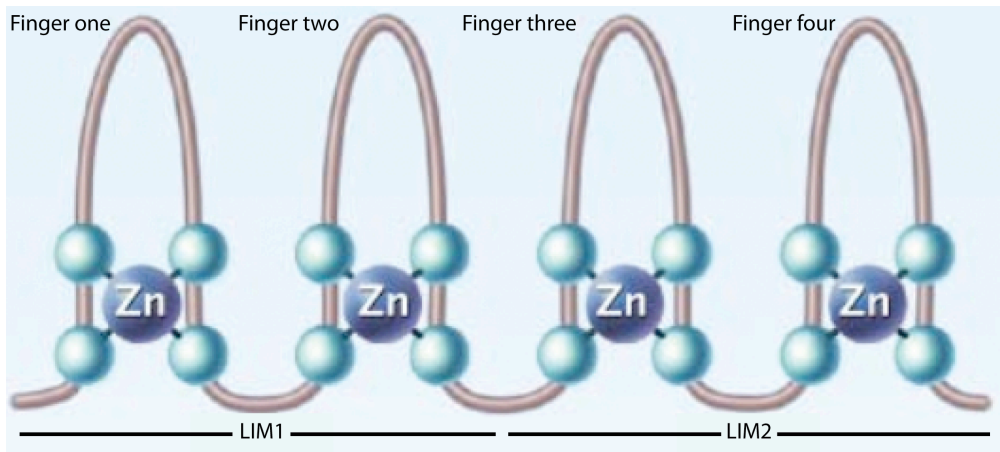
5.2.4 LMO2 interaction region

The binding site on the surface of LMO2 is also an important factor and for that reason, LMO2 mutagenesis data was also collected. This experiment was carried out by Dr. T. Tanaka (THR laboratory). Mutations were made to match the sequence of LMO4 or ISL-1 (both LIM domain proteins), which do not bind VH#576 (see Appendix B, figure 1, for further sequence information). Each LIM finger of LMO2 was mutated separately and the left and right sides of each LIM finger were also mutated individually. Again a mammalian two-hybrid luciferase reporter assay was used to measure the effect, of each mutation, on the VH#576 interaction (Figure 5.4). The results of the mutagenesis analysis indicate that the VH#576 binding site spans both LIM1 and LIM2 of LMO2 with finger two and three being involved in the interaction along with the left side of finger four. However, mutations to the right side of finger two had less of an effect on the interaction (Figure 5.4).

A:



B:



Data generated by Dr. T. Tanaka (THR laboratory)

Figure 5.4: Regions of LMO2 involved in the interaction with VH#576

A: Results of the mammalian two-hybrid luciferase reporter assay to analyse the interaction between VH#576 (dark purple bars) with various LIM finger mutants. Each of the LIM fingers (LF1I, LF2I, LF3I, LF4I) were mutated and the left and right side of each LIM finger were also mutated. According to this data, both LIM1 and LIM2 domains of LMO2 are involved in binding VH#576, with LIM finger two, three and the left side of finger four being the most interactive regions.

B: Schematic representation of LMO2. Each of the four zinc binding LIM fingers are shown and the two LIM domains and indicated. Adapted from (McCormack and Rabbitts, 2004).

5.3 VH#576/LMO2ΔN25 *in silico* structural model

At the time this *in silico* model was produced, no crystallography data was available due to the difficulties described in chapter three. Therefore, a program called HADDOCK (High Ambiguity Driven protein-protein Docking) was used in combination with the mutagenesis data to produce an *in silico* structural model of the VH#576/LMO2ΔN25 complex (Dominguez et al., 2003).

5.3.1 The use of docking software

The methods used to create the *in silico* structural model will be described in this section. Unbound structural models were built for LMO2 and VH#576 using M4T (Fernandez-Fuentes et al., 2007). PDB 1J2O (FLIN2: LIM1 and Ldb1_LID domain) and PDB 2DFY (FLINC4: LMO4 and Ldb1_LID domain) were used as templates for LMO2 and PDB 2UZI (anti-Ras VH#6) as a template for VH#576. The quality of the models was assessed using PROSA-II (Sippl, 1993) and PROCHECK (Laskowski et al., 1993). The models were then docked and scored using HADDOCK (Dominguez et al., 2003). The mutagenesis information was introduced as ambiguous interaction restraints to drive the docking process. Ambiguous interaction restraints were calculated from the ambiguous distance between all residues shown to be involved in the interaction. The HADDOCK score takes into account the intermolecular energy (electrostatic, van der Waals etc.) and as a general rule, the lower the score the more likely the two proteins will interact as predicted. This scoring system relies on asymmetry at the protein-protein interface due to the distribution of hydrophobic and hydrophilic residues and the shape complementarities of the proteins. Figure 5.5 shows one of ten of the lowest energy complexes. The *in silico* structural model was produced by Dr. N. Fernandez-Fuentes.

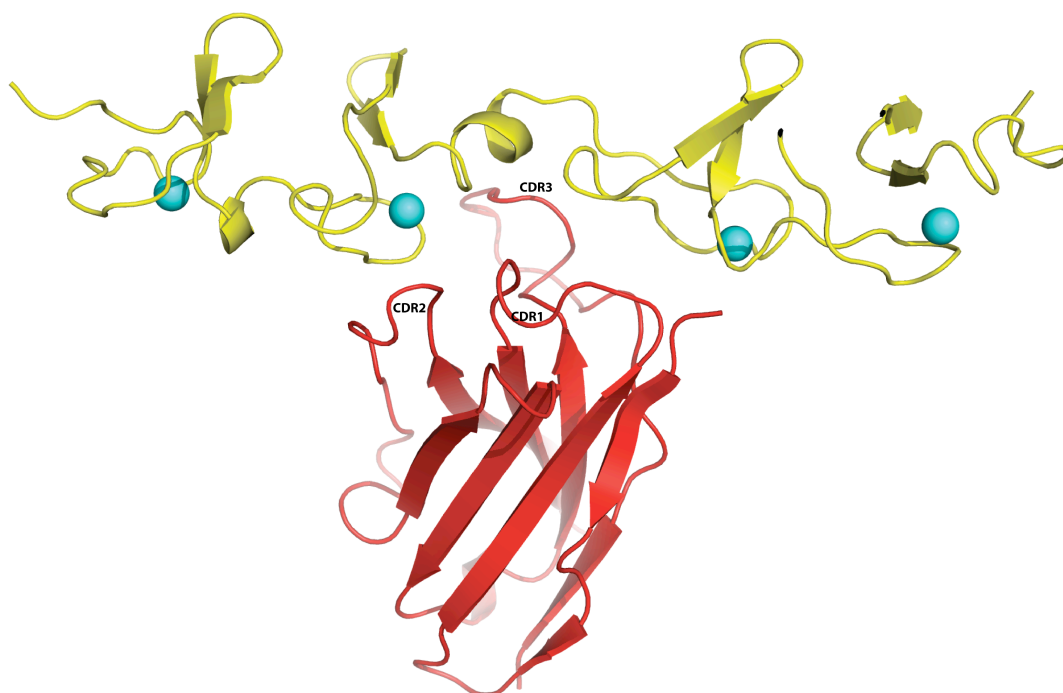


Figure 5.5: VH#576/LMO2ΔN25 *in silico* structural model

VH#576/LMO2ΔN25 *in silico* model. The overall shape of the complex is displayed as a ribbon diagram. LMO2ΔN25 is shown in yellow, with the zinc ions in cyan and VH#576 in red. The model aligns the CDR regions of VH#576 with LMO2. The structure has a HADDOCK score of -74.4 ± 10.6 and a buried surface area of $2207.4 \text{ \AA}^2 \pm 118.9$.

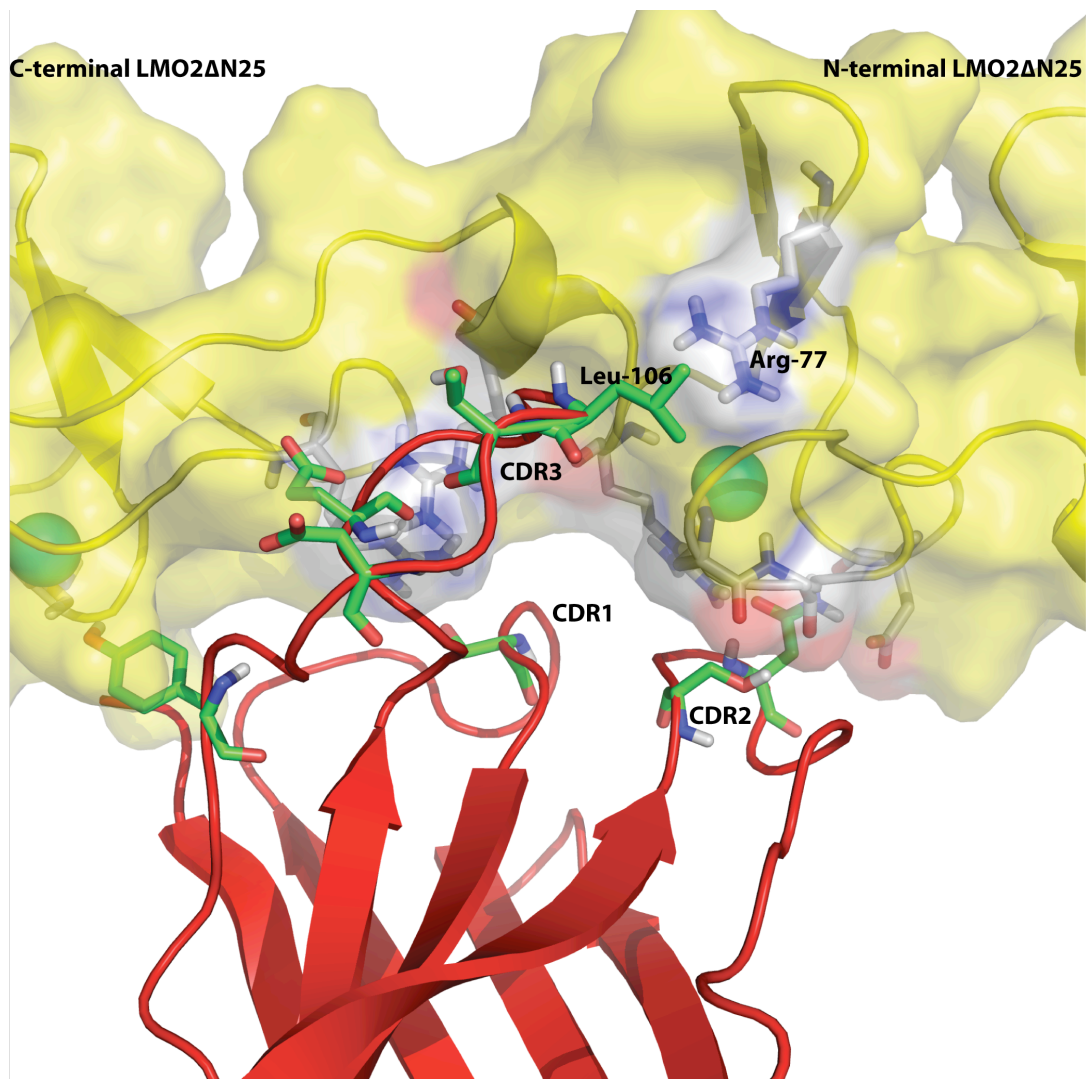


Figure 5.6: VH#576/LMO2 Δ N25 *in silico* structural model: The protein-protein interface
An enlargement of the protein-protein interaction interface with the CDR regions of VH#576 labeled. LMO2 is shown in yellow with surface rendering and VH#576 in red. Please note the orientation of LMO2 is C-terminus to N-terminus, as indicated. Residues involved in the interaction at the protein-protein interface are shown in ball-and-stick. A hydrophobic interaction between Leu 106 (VH#576) and Arg 77 (LMO2) is also indicated. Other interactions, not highlighted, include an electrostatic interaction between Glu 102 of VH#576 and Arg 86 of LMO2.

5.3.2 Interactions illustrated by the VH#576/LMO2 Δ N25 *in silico* model

The structural model shown in Figure 5.5 explains very well most of the mutagenesis data described in section 5.2. In particular, the hydrogen bond interactions between the side chain of Ser 32 in VH#576 and Arg 102 of LMO2. Electrostatic interactions are present between VH#576 residues Glu 102, Asp 109 and LMO2 residues Arg 86 and Arg 100 respectively. A hydrophobic interaction also exists between Leu 104 of VH#576 and Leu 59 of LMO2. At the end of this chapter (the discussion section) comparisons will be made between this *in silico* structural model and the VH#576/LMO2 Δ N7 Δ C11 crystal structure.

5.4 Solving the crystal structure of VH#576/LMO2 Δ N7 Δ C11 protein complex

5.4.1 Methods used to solve the structure

The following work was carried out by H. Sewell in collaboration with Dr. E. J. Mancini and Dr. K. El Omari. Omari *et al.* have recently solved the structure of LMO2 bound, through a flexible linker, to the LIM interacting domain (LID) of LDB1 (Omari *et al.*, 2010, manuscript in preparation). This structure will be referred to as FLINC2 (an abbreviation of ‘fusion of LDB1-LID and the N and C-terminal LIM domains of LMO2’).

5.4.1.1 Diffraction data collection

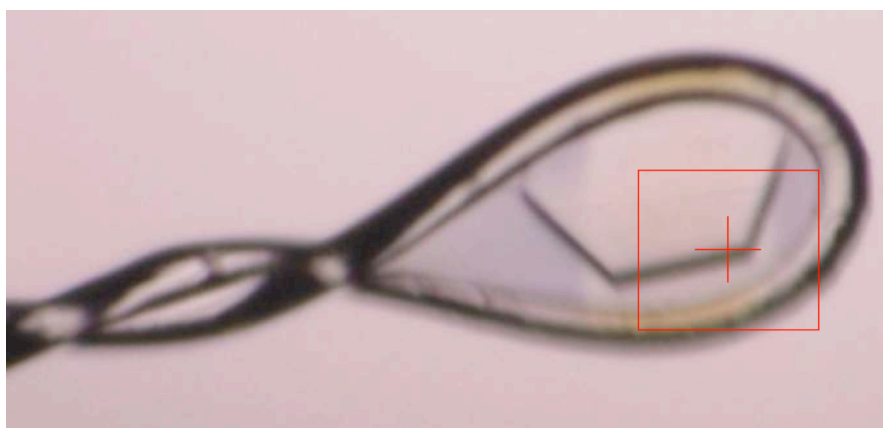
For data collection, crystals were cryo-protected with a solution containing 30% glycerol and flash frozen in liquid nitrogen. Single anomalous dispersion (SAD) data was collected at the peak wavelength of 1.28240 Å for the Zn K-edge, using macromolecular crystallography beamline IO2 (Diamond light source, Oxfordshire, UK). 360° of data were collected as a series of 1.0° oscillations and acquired on detector type ADSC Q315 CCD using an exposure time of 1second per frame (Figure 5.7).

5.4.1.2 Diffraction data processing

The diffraction data was indexed, scaled and merged using HKL2000 data processing software (Otwinowski and Minor, 1997). The unit cell was indexed as a primitive hexagonal Bravais lattice and the space group was chosen as P6, amongst the possible primitive hexagonal space groups, based on systematic absences and R_{merge} . The R_{merge} is used to assess the quality of the data and is a measure of the similarity of the recorded intensities for symmetry equivalent reflections. The data resolution was cut to 3.3Å, based on $I/\sigma(I)$ (signal to noise ratio of recorded intensities) and R_{merge} values in the highest resolution bin (see Table 5.1).

The recorded intensities were converted to amplitudes (while setting aside 5% of reflections as test set reflections in order to calculate an R_{free}) with the program SCALEPACK2MTZ from the CCP4i suite (Bailey, 1994). The number of VH#576/LMO2 Δ N7 Δ C11 molecules per asymmetric unit was determined to be 2 with a solvent content of 59.29%, using a Matthews volume of 3.02 Å³/Dalton of protein (Matthews, 1968).

A



B

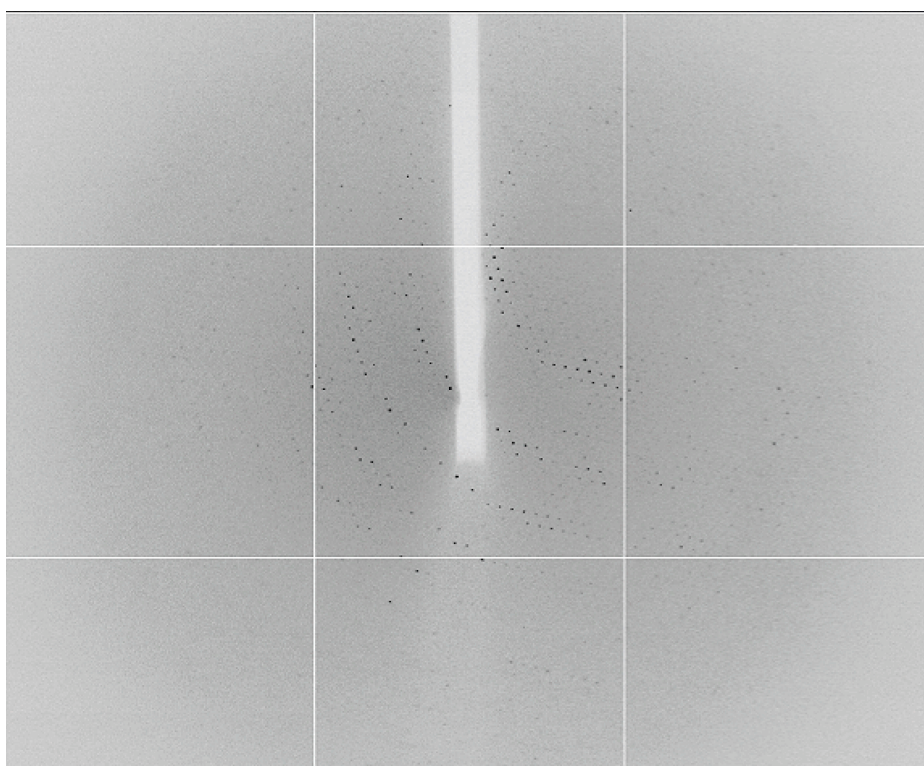


Figure 5.7: Diffraction pattern collected for a VH#576/LMO2 Δ N7 Δ C11 crystal

A: The crystal diffracted was grown in 0.8M Ammonium sulphate, 0.1M MES monohydrate pH 6 (diluted with H₂O to 88%) and additive 1, 6 Hexanediol with a drop ratio of 200nl of protein to 100nl of reagent. The crystal was loaded onto a fiber loop and mounted onto a goniometer head, which was adjusted to center the crystal in the X-ray beam and to allow rotation of the crystal while maintaining centering. The crystal was held in a stream of cold nitrogen gas.

B: The crystal was rotated through 360° and images collected as a series of 1.0° oscillations. This image was collected at an angle of $\psi=90^\circ$.

5.4.2 Structure Solution

The structure of VH#576/LMO2 Δ N7 Δ C11 was solved by a combination of molecular replacement and SAD phasing. The phases for VH#576 were determined by molecular replacement using the program PHASER from the CCP4i suite (Bailey, 1994). Anti-Ras FV heavy chain was used as a search model (PDB 2UZI). This search identified two VH molecules in the unit cell however one was deleted due to a lack of agreement with the electron density map.

Using the anomalous signal originating at the Zn peak anomalous wavelength, the program HKL2MAP (SHELX suite) (Pape and Schneider, 2004) was able to locate and refine the positions of the eight Zn atoms. The eight Zn atoms and the one previously located VH#576 molecule were imported into the program PHENIX autosol (Adams et al., 2002) which recalculated phases to produce an improved electron density map. The molecular replacement program MOLREP (Vagin and Teplyakov, 1997) was used to search within this map for two LIM1 domains (using the isolated LIM1 from FLINC2 as a search model) and successfully positioned the two molecules. PHENIX autosol was used again to recalculate phases, this time with the two additional LIM1 molecules, and also to produce an improved electron density map. The VH#576 and LIM1 for one VH#576/LMO2 Δ N7 Δ C11 complex in the asymmetric unit were superimposed onto the second LIM1 found in order to define the position of the remaining VH#576 and LIM1 for the second VH#576/LMO2 Δ N7 Δ C11 molecule. Again PHENIX autosol was used to improve the phases and calculate a new electron density map.

A series of molecular replacement experiments to search for the two missing LIM2 domains, either within the map (using MOLREP) or around the positions of the known, fixed molecules (using PHASER) failed. Therefore one molecule of LIM2 (from FLINC2) was manually positioned in the electron density map using the determined Zn positions as anchor points. Rigid body refinement performed by the program REFMAC5 (CCP4i suite) (Bailey, 1994, Murshudov et al., 1997) was used to refine the position of LIM2. Finally, one molecule each of LIM1, LIM2 and VH#576 from one VH#576/LMO2 Δ N7 Δ C11 complex were superimposed onto LIM1 and VH#576 of the second complex to define the position of the second LIM2. At this point all the components of the asymmetric unit were defined and rigid body refinement, followed by two fold-noncrystallographic symmetry (NCS) restrained refinement (REFMAC5, CCP4i), were used to refine the model.

5.4.2.1 Structure refinement

Once a model could be derived from the electron density map it was necessary to carry out structural refinement in order to adjust the structure to give the best possible fit to the crystallographic observations. The first stage in the refinement procedure was to rebuild the sections in the electron density map that were missing from the molecular replacement models such as the N-terminal of LMO2 (residues 9 to 26) and the CDR3 region of VH#576 (residues 101 to 109). This was done using the program COOT (Emsley and Cowtan, 2004), by adding terminal residues followed by real-space refinement of the added residue into the electron density map. Once a series of residues had been added, the “regularise zone” function in COOT was used to tighten the geometry of the created bonds, angles, planes and non-bonded contacts. The newly created model was then inputted into the REFMAC5 refinement program for cycles of refinement followed by more rebuilding.

The difference map ($F_{\text{obs}} - F_{\text{calc}}$), which is directly connected to the least-squares optimisation function and which has positive density where the model should have more electrons and negative density where the model should have fewer electrons, was continually visualised throughout rebuilding (Figure 5.8). Other indicators of accuracy (Table 5.1) including the Ramachandran plot, the root mean squared deviation for bond lengths and angles, the R_{factor} , the R_{free} and the average B factors were closely monitored during refinement. The Ramachandran plot shows the distribution of conformational angles (ϕ , ψ) along the protein backbone. C-N-C $_{\alpha}$ -C defines the torsion angle ϕ , whilst N-C $_{\alpha}$ -C-N defines ψ . The plot gives an indication of whether conformational angles are preferred, allowed or incorrect. Glycines have a different plot, as they are more flexible. Values for root mean squared deviation for bond lengths and bond angles reflect the deviation from an ideal geometry. The R_{factor} is a measure of how well the modified electron density fits to the observed structural amplitudes. The R_{free} or free R_{factor} is a residual function calculated during structure refinement in the same way as the conventional R_{factor} , but applied to a small subset of reflections that are not used in the refinement of the structural model. The purpose of this is to monitor the progress of refinement and to check that the R_{factor} is not being artificially reduced by the introduction of too many parameters. The values for R_{factor} and R_{free} following the last cycle of the current round of refinement are 29% and 33% respectively. The B-factor, or temperature factor, which describes the degree to which the electron density is spread out for each atom or group of atoms, is an indication of the dynamic mobility of an atom and can potentially highlight errors in model building.

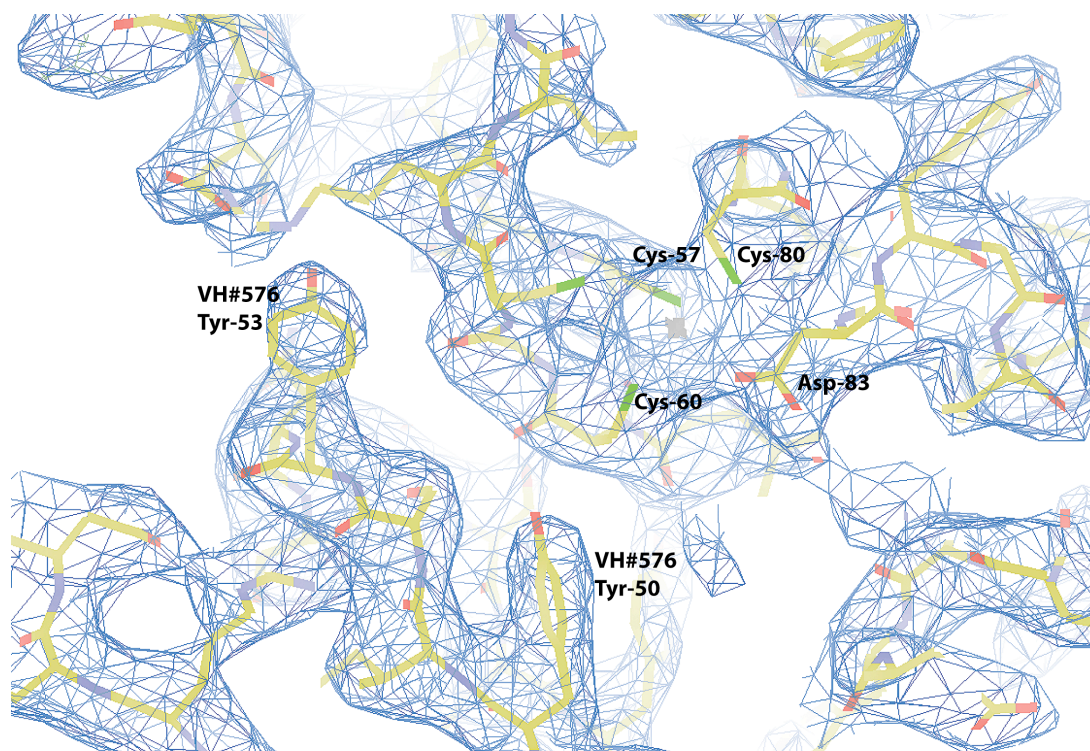


Figure 5.8: $2F_o-F_c$ electron density map of VH#576/LMO2 Δ N7 Δ C11

The $2F_o-F_c$ electron density map (blue) was contoured at 1.0σ . The map is shown for the LIM1 domain of LMO2 Δ N7 Δ C11, specifically zinc binding LIM finger two. The zinc ion, represented by a grey cross, is tetra-coordinated by LMO2 residues Cys 57, Cys 60, Cys 80, Asp 83. VH#576, CDR2 is also shown to the left of LMO2. The low resolution of the map (3.3 \AA) makes positioning of side chains difficult.

Table 5.1: Data Collection and refinement statistics for VH#576/LMO2ΔN7ΔC11 structure

Data collection statistics:	
Space group	P6
Number of molecules in the asymmetric unit	2
Unit cell Dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å) <i>a</i> , <i>β</i> , <i>γ</i> (°)	<i>a</i> = 124.5 <i>b</i> = 124.5 <i>c</i> = 81.7 <i>α</i> = 90 <i>β</i> = 90 <i>γ</i> = 120
Observations	237175
Unique reflections	10859 (1069)
Completeness (%)	100 (100)
I/σ(I)	24.2 (4.9)
R _{merge} ^a (%)	24.2 (98.9)
Refinement statistics:	
Resolution range (Å)	30 - 3.3 (3.42 – 3.30)
Number of reflections (working/test)	10333/496
R _{factor} ^b (%) / R _{free} ^c (%)	28.53/33.19
Number of atoms	4146
Rms Δ bond length (Å)	0.0093
Rms Δ bond angle (°)	1.53
Mean B-factor (Å ²): main chain/side chains and waters	43.19/43.29
Ramachandran Plot statistics:	
Residues in most favoured regions (%)	70.62
Residues in additionally allowed regions (%)	12.65
Outliers	16.73

The numbers in parentheses refer to the last (highest) resolution shell.

^a $R_{\text{merge}} = \frac{\sum_h \sum_i |I_i(h) - \langle I(h) \rangle|}{\sum_h \sum_i \langle I_i(h) \rangle}$, where $I_i(h)$ is the *i*th measurement and $\langle I(h) \rangle$ is the weighted mean of all measurements of $I_i(h)$.

^b R_{factor} and ^c $R_{\text{free}} = \frac{\sum_h ||F(h)_{\text{obs}}| - |F(h)_{\text{calc}}||}{\sum_h |F(h)_{\text{obs}}|}$ for reflection in the working set and test set respectively.

5.4.3 Crystal structure of VH#576/LMO2 Δ N7 Δ C11

The structure of VH#576/LMO2 Δ N7 Δ C11 shows VH#576 displays the common immunoglobulin fold, composed of two β sheets made up of nine anti-parallel β strands (Figure 5.9). Hydrogen bonds between adjacent strands stabilise the VH#576 structure. The LIM1 domain possesses the characteristic LIM domain structure of four β -strands and this is followed by a short α helix. Each zinc atom is positioned between a pair of anti-parallel β strands. Truncation of the eleven C-terminal residues of LMO2 has removed the C-terminal α helix of LMO2, which is present in the FLINC2 structure.

All four zinc ions are present in the structure; each zinc is coordinated in a tetrahedral geometry by the side chains of three cysteine residues and one histidine or aspartic acid residue. Specifically the zinc is coordinated by, the sulphur of cysteines, the nitrogen of the histidine or the oxygen of the aspartic acid. Zinc ions are important for stabilising the structure of LMO2, a role analogous to that of disulfide bonds, which are important in the folding, and stability of some proteins. Furthermore, zinc ions are stable in a cellular environment due to the lack of redox activity (McCall et al., 2000).

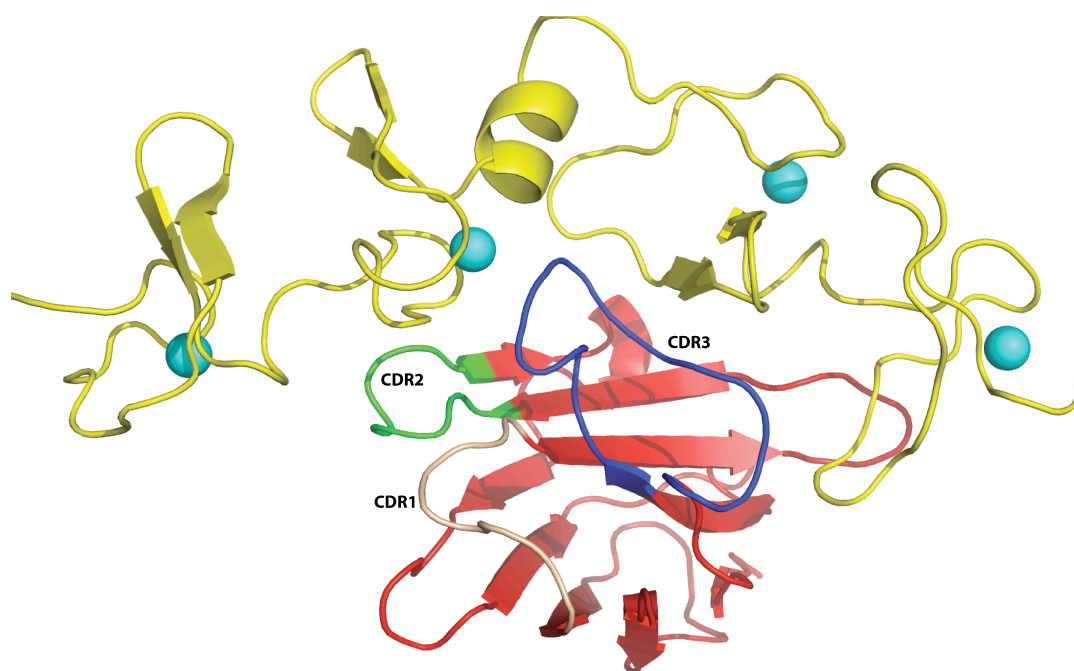


Figure 5.9: Crystal structure of VH#576/LMO2 Δ N7 Δ C11

The overall shape of VH#576/LMO2 Δ N7 Δ C11 is displayed as a ribbon diagram, where the VH#576 is shown in red, LMO2 in yellow and the zinc spheres in cyan. The architecture of VH#576 is that of the common immunoglobulin fold. CDR one, two and three of VH#576 are coloured in wheat, green and blue respectively.

VH#576/LMO2 Δ N7 Δ C11 forms a dimer in the asymmetric unit, where the dimerisation interface is solely between two LMO2 molecules (Figure 5.10). The N terminal tail of LMO2 (residue 7-26) can be seen to interact with the region comprising residues 40 to 70 of LIM1 in the neighbouring LMO2 molecule in the asymmetric unit. This interaction is intriguingly similar to the one seen between the LID domain of LDB1 and LMO2 in the FLINC2 structure; the N-terminal tail of LIM1 appears to mimic the LID domain of LDB1. Size exclusion chromatography and mass spectrometry analysis suggest VH#576/LMO2 Δ N7 Δ C11 is a monomer in solution and therefore it may be that the crystallographic dimer is a result of crystal packing rather than a reflection of a true dimer. Alternatively, dimer formation may be concentration dependent.

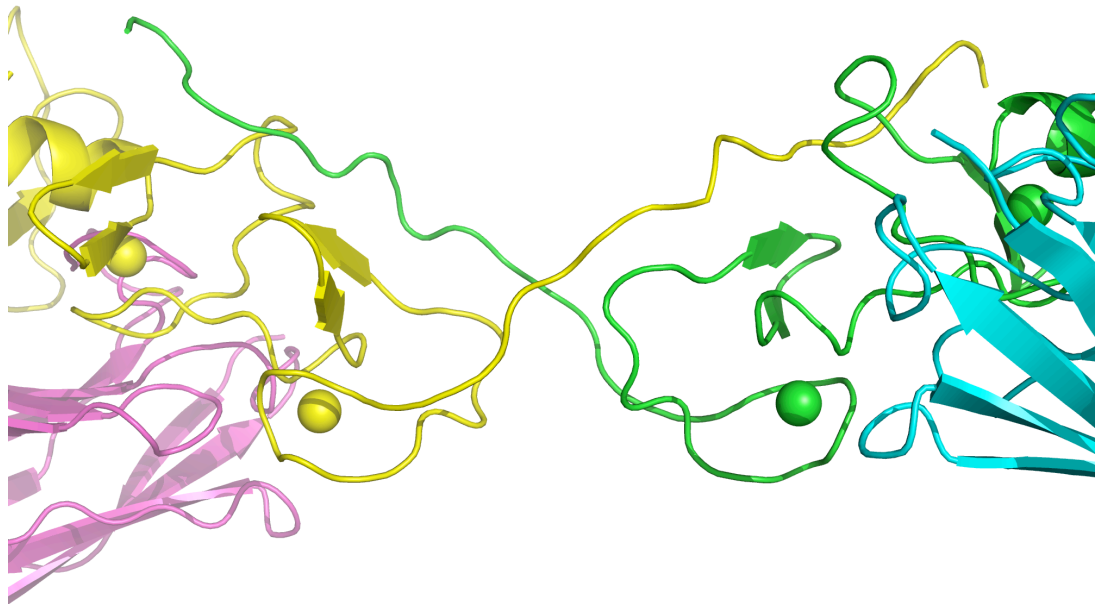


Figure 5.10: LMO2 forms a dimer in the asymmetric unit

The N-terminus of two LMO2 molecules, shown in yellow and green with respective VH#576 binding partners shown in pink and cyan. The long N-terminal tail of LMO2 interacts with LIM1 domain of another LMO2 molecule in the asymmetric unit to form a dimer.

The superimposition of the FLINC2 structure (Omari et al., 2010, manuscript in preparation) onto the LMO2 of the VH#576/LMO2 Δ N7 Δ C11, highlights major conformational differences, specifically in the relative positioning and angle between the two LIM domains. The conformation of LMO2 adopted in the VH#576/LMO2 Δ N7 Δ C11 structure may be influenced by the interactions with the VH#576 molecule. However, differences in crystal packing between the two structures could also give rise to differences in the conformation of the proteins.

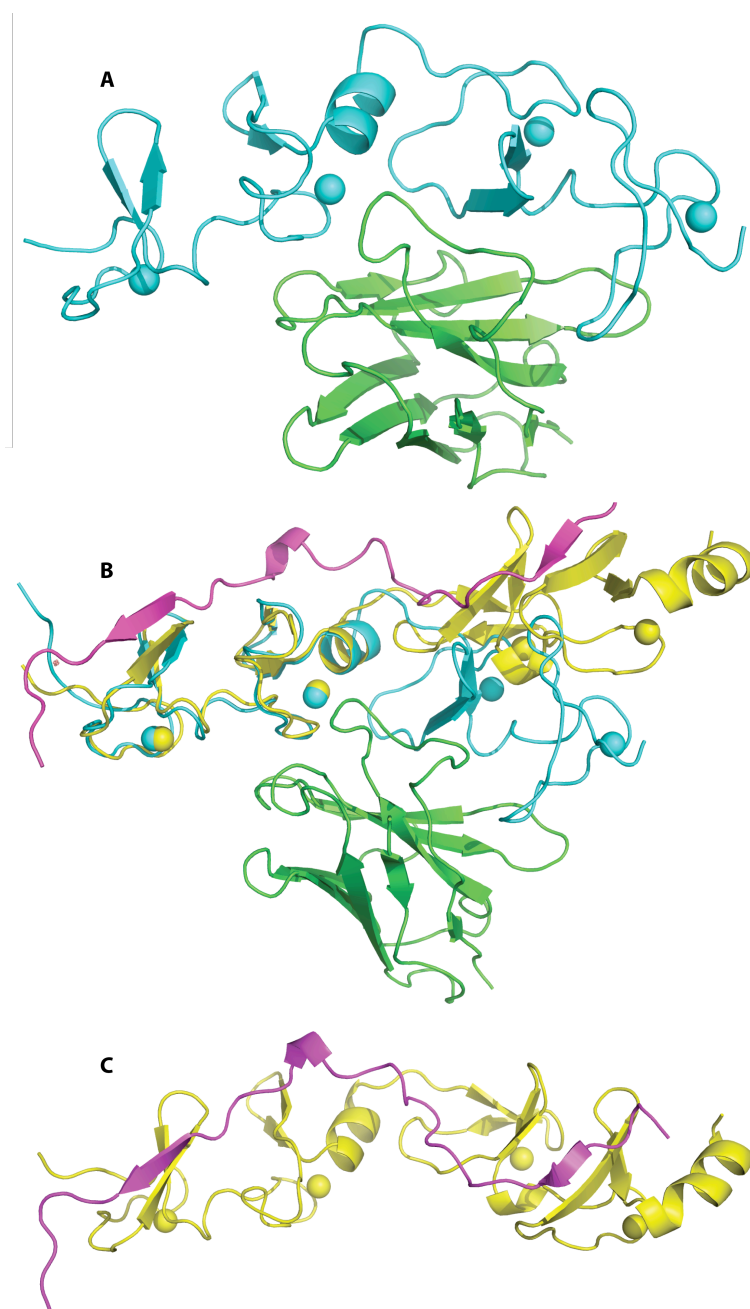


Figure 5.11: Superimposition of the FLINC2 and VH#576/LMO2 Δ N7 Δ C11 structures

A - The VH#576/LMO2 Δ N7 Δ C11 structure; LMO2 is coloured in cyan, VH#576 in green and the zinc ions in cyan.

B - The structures of FLINC2 (Omari et al., 2010, manuscript in preparation) and VH#576/LMO2 Δ N7 Δ C11 were superimposed with COOT, using the LIM1 domains as anchoring points.

C - The FLINC2 structure; LDB1_LID is coloured in magenta, LMO2 in yellow and the zinc ions in yellow.

5.4.4 Analysis of the interaction between VH#576 and LMO2 Δ N7 Δ C11

Atomic interaction between VH#576 and LMO2 were identified using Contacts of structural units (CSU) software (Sobolev et al., 1999). Any type of non-bonded atomic interactions, i.e. hydrogen bonds, aromatic-aromatic, hydrophobic-hydrophobic and polar contacts, were considered. Table 5.2 and Table 5.3 compile all the observed atomic interactions between VH#576 and LMO2 upon analysis of the solved structure. A hydrogen bonding network occurs between the β -sheet of LIM2 (LIM finger three) and CDR3 of VH#576, residues Leu 106 and Thr 107. Other interactions include electrostatic networks and hydrophobic contacts (Figure 5.13). A region of hydrophobic contacts lies at the former VH-VL interface of VH#576 that spans residues Val 37 to Tyr 50. The hydrophobic interactions occur between residues Leu 45 and Leu 117 and Trp 47 and Met 106 of VH#576 and LMO2, respectively (Figure 5.14). These hydrophobic framework interactions may explain why the orientation of VH#576 with respect to LMO2 appears to be distorted. The algorithm also highlighted the atomic interactions mediated by Trp 110 as shown in Table 5.3. This residue was not mutated during the alanine/glycine scan.

Table 5.2: Atomic contacts between VH#576 and LMO2 Δ N7 Δ C11

Antibody region	VH#576 residue	LMO2 residue	Distance (Å)	Contact surface area (Å ²)	Type of interaction
Framework	Leu45	Leu117	3.2	37.1	Hydrophobic-hydrophobic Polar contacts
Framework	Trp 47	Met 106	3.1	64.8	Hydrogen bond Hydrophobic-hydrophobic Polar contacts
Framework	Tyr 50	Cys 60	3.5	27.4	Hydrophobic-hydrophobic
CDR2	Ser 52	Cys 60	3.2	26.5	Polar contacts
CDR2	Asn 54	Asp 53	3.7	21.8	Polar contacts
CDR3	Leu 104	Arg 82	3.7	40.8	Hydrophobic-hydrophobic Polar contacts
CDR3	Glu 105	Asp 83	2.7	51.8	Hydrophobic-hydrophobic
		Arg 86	2.8	50.5	Polar contact (salt bridge)
CDR3	Leu 106	Thr 107	2.8	64.2	Hydrogen bond Hydrophobic-hydrophobic Polar contacts
CDR3	Thr 107	Arg 109	2.7	33.5	Hydrogen bond Polar contacts
CDR3	Asp 109	Tyr 135	2.7	43.1	Hydrophobic-hydrophobic

Table 5.3: Further atomic contacts between VH#576 and LMO2ΔN7ΔC11

Antibody region	VH#576 residue	LMO2 residue	Distance (Å)	Contact surface area (Å ²)	Type of interaction
CDR3	Trp 110	Phe 120	3.7	35.0	Aromatic-aromatic
		Cys 130	3.2	23.3	Hydrophobic-hydrophobic
		Gly 132	3.6	27.4	Hydrogen bond and polar contacts

1 10 20 30 40
 EVQLLESGGGLVQPGGSLRLSCAASGFSFSHSPMNVRQAPGKGLEWV

 50 60 70 80 90
SYISYNSSSIYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYC

 100 110 120
ARGLTESLELTADWFDYWGQTLVTVS

Figure 5.12: Important VH#576 residues for the interaction with LMO2

The amino acid sequence (single letter code) is shown for VH#576. The CDR regions are coloured in yellow. Critical residues for the interaction with LMO2, found at the protein-protein interface, are highlighted in red. Putative critical residues at the former VH/VL interface (coloured green) are highlighted in purple. Residues highlighted in blue were also found to be critical for the interaction, according to VH#576 mutagenesis data, however the crystal structure suggests they do not form atomic contacts with LMO2. This would imply the residues are important to maintain a specific conformation of VH#576.

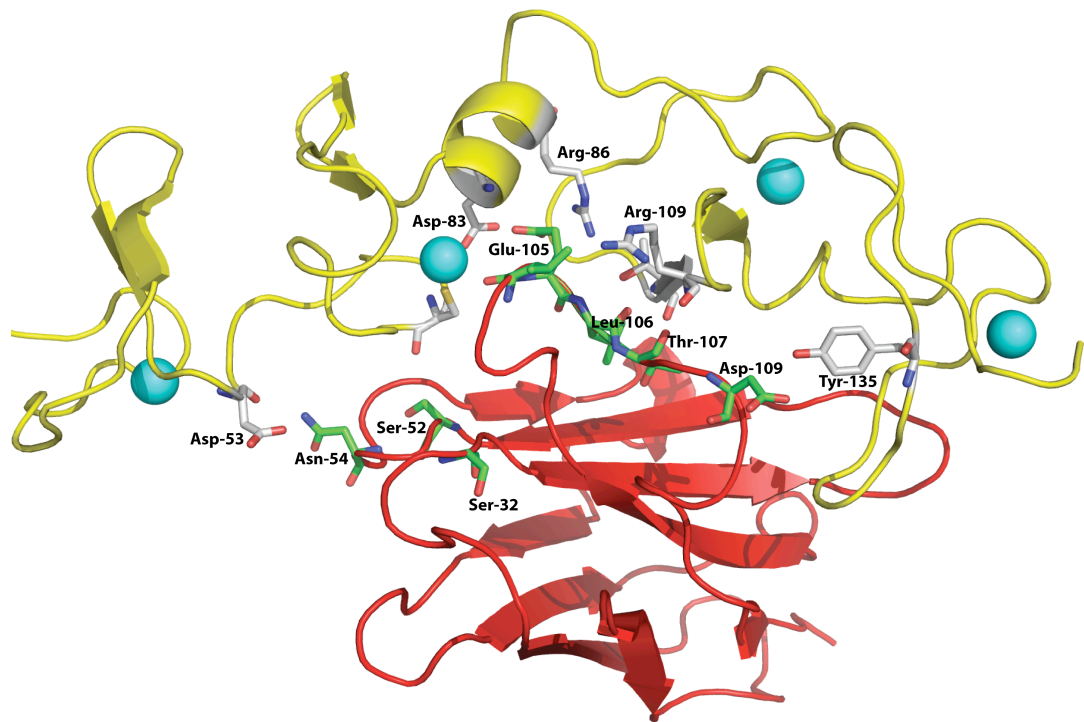


Figure 5.13: Key interactions between VH#576 and LMO2 Δ N7 Δ C11

Both VH#576 and LMO2 Δ N7 Δ C11 are displayed as ribbon diagrams and shown in red and yellow respectively. Key interacting residues are shown in ball-and-stick. The critical VH#576 binding site lays across LIM finger two and three of LMO2 Δ N7 Δ C11, LIM domain one and two respectively (the zinc can be seen as cyan spheres). CDR1 has no role in the interaction with LMO2 however, Ser 32 may contribute towards the overall conformation of VH#576. There are two critical residues shown in CDR2 (Ser 52 and Asn 54). Interactions can be seen at the α -helical region of LIM1 between residues Asp 83 and Arg 86 of LMO2 Δ N7 Δ C11 and Glu 105 of VH#576, CDR3. Other key interactions involving CDR3 residues are highlighted including that of Leu 106 and Thr 107 with the β -sheet of LIM2, finger three.

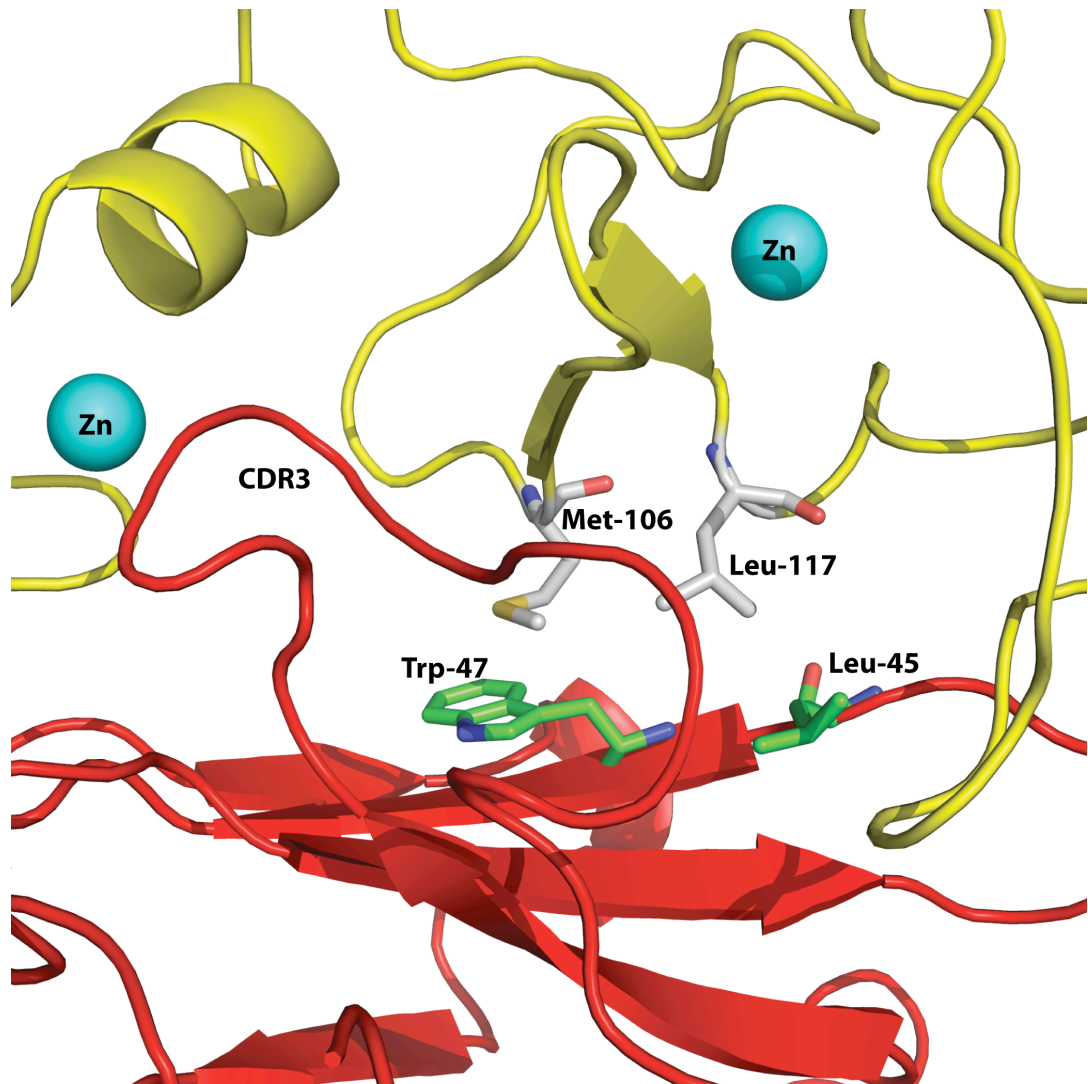


Figure 5.14: Important hydrophobic interactions between VH#576 and LMO2 Δ N7 Δ C11

LMO2 Δ N7 Δ C11 is shown in yellow and VH#576 in red. The anti parallel β -sheets of LMO2 Δ N7 Δ C11, LIM2 form part of LIM finger three. VH#576 residues, Trp 47 and Leu 45 at the former VH/VL interface are indicated as shown in ball-and-stick. Trp 47 and Leu 45 of VH#576 form hydrophobic interactions with Met 106 and Leu 117 of LMO2 Δ N7 Δ C11 respectively.

5.4.5 Interface area

The surface area of VH#576 which is critical for the interaction with LMO2 can be determined by calculating the absolute value of the difference between the solvent accessible surface area (SAS) in the bound state and the SAS in the unbound state. The calculation is shown in Table 5.4 along with the estimated size required for a small molecule to mimic the VH#576 interaction (Cheng et al., 2007).

Table 5.4: Surface area of VH#576 required for the interaction with LMO2 Δ N7 Δ C11

Surface area of VH#576 interacting region (\AA^2):		Represents a compound of molecular weight (Da)
Unbound	814	
Bound	302	
SAS (unbound) - SAS (bound) (buried surface area)	512	850

The surface area of VH#576 required for the interaction with LMO2 has been calculated as 512\AA^2 , which relates to a compound of molecular weight 850 Da (Table 5.4). As will be discussed later in this chapter, this is larger than the 500 Da classical Lipinski limit (Lipinski et al., 1997) of a typical drug like compound. This calculation also includes residues Leu 45 and Trp 47 that are not located in the CDR regions, however the structural data suggest they are critical for the interaction.

5.5 Discussion

This chapter presents the crystal structure of VH#576/LMO2 Δ N7 Δ C11. The structure corresponds strongly with the VH#576 mutagenesis data and has enabled the interactions between VH#576 and LMO2 Δ N7 Δ C11 to be defined.

Most of the VH#576 residues, shown to be critical by VH#576 mutagenesis studies, can be found at the interface of the crystal structure with the exception of Ser 32 (CDR1), Glu 102 (CDR3) and Tyr 113 (CDR3). Changes to these residues may alter the conformation of VH#576, enough to inhibit other vital interactions. These results highlight the effectiveness of alanine/glycine scanning to identify a subset of residues at the interface, which contribute to high affinity binding, and this region can be defined as the hotspot. Consequently, this

study represents alanine/glycine scanning as a key tool to focus and restrict the target area of the protein-protein interface for small molecule drug design.

Atomic contacts between VH#576 and LMO2 have been calculated using a computational geometry algorithm. Important interactions have been identified however, due to the low resolution of the crystal structure, some amino acid side chains may not have the correct orientation thus some interactions may not have been correctly identified. Some interactions predicted may be inaccurate or some may be missing due to incorrect geometry between the interacting residues.

5.5.1 A comparison between VH#576/LMO2 Δ N7 Δ C11 crystal structure and *in silico* structural model

Prior to the generation of X-ray crystallography data, VH#576 and LMO2 mutagenesis data were used to create an *in silico* structural model of the VH#576/LMO2 Δ N25 protein complex. Upon comparison of this model and the crystal structure errors in the *in silico* model have been identified such as the orientation of VH#576 with respect to LMO2. However, the CDR3 region is coordinating with the same region of LMO2 in both the structural model and crystal structure, albeit from a different angle. The *in silico* structural model is a structural prediction and there are many reasons for the differences between this model and the crystal structure. The *in silico* model is based on the VH#576 mutagenesis data which does not include analysis of the residues at the former VH/VL interface. The crystal structure shows these residues interact with LMO2 and thus this is one reason for the difference in orientation of the VH#576 and LMO2. The VH#576 mutagenesis data also suggests Ser 32, of CDR1 actively interacts with LMO2 however the crystal structure shows this is not the case. Ser 32 is actually important to maintain the conformation of VH#576. The allosteric bend of LMO2 seen in the crystal structure is not present in the *in silico* model. This illustrates a limitation of protein-protein docking algorithms, as these computational tools do not properly account for the intrinsic conformational flexibility of proteins.

5.5.2 Comparison with other LIM structures

A comparison between crystal structures VH#576/LMO2 Δ N7 Δ C11 and FLINC2 (Omari et al., 2010) shows a degree of flexibility in the LIM2 region of LMO2. Two possible explanations for this are considered here. Firstly, LMO2 interacts with many different proteins therefore it may be that LMO2 has intrinsically disordered regions to enable its interaction with different protein partners such as TAL1 and LDB1. If this is the case then LMO2 will fold with respect to the protein with which it is interacting. Secondly, it may be that the VH#576 interaction induces a conformational change in LMO2, which is different

from its native conformation. The hydrophobic interactions between residues located on the surface of the hydrophobic core of LMO2 and the former VH/VL interface may contribute to the change in the orientation of LMO2, LIM2.

Analysis of LMO4-LIM1:ldb1-LID complex (FLIN4) through NMR, ¹⁵N relaxation data, shows evidence of flexibility between the two zinc-binding modules. An overlay of FLIN4 and individual zinc binding modules shows a reduction in the root mean square deviation. This suggests the small loops between elements of defined secondary structure have increased mobility in comparison with the core regions of the domain (Deane et al., 2003).

The structures of other unbound LIM domains have been analysed and have shown different orientations and flexibility between the zinc binding fingers. Examples include the analysis of LIM2 of CRP2, a cysteine rich LIM-only protein. Comparison of ¹⁵N relaxation data for wild type CRP2-LIM2 and CRP2-LIM2_R122A (a mutation that disrupts zinc co-ordinating Cys 144) revealed a conformational change in the zinc binding site, alteration in the hydrophobic core and a subsequent change in the orientation of the two LIM fingers (Schuler et al., 2001). These differences were not confined to the site of mutation and shows that LIM domains are not rigid but their structure can be variable due to the breakage and formation of hydrogen bonds or electrostatic interactions. LIM domains characteristically interact with diverse protein-binding partners and the mode of conformational flexibility described by this group represents a possible mechanism of allosteric control mediating associations with the appropriate protein partners. It has also been proposed this mechanism allows for enthalpic compensation of entropy loss upon binding (Kontaxis et al., 1998).

5.5.3 Effect of VH#576 upon LMO2 activity within the cell

The FLINC2 structure (Omari et al., 2010, manuscript in preparation) may be perceived to be the more native conformation as it is bound to Ldb1_LID, a natural binding partner in the protein complex. This change in conformation, upon VH#576 interaction, may represent the mechanism of functional ablation. Results of a recent differentiation assay, (T. Tanaka, personal communication) has provided evidence that VH#576 interferes with LMO2 activity within the cell. The experiment can be briefly described as follows: Murine erythroleukaemia (MEL) cells were infected with retrovirus expressing GFP and either VH#576 or anti-LMO2 scFv (ALR3). Cells were sorted using FACS analysis and then treated with 4mM hexamethylene bisacetamide (HMBA). MEL cells undergo erythroid differentiation when treated with HMBA. After 4.5 days, the MEL cells were tested for erythroid differentiation by staining with 2,7-diaminofluorene (DAF), which sensitively stains haemoglobin blue. Approximately 15% of the MEL cells infected with retrovirus

expressing GFP and VH#576 differentiated into erythroid cells as oppose to 95% of uninfected MEL cells. This assay gave the same results for ALR3 (anti LMO2 scFv). A negative control of anti-RAS VH#6 was used to show that a VH antibody that does not bind to LMO2 cannot inhibit haemoglobinisation. This assay shows that erythroid cells expressing VH#576 do not undergo haemoglobinisation. This is a result of the interaction between VH#576 and LMO2. LMO2 is essential for haematopoietic development (Warren et al., 1994) and therefore the inhibition of LMO2 in MEL cell blocks their development.

Inhibition of LMO2 dependent leukaemia in a mouse T-cell tumorigenesis transplantation assay has been demonstrated for ALR3. *Rag1* null mice were injected with Thymoma T cells, which had been prepared from Lck-Lmo2 transgenic mice and infected with retrovirus expressing ALR3 and GFP. After 3 to 4 weeks neoplastic T cells were removed from the spleen for FACS analysis. Expression of the GFP after transplantation had decreased by approximately 70%, reflecting the inhibition of malignant mouse tumour growth by ALR3 (Nam et al., 2008). An equivalent transplantation assay replacing ALR3 for VH#576 could potentially demonstrate its ability to inhibit malignant mouse tumour growth.

The VH#576/LMO2 Δ N7 Δ C11 crystal structure described shows the binding of VH#576 locks LMO2 in a specific conformation. A possible mechanism for the inhibition of LMO2 activity by VH#576, as shown by the erythroid differentiation assay, is one of conformational change. It may be that the conformation of LMO2 when bound to VH#576 is changed to such a degree that the protein can no longer interact with its natural partners such as LDB1 and TAL1. Therefore design of a small molecule that could perfectly mimic VH#576 and stabilise LMO2 in this new conformation could potentially inhibit its protein interactions within the DNA binding protein complex (LMO2, LDB1, TAL1, and E47/E12), blocking its activity as a transcription factor.

5.6 Conclusion

In conclusion, the crystal structure of VH#576/LMO2 Δ N7 Δ C11 has been solved to a medium resolution of 3.3Å and concurs well with both the LMO2 and VH#576 mutagenesis data. Differences in the structure of LMO2 when bound to Ldb1_LID or VH#576 have been revealed upon comparison. This is most likely explained by the intrinsic flexibility of LIM domains mediating associations with the appropriate protein partners. The favourable interactions between VH#576 and LMO2 drive and stabilise LMO2 conformational changes. Thus, interaction with VH#576 changes the conformation of LMO2, particularly in the LIM2 region. Development of a small molecule, which could perfectly mimic VH#576

by altering and stabilising the conformation of LMO2 in the same way, could have potential as a therapeutic by blocking its activity within the cell. However, the resolution of the solved crystal structure is not optimal for structure-based drug design. Further refinement of the structure or collection of diffraction data at an improved resolution, will eventually lead to a superior structure, more suited for drug design. Furthermore, the structural information of the macro complex, i.e. LMO2, LDB1, TAL1, and E47/E12, is not yet available. Having this information would greatly aid the design of small molecule inhibitors, providing information on which area of LMO2 to target and the potential to design a dual therapy to target more than one interaction within the complex.

6 General discussion and future work

6.1 General discussion

Work detailed within this thesis has led to the crystal structure of LMO2 in complex with an intracellular antibody single domain, termed VH#576. It is the first time LMO2 has been crystallised in complex with a macrodrug. The crystal structure of the complex formed by LMO2 and VH#576 provides precise information about the nature of the interaction interface and this will be fundamental for structure based drug design. The structure was solved to a resolution of 3.3Å and reveals VH#576 binds across the two LIM domains of LMO2.

Comparison of the VH#576/LMO2ΔN7ΔC11 structure with FLINC2 (chapter five) demonstrates the high degree of flexibility between the two LIM domains represented by the differences in relative positioning and angle between the two domains. This demonstrates the flexibility of the loops between the defined secondary structure of LMO2 and may reflect a mechanism of allosteric control mediating associations with a broad range of proteins.

The VH#576/LMO2ΔN7ΔC11 crystal structure was analysed with regards to the crystal packing and an interaction was identified between the long N-terminal tail (residues 7 to 26) of LMO2 and LIM1 of a second LMO2 molecule. This interaction is intriguingly similar to the one seen between the LID domain of LDB1 and LMO2 in the FLINC2 structure (Omari et al., 2010, manuscript in preparation); the N-terminal tail of LIM1 appears to mimic the LID domain of LDB1. There is evidence to suggest that LMO2 can form a weak homodimer (Nam et al., 2008, Sanchezgarcia et al., 1995). Still, VH#576/LMO2ΔN7ΔC11 is certainly a monomeric complex prior to crystallography as shown by the size exclusion chromatography and mass spectrometry data presented in chapter three. The interaction seen between LMO2 molecules may be an artifact of crystal packing or it may reflect a true dimerisation pathway.

6.1.1 Medium resolution X-ray structure of VH#576/LMO2ΔN7ΔC11

The crystal structure of VH#576/LMO2ΔN7ΔC11 was solved to a resolution of 3.3Å due to the low resolution of the diffraction data collected and this is reflected in the electron density map. At a resolution of 3.3Å, the α-helices, β-sheets and main chain can clearly be identified. Most aromatic side-chains are clearly visible however, density for some of the smaller side chains is missing and the rotamers of some side chains are not clear. At this resolution, ordered water molecules cannot be placed. The crystals tested for diffraction had a high solvent content and this partly attributes to the low resolution data as the high level of

disordered solvent leads to a decrease in X-ray intensities with increasing diffraction angle. There are two proteins presents in the VH#576/LMO2 Δ N7 Δ C11 complex and this may confer a degree of flexibility and as such large atomic motilities may contribute to the low resolution diffraction data. High atomic motilities not only limit the resolution of the data that can be collected but also decrease the peak densities in the Fourier maps to such an degree that some regions of the map are difficult to fit (Jensen, 1997). Therefore diffraction data of a higher resolution may or may not be possible to collect but would decrease the errors in atom position and hence increase the accuracy and figure of merit of the model; an aspect which is critical for structure-based drug design.

Options for improving the crystal packing and hence diffraction capacity include optimisation of crystallisation conditions, which were bypassed in the initial stages. Alternatively, the kinetics of evaporation and hence crystallisation can be influenced by the use of an oil barrier over the vapour diffusion reservoir to slow down the rate of evaporation and potentially improve the order of the crystals grown (Chayen, 1997). Then again, perhaps the most likely solution, would be to examine the crystal contacts and packing of the molecules in the current crystal and design a construct with a higher chance of being more ordered and thus improve the X-ray diffraction properties (MacElrevey et al., 2007). Interestingly, upon initial inspection of the crystal contacts it appears residues 140 to 147 of LMO2, LIM finger four (right side) have very few intermolecular contacts and this may highlight flexibility within this region. The B-factors for residues within this region are also high again suggesting a high degree of mobility, which could have a negative impact on the crystal packing. Creating a construct with further C-terminal truncations would almost certainly interfere with the ability of LIM finger four to bind zinc. However, for the purpose of structure based drug design we are solely interested in the protein-protein interface between LMO2 and VH#576 and aim for an atomic resolution of 2.7Å or better.

6.1.2 The LMO2 DNA binding protein complex

Lmo2 is part of a DNA binding complex, originally identified in erythroid cells, comprising Tal1, Ldb1, E2a and Gata1 (Wadman et al., 1997). A distinct complex, which lacks Gata1, has been identified in T-cells from *Lmo2* transgenic mice (Grutz et al., 1998b). Ectopic expression of LMO2 has been observed in T-cell tumour cells from patients with T-cell acute lymphoblastic leukaemia (T-ALL) due to specific chromosomal translocations (Boehm et al., 1991). In this section, theories of possible mechanisms for the onset of T-ALL, as a result of the aberrant expression of LMO2, will be discussed. The prospects of designing a small molecule, which has the capacity to disrupt this LMO2 complex, will also be considered.

The increase in levels of LMO2 in T-cells, by enforced expression, may result in aberrant complex formation, which acts to sequester one or more proteins in the complex preventing their normal function, one such candidate is LMO4. There is no requirement for LMO2 in T-cells (McCormack et al., 2003), however LMO4 is highly expressed in T-cell precursors. Correspondingly, the displacement of LMO4, from LDB1, by LMO2 may contribute to the development of T-ALL although, LMO4 null mice have apparently normal populations of T-cells (Grutz et al., 1998a). The affinity of LMO2 for LDB1 has been shown to be lower than the affinity of LMO4 therefore an excess of LMO2 would be required to force binding equilibria to favour the formation of LMO2:LDB1 (Ryan et al., 2006). LIM1 of LMO2 has been shown to be more important than LIM2 for interaction with LDB1_LID. Based on point mutations analysed by yeast two-hybrid analysis, Ryan *et al.* demonstrate that residue I322 of LDB1 is an important residue for the interaction with LMO2-LIM1 and LMO4-LIM1 but is only critical for the interaction with LMO2-LIM1. Analysis of the NMR solution structure of LMO2-LIM1:LDB1_LID shows the side chain of this residue is buried in a hydrophobic pocket between the two LIM fingers of LIM1 and this can also be seen from the FLINC2 structure (Omari et al., 2010, manuscript in preparation). Thus, a small molecule, which has the ability to mimic I322, and surrounding residues, would have potential to specifically inhibit the interaction between LDB1 and LMO2 but not LMO4.

Aberrant expression, of LMO2, may play a role in the disruption of normal E2A function, contributing to the molecular pathway of T-ALL (Larson et al., 1996). Multiple competing equilibria experiments were used to characterise the assembly of the five component complex containing TAL1, LMO2, LDB1, E2A (E12 or E47) and DNA (Ryan et al., 2008). Due to a problem of protein solubility, TAL1_{bHLH} and E12_{bHLH} domains were purified as appose to full length proteins and LMO2 was purified as a chimera with the LID domain of LDB1. The main findings of Ryan *et al.* can be summarised in three points. Firstly, TAL1_{bHLH} and E12_{bHLH} preferentially form heterodimers rather than homodimers. Secondly, LMO2 has the capacity to interact with Tal1_{bHLH} however, preferentially binds TAL1_{bHLH}/E12_{bHLH} heterodimer. Finally, E12_{bHLH} is required for the complex to bind DNA. These results suggest the aberrant expression of TAL1, in T-cells, is likely to sequester E2A (E12/E47) protein and negatively affect E2A mediated gene expression and thus lymphocyte development as E2A activity is critical for this process (Bain et al., 1997). This is likely to be favoured in cells expressing both TAL1 and LMO2 as LMO2 has a higher affinity for the TAL1/E12 heterodimer. Hence, LMO2 promotes the formation of TAL1/E12 heterodimers (Ryan et al., 2008). As a result of these findings, Ryan *et al.* proposed that development of a small molecule to inhibit a single interaction, for example between LMO2 and TAL1 would have a limited effect on restoring E2A activity and that

two inhibitors (targeting LMO2-TAL1 and TAL1-E2A interactions) would be required as an effective treatment for T-ALL. However, there are likely to be additional LMO2-only effects contributing to T-ALL progression.

It is likely that LMO2 can interact with different sets of proteins to regulate the haematopoietic pathway (Yamada et al., 1998) and therefore, it is also likely that analogous interactions may ensue after ectopic LMO2 expression in T-cells. As LMO2 is not normally expressed in T-cells it may have affinity for proteins it does not normally interact with. These LMO2 interactions lead to the positive or negative regulation of gene transcription, which ultimately leads to a block in T-cell differentiation.

6.1.3 VH#576 has the potential to inhibit the activity of LMO2

As discussed previously (chapter five), the results of a recent erythroid differentiation assay demonstrated the ability of VH#576 to inhibit Lmo2, within the cell (T. Tanaka, personal communication). The VH#576/LMO2 Δ N7 Δ C11 crystal structure shows LMO2 in a conformation, which is different to that of LMO2 when bound to LID_LDB1 (FLINC2 PDB). This distinct LMO2 conformation may represent a biological mechanism for the inhibition of its downstream regulation of gene transcription. According to data, from a mammalian two-hybrid competition assay, VH#576 cannot compete with LDB1 for LMO2 binding, however LDB1 can compete with VH#576 (S. Waters, personal communication). Therefore inhibition of LMO2, as shown by the erythroid differentiation assay, must take place prior to formation of the complex.

The small size of LMO proteins suggests they should be capable of freely moving between the cytoplasm and the nucleus however they are found predominantly in the nucleus. The mechanism for nuclear retention is unclear as LMO proteins lack a nuclear localisation signal. LDB1, on the other hand, is a nuclear protein and contains a nuclear localisation sequence. Experimental evidence has suggested that one function of LDB1 is to maintain LMO proteins in the nucleus (Kenny et al., 1998). The interaction between VH#576 and LMO2 may occur in the cytoplasm post protein synthesis at the ribosome. This would prevent the interaction between LMO2 and LDB1 and as such LMO2 would remain in the cytoplasm, blocking its activity as a nuclear transcription factor.

TAL1 (36.5 KDa) is also predominantly a nuclear protein (Bernard et al., 1995). The interaction between Lmo2 and Tal1 seems to have a synergistic effect on tumour formation as well as on T-cell differentiation (Larson et al., 1996). Therefore, in terms of therapeutic design, this is an important interaction to target. Mammalian two-hybrid data suggests

TAL1 cannot compete with VH#576 for LMO2 binding and vice versa (S. Waters, personal communication) suggesting TAL1 and VH#576 interact with different regions of LMO2. Determination of the exact effect of the interaction between VH#576 and LMO2, upon formation of the DNA binding complex, requires further investigation using methods such as Electromobility shift assays. This would also provide information on the effect of VH#576 interaction on protein-nucleic acid interactions such as the interaction between E2A/TAL1 and DNA (E box motif). The pathogenic role of LMO2 is likely to be dependent on its ability to mediate protein interactions with DNA and this further highlights the importance of this experiment.

Previous data suggests other macrodrugs (e.g. anti-Lmo2 scFv termed ALR3), identified by intracellular antibody capture, bind to the same region of LMO2 as VH#576; the mid region composed of an α -helix from LIM1 followed by a loop and then a pair of β -sheets from LIM2. Therefore, one conclusion is that this region of LMO2 is particularly efficient at binding proteins, macrodrugs and perhaps even small molecules. Furthermore, this may be the only structured region of free LMO2, the loops either side of this interaction site may display intrinsic disorder until interaction with another protein. Moreover, ALR3 has been shown to inhibit Lmo2 in a mouse T-cell tumour transplantation assay by preventing Lmo2-dependent T-cell neoplasia (Appert et al., 2009, Nam et al., 2008). Therefore an experiment to test if VH#576 has the capacity to inhibit LMO2 dependent leukaemia in a mouse T-cell tumorigenesis transplantation assay would be key to determining its potential as a macrodrug.

6.1.4 VH#576 and small molecule mimetics

Monoclonal antibodies have interchain disulfide bonds which means they are almost always restricted to extracellular target antigens. VH#576 is a single domain intracellular antibody (iDab). iDabs were developed based on their ability to fold and interact with antigen in the reducing environment of the cell (Tanaka et al., 2003). In addition, VH#576 specifically binds LMO2 with nanomolar affinity and therefore has potential as a T-ALL therapeutic. The challenge associated with the use of antibody domains as therapeutics is delivery; the internalisation or expression inside the target cells. Options for delivery of VH#576 include protein transduction or virally mediated induction of antibody fragments into bone marrow cells *ex vivo* in association with transplantation in leukaemia patients (Lobato and Rabbitts, 2003). Nevertheless, further advances in delivery methods are required before clinical use is a possibility. Small molecules have several advantages over antibody single domains such as cell permeability, relative ease of manufacture, higher metabolic stability, lower cost and oral delivery (Lipinski et al., 1997).

The structure based drug design approach detailed within this thesis can be described in four stages of which progress has been made for the first three. Following identification of an anti-LMO2 antibody single domain, VH#576, the crystal structure of VH#576/LMO2 Δ N7 Δ C11 was solved. This allowed identification of a target region for LMO2 and the VH#576/LMO2 interface to be characterised. This was followed by identification of a VH#576 hotspot (a group of residues which make a significant energetic contribution to the interaction) thus reducing the target area for small molecule drug design. All this information can be utilised in stage four; design of a pharmacophore template to screen libraries of small molecules *in silico*.

The key VH#576 binding residues identified are mainly within CDR3 making it the most interactive of the hyper-variable loops. The crystal structure supports the mutagenesis data and shows the CDR3 residues interact with residues 82 to 135 located across LIM finger two (LIM1), three and four (LIM2) of LMO2 (Figure 6.1). The crystal structure and VH#576 mutagenesis data also demonstrate a critical interaction between VH#576 CDR2 residues (Ser 52 and Asn 54) and residues of LIM finger two (Cys 60, Asp 53). A full alanine/glycine scan of VH#576 residues 37 to 50, the former VH/VL interface, would identify precise residues that are critical for the high affinity binding. If residues in this region are also vital, the surface area of VH#576 required for the interaction with LMO2 would be 512Å², which relates to a compound of molecular weight 850 Da (Cheng et al., 2007) (Figure 6.1). This is larger than the typical 500 Da Lipinski limit (Lipinski et al., 1997). However, there are examples of larger compounds that have entered clinical trials such as ABT-263, a 974 Da small molecule inhibitor of anti-apoptotic proteins Bcl-2, Bcl-XL and Bcl-w. The small molecule acts to enhance the effect of cell death signals as shown by analysis of tumour regression in mice; a significant increase in the number of caspase three positive tumour cells was noted 24 hours post treatment (Shoemaker et al., 2008). The molecule (ABT-263) has successfully entered phase I/II clinical trials with acceptable oral bioavailability despite its large molecular weight (Park et al., 2008).

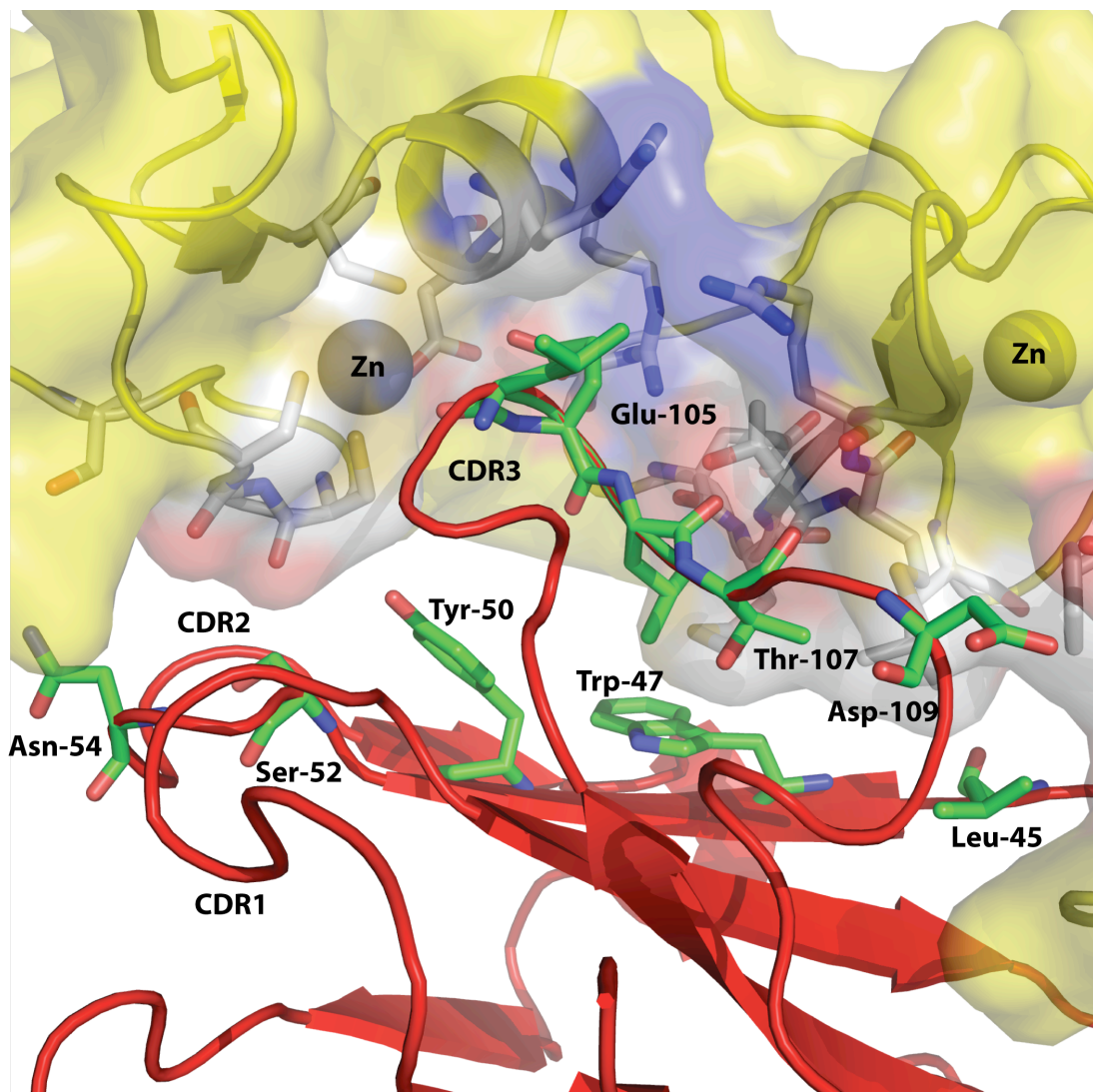


Figure 6.1: Key interaction region between LMO2 and VH#576

VH#576 is represented in red with the side chains of critical interaction residues represented in ball-and-stick. LMO2 is shown in yellow with zinc atoms represented as spheres. The critical VH#576 binding site lays across LIM finger two (LIM1), three and four (LIM2). CDR1 is not involved in the reaction, two residues from CDR2 and seven from CDR3 are vital for the interaction. The former VH-VL interface spans residues Val 37 to Tyr 50 and amino acids within this region form hydrophobic interactions with LMO2 that are potentially critical and therefore included in the surface area calculation. The buried surface area of VH#576, critical for interaction with LMO2, is 512\AA^2 .

As previously discussed, stage four of the strategy requires the production of a pharmacophore template and this forms the basis for the future work for this project. That is development of a small molecule lead structure (an initial starting point for medicinal chemistry efforts) based on the VH interaction with the target protein. An initial three-dimensional pharmacophore model can be derived from a protein structure based on the steric and electronic features of the ligand (VH#576) that are necessary to ensure optimal interactions with a target protein (LMO2) and in this case, block its biological activity (Pellecchia et al., 2008). General chemical features of hotspot regions, such as lipophilic groups, H-bond interactions and charge transfer are important to describe the binding mode (Wolber and Langer, 2005). The pharmacophore model must have a high degree of selectivity. Such a pharmacophore can then be used to search a database; molecules are screened on their ability to satisfy a certain number of pharmacophore features. This will produce a smaller subset of compounds for focused screens (Rella et al., 2006) such as docking.

Further *in silico* small molecule screens can reduce the chances of generating false positives and increase the probability of discovering genuine hits. In cases such as this, where the target protein structure is known, docking algorithms are commonly employed to further analyse the binding mode and interactions (Rella et al., 2006). SLIDE (Schnecke and Kuhn, 2000) and DOCK (Oshiro et al., 1995) are computer programs commonly used for this approach. DOCK software flexibly aligns the ligand molecule into a rigid macromolecule environment and then estimates the tightness of the interaction by different scoring functions. SLIDE produces many different conformers of one compound, which are docked individually. These programs are used to identify a set of compounds, which can be tested *in vitro*. One approach, which has been employed by the THR laboratory, is Surface plasmon resonance; the target protein (e.g. His-NusA-LMO2) is immobilised on the chip through interaction with anti-His antibody and a solution phase analyte (the compound) passed over the top (Appert et al., 2009). The changes in refractive index, caused by changes in the mass at the surface, are a measure of binding affinity. A set of lead compounds are then compiled based on how amenable they are for chemical optimisation considering factors such as good ADME (adsorption, distribution, metabolism and excretion) and membership to an established SAR series (structure activity relationship) (Oprea et al., 2001).

Recently a database of small molecules has been compiled, which have the potential to inhibit protein-protein interactions (Higueruelo et al., 2009) based on information from relevant scientific papers. The library, termed TIMBAL, has been analysed along side other

libraries and this highlighted significant differences; the molecules tend to form hydrophobic interactions and tend to be lipophilic and larger than the classical 500 Da Lipinski limit. Calculations shown in this chapter predict a small molecule with molecular weight 850 Da would be required to mimic VH#576 and this represents a common problem when targeting protein-protein interactions; the molecular weight of the compound required tends to be well above the classical Lipinski limit (Lipinski et al., 1997). Drug criteria for these types of targets may well have to vary from the Lipinski rules. The molecules in the TIMBAL database are not classical drug like molecules and may not be specific binders but they provide a platform for therapeutic development of protein-protein interaction inhibition.

6.1.5 LMO2 as a drug target

Lmo2 has a normal obligatory role in definitive haematopoiesis in mice (Yamada et al., 1998) implying that non-cell specific targeting of the protein complex may produce side effects such as anaemia, and possibly other haematopoietic defects. For this reason, it may be necessary to target an anti-LMO2 small molecule to T-cells. One idea, which has been the focus of a considerable amount of research over the past decade, is to direct liposomes to specific cancer cells. An approach considered was to attach an antibody to the liposome membrane, which binds cancer cell surface receptors so that the liposome is delivered along with the drug into the cell (Nielsen et al., 2002).

Further research into this problem has resulted in the development of Antibody-drug conjugates, which are designed to deliver a cytotoxic drug to tumour cells and release it after internalisation thereby activating the drug and thus restricting its toxicity to the diseased cells (Junutula et al., 2008). For this purpose, monoclonal antibodies have been designed to recognise antigens on the surface of tumour cells. Anti-CD33 conjugated to calicheamicin is one example of a successful application of this technology for the treatment of acute myeloid leukaemia. Cell surface markers on T-cells include CD44 and CD25, both of which are potential targets for development of antibody drug conjugates for the delivery of an anti-LMO2 small molecule.

6.2 Conclusion

In conclusion, the technologies described in this thesis offer an alternative strategy for the rational development of small molecule therapeutics. In this case, the strategy has been implemented with the aim of developing a new treatment for patients with T-ALL, associated with specific chromosomal translocations. The strategy of evolving a macrodrug, such as VH#576, into a small molecule is challenging. On the other hand it has the potential

General discussion and future work

to advance drug development for a wide variety of targets. Upon collection of higher resolution X-ray diffraction data the structure of VH#576/LMO2 will form the basis for the development of an anti-LMO2 small molecule inhibitor.

References

- ADAMS, P. D., GROSSE-KUNSTLEVE, R. W., HUNG, L. W., IOERGER, T. R., MCCOY, A. J., MORIARTY, N. W., READ, R. J., SACCHETTINI, J. C., SAUTER, N. K. & TERWILLIGER, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallographica Section D-Biological Crystallography*, 58, 1948-1954.
- AGARWAL, S., TAFEL, A. A. & KANAAR, R. (2006) DNA double-strand break repair and chromosome translocations. *DNA Repair*, 5, 1075-1081.
- AIFANTIS, I., RAETZ, E. & BUONAMICI, S. (2008) Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nature Reviews Immunology*, 8, 380-390.
- ALIZADEH, A. A., EISEN, M. B., DAVIS, R. E., MA, C., LOSSOS, I. S., ROSENWALD, A., BOLDRICK, J. G., SABET, H., TRAN, T., YU, X., POWELL, J. I., YANG, L. M., MARTI, G. E., MOORE, T., HUDSON, J., LU, L. S., LEWIS, D. B., TIBSHIRANI, R., SHERLOCK, G., CHAN, W. C., GREINER, T. C., WEISENBURGER, D. D., ARMITAGE, J. O., WARNKE, R., LEVY, R., WILSON, W., GREVER, M. R., BYRD, J. C., BOTSTEIN, D., BROWN, P. O. & STAUDT, L. M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403, 503-511.
- AOYAMA, M., OZAKI, T., INUZUKA, H., TOMOTSUNE, D., HIRATO, J., OKAMOTO, Y., TOKITA, H., OHIRA, M. & NAKAGAWARA, A. (2005) LMO3 interacts with neuronal transcription factor, HEN2, and acts as an oncogene in neuroblastoma. *Cancer Research*, 65, 4587-4597.
- APLAN, P. D. (2004) T-ALL: smoking guns and genes of interest. *Blood*, 103, 1568-1569.
- APLAN, P. D. (2006) Causes of oncogenic chromosomal translocation. *Trends in Genetics*, 22, 46-55.
- APPERT, A., NAM, C. H., LOBATO, N., PRIEGO, E., MIGUEL, R. N., BLUNDELL, T., DRYNAN, L., SEWELL, H., TANAKA, T. & RABBITTS, T. (2009) Targeting LMO2 with a Peptide Aptamer Establishes a Necessary Function in Overt T-Cell Neoplasia. *Cancer Research*, 69, 4784-4790.
- ARNAU, J., LAURITZEN, C., PETERSEN, G. E. & PEDERSEN, J. (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression and Purification*, 48, 1-13.
- AROUI, S., BRAHIM, S., DE WAARD, M. & KENANI, A. (2010) Cytotoxicity, intracellular distribution and uptake of doxorubicin and doxorubicin coupled to cell-penetrating peptides in different cell lines: A comparative study. *Biochemical and Biophysical Research Communications*, 391, 419-425.

References

- AUE, W. P., BARTHOLDI, E. & ERNST, R. R. (1976) 2-Dimensional spectroscopy - application to Nuclear magnetic-resonance. *Journal of Chemical Physics*, 64, 2229-2246.
- BAILEY, S. (1994) The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography*, 50, 760-763.
- BAIN, G., ENEL, I., MAANDAG, E. C. R., TERIELE, H. P. J., VOLAND, J. R., SHARP, L. L., CHUN, J., HUEY, B., PINKEL, D. & MURRE, C. (1997) E2A deficiency leads to abnormalities in alpha beta T-cell development and to rapid development of T-cell lymphomas. *Molecular and Cellular Biology*, 17, 4782-4791.
- BARTHELEMY, P. A., RAAB, H., APPLETON, B. A., BOND, C. J., WU, P., WIESMANN, C. & SIDHU, S. S. (2008) Comprehensive analysis of the factors contributing to the stability and solubility of autonomous human V-H domains. *Journal of Biological Chemistry*, 283, 3639-3654.
- BERNARD, M., DELABESSE, E., NOVAULT, S., HERMINE, O. & MACINTYRE, E. A. (1998) Antiapoptotic effect of ectopic TAL1/SCL expression in a human leukemic T-cell line. *Cancer Research*, 58, 2680-2687.
- BERNARD, M., SMIT, L., MACINTYRE, E., MATTHIEUMAHUL, D. & PULFORD, K. (1995) Nuclear-localization of the scl/tal1 basic helix-loop-helix protein is not dependent on the presence of the basic domain. *Blood*, 85, 3356-3357.
- BIJVOET, J. M. (1954) Structure of optically active compounds in the solid state. *Nature*, 173, 888-891.
- BLOW, D. (2005) *Outline of Crystallography for Biologists*, Oxford, Oxford University Press.
- BLUNDELL, T. L., JHOTI, H. & ABELL, C. (2002) High-throughput crystallography for lead discovery in drug design. *Nature Reviews Drug Discovery*, 1, 45-54.
- BOEHM, T., FORONI, L., KANEKO, Y., PERUTZ, M. F. & RABBITS, T. H. (1991) The Rhombotin Family of Cysteine-Rich Lim-Domain Oncogenes - Distinct Members Are Involved in T-Cell Translocations to Human Chromosome-11p15 and Chromosome-11p13. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 4367-4371.
- BOND, C. J., MARSTERS, J. C. & SIDHU, S. S. (2003) Contributions of CDR3 to VHH domain stability and the design of monobody scaffolds for naive antibody libraries. *Journal of Molecular Biology*, 332, 643-655.
- BONNET, D. & DICK, J. E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, 3, 730-737.
- BRAGG, W. L. (1913) The structure of some crystals as indicated by their diffraction of x-rays. *Proceedings of the Royal Society of London Series a-Containing Papers of a Mathematical and Physical Character*, 89, 248-277.

References

- BRAUN, P. & LABAER, J. (2004) High throughput protein production for functional proteomics. *Drug Discovery Today*, 9, S1-S7.
- CAI, M. L., HUANG, Y., SAKAGUCHI, K., CLORE, G. M., GRONENBORN, A. M. & CRAIGIE, R. (1998) An efficient and cost-effective isotope labeling protocol for proteins expressed in *Escherichia coli*. *Journal of Biomolecular Nmr*, 11, 97-102.
- CAVANAGH, J., FAIRBROTHER, W. J., III, A. G. P., RANCE, M. & SKELTON, N. J. (2007) *Protein NMR Spectroscopy*, Elsevier Academic Press.
- CHAYEN, N. E. (1997) A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *Journal of Applied Crystallography*, 30, 198-202.
- CHENG, A. C., COLEMAN, R. G., SMYTH, K. T., CAO, Q., SOULARD, P., CAFFREY, D. R., SALZBERG, A. C. & HUANG, E. S. (2007) Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotechnology*, 25, 71-75.
- CHENG, J. T., YANG, C. Y. C., HERNANDEZ, J., EMBREY, J. & BAER, R. (1990) The Chromosome-Translocation (11-14)(P13-Q11) Associated with T-Cell Acute-Leukemia - Asymmetric Diversification of the Translocation Junctions. *Journal of Experimental Medicine*, 171, 489-501.
- CRAZZOLARA, R. & BENDALL, L. (2009) Emerging Treatments in Acute Lymphoblastic Leukemia. *Current Cancer Drug Targets*, 9, 19-31.
- DAVE, U. P., AKAGI, K., TRIPATHI, R., CLEVELAND, S. M., THOMPSON, M. A., YI, M., STEPHENS, R., DOWNING, J. R., JENKINS, N. A. & COPELAND, N. G. (2009) Murine Leukemias with Retroviral Insertions at Lmo2 Are Predictive of the Leukemias Induced in SCID-X1 Patients Following Retroviral Gene Therapy. *Plos Genetics*, 5, 13.
- DAVIES, J. & RIECHMANN, L. (1994) Camelising human-antibody fragments - NMR - studies on VH domains. *Febs Letters*, 339, 285-290.
- DAWID, I. B., BREEN, J. J. & TOYAMA, R. (1998) LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends in Genetics*, 14, 156-162.
- DEANE, J. E., MACKAY, J. P., KWAN, A. H. Y., SUM, E. Y. M., VISVADER, J. E. & MATTHEWS, J. M. (2003) Structural basis for the recognition of Idb1 by the N-terminal LIM domains of LMO2 and LMO4. *Embo Journal*, 22, 2224-2233.
- DEANE, J. E., RYAN, D. P., SUNDE, M., MAHER, M. J., GUSS, J. M., VISVADER, J. E. & MATTHEWS, J. M. (2004) Tandem LIM domains provide synergistic binding in the LMO4 : Idb1 complex. *Embo Journal*, 23, 3589-3598.
- DEANE, J. E., SUM, E., MACKAY, J. P., LINDEMAN, G. J., VISVADER, J. E. & MATTHEWS, J. M. (2001) Design, production and characterization of FLIN2 and FLIN4: the engineering of intramolecular Idb1 : LMO complexes. *Protein Engineering*, 14, 493-499.
- DEAR, T. N., SANCHEZGARCIA, I. & RABBITTS, T. H. (1993) THE HOX11 GENE ENCODES A DNA-BINDING NUCLEAR TRANSCRIPTION FACTOR BELONGING

References

- TO A DISTINCT FAMILY OF HOMEBOX GENES. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 4431-4435.
- DELAGLIO, F., GRZESIEK, S., VUISTER, G. W., ZHU, G., PFEIFER, J. & BAX, A. (1995) Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of Biomolecular Nmr*, 6, 277-293.
- DIK, W. A., NADEL, B., PRZYBYLSKI, G. K., ASNAFI, V., GRABARCZYK, P., NAVARRO, J. M., VERHAAF, B., SCHMIDT, C. A., MACINTYRE, E. A., VAN DONGEN, J. J. M. & LANGERAK, A. W. (2007) Different chromosomal breakpoints impact the level of LMO2 expression in T-ALL. *Blood*, 110, 388-392.
- DOMINGUEZ, C., BOELEN, R. & BONVIN, A. M. J. J. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125, 1731-1737.
- DRUKER, B. J., TALPAZ, M., RESTA, D. J., PENG, B., BUCHDUNGER, E., FORD, J. M., LYDON, N. B., KANTARJIAN, H., CAPDEVILLE, R., OHNO-JONES, S. & SAWYERS, C. L. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *New England Journal of Medicine*, 344, 1031-1037.
- DRYNAN, L. F., HAMILTON, T. L. & RABBITTS, T. H. (2001) T cell tumorigenesis in Lmo2 transgenic mice is independent of V-D-J recombinase activity. *Oncogene*, 20, 4412-4415.
- DUBS, A., WAGNER, G. & WUTHRICH, K. (1979) Individual assignments of amide proton resonances in the proton NMR-spectrum of the basic pancreatic trypsin-inhibitor. *Biochimica Et Biophysica Acta*, 577, 177-194.
- DUDGEON, K., FAMM, K. & CHRIST, D. (2009) Sequence determinants of protein aggregation in human VH domains. *Protein Engineering Design & Selection*, 22, 217-220.
- EMSLEY, P. & COWTAN, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*, 60, 2126-2132.
- EYAL, E., GERZON, S., POTAPOV, V., EDELMAN, M. & SOBOLEV, V. (2005) The limit of accuracy of protein modeling: Influence of crystal packing on protein structure. *Journal of Molecular Biology*, 351, 431-442.
- FERNANDEZ-FUENTES, N., RAI, B. K., MADRID-ALISTE, C. J., FAJARDO, J. E. & FISER, A. (2007) Comparative protein structure modeling by combining multiple templates and optimizing sequence-to-structure alignments. *Bioinformatics*, 23, 2558-2565.
- FERRANDO, A. A., NEUBERG, D. S., STAUNTON, J., LOH, M. L., HUARD, C., RAIMONDI, S. C., BEHM, F. G., PUI, C. H., DOWNING, J. R., GILLILAND, D. G., LANDER, E. S., GOLUB, T. R. & LOOK, A. T. (2002) Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell*, 1, 75-87.

References

- FERREIRA, R., OHNEDA, K., YAMAMOTO, M. & PHILIPSEN, S. (2005) GATA1 function, a paradigm for transcription factors in hematopoiesis. *Molecular and Cellular Biology*, 25, 1215-1227.
- FIALKOW, P. J., SINGER, J. W., RASKIND, W. H., ADAMSON, J. W., JACOBSON, R. J., BERNSTEIN, I. D., DOW, L. W., NAJFELD, V. & VEITH, R. (1987) Clonal Development, Stem-Cell Differentiation, and Clinical Remissions in Acute Nonlymphocytic Leukemia. *New England Journal of Medicine*, 317, 468-473.
- FINGER, L. R., HARVEY, R. C., MOORE, R. C. A., SHOWE, L. C. & CROCE, C. M. (1986) A common mechanism of chromosomal translocation in T-cell and B-cell neoplasia. *Science*, 234, 982-985.
- FORONI, L., BOEHM, T., WHITE, L., FORSTER, A., SHERRINGTON, P., LIAO, X. B., BRANNAN, C. I., JENKINS, N. A., COPELAND, N. G. & RABBITTS, T. H. (1992) The Rhombotin Gene Family Encode Related Lim-Domain Proteins Whose Differing Expression Suggests Multiple Roles in Mouse Development. *Journal of Molecular Biology*, 226, 747-761.
- FUJIWARA, Y., CHANG, A. N., WILLIAMS, A. M. & ORKIN, S. H. (2004) Functional overlap of GATA-1 and GATA-2 in primitive hematopoietic development. *Blood*, 103, 583-585.
- GAIDARENKO, O. & XU, Y. (2009) Transcription activity is required for p53-dependent tumor suppression. *Oncogene*, 28, 4397-4401.
- GARCIA, I. S., KANEKO, Y., GONZALEZSARMIENTO, R., CAMPBELL, K., WHITE, L., BOEHM, T. & RABBITTS, T. H. (1991) A Study of Chromosome-11p13 Translocations Involving Tcr-Beta and Tcr-Delta in Human T-Cell Leukemia. *Oncogene*, 6, 577-582.
- GERING, M., YAMADA, Y., RABBITTS, T. H. & PATIENT, R. K. (2003) Lmo2 and Scf/Tal1 convert non-axial mesoderm into haemangioblasts which differentiate into endothelial cells in the absence of Gata1. *Development*, 130, 6187-6199.
- GIROUX, S., KAUSHIK, A. L., CAPRON, C., JALIL, A., KELAIDI, C., SABLITZKY, F., DUMENIL, D., ALBAGLI, O. & GODIN, I. (2007) *lyl-1* and *tal-1lscf*, two genes encoding closely related bHLH transcription factors, display highly overlapping expression patterns during cardiovascular and hematopoietic ontogeny. *Gene Expression Patterns*, 7, 215-226.
- GOODFORD, P. J. (1985) A Computational-Procedure for Determining Energetically Favorable Binding-Sites on Biologically Important Macromolecules. *Journal of Medicinal Chemistry*, 28, 849-857.
- GRATZINGER, D., ZHAO, S. C., WEST, R., ROUSE, R. V., VOGEL, H., GIL, E. C., LEVY, R., LOSSOS, I. S. & NATKUNAM, Y. (2009) The Transcription Factor LMO2 Is a

References

- Robust Marker of Vascular Endothelium and Vascular Neoplasms and Selected Other Entities. *American Journal of Clinical Pathology*, 131, 264-278.
- GRAUX, C., COOLS, J., MICHAUX, L., VANDENBERGHE, P. & HAGEMEIJER, A. (2006) Cytogenetics and molecular genetics of T-cell acute lymphoblastic leukemia: from thymocyte to lymphoblast. *Leukemia*, 20, 1496-1510.
- GREAVES, M. F. & WIEMELS, J. (2003) Origins of chromosome translocations in childhood leukaemia. *Nature Reviews Cancer*, 3, 639-649.
- GRUTZ, G., FORSTER, A. & RABBITS, T. H. (1998a) Identification of the LMO4 gene encoding an interaction partner of the LIM-binding protein LDB1/NLI1: a candidate for displacement by LMO proteins in T cell acute leukaemia. *Oncogene*, 17, 2799-2803.
- GRUTZ, G. G., BUCHER, K., LAVENIR, I., LARSON, T., LARSON, R. & RABBITS, T. H. (1998b) The oncogenic T cell LIM-protein Lmo2 forms part of a DNA-binding complex specifically in immature T cells. *Embo Journal*, 17, 4594-4605.
- GUNTERT, P. (1998) Structure calculation of biological macromolecules from NMR data. *Quarterly Reviews of Biophysics*, 31, 145-237.
- HACEIN-BEY-ABINA, S., GARRIGUE, A., WANG, G. P., SOULIER, J., LIM, A., MORILLON, E., CLAPPIER, E., CACCAVELLI, L., DELABESSE, E., BELDJORD, K., ASNAFI, V., MACINTYRE, E., DAL CORTIVO, L., RADFORD, I., BROUSSE, N., SIGAUX, F., MOSHOUS, D., HAUER, J., BORKHARDT, A., BELOHRADSKY, B. H., WINTERGERST, U., VELEZ, M. C., LEIVA, L., SORENSEN, R., WULFFRAAT, N., BLANCHE, S., BUSHMAN, F. D., FISCHER, A. & CAVAZZANA-CALVO, M. (2008) Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *Journal of Clinical Investigation*, 118, 3132-3142.
- HACEIN-BEY-ABINA, S., VON KALLE, C., SCHMIDT, M., MCCORMACK, M. P., WULFFRAAT, N., LEBOULCH, P., LIM, A., OSBORNE, C. S., PAWLIUK, R., MORILLON, E., SORENSEN, R., FORSTER, A., FRASER, P., COHEN, J. I., DE SAINT BASILE, G., ALEXANDER, I., WINTERGERST, U., FREBOURG, T., AURIAS, A., STOPPA-LYONNET, D., ROMANA, S., RADFORD-WEISS, I., GROSS, F., VALENSI, F., DELABESSE, E., MACINTYRE, E., SIGAUX, F., SOULIER, J., LEIVA, L. E., WISSLER, M., PRINZ, C., RABBITS, T. H., LE DEIST, F., FISCHER, A. & CAVAZZANA-CALVO, M. (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*, 302, 415-419.
- HAMERSCASTERMAN, C., ATARHOUCHE, T., MUYLDERMANS, S., ROBINSON, G., HAMERS, C., SONGA, E. B., BENDAHMAN, N. & HAMERS, R. (1993) Naturally-Occurring Antibodies Devoid of Light-Chains. *Nature*, 363, 446-448.
- HANNON, G. J. (2002) RNA interference. *Nature*, 418, 244-251.

References

- HARRIS, M. I., EASTMAN, R. C., COWIE, C. C., FLEGAL, K. M. & EBERHARDT, M. S. (1997) Comparison of diabetes diagnostic categories in the US population according to 1997 American Diabetes Association and 1980-1985 World Health Organization diagnostic criteria. *Diabetes Care*, 20, 1859-1862.
- HIGUERUELO, A. P., SCHREYER, A., BICKERTON, G. R. J., PITT, W. R., GROOM, C. R. & BLUNDELL, T. L. (2009) Atomic Interactions and Profile of Small Molecules Disrupting Protein-Protein Interfaces: the TIMBAL Database. *Chemical Biology & Drug Design*, 74, 457-467.
- HSU, H. L., CHENG, J. T., CHEN, Q. & BAER, R. (1991) Enhancer-binding activity of the Tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. *Molecular and Cellular Biology*, 11, 3037-3042.
- JANEWAY, C. A., TRAVERS, P., WALPORT, M. & SHLOMCHIK, M. J. (2005) *Immuno Biology: The immune system in health and disease*, New York, Garland Science Publishing.
- JATON, J. C., KLINMAN, N. R., GIVOL, D. & SELA, M. (1968) Recovery of Antibody Activity Upon Reoxidation of Completely Reduced Polyalaninyl Heavy Chain and Its Fd Fragment Derived from Anti-2,4-Dinitrophenyl Antibody. *Biochemistry*, 7, 4185-&.
- JENSEN, L. H. (1997) Refinement and reliability of macromolecular models based on X-ray diffraction data. *Macromolecular Crystallography, Pt B*, 277, 353-366.
- JESPER, L., SCHON, O., JAMES, L. C., VEPRINTSEV, D. & WINTER, G. (2004) Crystal structure of HEL4, a soluble, refoldable Human V-H single domain with a germ-line scaffold. *Journal of Molecular Biology*, 337, 893-903.
- JOHNSON, W. C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins-Structure Function and Genetics*, 35, 307-312.
- JUNUTULA, J. R., RAAB, H., CLARK, S., BHAKTA, S., LEIPOLD, D. D., WEIR, S., CHEN, Y., SIMPSON, M., TSAI, S. P., DENNIS, M. S., LU, Y. M., MENG, Y. G., NG, C., YANG, J. H., LEE, C. C., DUENAS, E., GORRELL, J., KATTA, V., KIM, A., MCDORMAN, K., FLAGELLA, K., VENOOK, R., ROSS, S., SPENCER, S. D., WONG, W. L., LOWMAN, H. B., VANDLEN, R., SLIWKOWSKI, M. X., SCHELLER, R. H., POLAKIS, P. & MALLET, W. (2008) Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nature Biotechnology*, 26, 925-932.
- KAY, L. E., IKURA, M., TSCHUDIN, R. & BAX, A. (1990) 3-DIMENSIONAL TRIPLE-RESONANCE NMR-SPECTROSCOPY OF ISOTOPICALLY ENRICHED PROTEINS. *Journal of Magnetic Resonance*, 89, 496-514.
- KENNY, D. A., JURATA, L. W., SAGA, Y. & GILL, G. N. (1998) Identification and characterization of LMO4, an LMO gene with a novel pattern of expression during

References

- embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 11257-11262.
- KIM, Y. C., QUARTEY, P., LI, H., VOLKART, L., HATZOS, C., CHANG, C., NOCEK, B., CUFF, M., OSIPIUK, J., TAN, K. M., FAN, Y., BIGELOW, L., MALTSEVA, N., WU, R. Y., BOROVILOS, M., DUGGAN, E., ZHOU, M., BINKOWSKI, T. A., ZHANG, R. G. & JOACHIMIAK, A. (2008) Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nature Methods*, 5, 853-854.
- KLEIN, G. (2000) Dysregulation of lymphocyte proliferation by chromosomal translocations and sequential genetic changes. *Bioessays*, 22, 414-422.
- KONTAXIS, G., KONRAT, R., KRAUTLER, B., WEISKIRCHEN, R. & BISTER, K. (1998) Structure and intramodular dynamics of the amino-terminal LIM domain from quail cysteine- and glycine-rich protein CRP2. *Biochemistry*, 37, 7127-7134.
- KOVARI, L. C., MOMANY, C. & ROSSMANN, M. G. (1995) The use of antibody fragments for crystallization and structure determinations. *Structure*, 3, 1291-1293.
- LANDRY, J. R., BONADIES, N., KINSTON, S., KNEZEVIC, K., WILSON, N. K., ORAM, S. H., JANES, M., PILTZ, S., HAMMETT, M., CARTER, J., HAMILTON, T., DONALDSON, I. J., LACAUD, G., FRAMPTON, J., FOLLOWS, G., KOUSKOFF, V. & GOTTGENS, B. (2009) Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors. *Blood*, 113, 5783-5792.
- LANDRY, J. R., KINSTON, S., KNEZEVIC, K., DE BRUIJN, M. F. T. R., WILSON, N., NOTTINGHAM, W. T., PEITZ, M., EDENHOFER, F., PIMANDA, J. E., OTTERSBAACH, K. & GOTTGENS, B. (2008) Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. *Blood*, 111, 3005-3014.
- LARSON, R. C., FISCH, P., LARSON, T. A., LAVENIR, I., LANGFORD, T., KING, G. & RABBITTS, T. H. (1994) T-CELL TUMORS OF DISPARATE PHENOTYPE IN MICE TRANSGENIC FOR RBTN-2. *Oncogene*, 9, 3675-3681.
- LARSON, R. C., LAVENIR, I., LARSON, T. A., BAER, R., WARREN, A. J., WADMAN, I., NOTTAGE, K. & RABBITTS, T. H. (1996) Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. *Embo Journal*, 15, 1021-1027.
- LARSON, R. C., OSADA, H., LARSON, T. A., LAVENIR, I. & RABBITTS, T. H. (1995) The Oncogenic Lim-Protein Rbtn2 Causes Thymic Developmental Aberrations That Precede Malignancy in Transgenic Mice. *Oncogene*, 11, 853-862.
- LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*, 26, 283-291.

References

- LAVAU, C., SZILVASSY, S. J., SLANY, R. & CLEARY, M. L. (1997) Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced HRX-ENL. *Embo Journal*, 16, 4226-4237.
- LEE, S. Y., KUMANO, K., MASUDA, S., HANGAISHI, A., TAKITA, J., NAKAZAKI, K., KUROKAWA, M., HAYASHI, Y., OGAWA, S. & CHIBA, S. (2005) Mutations of the Notch1 gene in T-cell acute lymphoblastic leukemia: analysis in adults and children. *Leukemia*, 19, 1841-1843.
- LEFRANC, M. P. (2001) Nomenclature of the human immunoglobulin heavy (IGH) genes. *Experimental and Clinical Immunogenetics*, 18, 100-116.
- LEPRE, C. A. & MOORE, J. M. (1998) Microdrop screening: A rapid method to optimize solvent conditions for NMR spectroscopy of proteins. *Journal of Biomolecular Nmr*, 12, 493-499.
- LICHT, J. D. (2001) AML1 and the AML1-ETO fusion protein in the pathogenesis of t(8;21) AML. *Oncogene*, 20, 5660-5679.
- LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23, 3-25.
- LIU, J., ANDYA, J. D. & SHIRE, S. J. (2006) A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. *Aaps Journal*, 8, E580-E589.
- LOBATO, M. N. & RABBITTS, T. H. (2003) Intracellular antibodies and challenges facing their use as therapeutic agents. *Trends in Molecular Medicine*, 9, 390-396.
- LOOK, A. T. (1997) Oncogenic transcription factors in the human acute leukemias. *Science*, 278, 1059-1064.
- LOSSOS, I. S., CZERWINSKI, D. K., ALIZADEH, A. A., WECHSER, M. A., TIBSHIRANI, R., BOTSTEIN, D. & LEVY, R. (2004) Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *New England Journal of Medicine*, 350, 1828-1837.
- MA, S., GUAN, X. Y., BEH, P. S. L., WONG, K. Y., CHAN, Y. P., YUEN, H. F., VIELKIND, J. & CHAN, K. W. (2007) The significance of LMO2 expression in the progression of prostate cancer. *Journal of Pathology*, 211, 278-285.
- MACELREVEY, C., SPITALE, R. C., KRUCINSKA, J. & WEDEKIND, J. E. (2007) A posteriori design of crystal contacts to improve the X-ray diffraction properties of a small RNA enzyme. *Acta Crystallographica Section D-Biological Crystallography*, 63, 812-825.
- MANSOUR, M. R., DUKE, V., FORONI, L., PATEL, B., ALLEN, C. G., ANCLIFF, P. J., GALE, R. E. & LINCH, D. C. (2007) Notch-1 mutations are secondary events in some

References

- patients with T-Cell acute lymphoblastic leukemia. *Clinical Cancer Research*, 13, 6964-6969.
- MARLEY, J., LU, M. & BRACKEN, C. (2001) A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, 20, 71-75.
- MARTIN-CORDERO, C., LOPEZ-LAZARO, M., GALVEZ, M. & AYUSO, M. J. (2003) Curcumin as a DNA topoisomerase II poison. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 18, 505-509.
- MATTHEWS, B. W. (1968) Solvent Content of Protein Crystals. *Journal of Molecular Biology*, 33, 491-&.
- MATTHEWS, J. M., VISVADER, J. E. & MACKAY, J. P. (2001) Letter to the Editor: H-1, N-15 and C-13 assignments of FLIN2, an intramolecular LMO2 : ldb1 complex. *Journal of Biomolecular Nmr*, 21, 385-386.
- MCKELVIE, J. F., VANGENT, D. C., RAMSDEN, D. A., ROMEO, C., CUOMO, C. A., GELLERT, M. & OETTINGER, M. A. (1995) Cleavage at a V(D)J Recombination Signal Requires Only Rag1 and Rag2 Proteins and Occurs in 2 Steps. *Cell*, 83, 387-395.
- MCCAFFREY, A. P., MEUSE, L., PHAM, T. T. T., CONKLIN, D. S., HANNON, G. J. & KAY, M. A. (2002) Gene expression - RNA interference in adult mice. *Nature*, 418, 38-39.
- MCCALL, K. A., HUANG, C. C. & FIERKE, C. A. (2000) Function and mechanism of zinc metalloenzymes. *Journal of Nutrition*, 130, 1437S-1446S.
- MCCORMACK, M. P., FORSTER, A., DRYNAN, L., PANNELL, R. & RABBITTS, T. H. (2003) The LMO2 T-cell oncogene is activated via chromosomal translocations or retroviral insertion during gene therapy but has no mandatory role in normal T-cell development. *Molecular and Cellular Biology*, 23, 9003-9013.
- MCCORMACK, M. P. & RABBITTS, T. H. (2004) Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency. *New England Journal of Medicine*, 350, 913-922.
- MCGREGOR, M. J. (2007) A pharmacophore map of small molecule protein kinase inhibitors. *Journal of Chemical Information and Modeling*, 47, 2374-2382.
- MEIER, N., KRPIC, S., RODRIGUEZ, P., STROUBOULIS, J., MONTI, M., KRIJGSVELD, J., GERING, M., PATIENT, R., HOSTERT, A. & GROSVELD, F. (2006) Novel binding partners of Ldb1 are required for haematopoietic development. *Development*, 133, 4913-4924.
- MENG, Y. S., KHOURY, H., DICK, J. E. & MINDEN, M. D. (2005) Oncogenic potential of the transcription factor LYL1 in acute myeloblastic leukemia. *Leukemia*, 19, 1941-1947.
- MIROUX, B. & WALKER, J. E. (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, 260, 289-298.

References

- MISTRY, A. R., FELIX, C. A., WHITMARSH, R. J., MASON, A., REITER, A., CASSINAT, B., PARRY, A., WALZ, C., WIEMELS, J. L., SEGAL, M. R., ADES, L., BLAIR, I. A., OSHEROFF, N., PENIKET, A. J., LAFAGE-POCHITALOFF, M., CROSS, N. C. P., CHOMIENNE, C., SOLOMON, E., FENAUX, P. & GRIMWADE, D. (2005) DNA topoisomerase II in therapy-related acute promyelocytic leukemia. *New England Journal of Medicine*, 352, 1529-1538.
- MITELMAN, F., JOHANSSON, B. & MERTENS, F. (2007) The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, 7, 233-245.
- MONTANEZ-WISCOVICH, M. E., SEACHRIST, D. D., LANDIS, M. D., VISVADER, J., ANDERSEN, B. & KERI, R. A. (2009) LMO4 is an essential mediator of ErbB2/HER2/Neu-induced breast cancer cell cycle progression. *Oncogene*, 28, 3608-3618.
- MULLOY, J. C., CAMMENGA, J., MACKENZIE, K. L., BERGUIDO, F. J., MOORE, M. A. S. & NIMER, S. D. (2002) The AML1-ETO fusion protein promotes the expansion of human hematopoietic stem cells. *Blood*, 99, 15-23.
- MURSHUDOV, G. N., VAGIN, A. A. & DODSON, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography*, 53, 240-255.
- NAKATA, K., NAGAI, E., OHUCHICLA, K., MIYASAKA, Y., HAYASHI, A., MIZUMOTO, K., TSUNEYOSHI, M. & TANAKA, M. (2008) Overexpression of Lmo2 in pancreatic cancer and its clinical implication as a therapeutic target. *Gastroenterology*, 134, A694-A694.
- NAKATA, K., OHUCHIDA, K., NAGAI, E., HAYASHI, A., MIYASAKA, Y., KAYASHIMA, T., YU, J., AISHIMA, S., ODA, Y., MIZUMOTO, K., TANAKA, M. & TSUNEYOSHI, M. (2009) LMO2 Is a Novel Predictive Marker for a Better Prognosis in Pancreatic Cancer. *Neoplasia*, 11, 712-719.
- NAM, C. H., LOBATO, M. N., APPERT, A., DRYNAN, L. F., TANAKA, T. & RABBITTS, T. H. (2008) An antibody inhibitor of the LMO2-protein complex blocks its normal and tumorigenic functions. *Oncogene*, 27, 4962-4968.
- NAM, C. H. & RABBITTS, T. H. (2006) The role of LMO2 in development and in T cell leukemia after chromosomal translocation or retroviral insertion. *Molecular Therapy*, 13, 15-25.
- NATKUNAM, Y., FARINHA, P., HSI, E. D., HANS, C. P., TIBSHIRANI, R., SEHN, L. H., CONNORS, J. M., GRATZINGER, D., ROSADO, M., ZHAO, S., POHLMAN, B., WONGCHAOWART, N., BAST, M., AVIGDOR, A., SCHIBY, G., NAGLER, A., BYRNE, G. E., LEVY, R., GASCOYNE, R. D. & LOSSOS, I. S. (2008) LMO2 protein expression predicts survival in patients with diffuse large B-Cell lymphoma treated with

References

- anthracycline-based chemotherapy with and without rituximab. *Journal of Clinical Oncology*, 26, 447-454.
- NATKUNAM, Y., ZHAO, S. C., MASON, D. Y., CHEN, J., TAIDI, B., JONES, M., HAMMER, A. S., DUTOIT, S. H., LOSSOS, I. S. & LEVY, R. (2007) The oncoprotein LMO2 is expressed in normal germinal-center B cells and in human B-cell lymphomas. *Blood*, 109, 1636-1642.
- NETTLESHIP, J. E., REN, J., RAHMAN, N., BERROWA, N. S., HATHERLEY, D., BARCLAY, A. N. & OWENS, R. J. (2008) A pipeline for the production of antibody fragments for structural studies using transient expression in HEK 293T cells. *Protein Expression and Purification*, 62, 83-89.
- NEVES, H., RAMOS, C., DA SILVA, M. G., PARREIRA, A. & PARREIRA, L. (1999) The nuclear topography of ABL, BCR, PML, and RAR alpha genes: Evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. *Blood*, 93, 1197-1207.
- NIELSEN, U. B., KIRPOTIN, D. B., PICKERING, E. M., HONG, K. L., PARK, J. W., SHALABY, M. R., SHAO, Y., BENZ, C. C. & MARKS, J. D. (2002) Therapeutic efficacy of anti-ErbB2 immunoliposomes targeted by a phage antibody selected for cellular endocytosis. *Biochimica Et Biophysica Acta-Molecular Cell Research*, 1591, 109-118.
- NISHIDA, M., HARADA, S., NOGUCHI, S., SATOW, Y., INOUE, H. & TAKAHASHI, K. (1998) Three-dimensional structure of *Escherichia coli* glutathione S-transferase complexed with glutathione sulfonate: Catalytic roles of Cys10 and His106. *Journal of Molecular Biology*, 281, 135-147.
- O'NEIL, J., SHANK, J., CUSSON, N., MURRE, C. & KELLIHER, M. (2004) TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB. *Cancer Cell*, 5, 587-596.
- OKUDA, T., CAI, Z. L., YANG, S. L., LENNY, N., LYU, C. J., VAN DEURSEN, J. M. A., HARADA, H. & DOWNING, J. R. (1998) Expression of a knocked-in AML1-ETO leukemia gene inhibits the establishment of normal definitive hematopoiesis and directly generates dysplastic hematopoietic progenitors. *Blood*, 91, 3134-3143.
- OMARI, K. E., KARIA, D., PORCHER, C. & MANCINI, E. J. (2010) Structure of the leukemia oncogene LMO2 in complex with the LID domain of nuclear adaptor protein LDB1.
- OPREA, T. I., DAVIS, A. M., TEAGUE, S. J. & LEESON, P. D. (2001) Is there a difference between leads and drugs? A historical perspective. *Journal of Chemical Information and Computer Sciences*, 41, 1308-1315.
- OSADA, H., GRUTZ, G., AXELSON, H., FORSTER, A. & RABBITTS, T. H. (1995) Association of Erythroid Transcription Factors - Complexes Involving the Lim Protein

References

- Rbtn2 and the Zinc-Finger Protein Gata1. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 9585-9589.
- OSHIRO, C. M., KUNTZ, I. D. & DIXON, J. S. (1995) Flexible Ligand Docking Using a Genetic Algorithm. *Journal of Computer-Aided Molecular Design*, 9, 113-130.
- OTWINOWSKI, Z. & MINOR, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A*, 276, 307-326.
- PABO, C. O., PEISACH, E. & GRANT, R. A. (2001) Design and selection of novel Cys(2)His(2) zinc finger proteins. *Annual Review of Biochemistry*, 70, 313-340.
- PAPE, T. & SCHNEIDER, T. R. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *Journal of Applied Crystallography*, 37, 843-844.
- PARK, C. M., BRUNCKO, M., ADICKES, J., BAUCH, J., DING, H., KUNZER, A., MARSH, K. C., NIMMER, P., SHOEMAKER, A. R., SONG, X., TAHIR, S. K., TSE, C., WANG, X. L., WENDT, M. D., YANG, X. F., ZHANG, H. C., FESIK, S. W., ROSENBERG, S. H. & ELMORE, S. W. (2008) Discovery of an Orally Bioavailable Small Molecule Inhibitor of Prosurvival B-Cell Lymphoma 2 Proteins. *Journal of Medicinal Chemistry*, 51, 6902-6915.
- PASSEGUE, E., JAMIESON, C. H. M., AILLES, L. E. & WEISSMAN, I. L. (2003) Normal and leukemic hematopoiesis: Are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proceedings of the National Academy of Sciences of the United States of America*, 100, 11842-11849.
- PATTERSON, L. J., GERING, M., ECKFELDT, C. E., GREEN, A. R., VERFAILLIE, C. M., EKKER, S. C. & PATIENT, R. (2007) The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. *Blood*, 109, 2389-2398.
- PELLECCHIA, M., BERTINI, I., COWBURN, D., DALVIT, C., GIRALT, E., JAHNKE, W., JAMES, T. L., HOMANS, S. W., KESSLER, H., LUCHINAT, C., MEYER, B., OSCHKINAT, H., PENG, J., SCHWALBE, H. & SIEGAL, G. (2008) Perspectives on NMR in drug discovery: a technique comes of age. *Nature Reviews Drug Discovery*, 7, 738-745.
- PETROS, A. M., DINGES, J., AUGERI, D. J., BAUMEISTER, S. A., BETEBENNER, D. A., BURES, M. G., ELMORE, S. W., HAJDUK, P. J., JOSEPH, M. K., LANDIS, S. K., NETTESHEIM, D. G., ROSENBERG, S. H., SHEN, W., THOMAS, S., WANG, X. L., ZANZE, I., ZHANG, H. C. & FESIK, S. W. (2006) Discovery of a potent inhibitor of the antiapoptotic protein Bcl-x(L) from NMR and parallel synthesis. *Journal of Medicinal Chemistry*, 49, 656-663.

References

- PEVNY, L., LIN, C. S., DAGATI, V., SIMON, M. C., ORKIN, S. H. & COSTANTINI, F. (1995) Development of hematopoietic-cells lacking transcription factor GATA-1. *Development*, 121, 163-172.
- PICKSLEY, S. M., VOJTESEK, B., SPARKS, A. & LANE, D. P. (1994) Immunochemical Analysis of the Interaction of P53 with Mdm2 - Fine Mapping of the Mdm2 Binding-Site on P53 Using Synthetic Peptides. *Oncogene*, 9, 2523-2529.
- PORCHER, C., SWAT, W., ROCKWELL, K., FUJIWARA, Y., ALT, F. W. & ORKIN, S. H. (1996) The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell*, 86, 47-57.
- PRONGAY, A. J., SMITH, T. J., ROSSMANN, M. G., EHRLICH, L. S., CARTER, C. A. & MCCLURE, J. (1990) Preparation and Crystallization of a Human-Immunodeficiency-Virus P24-Fab Complex. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 9980-9984.
- PUI, C. H. (2009) T Cell Acute Lymphoblastic Leukemia: NOTCHing the Way toward a Better Treatment Outcome. *Cancer Cell*, 15, 85-87.
- PUI, C. H. & JEHA, S. (2007) New therapeutic strategies for the treatment of acute lymphoblastic leukaemia. *Nature Reviews Drug Discovery*, 6, 149-165.
- RABBITTS, T. H. (1991) Translocations, Master Genes, and Differences between the Origins of Acute and Chronic Leukemias. *Cell*, 67, 641-644.
- RABBITTS, T. H., KORSMEYER, S. & SHARP, P. (1999) The effect of chromosomal translocations in acute leukemias: The LMO2 paradigm in transcription and development - Discussion. *Cancer Research*, 59, 1798S-1798S.
- RELLA, M., RUSHWORTH, C. A., GUY, J. L., TURNER, A. J., LANGER, T. & JACKSON, R. M. (2006) Structure-based pharmacophore design and virtual screening for novel angiotensin converting enzyme 2 inhibitors. *J Chem Inf Model*, 46, 708-16.
- RHODES, G. (2006) *Crystallography Made Crystal Clear*, Academic Press.
- RIECHMANN, L. & DAVIES, J. (1995) Backbone assignment, secondary structure and protein-A binding of an isolated, human-antibody VH domain. *Journal of Biomolecular NMR*, 6, 141-152.
- RIECHMANN, L. & MUYLDERMANS, S. (1999) Single domain antibodies: comparison of camel VH and camelised human VH domains. *Journal of Immunological Methods*, 231, 25-38.
- RIGGI, N., CIRONI, L., PROVERO, P., SUVA, M. L., KALOULIS, K., GARCIA-ECHEVERRIA, C., HOFFMANN, F., TRUMPP, A. & STAMENKOVIC, I. (2005) Development of Ewing's sarcoma from primary bone marrow-derived mesenchymal progenitor cells. *Cancer Research*, 65, 11459-11468.

References

- RIZZO, S., ATTARD, G. & HUDSON, D. L. (2005) Prostate epithelial stem cells. *Cell Proliferation*, 38, 363-374.
- ROBB, L., LYONS, I., LI, R. L., HARTLEY, L., KONTGEN, F., HARVEY, R. P., METCALF, D. & BEGLEY, C. G. (1995) ABSENCE OF YOLK-SAC HEMATOPOIESIS FROM MICE WITH A TARGETED DISRUPTION OF THE SCL GENE. *Proceedings of the National Academy of Sciences of the United States of America*, 92, 7075-7079.
- ROSSMANN, M. G. & BLOW, D. M. (1962) Detection of sub-units within crystallographic asymmetric unit. *Acta Crystallographica*, 15, 24-&.
- ROYERPOKORA, B., ROGERS, M., ZHU, T. H., SCHNEIDER, S., LOOS, U. & BOLITZ, U. (1995) The TTG-2/RBTN2 T-cell oncogene encodes 2 alternative transcripts from 2 promoters - the distal promoter is removed by most 11p13 translocations in acute T-cell leukemias (T-ALL). *Oncogene*, 10, 1353-1360.
- RUBNITZ, J. E., CAMITTA, B. M., MAHMOUD, H., RAIMONDI, S. C., CARROLL, A. J., BOROWITZ, M. J., SHUSTER, J. J., LINK, M. P., PULLEN, D. J., DOWNING, J. R., BEHM, F. G. & PUI, C. H. (1999) Childhood acute lymphoblastic leukemia with the MLL-ENL fusion and t(11;19)(q23;p13.3) translocation. *Journal of Clinical Oncology*, 17, 191-196.
- RUBNITZ, J. E. & CRIST, W. M. (1997) Molecular genetics of childhood cancer: Implications for pathogenesis, diagnosis, and treatment. *Pediatrics*, 100, 101-108.
- RYAN, D. P., DUNCAN, J. L., LEE, C., KUCHEL, P. W. & MATTHEWS, J. M. (2008) Assembly of the oncogenic DNA-binding complex LMO2 center dot Ldb1 center dot TAL1 center dot E12. *Proteins-Structure Function and Bioinformatics*, 70, 1461-1474.
- RYAN, D. P., SUNDE, M., KWAN, A. H. Y., MARIANAYAGAM, N. J., NANCARROW, A. L., HOVEN, R. N. V., THOMPSON, L. S., BACA, M., MACKAY, J. P., VISVADER, J. E. & MATTHEWS, J. M. (2006) Identification of the key LMO2-binding determinants on Ldb1. *Journal of Molecular Biology*, 359, 66-75.
- SANCHEZGARCIA, I., AXELSON, H. & RABBITTS, T. H. (1995) FUNCTIONAL DIVERSITY OF LIM PROTEINS - AMINO-TERMINAL ACTIVATION DOMAINS IN THE ONCOGENIC PROTEINS RBTN1 AND RBTN2. *Oncogene*, 10, 1301-1306.
- SCHNECKE, V. & KUHN, L. A. (2000) Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design*, 20, 171-190.
- SCHULER, W., KLOIBER, K., MATT, T., BISTER, K. & KONRAT, R. (2001) Application of cross-correlated NMR spin relaxation to the zinc-finger protein CRP2(LIM2): Evidence for collective motions in LIM domains. *Biochemistry*, 40, 9596-9604.
- SHAH, N. P., TRAN, C., LEE, F. Y., CHEN, P., NORRIS, D. & SAWYERS, C. L. (2004) Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science*, 305, 399-401.

References

- SHAPIRO, L., FANNON, A. M., KWONG, P. D., THOMPSON, A., LEHMANN, M. S., GRUBEL, G., LEGRAND, J. F., ALSNIELSEN, J., COLMAN, D. R. & HENDRICKSON, W. A. (1995) Structural Basis of Cell-Cell Adhesion by Cadherins. *Nature*, 374, 327-337.
- SHIVDASANI, R. A., MAYER, E. L. & ORKIN, S. H. (1995) ABSENCE OF BLOOD FORMATION IN MICE LACKING THE T-CELL LEUKEMIA ONCOPROTEIN TAL-1/SCL. *Nature*, 373, 432-434.
- SHOEMAKER, A. R., MITTEN, M. J., ADICKES, J., ACKLER, S., REFICI, M., FERGUSON, D., OLEKSIJEW, A., O'CONNOR, J. M., WANG, B. L., FROST, D. J., BAUCH, J., MARSH, K., TAHIR, S. K., YANG, X. F., TSE, C., FESIK, S. W., ROSENBERG, S. H. & ELMORE, S. W. (2008) Activity of the Bcl-2 family inhibitor ABT-263 in a panel of small cell lung cancer xenograft models. *Clinical Cancer Research*, 14, 3268-3277.
- SIMONIS, M., KLOUS, P., HOMMINGA, I., GALJAARD, R. J., RIJKERS, E. J., GROSVELD, F., MEIJERINK, J. P. P. & DE LAAT, W. (2009) High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nature Methods*, 6, 837-U79.
- SIPPL, M. J. (1993) Recognition of Errors in 3-Dimensional Structures of Proteins. *Proteins-Structure Function and Genetics*, 17, 355-362.
- SIVASHANMUGAM, A., MURRAY, V., CUI, C. X., ZHANG, Y. H., WANG, J. J. & LI, Q. Q. (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science*, 18, 936-948.
- SMYTH, D. R., MROZKIEWICZ, M. K., MCGRATH, W. J., LISTWAN, P. & KOBE, B. (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Science*, 12, 1313-1322.
- SOBOLEV, V., SOROKINE, A., PRILUSKY, J., ABOLA, E. E. & EDELMAN, M. (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics*, 15, 327-332.
- SPERA, S. & BAX, A. (1991) Empirical Correlation between Protein Backbone Conformation and C-Alpha and C-Beta C-13 Nuclear-Magnetic-Resonance Chemical-Shifts. *Journal of the American Chemical Society*, 113, 5490-5492.
- STEGMAIER, K., PENDSE, S., BARKER, G. F., BRAYWARD, P., WARD, D. C., MONTGOMERY, K. T., KRAUTER, K. S., REYNOLDS, C., SKLAR, J., DONNELLY, M., BOHLANDER, S. K., ROWLEY, J. D., SALLAN, S. E., GILLILAND, D. G. & GOLUB, T. R. (1995) FREQUENT LOSS OF HETEROZYGOSITY AT THE TEL GENE LOCUS IN ACUTE LYMPHOBLASTIC-LEUKEMIA OF CHILDHOOD. *Blood*, 86, 38-44.
- SUM, E. Y. M., PENG, B., YU, X., CHEN, J. J., BYRNE, J., LINDEMAN, G. J. & VISVADER, J. E. (2002) The LIM domain protein LMO4 interacts with the cofactor CtIP

References

and the tumor suppressor BRCA1 and inhibits BRCA1 activity. *Journal of Biological Chemistry*, 277, 7849-7856.

TANAKA, T., LOBATO, M. N. & RABBITTS, T. H. (2003) Single domain intracellular antibodies: A minimal fragment for direct in vivo selection of antigen-specific intrabodies. *Journal of Molecular Biology*, 331, 1109-1120.

TANAKA, T. & RABBITTS, T. H. (2003) Intrabodies based on intracellular capture frameworks that bind the RAS protein with high affinity and impair oncogenic transformation. *Embo Journal*, 22, 1025-1035.

TANAKA, T. & RABBITTS, T. H. (2008) Functional intracellular antibody fragments do not require invariant intra-domain disulfide bonds. *Journal of Molecular Biology*, 376, 749-757.

TANAKA, T. & RABBITTS, T. H. (2009) Selection of complementary single-variable domains for building monoclonal antibodies to native proteins. *Nucleic Acids Research*, 37.

TANAKA, T. & RABBITTS, T. H. (2010) Protocol for the selection of single-domain antibody fragments by third generation intracellular antibody capture. *NATURE PROTOCOLS*, 5, 67-92.

TANAKA, T., WILLIAMS, R. L. & RABBITTS, T. H. (2007) Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. *Embo Journal*, 26, 3250-3259.

TERANO, T., ZHONG, Y., TOYOKUNI, S., HIAI, H. & YAMADA, Y. (2005) Transcriptional control of fetal liver hematopoiesis: dominant negative effect of the overexpression of the LIM domain mutants of LMO2. *Experimental Hematology*, 33, 641-651.

TERESHKO, V., UYSAL, S., KOIDE, A., MARGALEF, K., KOIDE, S. & KOSSIAKOFF, A. A. (2008) Toward chaperone-assisted crystallography: Protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. *Protein Science*, 17, 1175-1187.

TIE, Y. F., BOROSS, P. I., WANG, Y. F., GADDIS, L., HUSSAIN, A. K., LESHCHENKO, S., GHOSHL, A. K., LOUIS, J. M., HARRISON, R. W. & WEBER, I. T. (2004) High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *Journal of Molecular Biology*, 338, 341-352.

TRIPIC, T., DENG, W., CHENG, Y., ZHANG, Y., VAKOC, C. R., GREGORY, G. D., HARDISON, R. C. & BLOBEL, G. A. (2009) SCL and associated proteins distinguish active from repressive GATA transcription factor complexes. *Blood*, 113, 2191-2201.

VAGIN, A. & TEPLYAKOV, A. (1997) MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography*, 30, 1022-1025.

References

- VAN MIERLO, C. P. M. & STEENSMA, E. (2000) Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story. *Journal of Biotechnology*, 79, 281-298.
- VAN VLIERBERGHE, P., BEVERLOO, H. B., BUIJS-GLADDINES, J., VAN WERING, E. R., HORSTMANN, M., PIETERS, R. & MEIJERINK, J. P. P. (2008) Monoallelic or biallelic LMO2 expression in relation to the LMO2 rearrangement status in pediatric T-cell acute lymphoblastic leukemia. *Leukemia*, 22, 1434-1437.
- VAN VLIERBERGHE, P., PIETERS, R., BEVERLOO, H. B. & MEIJERINK, J. P. P. (2008) Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. *British Journal of Haematology*, 143, 153-168.
- VASSILEV, L. T., VU, B. T., GRAVES, B., CARVAJAL, D., PODLASKI, F., FILIPOVIC, Z., KONG, N., KAMMLOTT, U., LUKACS, C., KLEIN, C., FOTOUHI, N. & LIU, E. A. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*, 303, 844-848.
- VENTERS, R. A., HUANG, C. C., FARMER, B. T., TROLARD, R., SPICER, L. D. & FIERKE, C. A. (1995) High-Level H-2/C-13/N-15 Labeling of Proteins for Nmr-Studies. *Journal of Biomolecular Nmr*, 5, 339-344.
- VISVADER, J. E., VENTER, D., HAHM, K., SANTAMARIA, M., SUM, E. Y. M., O'REILLY, L., WHITE, D., WILLIAMS, R., ARMES, J. & LINDEMAN, G. J. (2001) The LIM domain gene LMO4 inhibits differentiation of mammary epithelial cells in vitro and is overexpressed in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 14452-14457.
- VRANKEN, W., TOLKATCHEV, D., XU, P., TANHA, J., CHEN, Z. G., NARANG, S. & NI, F. (2002) Solution structure of a llama single-domain antibody with hydrophobic residues typical of the VH/VL interface. *Biochemistry*, 41, 8570-8579.
- VRANKEN, W. F., BOUCHER, W., STEVENS, T. J., FOGH, R. H., PAJON, A., LLINAS, P., ULRICH, E. L., MARKLEY, J. L., IONIDES, J. & LAUE, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins-Structure Function and Bioinformatics*, 59, 687-696.
- WADMAN, I. A., OSADA, H., GRUTZ, G. G., AGULNICK, A. D., WESTPHAL, H., FORSTER, A. & RABBITTS, T. H. (1997) The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *Embo Journal*, 16, 3145-3157.
- WALL, J. G. & PLUCKTHUN, A. (1995) Effects of Overexpressing Folding Modulators on the in-Vivo Folding of Heterologous Proteins in Escherichia-Coli. *Current Opinion in Biotechnology*, 6, 507-516.

References

- WALTER, T. S., DIPROSE, J., BROWN, J., PICKFORD, M., OWENS, R. J., STUART, D. I. & HARLOS, K. (2003) A procedure for setting up high-throughput nanolitre crystallization experiments. I. Protocol design and validation. *Journal of Applied Crystallography*, 36, 308-314.
- WALTER, T. S., DIPROSE, J. M., MAYO, C. J., SIEBOLD, C., PICKFORD, M. G., CARTER, L., SUTTON, G. C., BERROW, N. S., BROWN, J., BERRY, I. M., STEWART-JONES, G. B. E., GRIMES, J. M., STAMMERS, D. K., ESNOUF, R. M., JONES, E. Y., OWENS, R. J., STUART, D. I. & HARLOS, K. (2005) A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallographica Section D-Biological Crystallography*, 61, 651-657.
- WALTER, T. S., MEIER, C., ASSENBERG, R., AU, K. F., REN, J. S., VERMA, A., NETTLESHIP, J. E., OWENS, R. J., STUART, D. I. & GRIMES, J. M. (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure*, 14, 1617-1622.
- WANG, Q., ZHANG, M., WANG, X., YUAN, W., CHEN, D., ROYER-POKORA, B. & ZHU, T. (2007) A novel transcript of the LMO2 gene, LMO2-c, is regulated by GATA-1 and PU.1 and encodes an antagonist of LMO2. *Leukemia*, 21, 1015-1025.
- WARKE, A. & MOMANY, C. (2007) Addressing the protein crystallization bottleneck by cocrystallization. *Crystal Growth & Design*, 7, 2219-2225.
- WARREN, A. J., COLLEDGE, W. H., CARLTON, M. B. L., EVANS, M. J., SMITH, A. J. H. & RABBITS, T. H. (1994) THE ONCOGENIC CYSTEINE-RICH LIM DOMAIN PROTEIN RBTN2 IS ESSENTIAL FOR ERYTHROID DEVELOPMENT. *Cell*, 78, 45-57.
- WELLS, J. A. & MCCLENDON, C. L. (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, 450, 1001-1009.
- WHITFORD, D. (2005) *Proteins structure and function*, Chichester, John Wiley and Sons Ltd.
- WHITMORE, L. & WALLACE, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, 32, W668-W673.
- WILKINSON, I. C., HALL, C. J., VEVERKA, V., SHI, J. Y., MUSKETT, F. W., STEPHENS, P. E., TAYLOR, R. J., HENRY, A. J. & CARR, M. D. (2009) High Resolution NMR-based Model for the Structure of a scFv-IL-1 beta Complex POTENTIAL FOR NMR AS A KEY TOOL IN THERAPEUTIC ANTIBODY DESIGN AND DEVELOPMENT. *Journal of Biological Chemistry*, 284, 31928-31935.
- WILLIAMS, P. A., COSME, J., SRIDHAR, V., JOHNSON, E. F. & MCREE, D. E. (2000) Mammalian microsomal cytochrome P450 monooxygenase: Structural adaptations for membrane binding and functional diversity. *Molecular Cell*, 5, 121-131.

References

- WOLBER, G. & LANGER, T. (2005) LigandScout: 3-d pharmacophores derived from protein-bound Ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling*, 45, 160-169.
- XU, Z. X., MENG, X. Z., CAI, Y., LIANG, H., NAGARAJAN, L. & BRANDT, S. J. (2007) Single-stranded DNA-binding proteins regulate the abundance of LIM domain and LIM domain-binding proteins. *Genes & Development*, 21, 942-955.
- YAMADA, K., TSUKAHARA, T., YOSHINO, K., KOJIMA, K., AGAWA, H., YAMASHITA, Y., AMANO, Y., HATTA, M., MATSUZAKI, Y., KUROTORI, N., WAKUI, K., FUKUSHIMA, Y., OSADA, R., SHIOZAWA, T., SAKASHITA, K., KOIKE, K., KUMAKI, S., TANAKA, N. & TAKESHITA, T. (2009) Identification of a high incidence region for retroviral vector integration near exon 1 of the LMO2 locus. *Retrovirology*, 6.
- YAMADA, Y., PANNELL, R., FORSTER, A. & RABBITTS, T. H. (2000) The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 320-324.
- YAMADA, Y., PANNELL, R., FORSTER, A. & RABBITTS, T. H. (2002) The LIM-domain protein Lmo2 is a key regulator of tumour angiogenesis: a new anti-angiogenesis drug target. *Oncogene*, 21, 1309-1315.
- YAMADA, Y., WARREN, A. J., DOBSON, C., FORSTER, A., PANNELL, R. & RABBITTS, T. H. (1998) The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 3890-3895.
- YANG, Z. R., THOMSON, R., MCNEIL, P. & ESNOUF, R. M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 21, 3369-3376.
- YANO, J., HIRABAYASHI, K., NAKAGAWA, S., YAMAGUCHI, T., NOGAWA, M., KASHIMORI, I., NAITO, H., KITAGAWA, H., ISHIYAMA, K., OHGI, T. & IRIMURA, T. (2004) Antitumor activity of small interfering RNA/cationic liposome complex in mouse models of cancer. *Clinical Cancer Research*, 10, 7721-7726.
- ZHANG, Y. M. & ROWLEY, J. D. (2006) Chromatin structural elements and chromosomal translocations in leukemia. *DNA Repair*, 5, 1282-1297.
- ZHANG, Z. & PALZKILL, T. (2004) Dissecting the protein-protein interface between beta-lactamase inhibitory protein and class A beta-lactamases. *Journal of Biological Chemistry*, 279, 42860-42866.

References

ZHONG, Y., JIANG, L., HIAI, H., TOYOKUNI, S. & YAMADA, Y. (2007) Overexpression of a transcription factor LYL1 induces T- and B-cell lymphoma in mice. *Oncogene*, 26, 6937-6947.

ZHU, Y. M., ZHAO, W. L., FU, J. F., SHI, J. Y., PAN, Q., HU, J., GAO, X. D., CHEN, B., LI, J. M., XIONG, S. M., GU, L. J., TANG, J. Y., LIANG, H., JIANG, H., XUE, Y. Q., SHEN, Z. X., CHEN, Z. & CHEN, S. J. (2006) NOTCH1 mutations in T-cell acute lymphoblastic leukemia: Prognostic significance and implication in multifactorial leukemogenesis. *Clinical Cancer Research*, 12, 3043-3049.

Appendix A

Size exclusion chromatography

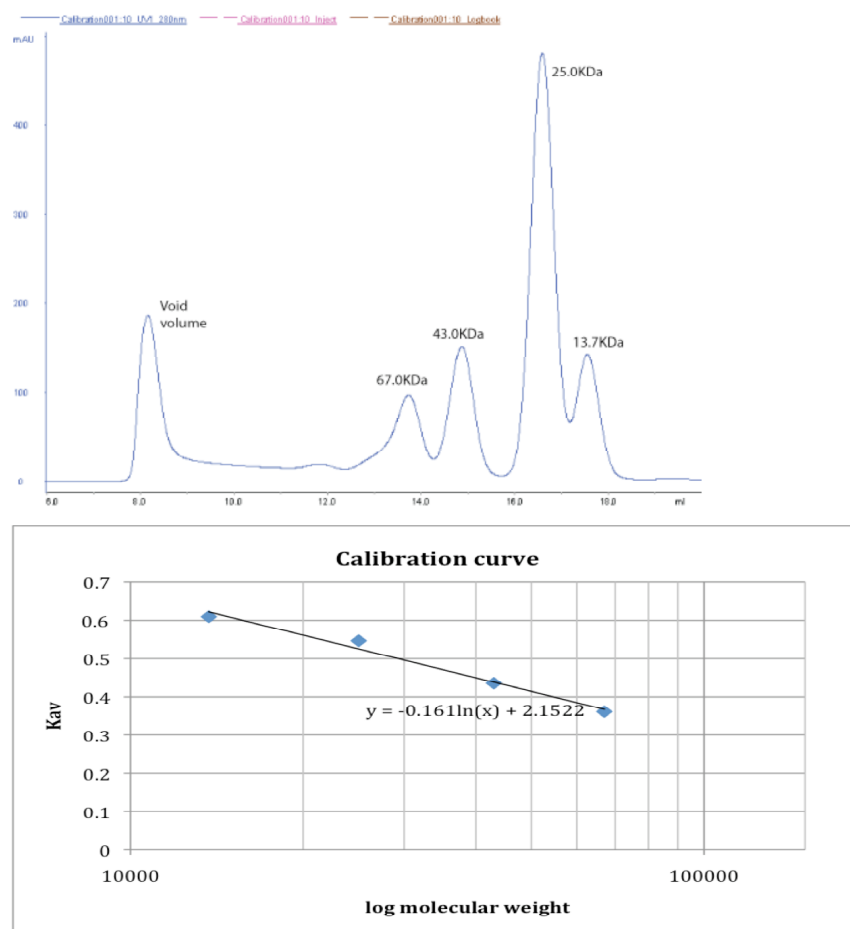
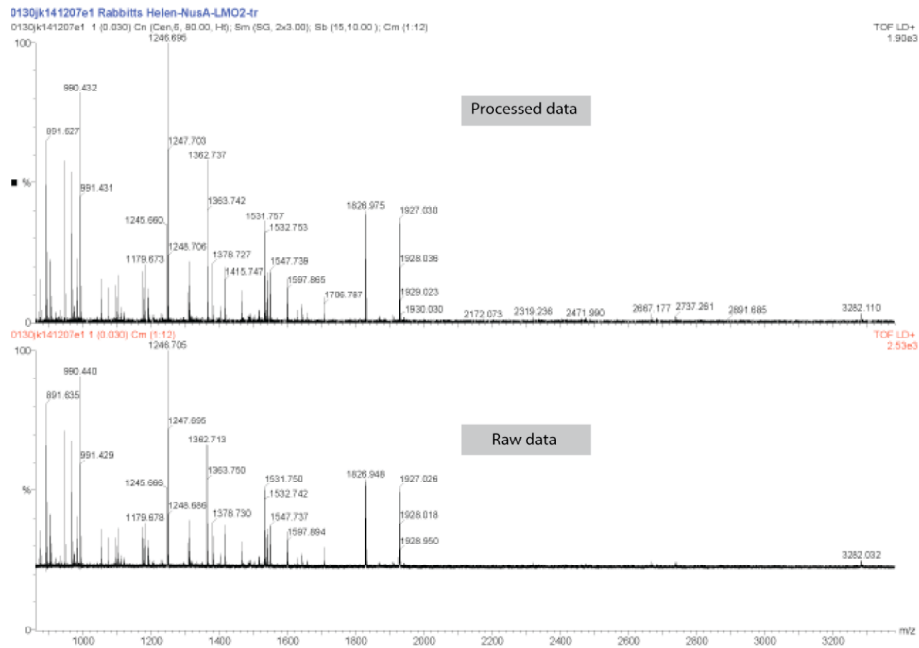


Figure 1: Calibration of Superdex 200 10/300 GL

A low molecular weight gel filtration kit (Amersham) was used to calibrate the Superdex 200 10/300 GL. A calibration curve was prepared by measuring the elution volumes of several protein standards, calculating their corresponding gel phase distribution co-efficient (K_{av}) and plotting the K_{av} values versus the logarithm of their molecular weight. The K_{av} was calculated for the protein of interest and the corresponding molecular weight extrapolated from the logarithmic scale.

Mass spectrometry

A:



B:

Sequence coverage:

```

MNKEILAVVEAVSNEKALPREKIFEALESALATATKKYEQEIDVRVQIDRKSGDFDFRRW
LVVDEVTQPTKEITLEAARYEDESINLGDYVEDQIESVTFDRITTTQAKQVIVQKVREAERA
MVVDQFREHEGEIITGVVKKVNRDNISLDLGNNAEAVILREDMLPRENFRPGDRVGVLYSV
RPEARGAQLFVTRSKPEMLIELFRIEVPEIGEEVIEIKAAARDPGSRAKIAVKTNDKRIDPV
GACVGMRGARVQAVSTELGGERIDIVLWDDNPAQFVINAMAPADVASIVVDEDKHTMDIAVE
AGNLAQAIGRNQGNVRLASQLSGWELNVMTVDDLQAKHQAAHAADITFTKYLDIDEDFATV
LVEEGFSTLEELAYVPMELLEIEGLDEPTVEALREKRNALATIAQAQEESELGDNKPADDL
LNLEGVDRDLAFKLAARGVCTLEDLAEQGIDDLADI EGLTDEKAGALIMAARNICWFGDEAT
SGSGHHHHHSAGKETAAAKFERQHMDSPPTGLVPRGSAGSGTIDDDDKMSSAIERKSLDP
SEEPVDEVLQIIPPSLLTCGGCCQNI GDRYFLKAIDQYWHEDCLSCDLGCRLGEVGRRLYYK
LGRKLCRRDYLRLFQDGLCASCDKRIAYEMTMRVKDKVVYHLECFRCAACQRHFCVGDRYL
LINSDIVCEQDIYEWTKINGIT
  
```

Peptide sequences covering 343/704 = 49% of the sequence

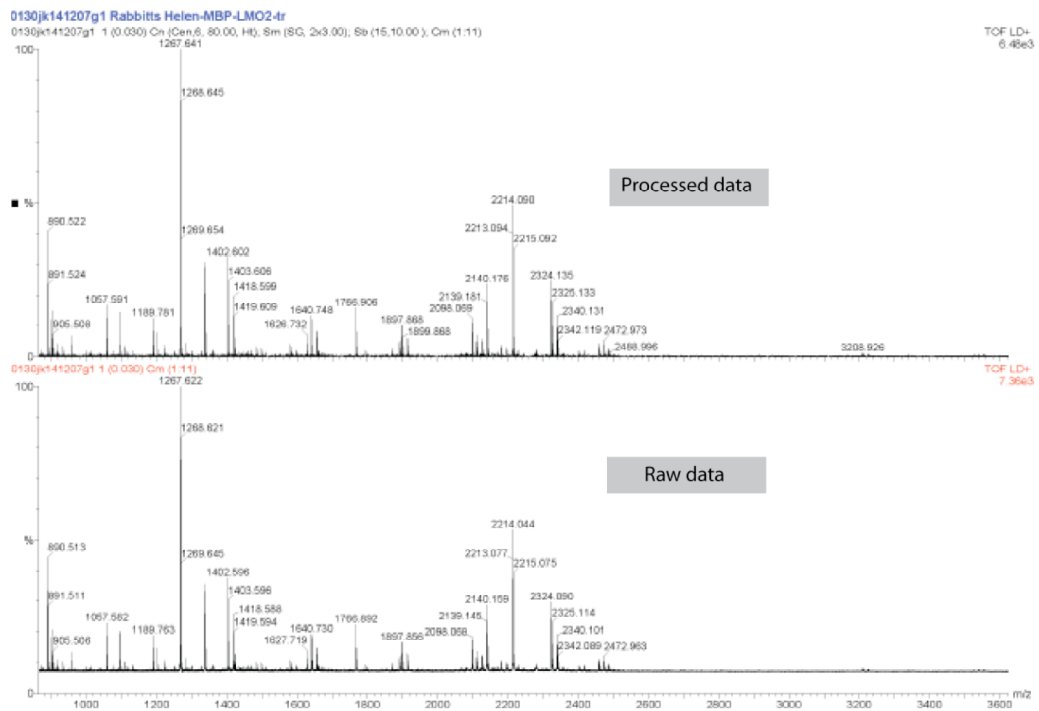
Figure 2: NusA-LMO2 MALDI-MS peptide mass fingerprinting

A: At least 100 laser shots were accumulated and combined to produce a raw spectrum (bottom). Spectra were processed (background subtraction, smoothing and peak centroiding) and calibrated externally using a tryptic digest of alcohol dehydrogenase.

B: Peptide peaks matched between predicted (Peptide Mass program) and the actual spectrum are shown in bold and underlined; they cover 49% of the sequence thus confirming the identity of NusA-LMO2.

Appendix A

A:



B:

Sequence coverage:

```

MAHHHHHSSGMKTEEGKLVIIWINGDKGYNGLAEVGGKKEKDTGIIKVTVEHFDKLEEKFPQVAATGGDPDIIFWA
HDFGGYAQSGLLAEITPDKAEQDKLYPFTWDVAVRYNGKLIAYPIAVEALSLEYNKDLLPNPPKRWREIPALDKE
LKAQKGSALMFNLQEPYFTWFLIADGGYAFKYENKDYIKDVGVDNAGAKAGLTFVLVDLIRKKNHMNADTDYSIA
EAAFNKGETAMTINGPFAWNSHIDTSKVNYGVTVLPTFKGQPSKFFVGVLSAGINAASFNKELAKEFLENYLLTDE
GLEAVNKDKPLGAVALKSYEELAKDPRIATMENAQKGEIMPNI PQMSAFYAVRTAVINAASGRQTVDEALKD
AQTSSGLEVLVFGTMSAIERKSLDPSEFPVDEVLIQIPPSLLTCGGCQQNIGDRYFLKAIIDQYWHEDCLSCDLGG
CRLGEVGRRLYYKLRKLCRRDYLRLFGDGLCASCDRIRAYEMMRVKDKVYHLECFKCAACQKHFCVGDRYL
LINSDIVCBQDIYEWTKINGII
    
```

Peptide sequences covering 327/547 = 60% of the sequence

Figure 3: MBP-LMO2 MALDI-MS peptide mass fingerprinting

A: Raw and processed spectra.

B: Peptide peaks matched between predicted (Peptide Mass program) and the actual spectrum are shown in bold and underlined; they cover 60% of the sequence and thus confirm the identity of MBP-LMO2.

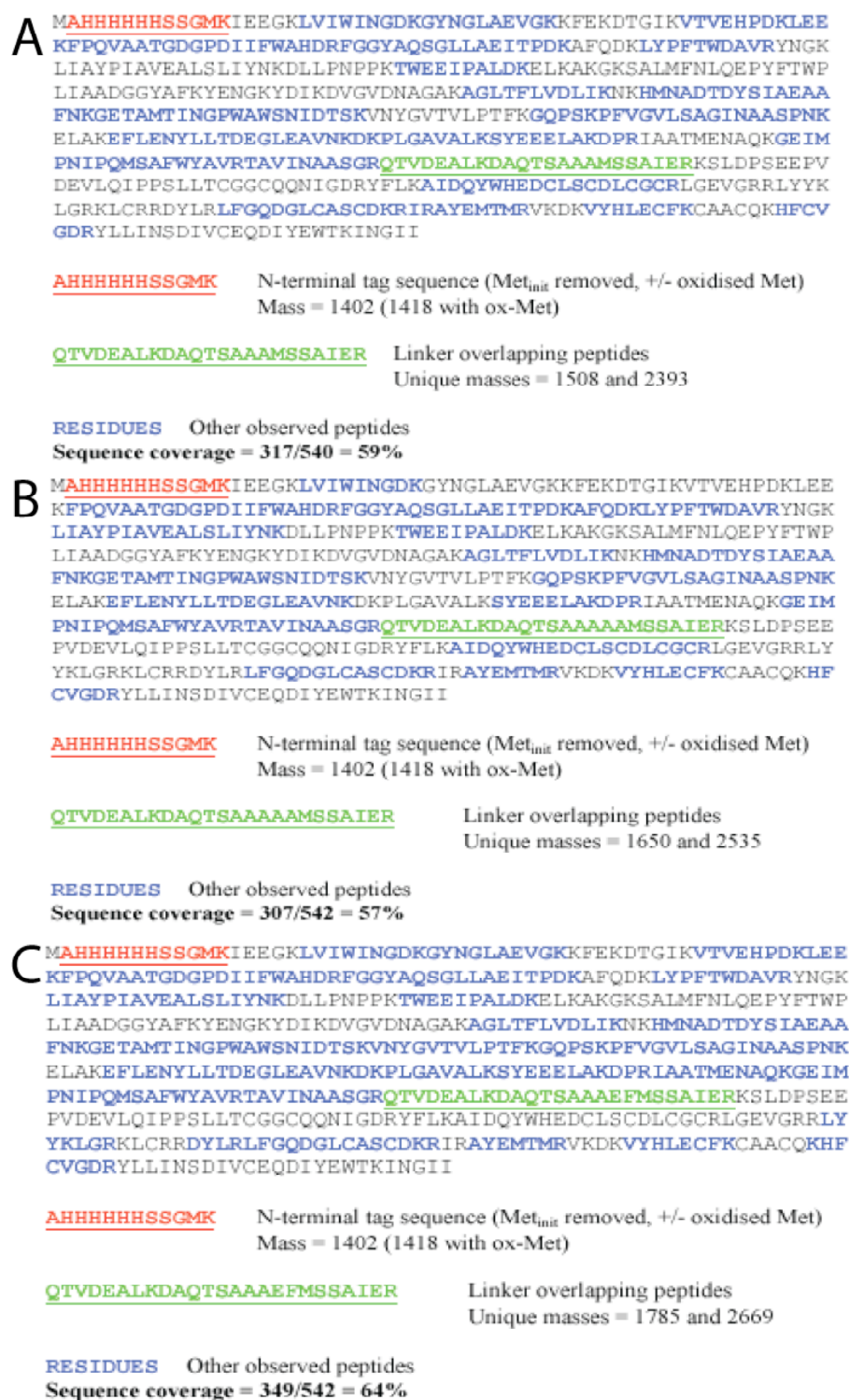
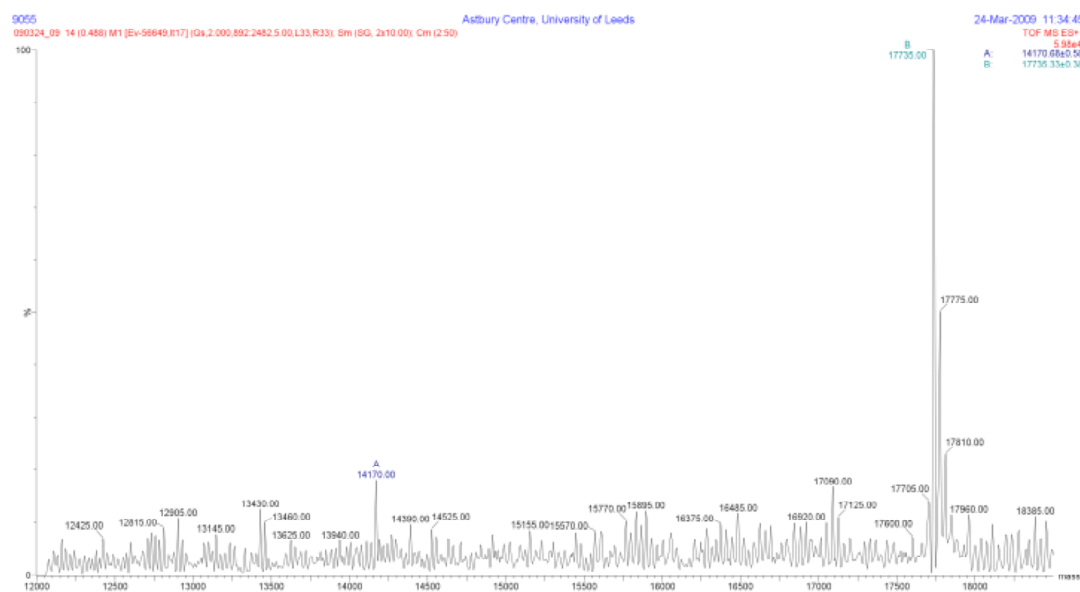


Figure 4: MBP-LMO2 rigid linker MALDI-MS fingerprinting

Identity of the MBP-LMO2 rigid linker constructs was confirmed by MALSI-MS fingerprinting. Peptide peaks matched between predicted (Peptide Mass program) and spectrum are shown in colour and the sequence coverage indicated in bold.

Appendix A

A:



B:

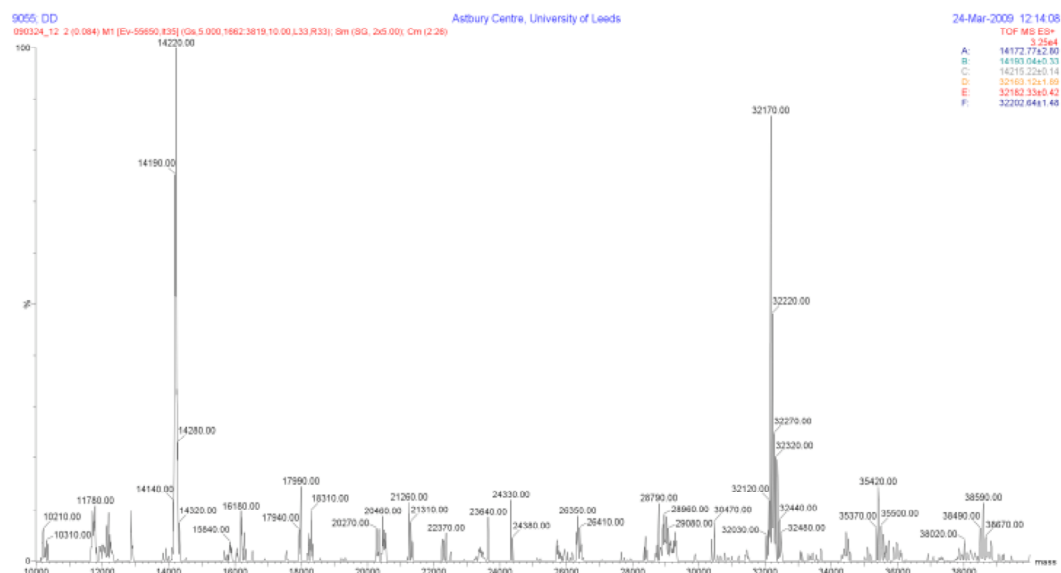
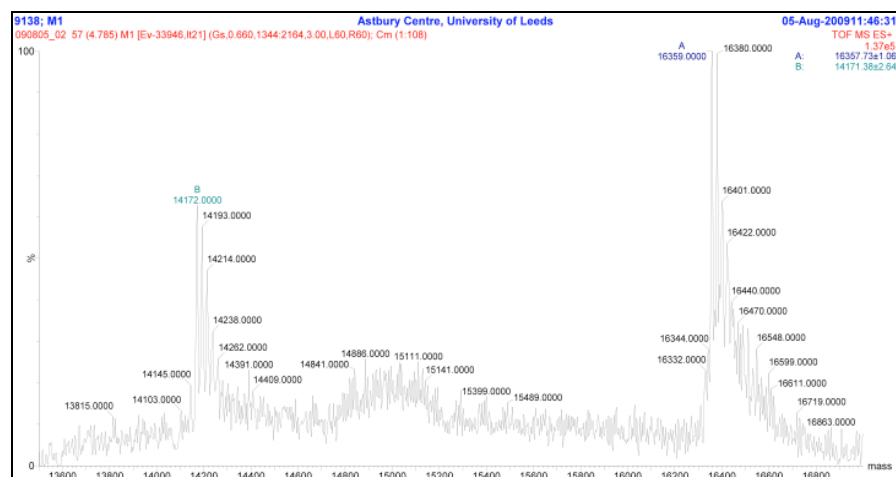


Figure 5: Positive electrospray mass spectrometry analysis of lysine methylated VH#576/LMO2ΔN7

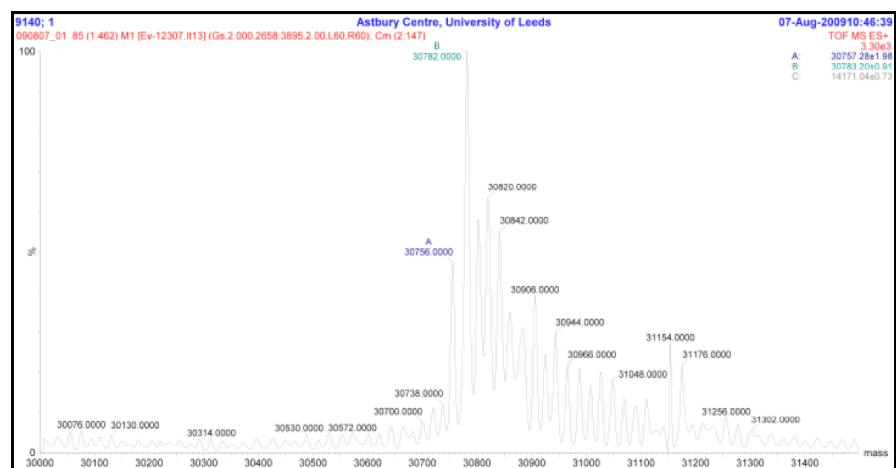
Mass spectrometry was used to analyse methylated VH#576/LMO2ΔN7 under denaturing (A) and native (B) conditions. For each lysine methylated the molecular mass is increased by 28 Da therefore calculations indicated VH#576 had four lysines methylated and LMO2 had ten. This indicated that all lysine residues in the complex had been methylated plus the N-terminal amine of both proteins.

Appendix A

A:



B:



C:

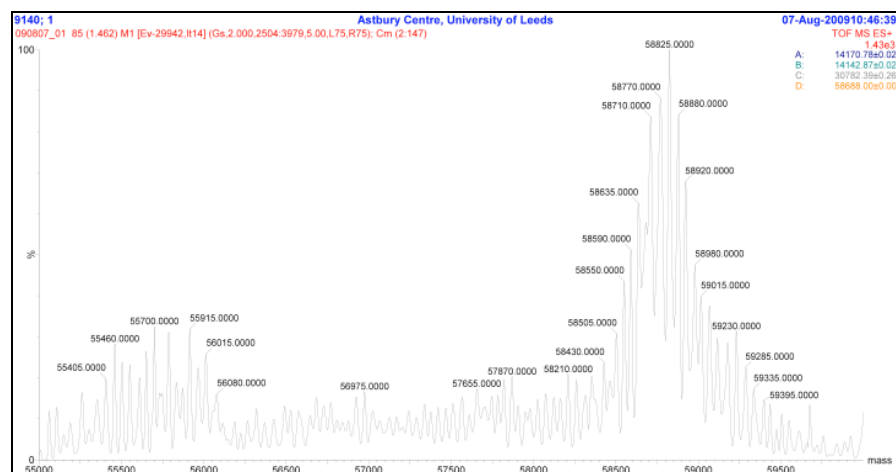
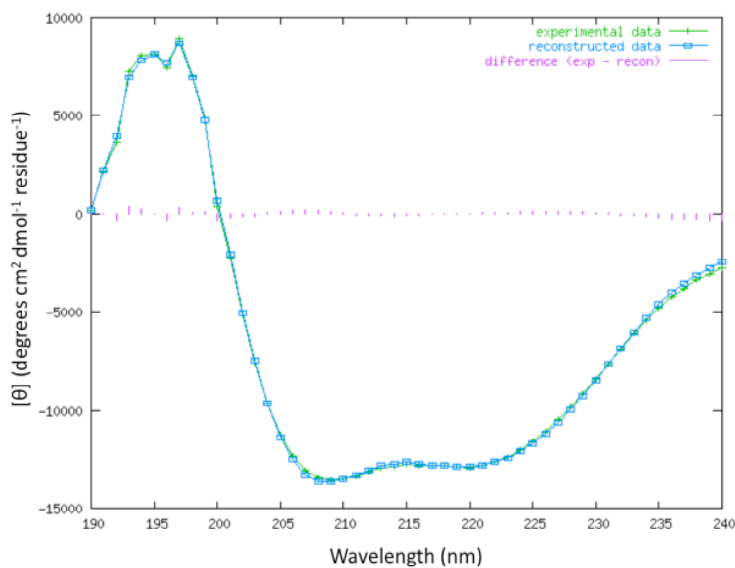


Figure 6: Positive electrospray mass spectrometry analysis of lysine methylated VH#576/LMO2AN7AC11 under denaturing (A) and native (B and C) conditions.

Circular Dichroism spectropolarimetry

A:



B:

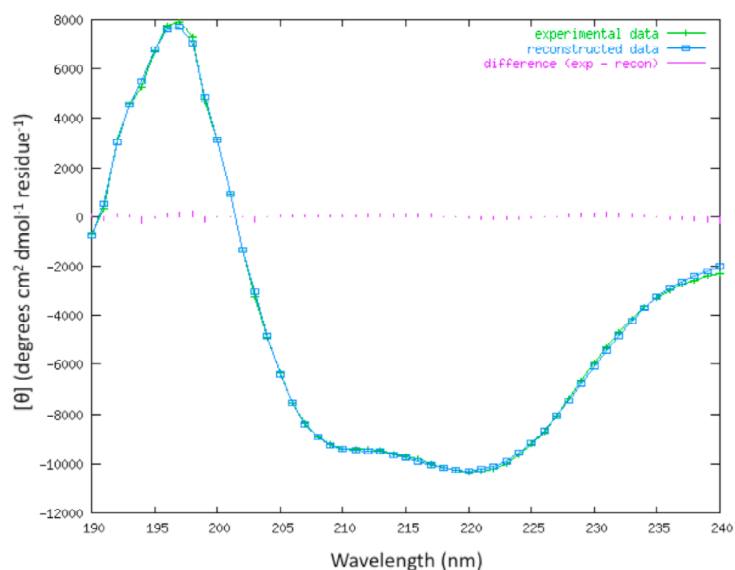


Figure 7: Circular Dichroism spectropolarimetry analysis of NusA-LMO2 and MBP-LMO2

NusA-LMO2 is represented by the top spectrum and MBP-LMO2 the bottom spectrum. Spectra were analysed by CDSSTR (Johnson, 1999) obtained from DICHROWEB (Whitmore and Wallace, 2004). Experimental data is shown in green and analysis by CDSSTR method is shown in blue. The characteristic minimum at around 215nm was found for both recombinant proteins, which is typical for a structure with considerable beta sheet content. From this data, we can confirm that both proteins are folded as the spectra of both are negative at more than 208 nm.

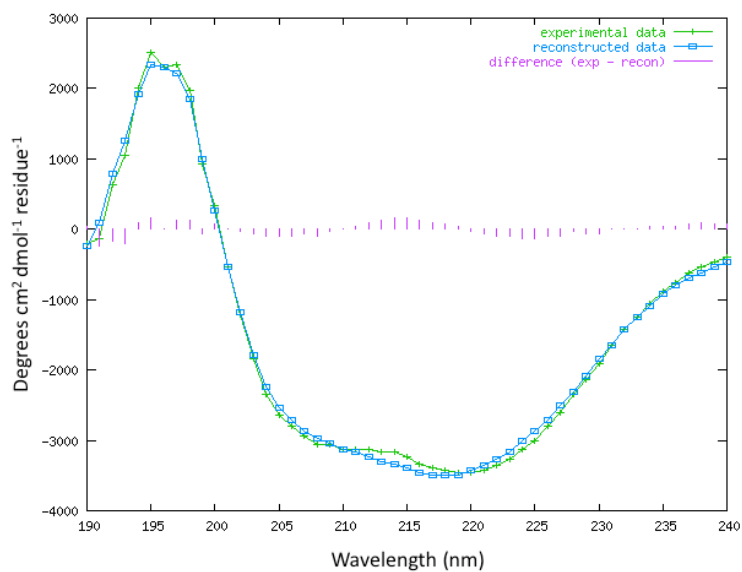


Figure 8: Circular Dichroism spectropolarimetry of VH/LMO2ΔN7ΔC11 complex

The spectrum was analysed by CDSSTR obtained from DICHROWEB. Experimental data is shown in green and analysis by CDSSTR method is shown in blue. The VH#576/LMO2ΔN7ΔC11 spectrum shows the characteristic minimum at around 215nm, which is typical for a structure with considerable beta sheet content. From this data, we can confirm that the protein complex is folded as the spectrum is negative at more than 208 nm.

Crystallisation reagents optimised

Table 1: Optimisation of crystallisation conditions

Plate Number	Well	Stars 1=good 3=excellent
298955 MPD no. A2 3 row dilution Original condition: MPD 10%, 100mM Na acetate pH 5.0	H11	**
	H12	**
298962 Index no. 56 3 row pH	Nothing interesting	
298948 Ammonium sulphate A3 additive screen Original condition: 0.8M Ammonium sulphate, MES pH 6	A8	*
	F2	**
	H4	**
	H7	***
	H8	***
	H9	*
298979 Natrx no. 38 3 row dilution Original condition: PEG 4000 5%, 200mM Ammonium acetate, 150mM Mg acetate, 50mM HEPES-Na pH 7.0	Rod cluster (perform additive screen)	
	H10	**
298986 Natrx no. 3 3 row dilution Original condition: MPD 20% v/v, 100mM Mg acetate tetrahydrate, 50mM MES monohydrate pH 5.6 (acts as cryo-protectant)	H9	**
	H10	***
298917 Natrx no.5 additive screen Original condition: PEG 8000 5% w/v, 10mM MgCl hexahydrate, 200mM KCl, 50mM MES monohydrate pH 5.6	B9	*
	C10	*
	E6	**
	F4	**
	G3	**
	H4	**
298023 Natrx no.5 Original screen Original condition as above	A5	***
298931 Natrx no.5 3 row pH Original condition as above	Nothing interesting	
298924 Natrx no.5 3 row dilution Original condition as above	Nothing interesting	

Appendix B

Alignment of VH#576 with anti-Ras antibody single domain, VH#6

```
1          10          20          30          40
EVQLLESGGGLVQPGGSLRLSCAASGFSFSHSPMNWVRQAPGKGLEWV

EVQLLESGGGLVQPGGSLRLSCAASGFTFSTFSMNWVRQAPGKGLEWV

50          60          70          80          90
SYISYNSSSIYYADSVKGRFTISRDN SKNTLYLQMNSLRAEDTAVYYC

SYISRTSKTIYYADSVKGRFTISRDN SKNTLYLQMNSLRAEDTAVYYC

100          110          120
ARGLTESLELTADWFDYWGQGTLLVTVS

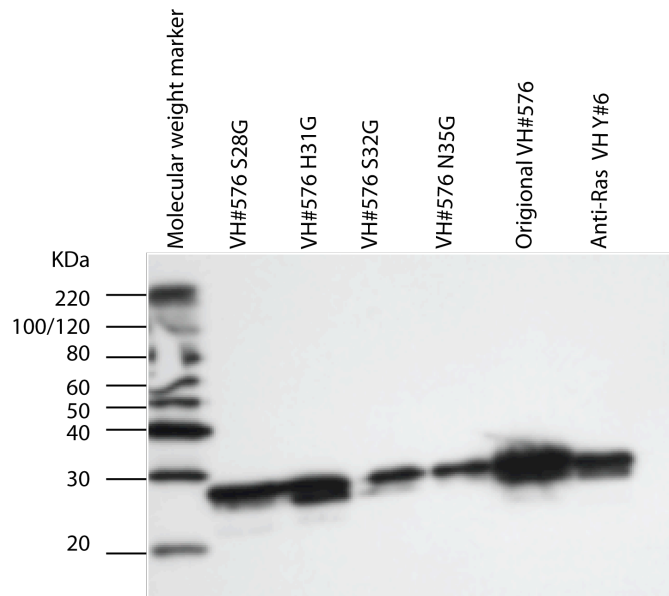
ARGRFDY*****WGQGTLLVTVS
```

Figure 1: Alignment of anti-Ras VH#6 and anti-LMO2 VH#576

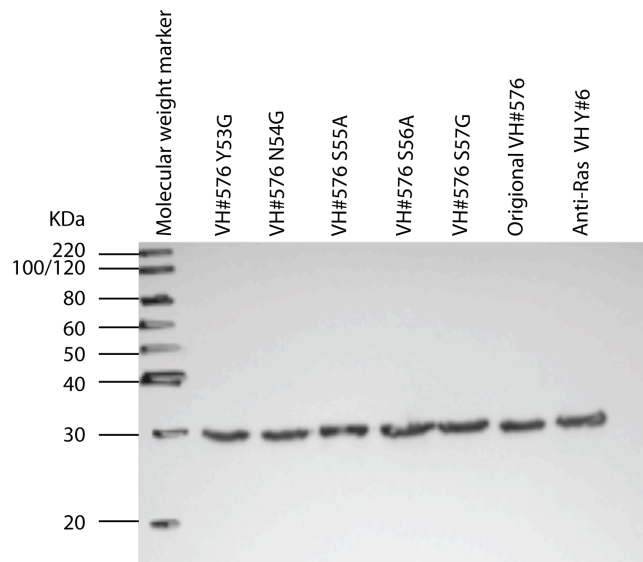
The amino acid sequence (in single letter code) of anti-Ras, Y#6 (top), and anti-LMO2, VH#576 (bottom). The CDRs are coloured in yellow and differences in the amino acid sequences are highlighted in red.

Western blots

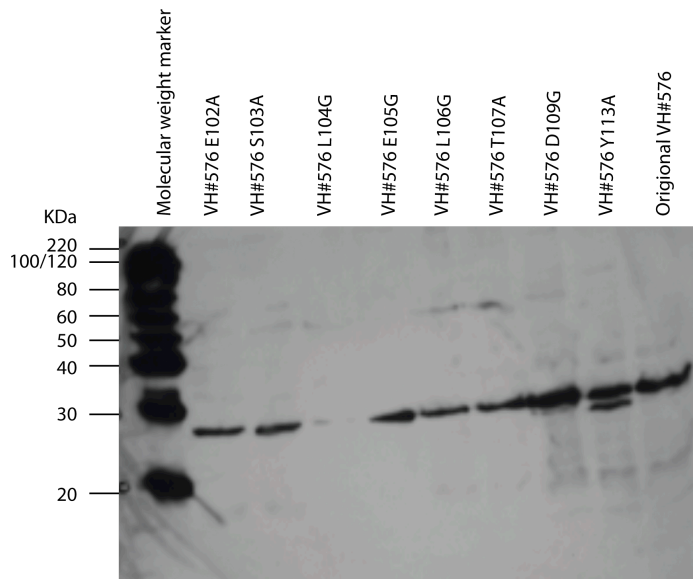
CDR1; round one VH#576 mutations



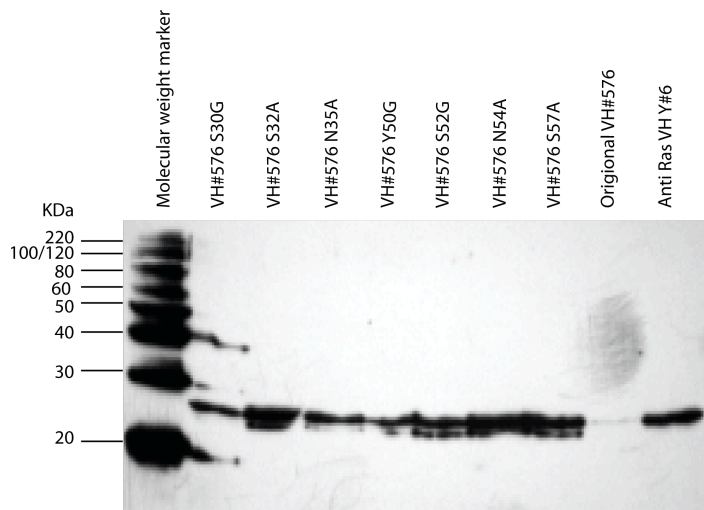
CDR2; round two VH#576 mutations



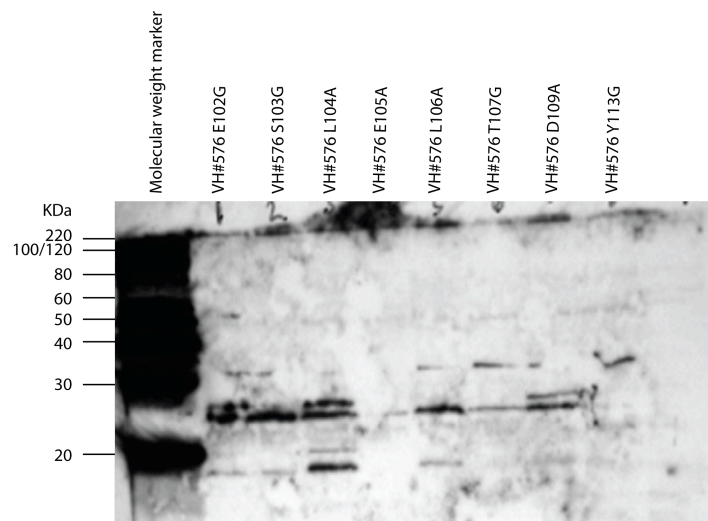
CDR3; round three VH#576 mutations



CDR1 and 2; round four VH#576 mutations



CDR3; round five VH#576 mutations



Appendix C

Expression vector maps

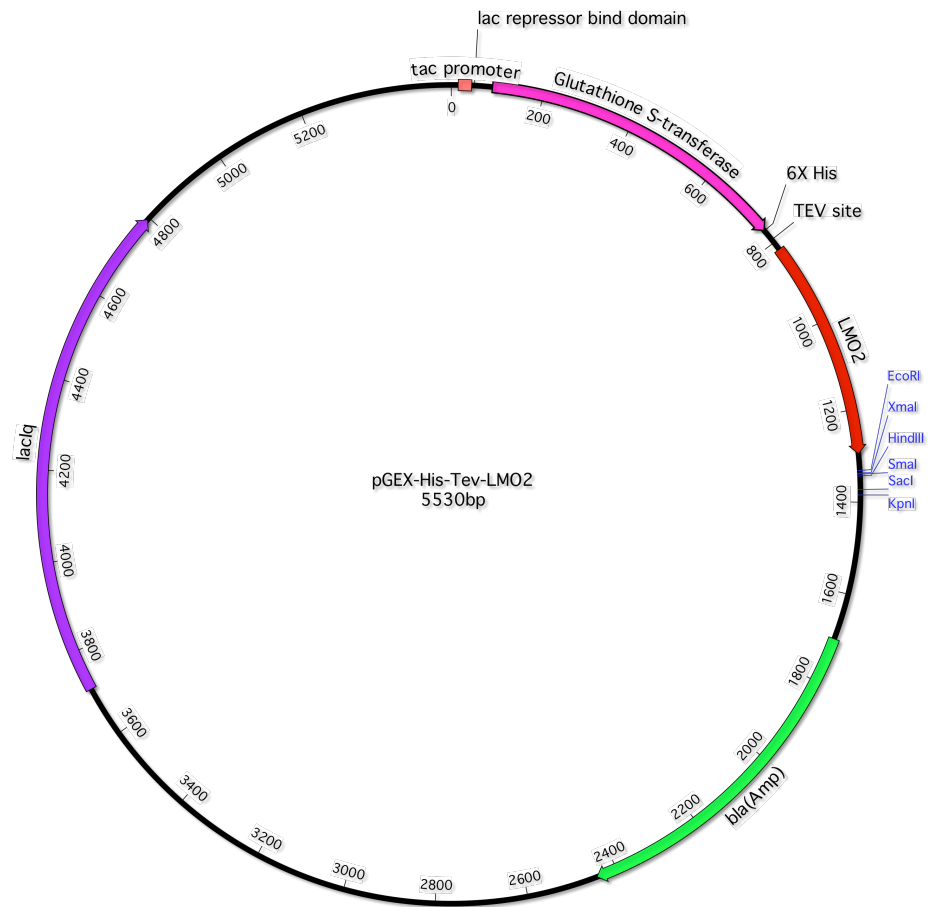


Figure 1: Map of *E. coli* expression vector pGEX-His-Tev-LMO2

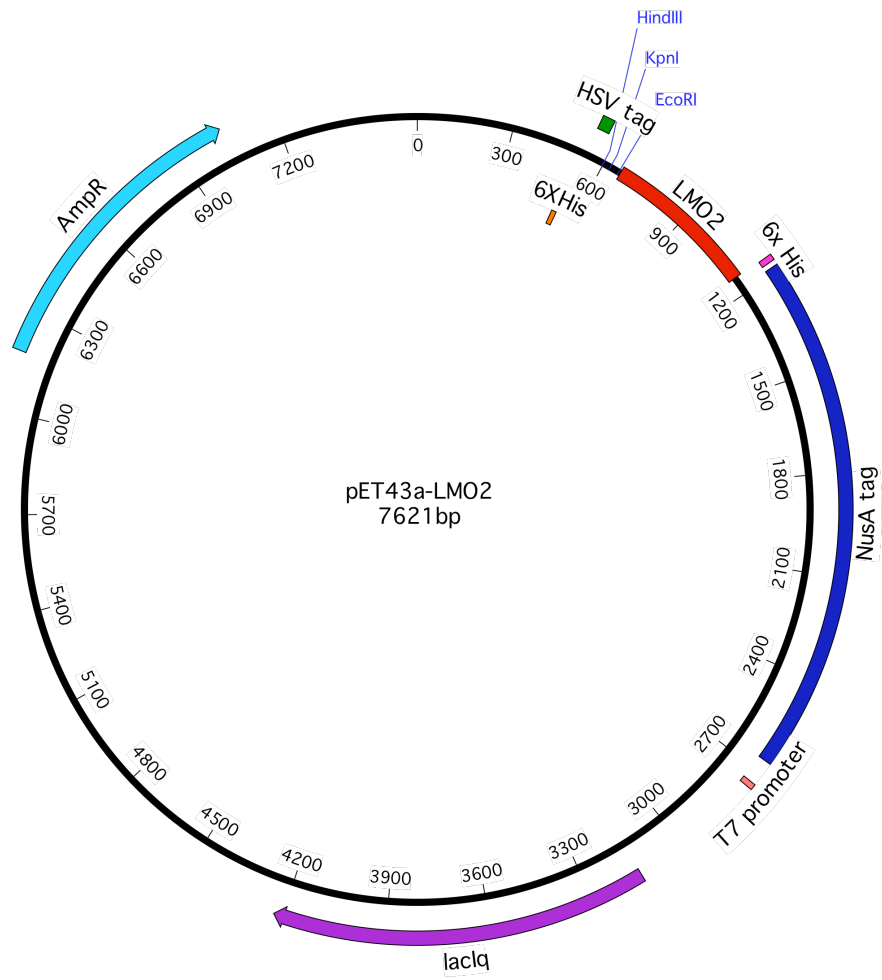


Figure 2: Map of *E. coli* expression vector pET43a-LMO2

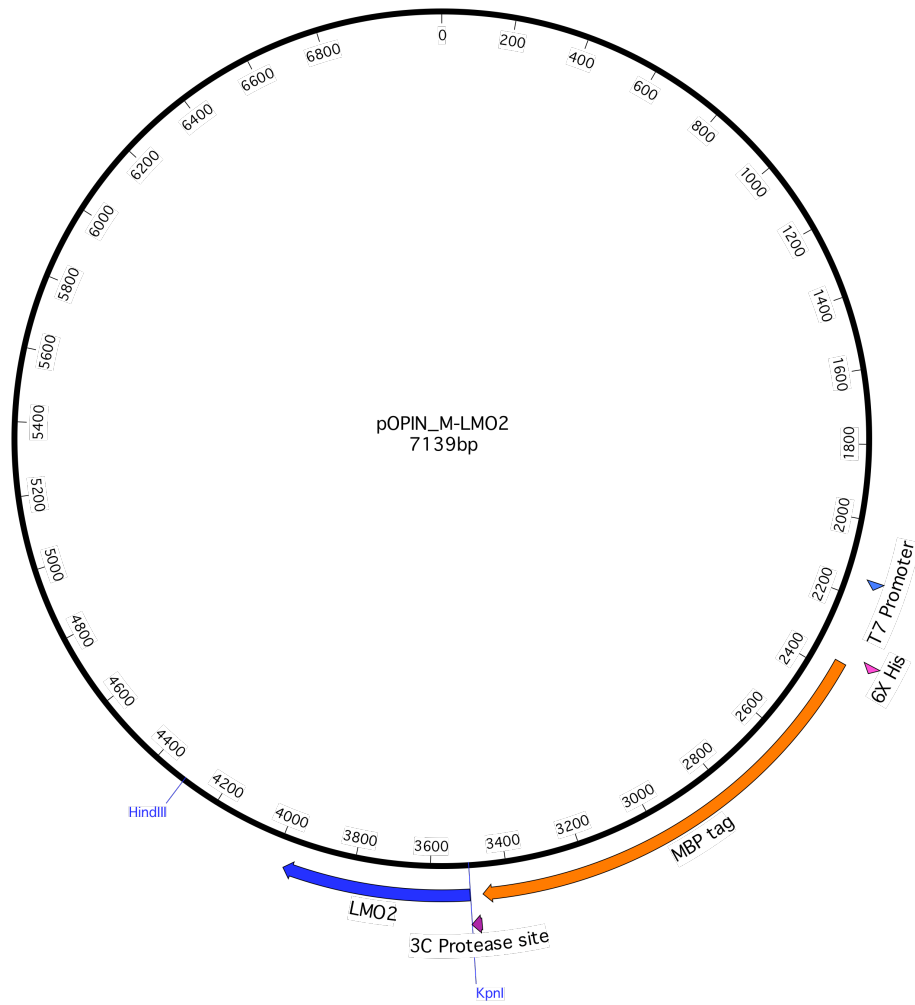


Figure 3: Map of *E. coli* expression vector pOPIN_M-LMO2

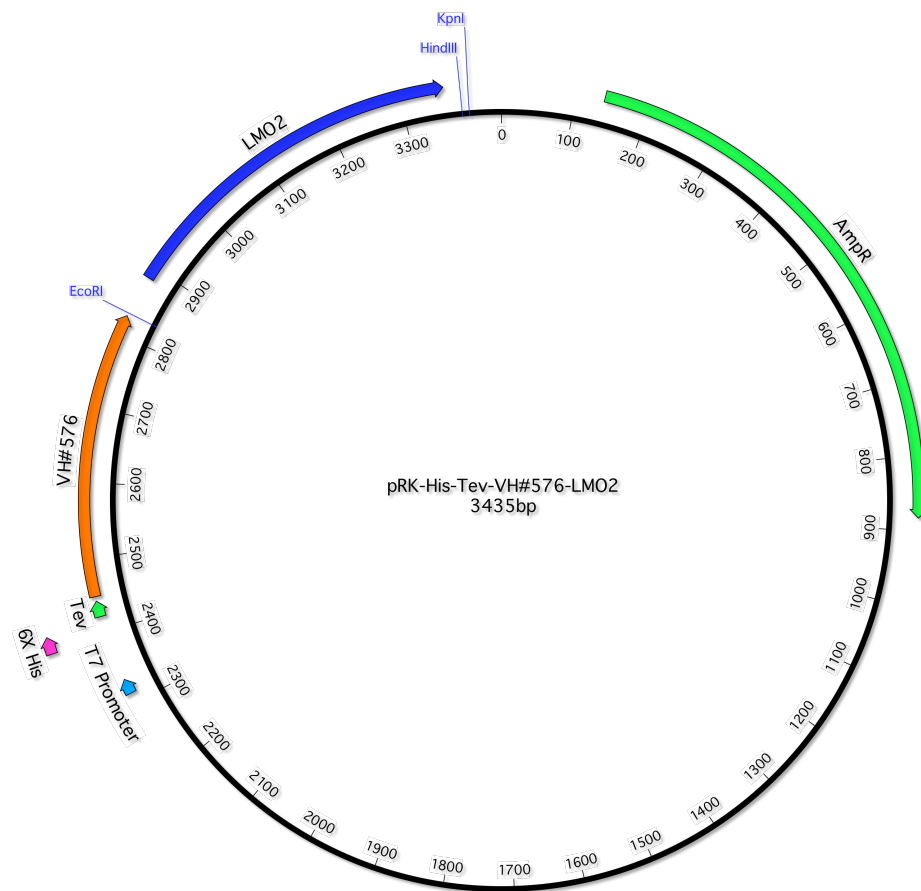


Figure 4: Map of *E. coli* expression vector pRK-His-Tev-VH#576-LMO2

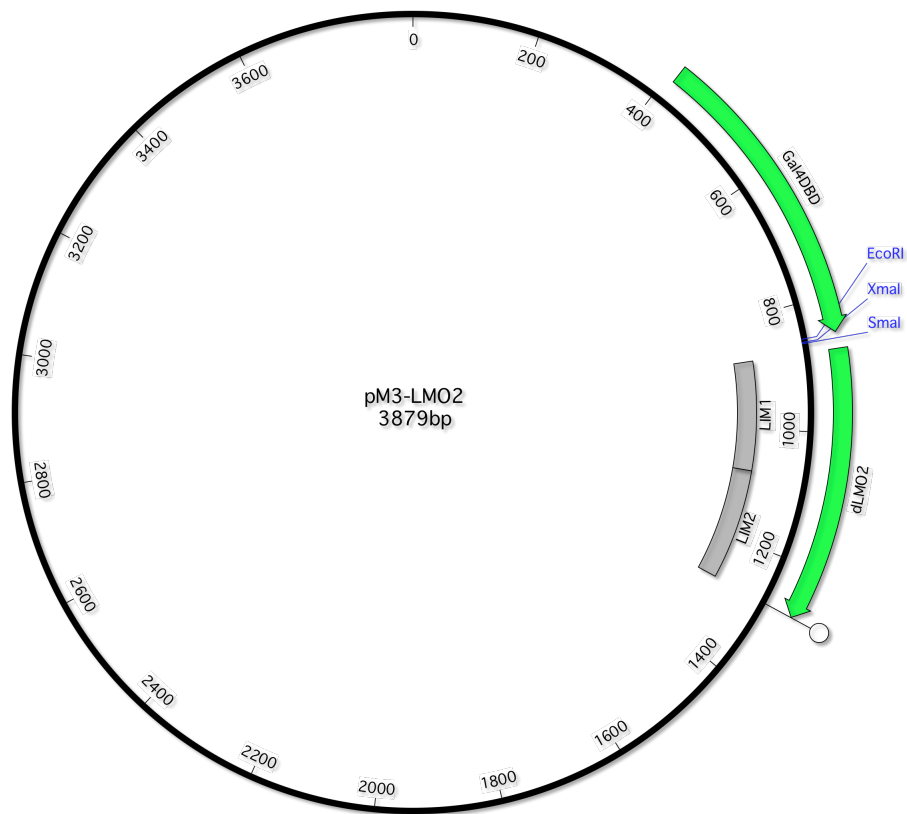


Figure 5: Map of mammalian expression vector PM3-LMO2 Δ N24

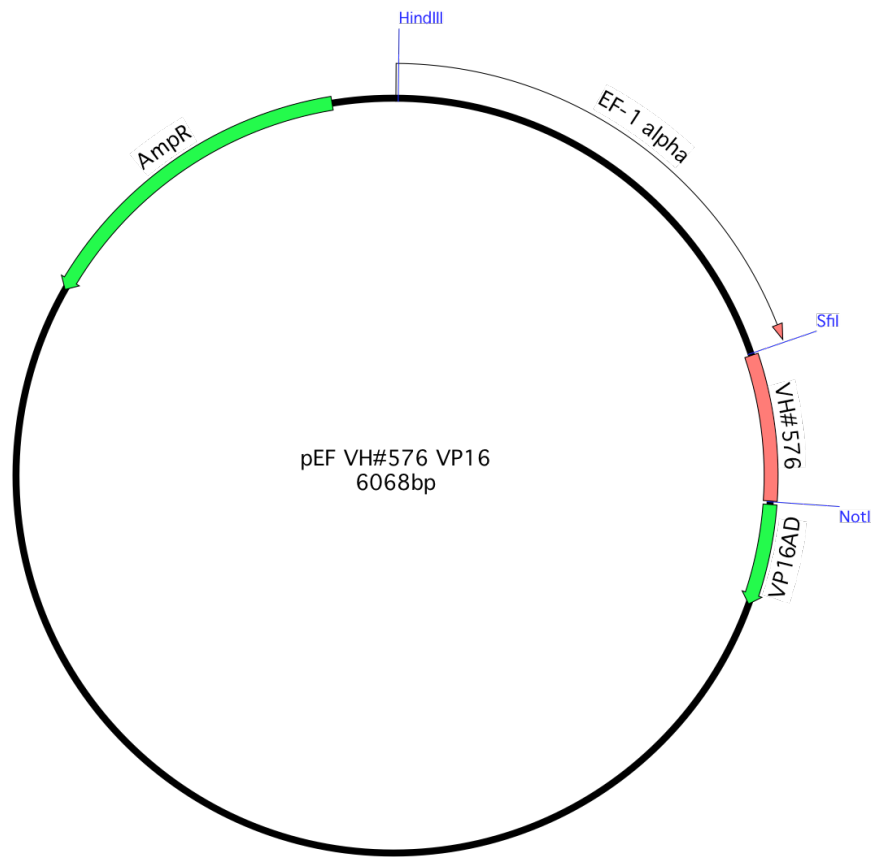


Figure 6: Map of mammalian expression vector pEF-VH#576-VP16

List of primers

Chapter three

Domain	N-terminal fusion tag	C-terminal fusion tag	Forward primer	Reverse primer	PCR product (bp)	Resistance
LMO2		KHis6	AGGAGATATACCATGATGTCCTCGGCCATCGAA	GTGATGGTGATGTTTCTAGATGATCCCATTGATCTTGG	507	Carbenicillin
LMO2	MAHHHHHHSSGLEVLFGGP		AAGTTCTGTTTCAGGGCCCCGATGTCCTCGGCCATCGAA	ATGGTCTAGAAAGCTTTACTAGATGATCCCATTGATCT	515	Carbenicillin
LMO2	HIS-GST3C		AAGTTCTGTTTCAGGGCCCCGATGTCCTCGGCCATCGAA	ATGGTCTAGAAAGCTTTACTAGATGATCCCATTGATCT	515	Carbenicillin
LMO2	HIS-MBP-3C		GCGAACAGATCGGTGGTATGTCCTCGGCCATCGAA	ATGGTCTAGAAAGCTTTACTAGATGATCCCATTGATCT	515	Carbenicillin
LMO2	HIS-SUMO		GCGAACAGATCGGTGGTATGTCCTCGGCCATCGAA	ATGGTCTAGAAAGCTTTACTAGATGATCCCATTGATCT	512	Carbenicillin
Tal1		KHis6	AGGAGATATACCATGACCAAAGTTGTGCGGCGTAT	GTGATGGTGATGTTTGTAGCAGCTTGGCCAAGAA	210	Carbenicillin
Tal1	MAHHHHHHSSGLEVLFGGP		AAGTTCTGTTTCAGGGCCCCGACCAAAGTTGTGCGGCGTAT	ATGGTCTAGAAAGCTTTAGAGCAGCTTGGCCAAGAA	218	Carbenicillin
Tal1	HIS-GST3C		AAGTTCTGTTTCAGGGCCCCGACCAAAGTTGTGCGGCGTAT	ATGGTCTAGAAAGCTTTAGAGCAGCTTGGCCAAGAA	218	Carbenicillin
Tal1	HIS-MBP-3C		AAGTTCTGTTTCAGGGCCCCGACCAAAGTTGTGCGGCGTAT	ATGGTCTAGAAAGCTTTAGAGCAGCTTGGCCAAGAA	218	Carbenicillin
Tal1	HIS-SUMO		GCGAACAGATCGGTGGTACCAAAGTTGTGCGGCGTAT	ATGGTCTAGAAAGCTTTAGAGCAGCTTGGCCAAGAA	215	Carbenicillin
Ldb1		KHis6	AGGAGATATACCATGATGCTGGATCGGGATGTG	GTGATGGTGATGTTTCACTGGGAAGCCTGTGAC	1158	Carbenicillin
Ldb1	MAHHHHHHSSGLEVLFGGP		AAGTTCTGTTTCAGGGCCCCGATGCTGGATCGGGATGTG	ATGGTCTAGAAAGCTTTATCACTGGGAAGCCTGTGAC	1166	Carbenicillin
Ldb1	HIS-GST3C		AAGTTCTGTTTCAGGGCCCCGATGCTGGATCGGGATGTG	ATGGTCTAGAAAGCTTTATCACTGGGAAGCCTGTGAC	1166	Carbenicillin
Ldb1	HIS-MBP-3C		AAGTTCTGTTTCAGGGCCCCGATGCTGGATCGGGATGTG	ATGGTCTAGAAAGCTTTATCACTGGGAAGCCTGTGAC	1166	Carbenicillin
Ldb1	HIS-SUMO		GCGAACAGATCGGTATGCTGGATCGGGATGTG	ATGGTCTAGAAAGCTTTATCACTGGGAAGCCTGTGAC	1163	Carbenicillin
LMO2	HIS-GST3C		AAGTTCTGTTTCAGGGCCCCGATGTCCTCGGCCATCGAA	ATGGTCTAGAAAGCTTTACTAGATGATCCCATTGATCT	515	Kanomycin
Tal1	HIS-GST3C		AAGTTCTGTTTCAGGGCCCCGACCAAAGTTGTGCGGCGTAT	ATGGTCTAGAAAGCTTTAGAGCAGCTTGGCCAAGAA	218	Kanomycin
Ldb1	HIS-GST3C		AAGTTCTGTTTCAGGGCCCCGATGCTGGATCGGGATGTG	ATGGTCTAGAAAGCTTTATCACTGGGAAGCCTGTGAC	1166	Kanomycin
	T7 forward primer		TAATACGACTCACTATAGG			

Appendix C

pOPIN_M NotI mutagenesis	
Forward (5' to 3')	ACG CGC AGA CTA GCG CGG CCG CCT GGA ACT
Reverse (5' to 3')	CAG AAC TTC CAG GCG GCC GCG CTA GTC TGC
Assembly PCR primers for the previous construct	
Forward (5' to 3')	GAG ATA CCA TGG CAC ACC ATC
Reverse (5' to 3')	CCA GTC ACG ACG TTG TAA AAC
MBP-AAA-LMO2	
Forward (5' to 3')	GCG GCC GCC ATG TCC TCG GCC ATC
MBP-AAAAA-LMO2	
Forward (5' to 3')	GCG GCC GCC GCC GCC ATG TCC TCG GCC ATC
MBP-AAAEF-LMO2	
Forward (5' to 3')	GCG GCC GCC GAA TTT ATG TCC TCG GCC ATC
pRK-His-Tev-VH#576-LMO2-7N mutagenesis	
Forward (5' to 3')	GGA GAT ATA CAT ATG AGC CTG GAC CCT TCA
Reverse (5' to 3')	TGA AGG GTC CAG GCT CAT ATG TAT ATC TCC
pRK-His-Tev-VH#576-LMO2-11N mutagenesis	
Forward (5' to 3')	GGA GAT ATA CAT ATG TCA GAG GAA CCA GTG
Reverse (5' to 3')	CAC TGG TTC CTC TGA CAT ATG TAT ATC TCC
pRK-His-Tev-VH#576-LMO2-13N mutagenesis	
Forward (5' to 3')	GGA GAT ATA CAT ATG GAA CCA GTG GAT GAG
Reverse (5' to 3')	CTC ATC CAC TGG TTC CAT ATG TAT ATC TCC
pRK-His-Tev-VH#576-LMO2-16N mutagenesis	
Forward (5' to 3')	GGA GAT ATA CAT ATG GAT GAG GTC CTG CAG
Reverse (5' to 3')	CTG CAG GAC CTC ATC CAT ATG TAT ATC TCC
pRK-His-Tev-VH#576-LMO2-20N mutagenesis	
Forward (5' to 3')	GGA GAT ATA CAT ATG CTG CAG ATC CCC CCA TCC CTG CTG ACA TGC GGC GGC TGC CAG
Reverse (5' to 3')	CTG GCA GCC GCC GCA TGT CAG CAG GGA TGG GGG GAT CTG CAG CAT ATG TAT ATC TCC
pRK-His-Tev-VH#576-LMO2-25N mutagenesis	
Forward (5' to 3')	GGA GAT ATA CAT ATG TCC CTG CTG ACA TGC GGC GGC TGC CAG
Reverse (5' to 3')	CTG GCA GCC GCC GCA TGT CAG CAG GGA CAT ATG TAT ATC TCC
pRK-His-Tev-VH#576-LMO2-29N mutagenesis	

Forward (5' to 3')	ATA CAT ATG TGC GGC GGC TGC CAG CAG AAC ATT
Reverse (5' to 3')	CGC GTG GTA CCA AGC TTA C
Assembly PCR primers for the previous constructs	
Forward (5' to 3')	CAT ATG AGA GGA TCG CAT CAC CA
Reverse (5' to 3')	GCG CGC ACG CGT GGT A
pRK-His-Tev-VH#576-LMO2-7N-2C mutagenesis	
Forward (5' to 3')	ACT AAG ATC AAT GGG TAG ATA TAG TGT ACA GGA
Reverse (5' to 3')	TCC TGT ACA CTA TAT CTA CCC ATT GAT CTT AGT
pRK-His-Tev-VH#576-LMO2-7N-7C mutagenesis	
Forward (5' to 3')	GAC ATT TAC GAG TGG TAG AAG ATC AAT GGG ATG
Reverse (5' to 3')	CAT CCC ATT GAT CTT CTA CCA CTC GTA AAT GTC
pRK-His-Tev-VH#576-LMO2-7N-8C mutagenesis	
Forward (5' to 3')	CAG GAC ATT TAC GAG TAG ACT AAG ATC AAT GGG
Reverse (5' to 3')	CCC ATT GAT CTT AGT CTA CTC GTA AAT GTC CTG
pRK-His-Tev-VH#576-LMO2-7N-11C mutagenesis	
Forward (5' to 3')	GTG TGC GAA CAG GAC TAG TAC GAG TGG ACT AAG
Reverse (5' to 3')	CTTAGTCCACTCGTACTAGTCCTGTTCGCACAC
Assembly PCR primers for the previous constructs	
Forward (5' to 3')	CCG TCA TCA CCG AAA CG
Reverse (5' to 3')	TCC TTT CGG GCT TTG TTA GA

Chapter four

NusA-LMO2 expression vector	
pET43a-LMO2 Forward	CGCGACTAGTGGGTCTGAAAACCTGTATTTC
pET43a-LMO2 Reverse	CGCGCCTAGGGAATTCCTAGATGATCCCATGA

Chapter five

Primer	Sequence (5'to3')
CDR1	
pEF-VP16-VH#576 Mutagenesis	
S28G Forward	GCAGCCTCTGGATTTCGGCTTCAGTCATAGT

Appendix C

S28G Reverse	ACTATGACTGAAGCCGAATCCAGAGGCTGC
H31G Forward	TCTGGATTCAGCTTCAGTGGTAGTCCTATGAAT
H31G Reverse	ATTCATAGGACTACCACTGAAGCTGAATCCAGA
S32G Forward	TTCAGCTTCAGTCATGGTCCTATGAATTGGGTC
S32G Reverse	GACCCAATTCATAGGACCATGACTGAAGCTGAA
N35G Forward	AGTCATAGTCCTATGGGTTGGGTCCGCCAGGCT
N35G Reverse	AGCCTGGCGGACCCAACCCATAGGACTATGACT
CDR2	
pEF-VP16-VH#576 Mutagenesis	
Y53G Forward	GTTTCATACATTAGTGGTAATTCTTCGAGTATA
Y53G Reverse	TATACTCGAAGAATTACCACTAATGTATGAAAC
N54G Forward	TCATACATTAGTTATGGTTCTTCGAGTATATAC
N54G Reverse	GCATATACTCGAAGAACCATAACTAATGTATGA
S55A Forward	TACATTAGTTATAATGCTTCGAGTATATACTAT
S55A Reverse	ATAGTATATACTCGAAGCATTATAACTAATGTA
S56A Forward	ATTAGTTATAATTCTGCGAGTATATACTATGCA
S56A Reverse	TGCATAGTATATACTCGCAGAATTATAACGAAT
S57G Forward	AGTTATAATTCTTCGGGTATATACTATGCAGAC
S57G Reverse	GCTTGCATAGTATATACCCGAAGAATTATAACT
Assembly PCR primers for the previous constructs	
VH#576 Forward	TGAACACGTGGCCCA
VH#576 Reverse	AGCTTCATTGCGGCC
CDR3	
pEF-VP16-VH#576 Mutagenesis	
E102A Forward	GCGAGAGGGTTGACGGCGTCTCTTGAGTTGACG
E102A Reverse	CGTCAACTCAAGAGACGCCGTCAACCCTCTCGC
S103G Forward	AGAGGGTTGACGGAGGCTCTTGAGTTGACGGCG
S103G Reverse	CGCCGTCAACTCAAGAGCCTCCGTCAACCCTCT
L104G Forward	GGGTTGACGGAGTCTGGTGAGTTGACGGCGGAT
L104G Reverse	ATCCGCCGTCAACTCACCAGACTCCGTCAACCC
E105G Forward	TTGACGGAGTCTCTTGGGTGGACGGCGGATTGG
E105G Reverse	CCAATCCGCCGTCAACCCAAGAGACTCCGTCAA
L106G Forward	ACGGAGTCTCTTGAGGGGACGGCGGATTGGTTT
L106G Reverse	AAACCAATCCGCCGTCCCCTCAAGAGACTCCGT
T107A Forward	GAGTCTCTTGAGTTGGCGGCGGATTGGTTTGAT

Appendix C

T107A Reverse	ATCAAACCAATCCGCCGCAACTCAAGAGACTC
D109G Forward	CTTGAGTTGACGGCGGGTTGGTTTGATTACTGG
D109G Reverse	CCAGTAATCAAACCAACCCGCCGTCAACTCAAG
Y113A Forward	GCGGATTGGTTTGATGCCTGGGGCCAGGGAACC
Y113A Reverse	GGTTCCTGGCCCCAGGCATCAAACCAATCCGC
Assembly PCR primers for the previous constructs	
VH#576 Forward	TGAACACGTGGCCCA
VH#576 Reverse	CGAGGCTGATCAGCGA
pEF-VP16-VH#576 Mutagenesis	
CDR1 and CDR2 Second round	
S30G Forward	TCTGGATTGAGCTTCGGTCATAGTCCTATGAAT
S30G Reverse	ATTCATAGGACTATGACCGAAGCTGAATCCAGA
S32A Forward	TTCAGCTTCAGTCATGCTCCTATGAATTGGGTC
S32A Reverse	GACCCAATTCATAGGAGCATGACTGAAGCTGAA
N35A Forward	AGTCATAGTCCTATGGCTTGGGTCCGCCAGGCT
N35A Reverse	AGCCTGGCGGACCCAAGCCATAGGACTATGACT
Y50A Forward	CTGGAGTGGGTTTCAGCCATTAGTTATAATTCT
Y50A Reverse	AGAATTATAACTAATGGCTGAAACCCACTCCAG
S52G Forward	TGGGTTTCATACATTGGTTATAATTCTTCGAGT
S52G Reverse	ACTCGAAGAATTATAACCAATGTATGAAACCCA
N54A Forward	TCATACATTAGTTATGCTTCTTCGAGTATATAC
N54A Reverse	GTATATACTCGAAGAAGCATAACTAATGTATGA
S57A Forward	AGTTATAATTCTTCGGCTATATACTATGCAGAC
S57A Reverse	GTCTGCATAGTATATAGCCGAAGAATTATAACT
Assembly PCR primers for the previous constructs	
VH#576 Forward	TTTCCAGGGCGGATCC
VH#576 Reverse	GCCGCACATATGTATATCTCCTTCTTAAAGTTAA ACAAAA
pEF-VP16-VH#576 Mutagenesis	
CDR3 Second round	
E102G Forward	GCGAGAGGGTTGACGGGGTCTCTTGAGTTGACG
E102G Reverse	CGTCAACTCAAGAGACCCCGTCAACCCTCTCGC
S103G Forward	AGAGGGTTGACGGAGGGTCTTGAGTTGACGGCG
S103G Reverse	CGCCGTCAACTCAAGACCCTCCGTCAACCCTCT
L104A Forward	GGGTTGACGGAGTCTGCTGAGTTGACGGCGGAT

Appendix C

L104A Reverse	ATCCGCCGTCAACTCAGCAGACTCCGTCAACCC
E105A Forward	TTGACGGAGTCTCTTGCGTTGACGGCGGATTGG
E105A Reverse	CCAATCCGCCGTCAACGCAAGAGACTCCGTCAA
L106A Forward	ACGGAGTCTCTTGAGGCGACGGCGGATTGG
L106A Reverse	CCAATCCGCCGTGCGCTCAAGAGACTCCGT
T107G Forward	GAGTCTCTTGAGTTGGGGGCGGATTGGTTTGAT
T107G Reverse	ATCAAACCAATCCGCCCCCAACTCAAGAGACTC
D109A Forward	CTTGAGTTGACGGCGGCTTGGTTTGATTACTGG
D109A Reverse	CCAGTAATCAAACCAAGCCGCCGTCAACTCAAG
Y113G Forward	GCGGATTGGTTTGATGGCTGGGGCCAGGGAACC
Y113G Reverse	GGTTCCTGGCCCCAGCCATCAAACCAATGCGC
pEF-VH#576 Framework mutations	
Forward	TGA ACA CGT GGC CCA
Reverse	CGA GGC TGA TCA GCG A
Assembly PCR primers for the previous constructs	
Forward	TTTCCAGGGCGGATCC
Reverse	GCCGCACATATGTATATCTCCTTCTTAAAGTTAA ACAAAA
pRK-VH#576 Framework mutations	
Forward	TCC AGG GAA GAG AGG GAG ATA GTT TCA TAC ATT AGT
Reverse	ACT AAT GTA TGA AAC TAT CTC CCT CTC TTC CCT GGA
Assembly PCR primers for the previous constructs	
Forward	ATA GTT AAG CCA GTA TAC ACT CCG CT
Reverse	CGC GCG AAG CTT GCG AAT TCT CAG CTC GAG AC

Appendix D