Engineering and characterisation of anti-progesterone OBodies

A thesis submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy

at

The University of Waikato

by

Vikas Kariyappa Chonira

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2018

Abstract

Molecular interactions are fundamental to communication between different parts of the cell or of an organism. These interactions can be weak and transient or strong and semi-permanent. In the case of the adaptive immune system, strong interactions between foreign antigens and specific antibodies lead to a cascade of events comprising the immune response. This phenomenon has been exploited by industry to produce high affinity binding molecules that have been used as therapeutics or diagnostics and engineered antibodies have been at the forefront of these industries. More recently, novel non-immunoglobulin binding proteins, have been similarly engineered to produce high affinity binding proteins that can potentially replace the binding function(s) of antibodies or even surpass them. The OB-fold is a high affinity binding protein domain which has previously been engineered to bind to Hen Egg-white Lysozyme (HEL) with nanomolar affinity as a proof-of-concept technology and has been given the name OBodies. This thesis explores this concept further with an OBody (D7) engineered to bind to the small molecule progesterone (P4) with potential applications to detect P4 in cow's milk to evaluate pregnancy. Obody-P4 binding was characterised with an optimized ELISA system. This was followed by the engineering of improved versions of this OBody using phage display technology and structural characterisation of one such Obody-P4 complex using X-ray Crystallography. The three-dimensional structure provided surprising insights into the nature of the molecular interactions between P4 and the OBody. During phage display selection a new signal sequence was fortuitously discovered that provided an advantage during phage display. The new signal sequence was investigated and characterised using mutants of GFP and a DARPin sequence to uncover the nature of the selective advantage. The combination of this new signal sequence and thermostable OBody libraries further demonstrates the potential of this system to produce robust bio-sensors for diagnostic and therapeutic applications.

Acknowledgements

I would like to thank Prof. Vic Arcus for his belief in my stubbornness to complete my PhD and supporting me in other ways. A special thanks to Dr. Judith Burrows for keeping me on my toes to weather the tough times and distracting me with her sense of humor.

I cannot thank my lab mates Chelsea, Tiffany, Jo H, Emma A, Emma S, Heng, Erica, Claire, Joanna H, Kirsty and other invaluable members of the C2 labs enough for supporting me through this journey and answering my silly queries.

I would also like to thank Dr. Magda Bereza, Dr. Mark Liddament and Dr. Ray Cursons for helping me with their expertise. My gratitude goes out to University of Waikato for supporting my PhD with the International Doctoral Scholarship.

I would also like to thank my parents and family back in India for their belief in my commitment to make a difference. My gratitude goes out to all those people who went in and out of my life during the most important part of this journey.

Table of Contents

Abstract ii

Acknowledgements iii

List of Figures x

List of Tables xiv

List of Abbreviations i

1 Introduction 5

1.1 Proteins and specific molecular interactions 5

1.2 Antibodies 6

1.3 Non-immunoglobulin scaffolds 9

1.3.1 DARPins 9

1.3.2 Monobodies 11

1.3.3 Anticalins 13

1.3.4 Affibodies 15

1.4 OBodies 17

1.5 Techniques in protein engineering 23

1.6 Display technology 24

1.6.1 Ribosome display 26

1.6.2 mRNA display 27

1.6.3 Yeast display 27

1.7 Phage display 28

1.7.1 pIII as display protein in phage display 36

1.7.2 Protein translocation systems for phage display using pIII coat protein 37

1.7.3 Phage display using pVIII coat protein 41

1.7.4 T7 phage display 41

1.8 Thesis Outline 42

2 Materials and methods 43
2.1 General molecular biology 43
2.1.1 Media and agar plates 43
2.1.2 *E. coli* strains 43
2.1.3 Induction of culture 43
2.1.4 Plasmids 44
2.2 Protein expression and purification in *E.coli* cytoplasm 45
2.2.1 Protein expression and extraction 45
2.2.2 Protein purification 45
2.2.3 Protein Analysis by SDS-PAGE 47
2.2.4 Protein-ligand incubation 47
2.3 Protein crystallography 48
2.3.1 Crystal trials 48
2.3.2 X-ray diffraction 49
2.3.3 Data Analysis and solving of structure 50
2.4 Functional characterisation 51
2.4.1 ELISA of protein-ligand binding 51
2.4.2 SPR for protein-ligand binding 55
2.4.3 FTSA for thermo-stability of OBody 56
2.5 DNA manipulation and analysis 58
2.5.1 Cloning of Obody DNA sequences into *E. coli* 58
2.5.2 Confirmation and storage of clones 58
2.5.3 Primer design 59
2.5.4 Polymerase Chain Reaction (PCR) 59
2.5.5 DNA Agarose gel electrophoresis 62
2.5.6 Restriction enzyme digestion of DNA fragments 62
2.5.7 Ligation of DNA fragments 62
2.5.8 Transformation of ligated plasmid 63

2.5.9 Bioinformatic tools....64
2.6 Phage library and selection 65
2.6.1 Construction of Affinity Maturation (AM) phage library....65
2.6.2 Phage selection of Affinity maturation library....68
2.6.3 Phage ELISA....70
3 Characterisation of anti-progesterone OBody (D7)..... 72
3.1 Introduction 72
3.1.1 Structure of Progesterone (P4) 72
3.1.2 Differences between periplasmic and cytoplasmic expression 73
3.1.3 Structure of an OB-fold....78
3.1.4 Preliminary OBodies work....79
3.1.5 Re-design of OBody library for P4 82
3.2 Results 85
3.2.1 Large scale expression and purification of D7....85
3.2.2 Strategies to improve protein stability 89
3.2.3 Construction of D7 mutants 91
3.2.4 Purification of D7 mutants 92
3.2.5 Functional characterisation of D7 mutants: *FTSA* 95
3.2.6 Functional characterisation of D7 mutants: *ELISA*....97
3.2.7 MBP fusion to D7 for crystallization 106
3.3 Discussion 110
4 Structural and functional characterisation of improved anti-progesterone (P4) OBodies..... 113
4.1 Introduction 113
4.1.1 Directed evolution with phage display....113
4.1.2 Affinity maturation with focused mutagenesis and phage display ...114
4.1.3 Construction of pF and L4 based (AM) libraries 114
4.2 Results 115

4.2.1 Construction of AM library 115
4.2.2 Phage selections of AM libraries 118
4.2.3 Analysis of P4 binding OBodies 123
4.2.4 Surface Plasmon Resonance (SPR) analysis of P4-binding OBodies 124
4.2.5 Large scale expression and purification of B7 127
4.2.6 Fusion of crystallization chaperone MBP to OBody B7 (MBP-B7). 130
4.2.7 Crystal structure of fusion protein (MBP-B7) with progesterone (P4) 133
4.2.8 Structure of OBody B7 and interaction with P4 138
4.3 Discussion 151
4.3.1 Affinity maturation of OBodies 151
4.3.2 Randomization of loop (L4) and P4 binding 151
4.3.3 Structure of OBody bound to P4 152
4.3.4 Future experiments 153
5 Characterisation of YscC for phage display 156
5.1 Introduction 156
5.1.1 The YscC fragment and phage display 156
5.1.2 The phagemid (YscC) for AM library construction 158
5.1.3 Phage display and inner membrane translocation pathways 161
5.1.4 Strategies for highly stable protein translocation 162
5.2 Results 164
5.2.1 Characterisation of YscC fragment 164
5.2.2 Testing YscC for better phage display of other stable (or fast folding) proteins 169
5.3 Discussion 174
6 Discussion 177
6.1 Characterisation of OBody D7 from naïve library 178
6.2 Construction of AM libraries and characterisation of OBody B7 179

6.3 Characterisation of YscC 180

6.4 Future experiments 181

References 184

7 Appendices 200

7.1 General Molecular Biology 200

7.1.1 Media composition 200

7.1.2 Antibiotics 201

7.1.3 Buffer composition 201

7.1.4 SDS-PAGE reagents 202

7.1.5 Agarose gel ingredients 203

7.1.6 Electrophoresis standards 204

7.1.7 Bacterial strains 204

7.1.8 Plasmids 205

7.1.9 Primers 209

7.2 OBody sequences 211

7.2.1 Naïve library OBody D7 in phagemid (pOB2) 211

7.2.2 Naïve library OBody D7 in cytoplasm expression plasmid (pET28b) 212

7.2.3 MBP fusion of OBody D7 213

7.2.4 Affinity maturation (AM) library OBody B7 214

7.2.5 AM library OBody B7 in cytoplasm expression plasmid (pPROEX) 215

7.2.6 MBP fusion of OBody B7 216

7.2.7 AM library clones 217

7.3 FTSA of D7 mutants 218

7.4 ELISA of D7 and mutants 218

7.4.1 P4 immobilized ELISA: pH 218

7.4.2 Antigen immobilized ELISA: P4 and E2 220

7.4.3 ELISA of D7 mutants 222

7.4.4 Antibody immobilized ELISA 225
7.5 SPR..... 226
7.5.1 SPR of D7.....226
7.5.2 SPR of B7.....227
7.5.3 SPR of MBP-B7.....229
7.5.4 SPR of other OBodies 230
7.6 Diffraction datasets 231

List of Figures

Figure 1.1 Antibody generation in the immune system 7

Figure 1.2 Cartoon representation of popular non-immunoglobulin scaffolds.....9

Figure 1.3 Cartoon representation of a DARPin.....10

Figure 1.4 Cartoon representation of a Monobody 12

Figure 1.5 Cartoon representation of an Anticalin.....14

Figure 1.6 Cartoon representation of an Affibody structure 15

Figure 1.7 Schematic of OB-fold domain model.....17

Figure 1.8 Cartoon representation of an OB-fold 18

Figure 1.9 Examples of OB-fold domains 20

Figure 1.10 Engineered OB-fold (OBody) to lysozyme 21

Figure 1.11 Comparison between non-immunoglobulin scaffolds.....22

Figure 1.12 Hybridoma technology 23

Figure 1.13 Phage display cycle.....25

Figure 1.14 Ribosome and mRNA display 26

Figure 1.15 Schematic of filamentous phage.....29

Figure 1.16 Structure of filamentous phage coat proteins pVIII and pIII.....31

Figure 1.17 Illustration of pIII mediated infection.....33

Figure 1.18 Filamentous phage life cycle 35

Figure 1.19 Different formats of phage display using pIII 37

Figure 1.20 Inner membrane translocation pathways in bacteria 38

Figure 1.21 Structures of Tat and Sec signal sequences 39

Figure 2.1 Schematic comparing different types of molecule immobilization in ELISA 52

Figure 2.2 Binding curve of protein-ligand binding from ELISA 54

Figure 2.3 SPR sensorgram from protein-ligand binding.....56

Figure 2.4 Graph of Fluorescent Thermal Shift Assay (FTSA) melt curve.....57

Figure 2.5 Illustration of Overlap-Extension PCR (OE-PCR).....61

Figure 2.6 Sequence of randomisation scheme for loop (L4) on DM (stop) template 66

Figure 2.7 Cartoon of Phage ELISA system 71

Figure 3.1 Comparison of Estradiol (E2) and Progesterone(P4) bound by their receptors ER and PR 72

Figure 3.2 Protein folding in reducing cytoplasmic environment 74

Figure 3.3 Protein folding in oxidising periplasmic environment 75

Figure 3.4 Illustration of redox reactions 76

Figure 3.5 Reducing environment in cytoplasm 77

Figure 3.6 Model of OB-fold showing active side 78

Figure 3.7 Model showing OBody NL8 interacting with HEL in comparison to YkfE (HEL inhibitor) 80

Figure 3.8 Animated electrostatic models of HEL and AM3L09 81

Figure 3.9 Comparison of OBodies from 1GPATS with 2GPATS model 83

Figure 3.10 Gene sequence of 2GPATS naïve library 84

Figure 3.11 Comparison of D7 (2GPATS) with AM3L15 (1GPATS) 85

Figure 3.12 UV elution profile and SDS-PAGE image of IMAC purified D7 86

Figure 3.13 Purification of D7 on S75 16/600 column 87

Figure 3.14 SDS-PAGE image of S75 16/600 purified D7 87

Figure 3.15 SEC analysis of D7 88

Figure 3.16 SDS-PAGE image of low yield D7 89

Figure 3.17 SDS-PAGE image of D7 with DTT 90

Figure 3.18 D7 model showing position of cysteines 91

Figure 3.19 DNA sequencing of D7 mutants 92

Figure 3.20 SDS-PAGE images of D7 mutants 93

Figure 3.21 SDS-PAGE image of β-sheet mutants 94

Figure 3.22 SDS-PAGE image of DM at high concentrations 94

Figure 3.23 SDS-PAGE image of D7 (β-sheet) alanine mutants 95

Figure 3.24 Fluorescent Thermal Shift Assays of D7 and mutants 96

Figure 3.25 Equilibrium binding curves of D7 to P4 from pH 4 to 9 98

Figure 3.26 Equilibrium binding curves of D7 to E2.....99

Figure 3.27 Equilibrium binding curves of D7 to P4 and E2 with high salt.....100

Figure 3.28 Equilibrium binding curves of D7 mutants to P4 103

Figure 3.29 Equilibrium binding curves of D7 mutants to P4 (reverse ELISA)..104

Figure 3.30 Equilibrium binding curve of anti-P4 antibody (Ab) to P4 106

Figure 3.31 UV elution profile and SDS-PAGE: IMAC purified MBP-D7.....108

Figure 3.32 UV elution profile and SDS-PAGE: SEC purified MBP-D7 109

Figure 3.33 Protein folding in the periplasm 110

Figure 4.1 DNA sequencing of stop codon insertion into D7 for AM library construction 116

Figure 4.2 Sequence of randomisation scheme for β-sheet 3 (pF_35) and loop (L4_DM) on D7 stop mutants 117

Figure 4.3 DNA sequencing of pF_35 and L4_DM 118

Figure 4.4 Bar graph showing phage selection of AM library.....121

Figure 4.5 Phage ELISA of selections 123

Figure 4.6 DNA sequencing of P4 binding OBodies from AM library 123

Figure 4.7 Amino acid alignment of P4 binding OBodies (AM library).....124

Figure 4.8 Equilibrium binding curves of naïve library OBodies (SPR).....125

Figure 4.9 Equilibrium binding curves of AM library OBodies (SPR) 126

Figure 4.10 Purification of B7 on Histrap column.....128

Figure 4.11 SDS-PAGE image of Histrap purified B7 128

Figure 4.12 Purification of B7 on S75 16/60 column 129

Figure 4.13 SDS-PAGE image of S75 16/600 purified B7 129

Figure 4.14 Purification of MBP-B7 on Histrap column.....131

Figure 4.15 Purification of MBP-B7 on S75 16/60 column 132

Figure 4.16 Equilibrium binding curve of MBP-B7 using SPR 133

Figure 4.17 Crystals of MBP-B7 (with P4) under X-ray diffraction 134

Figure 4.18 Model of AM OBody (B7) with progesterone (P4).....138

Figure 4.19 Overlap of 2GPATS OBody (B7) and 1GPATS Obody (AM3L15)139
Figure 4.20 Crystal packing of MBP and OBody B7 141
Figure 4.21 Crystal packing of MBP molecules 145
Figure 4.22 Interaction surface between OBody B7 and P4.....147
Figure 4.23 Surface representation of B7 with P4 150
Figure 4.24 Improved binding model of B7 with P4 154
Figure 5.1 Insertion of foreign YscC fragment in (D7) phagemid 157
Figure 5.2 Structure of foreign YscC fragment in (D7) phagemid.....159
Figure 5.3 Alternative strategies for stable protein translocation 163
Figure 5.4 DNA sequencing of B7 gene without YscC fragment.....165
Figure 5.5 Comparison of YscC and pOB2 for phage display of B7 165
Figure 5.6 Phage display of B7 with signal sequence- ssYscC 167
Figure 5.7 Structure and comparison of ssYscC.....168
Figure 5.8 Phage display of GFPuv with YscC fragment.....171
Figure 5.9 Phage display of DARPin with YscC.....174

List of Tables

Table 3.1 ELISA of OBody D7 at different pH 98

Table 3.2 ELISA of OBody D7 with different targets 102

Table 3.3 ELISA of OBody D7 mutants to P4.....105

Table 4.1 summary of AM library affinity-screening 120

Table 4.2 SPR of P4 binding OBodies.....127

Table 4.3 Data collection statistics of OBody B7 135

Table 4.4 Data refinement statistics of OBody B7 137

Table 4.5 Interactions between MBP and Obody B7.....142

Table 4.6 Interactions at binding interface of Obody B7.....148

List of Abbreviations

2GPATS	2nd gen *P. aerophilum* Asp. t-RNA synthetase
2xYT	2x yeast extract tryptone
3D	three-dimensional
Ab	antibody
Abs_{450}	absorbance at 450 nm
AM	affinity maturation
Amp	ampicillin
APS	ammonium persulfate
ASU	asymmetric unit
β2/3	β sheet 2/3
βme	β-mercaptoethanol
B7	OBody B7 from affinity maturation library
C100	carbenicillin 100 ug/ul
C35S	cysteine replaced by serine at position 35
CDR	complementarity determining region
cfu	colony forming units
C-terminal	carboxy-terminus
D7	original OBody D7
Da	Daltons
DARPin	Designed ankyrin repeat protein
DM	double mutant of OBody D7 (C35S:C48S)
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double stranded DNA
D7	OBody D7
DARPins	Designed ankyrin repeat proteins
DTT	Dithiothreitol
E2	17β-Estradiol
EC	electro-competent
EDTA	ethylene-diamine-tetraacetic acid (disodium salt)
ELISA	enzyme-linked immunosorbent assay
Fab	antibody “antibody” fragment
FN3	Fibronectin type-III domain

FPLC	fast protein liquid chromatography
FTSA	fluorescent thermal shift assay
Fv	antibody “variable” fragment
G	*Gibbs free energy*
gIII	gene three from M13 bacteriophage
GFP	green fluorescent protein
H	*Enthalpy*
H-bond	hydrogen bond
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEWL	hen egg white lysozyme
His-tag	poly-histidine tag
His-trap	his-tag binding nickel sepharose column
Ig	immunoglobulin
IMAC	immobilised metal affinity chromatography
IPA	isopropan-2-ol
IPTG	isopropylthio-β-D-galactosidase
K50	kanamycin 50 ug/ul
K_a	*equilibrium association constant*
Kan	kanamycin
Kb	kilobase
K_d	*equilibrium dissociation constant*
kDa	kilo Dalton
KM13	helper phage
k_{off}	*rate of dissociation*
k_{on}	*rate of association*
L4	loop 4 of OBody
LB	luria bertani (media)
LH	luteinizing hormone
MK13	mitogen activated kinase 13
MBP	maltose binding protein
Moi	multiplicity of infection
MPBS	skim milk in PBS
mRNA	messenger ribose nucleic acid
MW	molecular weight
MQ	molecular biology grade water
NCS	non-crystallographic symmetry

NNK/NRK	N= A, C, T or G; K= G or T; R = G or A
N-terminal	amino terminus
OB-fold	Oligonucleotide/oligosaccharide binding fold
OD_{600}	optical density at 600 nm
OE-PCR	Overlap extension polymerase chain reaction
pIII	protein translated from gIII of M13 bacteriophage
P4	Pregn-4-ene-3,20-dione (progesterone)
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pF	proximal binding face of OBody (β-sheet 2 and 3)
pH	potential of hydrogen
pI	isoelectric point
pKa	acid dissociation constant
pOB2	OBody displaying phagemid- 2^{nd} generation
POI	protein of interest
R_{eq}	refractive index at equilibrium
r.m.s.d.	root mean squared deviation
rpm	revolutions per minute
rTEV	recombinant tobacco etch virus
RU	response unit
s	second
S	*entropy*
S75	superdex 75 prep grade SEC column
SAP	shrimp alkaline phosphatase
Scfv	single chain Fv fragment (antibody)
SD	standard deviation
SDS	sodium dodecyl sulfate
Sec	general secretory pathway
SEC	size exclusion chromatography
SOC	super optimal broth with catabolite repression
SPR	surface plasmon resonance
ssDNA	single stranded DNA
ssYscC	signal sequence of YscC

T	*temperature*
TAE	tris-acetate-EDTA
Tat	Twin-arginine translocation pathway
TBS	tris (hydroxymethyl) aminomethane buffered saline
TE	tris ethylene diamine tetraacetic acid
TEMED	N, N, N', N'-tetramethylethylenediamine
T_m	temperature mid-point for unfolding
tRNA	transfer ribose nucleic acid
UP H_2O	ultra-pure water
UV_{280}	ultraviolet (280 nm)
V	volts
v/v	volume per volume
w/v	weight per volume
WT	wild type
YscC	Yersinia secretory chaperone C

1 Introduction

1.1 Proteins and specific molecular interactions

Proteins have evolved to play many different roles in organisms such as enzymes, structural components, molecular transport, regulation, energy storage, molecular motors, signalling and immune proteins, to name a few. The signalling and immune systems primarily involve specific molecular interactions with signalling pathways permitting specific protein-protein interactions that are accurate enough to differentiate between phosphorylation status of a single amino acid on the interacting surface of the protein(s). Proteins can also interact with other proteins to form multi-complex structures that transmit an extracellular signal to produce an intracellular response. Similarly proteins can produce an immune response with the help of antibodies which are specific in recognising antigens but ignore intrinsic proteins of the host organism (Garrett & Grisham, 2004).

Molecular specificity primarily depends on binding through non-covalent interactions with another protein or ligand (Sali, Glaeser, Earnest, & Baumeister, 2003). In general, tight interactions are involved in immune functions whereas weaker interactions are involved in signalling functions. This can be the result of the adaptation of the interacting surfaces during protein evolution (Steemson, Baake, Rakonjac, Arcus, & Liddament, 2014). Antibodies are a good example of proteins involved in this kind of molecular specificity. They have a modifiable interacting surface which supports specific binding to a variety of ligands. These ligands are often called antigens in an immune system context.

There have been some fundamental features observed in high binding affinity between protein and ligand. They have to have geometric complementarity of interacting surfaces (Mitchell, Kerr, & Ten Eyck, 2001). These interacting surfaces are made up of polar or hydrophobic groups which bind only to their respective groups in the interacting partner protein or ligand. Highly ordered water molecules at the binding surface have also been found to contribute to protein-ligand binding (Wilson & Finlay, 1998). The binding of protein and ligand is generally thermodynamically favourable for the stability of the protein (Böhm, 2005). Molecular recognition can be exploited to design *in vitro* diagnostics to detect antigens via specific binding and to design capturing agents

for affinity chromatography in columns that can be routinely regenerated (Lofblom et al., 2010). Antibodies have a binding surface formed of variable regions that are sequence independent and obey all features of good protein-ligand interactions. The immune system, responsible for producing antibodies in vertebrates to external biochemical threats (Binz, Amstutz, & Pluckthun, 2005) has been exploited by hybridoma technology to produce tight binding and highly specific antibodies against certain ligands and these have been used in clinical diagnostics for example. This was done by the fusion of antibody producing B-cells with an immortal myeloma cancer cell line and the resulting hybridoma cells being propagated indefinitely in tissue culture (Lee, Iorno, Sierro, & Christ, 2007). Antibodies have also risen to prominence as pharmaceutical agents due to their capacity to be engineered to bind specific targets in the body (Hoogenboom, 2005; Kim, Park, & Hong, 2005; A. J. Smith, 2015).

1.2 Antibodies

The dominant class of proteins involved in molecular specificity is currently immunoglobulin antibodies and antibody fragments. Libraries of DNA constructs encoding variable regions of antibodies (greater than 10^{11} different clones) have been made *in vitro*. In combination with display technology, specific clones may be isolated by displaying their respective proteins on the surface of a cell or phage. This approach has led to efficient production of large antibody libraries and isolation of tight binding and specific clones. The binding surfaces of antibodies are formed from six hypervariable regions also called Complementarity Determining Regions (CDR) constituted by three variable loops from the heavy chain and three from the light chain. They are encoded by independent germline segments which are shuffled and come together to produce a variable binding surface (Hoogenboom, 2005).

When an antigen is encountered, the immune system produces a variety of these antibodies (Figure 1.1). The antibody clones with the closest match are selected by the immune system via specific binding. These specific clones are randomly mutated and exposed to the same antigen resulting in selection of high affinity and specific antibody clones. Display technology makes good use of this system *in vitro*, i.e. it does not require the use of a mammalian host.

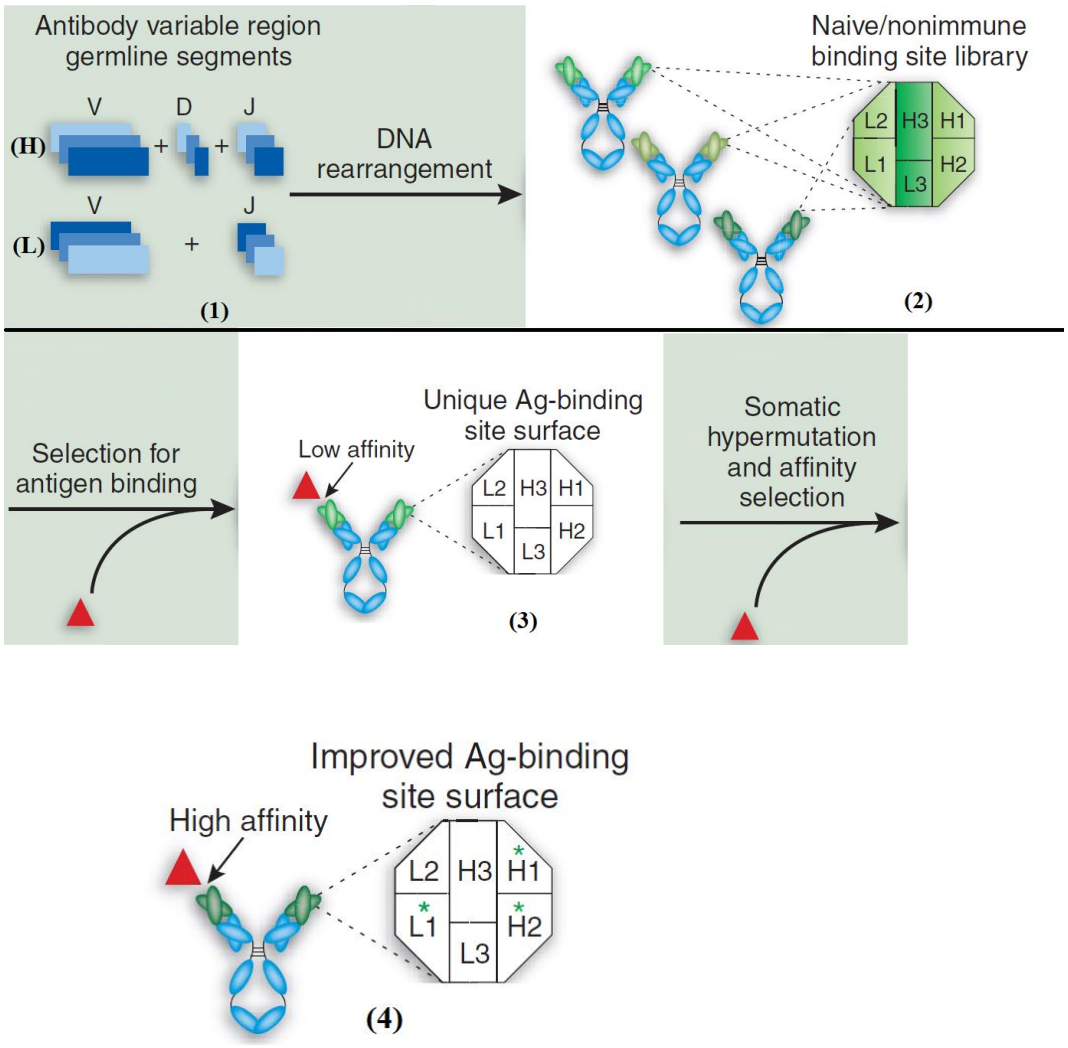


Figure 1.1 Antibody generation in the immune system

(1) Antibody binding surface is made up of variable regions of heavy (H) and light (L) chains, which are the result of combinations of V, D and J gene segments. (2) Each binding surface is made up of different combinations of three H and L variable regions (CDR) (3) When a collection of such clones (library) is exposed to an antigen (red triangle, Ag), a specific clone with a weak binding (low affinity) complementary surface is selected by the immune system 4) This particular clone is randomly mutated (somatic hypermutation) to introduce another round of selection with the antigen, resulting in a high affinity antibody with a unique binding surface (green asterisk) [adapted with permission from (Hoogenboom, 2005)].

However, antibodies have their limitations: they are bivalent and multidomain, dependent on intrachain disulphide bonds, are relatively large in size and typically need to be glycosylated. This can give them unfavourable solution properties such as a tendency to aggregate (Lofblom et al., 2010). Single domain antibodies, with half the interface area comprised of three hypervariable loops are alternatives, though they lack the stability of multi-domain antibodies (Nian et al., 2016). There is a clear need for molecules which expand the range of applications over and above those that are currently available with antibodies. To overcome these

limitations and to avoid intellectual property issues, work on non-immunoglobulin protein scaffolds has taken off rapidly in the last decade (Gebauer & Skerra, 2009; Jost & Pluckthun, 2014; Nygren & Skerra, 2004).

Scaffolds are protein domains with binding surfaces that support a wide range of amino acids that confer to it the ability to bind to specific ligands (Binz et al., 2005). These binding domains are produced by site directed random mutagenesis or phage display technology from soluble monomeric proteins or extramembrane domains of cell surface receptors. Some change conformation on binding and others follow a more rigid lock and key model. Ideally, they can be selected for specific properties and expressed in large quantities in simple microbial expression systems. They are ideal for *in vitro* diagnostics as they can demonstrate properties that are advantageous for "reagents" such as high stability and/or low production costs.

Four popular non-immunoglobulin scaffolds (Figure 1.2) will be covered in detail. Designed Ankyrin Repeat Proteins (DARPin) are modified ankyrin proteins involved in protein-protein interactions in various parts of the cell (Binz et al., 2004). Monobodies are derived from the extracellular domain of human fibronectin III (Koide, Wojcik, Gilbreth, Hoey, & Koide, 2012). Anticalins are modified lipocalin proteins which are small monomeric globular proteins involved in transport of poorly water soluble molecules (Arne Skerra, 2001). Affibodies are derived from three-helix bundle Z domain of Staphylococcal protein A (Nygren, 2008). Nanofitins and engineered lymphocyte receptors called Repebodies are other examples (Steemson et al., 2014).

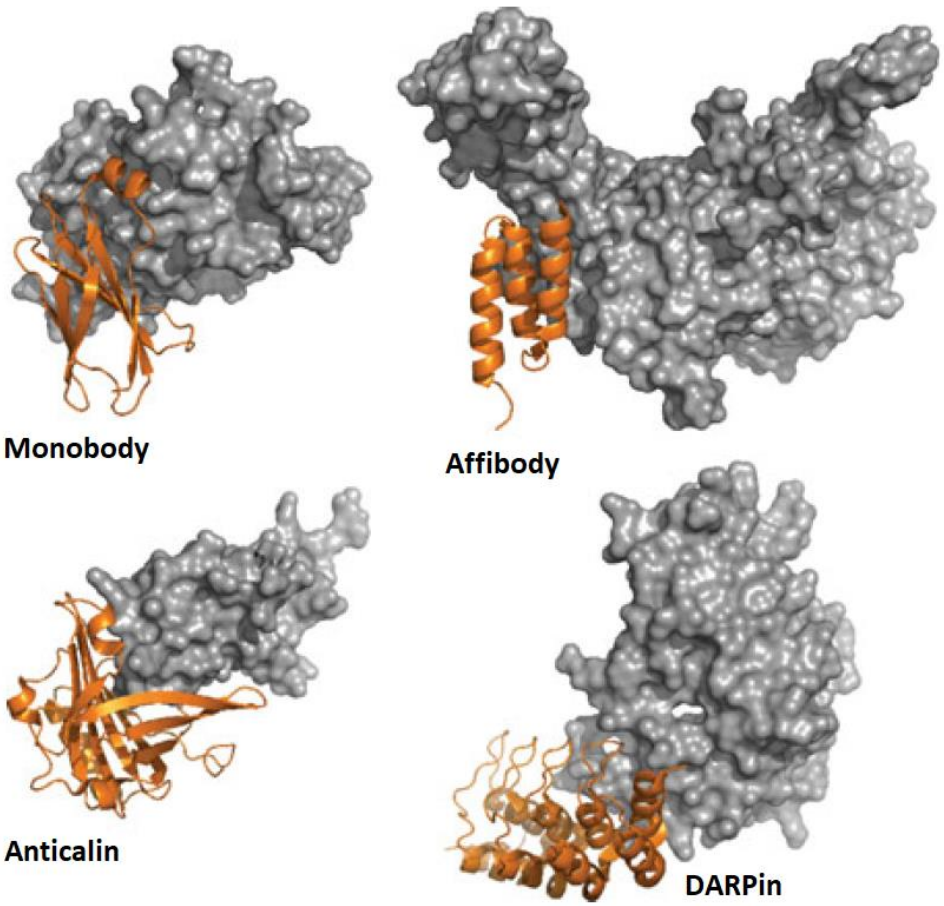


Figure 1.2 Cartoon representation of popular non-immunoglobulin scaffolds

A cartoon representation of the labelled non-immunoglobulin scaffolds (orange) are shown bound to respective protein ligands (grey) [used with permission from (Löfblom & Frejd, 2011)].

1.3 Non-immunoglobulin scaffolds

1.3.1 DARPins

DARPins are repeating modular protein regions (domains) with adjacent loops that can be modified by point mutations, insertions, deletions or repeat domain shuffling and as such are capable of binding different antigens (Binz et al., 2004) (Figure 1.3).

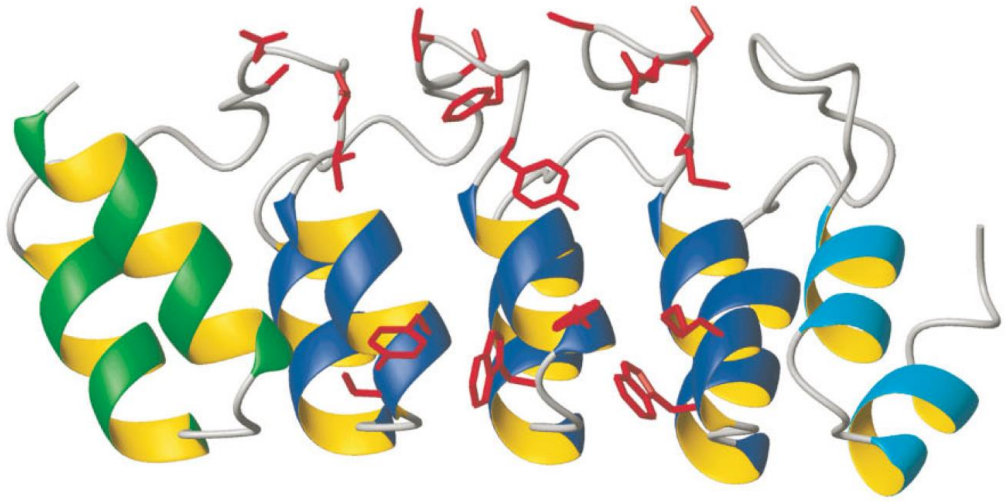


Figure 1.3 Cartoon representation of a DARPin

N-terminal (green) and C-terminal (cyan) capping domains with repeat domains (blue) showing randomised residues (red sticks) [used with permission from (Binz et al., 2004)].

Each domain consists of 33 amino acid residues starting with a short peptide chain called the β-turn and forming two anti-parallel α-helices ending as a loop connecting the β-turn of the adjacent domain, with the binding surface (7 residues) comprising of the β-turn and the first α-helix of every domain. DARPins exploit the modular nature (2 to 4 modules) of scaffolds which may assist in attaining low nanomolar affinity to ligands (Binz et al., 2004). The capping repeats are special terminal domains protecting the hydrophobic core and a distinct number of thermostable, monomeric domains can be designed between them with randomised positions occurring in the β-turn and the first α-helix of every domain. This gives a DARPin library very large theoretical diversity of 10^{15} to 10^{23} distinct clones (Binz et al., 2004).

A DARPin is one-tenth the size of an antibody. It is highly thermostable with melting temperatures close to 100°C and few flexible regions making it resistant to proteolytic damage. DARPins interact with their targets with a rigid body conformation that often shows high affinity and specificity. The large interaction surface, the diversity from addition and shuffling of domains makes it an ideal candidate as a non-immunoglobulin scaffold with functions that range from enzyme inhibition to anchoring proteins together.

Both phage and ribosome display technologies (using a consensus sequence-based approach) have been used to engineer DARPins. DARPins have some advantages

over antibodies and antibody fragments as they can be fused to other chemical agents or proteins of varying function. With regards to therapeutics, their smaller size has an advantage in intracellular penetration for precise targeting and also the lack of extra domains avoids interaction with the host immune system. Rapid renal clearance and allosteric inhibition where inhibition occurs by binding of a scaffold at a different non-active site on the target protein, are other advantages. DARPins lack cysteines in contrast to antibodies which need cysteine based interchain disulphide bonds to function. Instead, with DARPins cysteines can be used for fusion of its thiol groups to cytotoxins or radioactive isotopes for cancer treatment and imaging. Two or more DARPins of affinities to different targets or different epitopes of the same target can be engineered in the same molecule extending multi-functionality with affinity (Stumpp, Binz, & Amstutz, 2008).

A proof-of-concept example for the application of DARPins was reported using Maltose Binding Protein (MBP) as a target. The binding surface of the DARPin resulted in a complementary concave shape with its randomised binding residues enriched in aromatic amino acids, with tyrosines contributing to both hydrogen bonds and hydrophobic contacts with lysines of MBP, indicating a good fit with respect to shape and charge. The binding affinity (K_d) was 4.4 nM and consisted of 6 hydrogen bonds and 3 hydrophobic interactions between MBP and the specific DARPin with tyrosine contributing to four hydrogen bonds. The remnant hydrogen bonds were contributed by aspartic acid and tryptophan, the latter amino acid sharing hydrophobic interactions with Leucine and Phenylalanine to MBP (Binz et al., 2004).

1.3.2 Monobodies

Monobodies are modified Fibronectin Type III domains (FN3) which are small globular protein domains of 90 to 100 amino acids. Monobodies are highly stable like DARPins and are made up of 7 anti-parallel β-strands that come together as a β-sandwich. The loops connecting these strands (B- G) show variability of amino acids as seen in other scaffolds (Figure 1.4) (Jacobs et al., 2012).

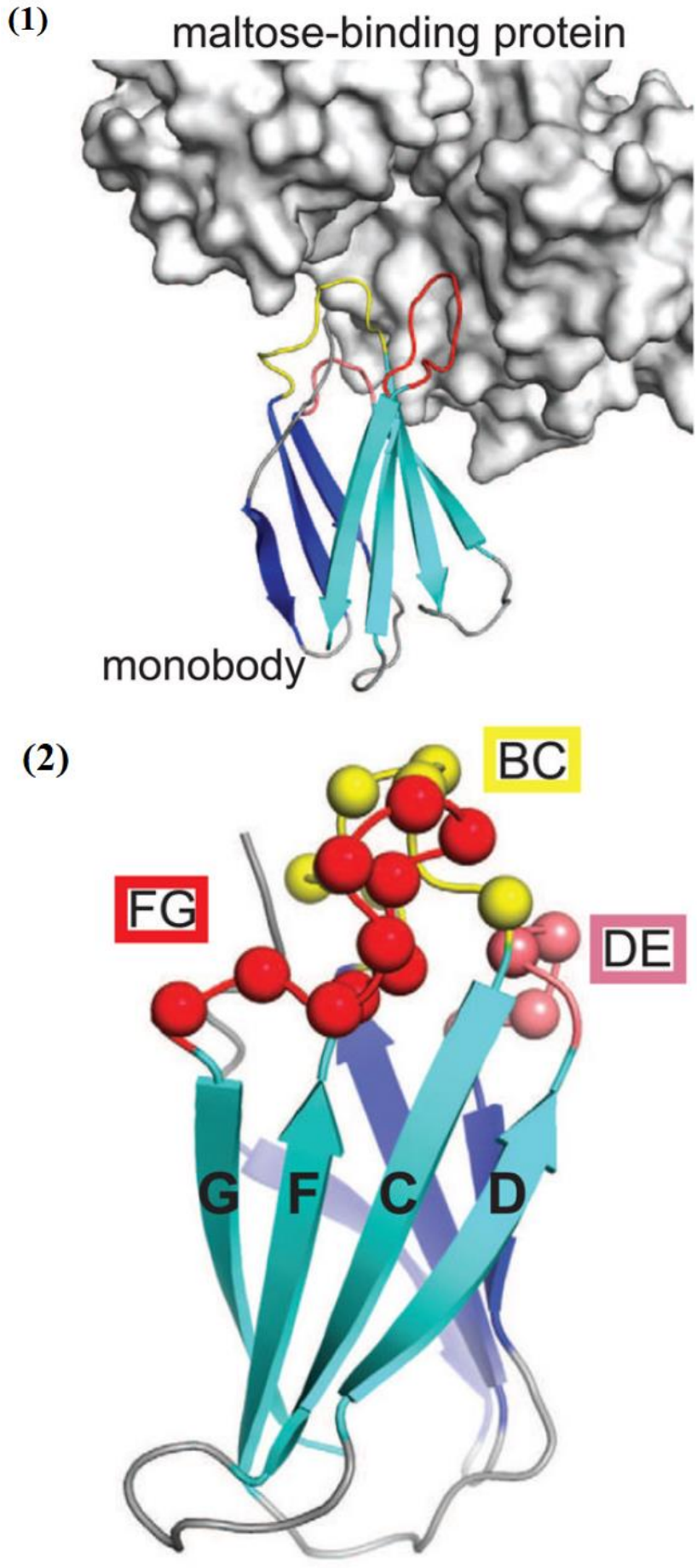


Figure 1.4 Cartoon representation of a Monobody

(1) Monobody model showing MBP specific Monobody (cyan/blue) binding to target MBP (grey) via its three loops (red/pink/yellow) (2) Detailed model of Monobody showing β-sheets (G, F, C and D) and mutated loops (FG, BC and DE). β-sheets (A, B and E) are hidden from view. [used with permission from (Koide et al., 2012)].

In the Monobody, β-strands A, B and E form one β-sheet and β-strands C, D, F and G form another β-sheet. The loops connecting these strands are structurally homologous to the CDR loops of antibodies. They come together to form a contiguous binding surface at the binding site of the target protein. Monobodies are not dependent on intradomain disulphide bonds in contrast to their antibody counterparts (Koide et al., 2012).

Display technology has been used to select tight binding Monobodies. A picomolar affinity clone has been selected to vascular endothelial growth factor receptor-2 (VEGFR2) protein using a type of display technology called mRNA display. A library of Monobodies was made with 21 residues mutated in three loops BC, DE and FG (named after adjoining β-strands) providing a theoretical diversity of up to 10^{27} clones. Monobodies have been selected for a variety of targets both intracellular and extracellular(Jacobs et al., 2012).

Monobody binding surface diversity has also been reported to be introduced into the β-sheet G-F-C-D obtaining stronger Monobodies than 'loop only' Monobodies. The buried surface area contributed by a Monobody binding another protein domain called Abl SH2 is 880 $Å^2$ (Koide et al., 2012).

1.3.3 Anticalins

Anticalins are modified lipocalin proteins which are small monomeric proteins (like Monobodies) of 160 to 180 residues made up of 8 anti-parallel β-strands coming together to form a β-barrel. The four loops that join these strands on one end of the β-barrel form a binding pocket (Figure 1.5) (Arne Skerra, 2001).

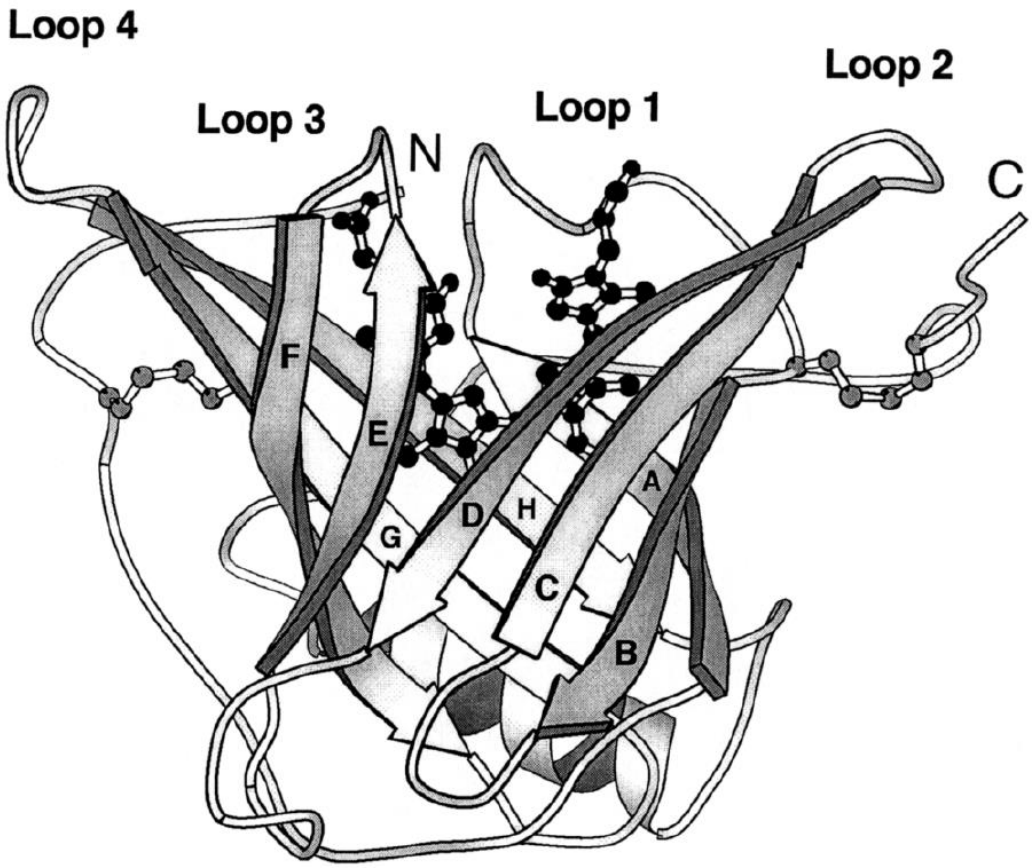


Figure 1.5 Cartoon representation of an Anticalin

Anticalin model showing how the β-barrel is made up of β-sheets (A-H) and four numbered loops (1-4) on one end forming the binding pocket. The ligand biliverdin IX is shown as a ball and stick Figure in the binding pocket [used with permission from (Arne Skerra, 2001)].

The β-barrel of the Anticalin is conserved in most lipocalins whereas the four loops forming the binding surface show high variability in their amino acid composition and conformations. The other end of the barrel is closed off by short loops forming a hydrophobic core. The β-sheets are hydrogen bonded with strand H binding with strand A to close the barrel (Arne Skerra, 2001).

Anticalins are highly stable often with melting temperatures over 70°C. Picomolar affinity Anticalins have been produced and the very high affinity is thought to be due to the ability of the ligand to bury itself deep in the binding pocket aided by aromatic side chains and hydrogen bonds. Anticalins are excreted rapidly through the kidneys making them useful in radioisotope-based imaging. Two Anticalins can be fused to each other to improve affinity by binding to two different areas on the same target or two separate targets. Anticalins have been developed for two cancer drug targets, cytotoxic T-lymphocyte antigen-4 (CTLA-4) and vascular endothelial growth factor (VEGF) (A. Skerra, 2008).

1.3.4 Affibodies

Affibodies are engineered protein scaffolds made up of three α-helices based on the B-domain in the immunoglobulin-binding region of Staphylococcal protein A. Affibodies were one of the first non-immunoglobulin scaffolds to be engineered to bind chosen ligands (Fiedler & Skerra, 2014). Affibodies have fast folding kinetics, are thermostable and have no cysteines (Figure 1.6).

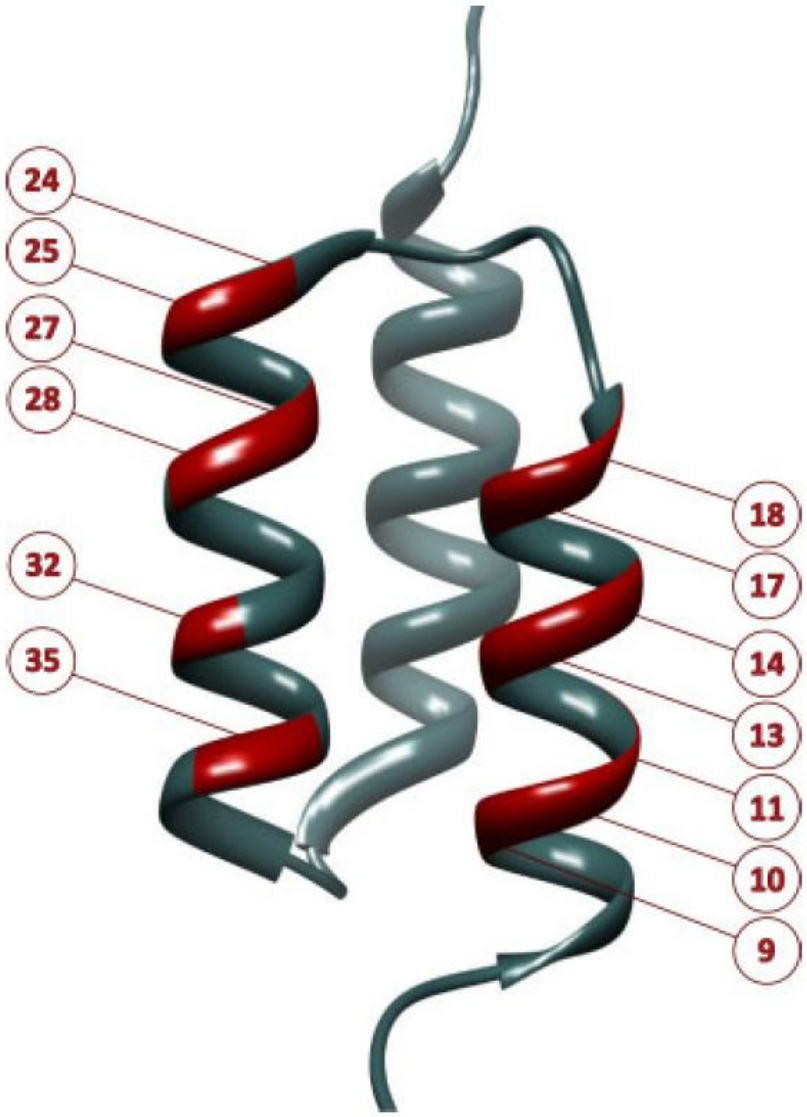


Figure 1.6 Cartoon representation of an Affibody structure

Affibody Structure showing binding surface. The binding surface consists of 13 mutated positions numbered according to residue and located in the red coloured region on two of the three α-helices in an Affibody. [used with permission from (Ståhl, Kronqvist, Jonsson, & Löfblom, 2013)].

The binding surface of an Affibody comprises of helices one and two which provide 13 amino acid positions for randomisation using phage display technology to create Affibody libraries (Lofblom et al., 2010). An Affibody is

relatively small (6 kDa in size) and comprises only 58 amino acids. An Affibody phage library can have a diversity of up to 20^{13}- 10^{17} unique clones. The interaction surface of Affibodies made up of α-helices is different when compared to the loops used by antibody formats and other non-immunoglobulin scaffolds. The interaction surface is generally non-polar in nature and buries up to 900 $Å^2$ per Affibody. There is often conformational change during binding (induced fit). An engineered Affibody has a melting temperature ~57°C and is capable of functional blocking similar to DARPins (Nygren, 2008).

Affibodies have affinities ranging from micromolar to picomolar values. Their affinity has been improved by helix shuffling or sequence alignment with directed combinatorial mutagenesis. Affibodies can also be synthesized using chemical methods as an alternative to recombinant bacterial methods, making it easier to attach other chemicals. Two Affibodies with different binding properties can be fused to each other in a modular fashion like DARPins. Affibodies are also compatible with a variety of display systems, from cell based bacterial display to cell-free ribosome display methods. The B- domain used to construct affibodies was modified for chemical stability (named Z- domain) thus losing affinity to the Fab component of antibodies making it less immunogenic and opening up the possibility of *in vivo* applications. Affibodies designed to target human epidermal growth factor receptor (HER2) have shown rapid biodistribution and blood clearance allowing quality imaging of tumours in mouse models (Lofblom et al., 2010).

Affibodies have been fused to foreign agents ranging from radionucleotides to cytotoxins for various diagnostic and therapeutic applications. Affibodies have been fused to colorimetric enzymes or used as capture agents in ELISA, fused to fluorophores in fluorescent based *in vitro* diagnostics or used as polymerase inhibitors in PCR when designed to have high affinity to DNA polymerases (Lofblom et al., 2010).

1.4 OBodies

Oligosaccharide-oligonucleotide Binding folds (OB-fold) are small, single protein domains found in diverse organisms in nature from humans to thermophilic organisms. The wide distribution of this domain suggests the potential for both therapeutic and diagnostic applications due to its ubiquity. Many OB-fold domains lack disulphide bonds and use both the binding surface and loops for intermolecular interactions unlike loop dependent antibodies (Steemson et al., 2014) (Figure 1.7).

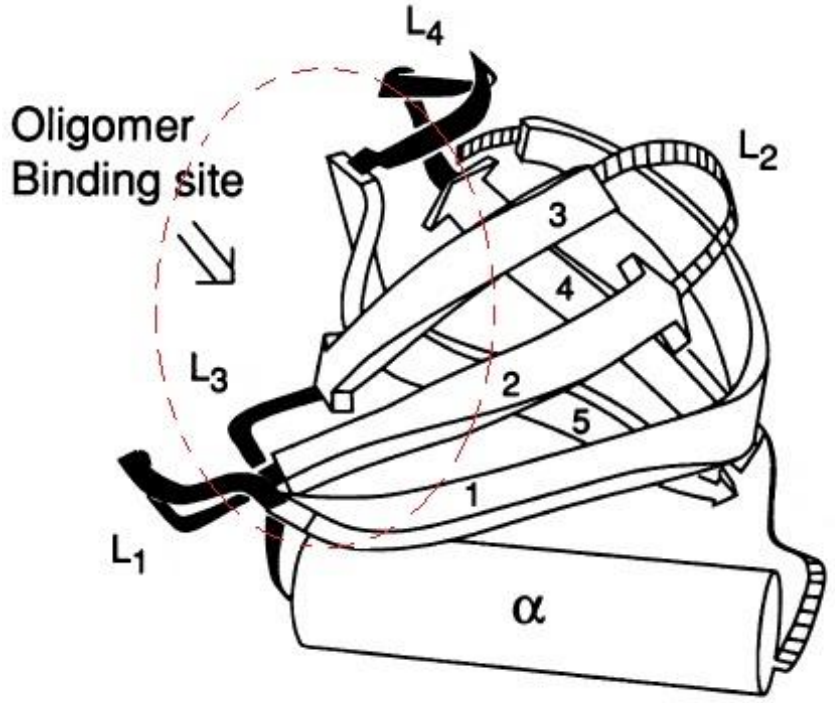


Figure 1.7 Schematic of OB-fold domain model

Oligomer Binding fold (OB-fold) consisting of five numbered β-sheets coming together to form a β-barrel with binding surface (dotted circle) surrounded by loops L1, L3 and L4 (black) [used with permission from (Murzin, 1993)].

OB-fold domains are non-immunoglobulin scaffolds made up of a five-stranded closed β-barrel (capped by an α-helix between β3 and β4 strands) with Greek key topology. OB-folds have been found to bind to ligands or proteins at the same location on one side of the β-barrel and binding is facilitated by residues from loops 1, 3 and 4 (Murzin, 1993). This key feature seems to be conserved in OB-folds found in all three kingdoms. Therefore, it appears that OB-folds evolved a structure tolerant to mutation and a binding face which could support a wide variety of sequences. A detailed representation of an OB-fold is shown in Figure 1.8 (2).

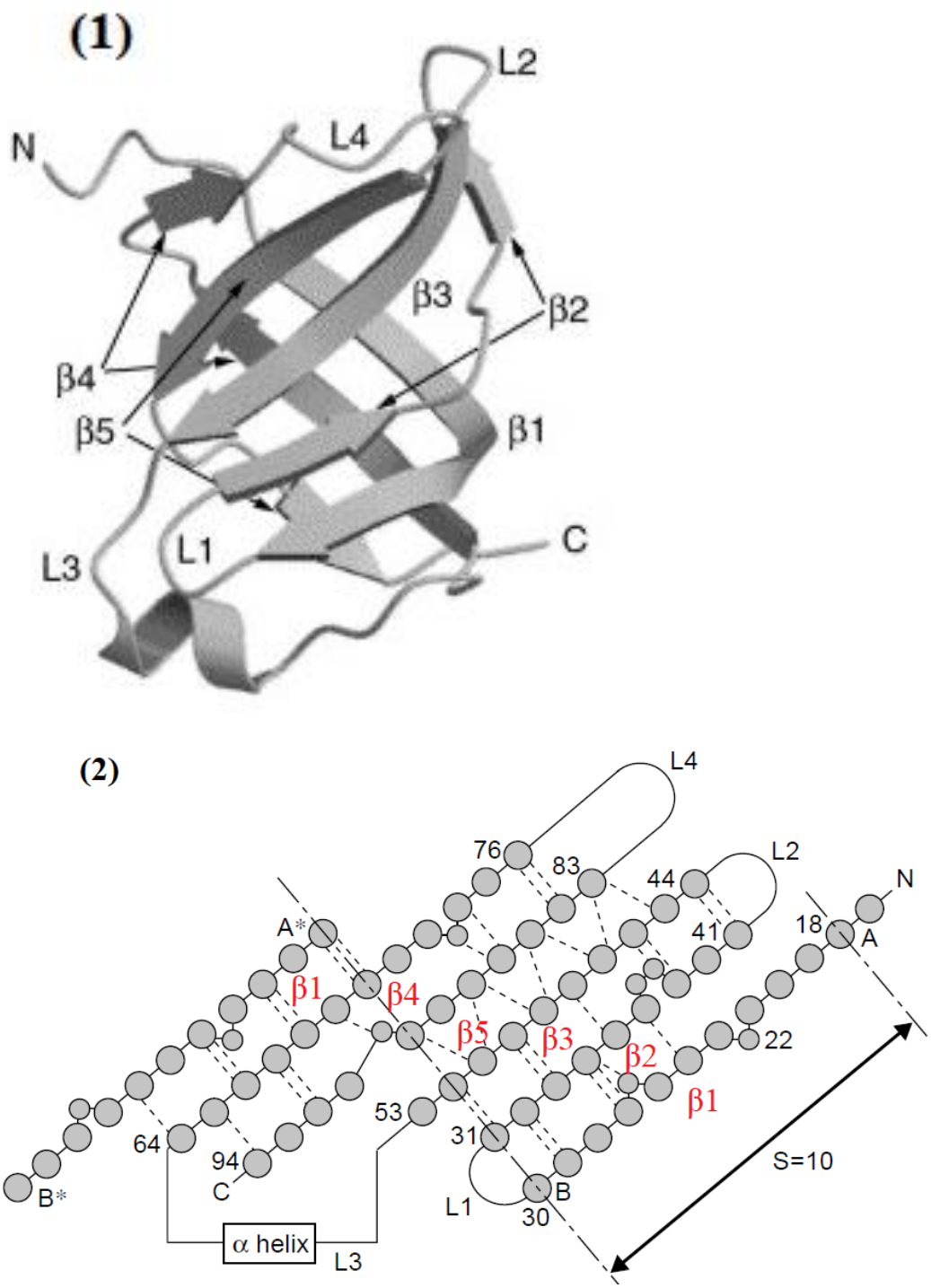


Figure 1.8 Cartoon representation of an OB-fold

(1) Model of an OB-fold, starting with (N)-terminus followed by numbered (β)-sheets and (L)oops ending with the (C)-terminus. (2) topology of β-barrel spread out with β-sheets (marked in red) placed side to side showing hydrogen-bonds holding them together (β1:18-30, β2:31-41, β3:44-53, β4:64-76, β5:83-94 amino acids). β1(AB,18-30) has been repeated on the other side (A*B*) to show closing side of β-barrel (β1-β4); S-Shear number. [adapted with permission from (Arcus, 2002)].

The β-barrel is described by the number of strands (n) and the shear number (S), the latter being the number of residues that are displaced to enclose the β-barrel. S=10 refers to 10 amino acids from A to A* [Figure 1.8 (2)]. β2, β4 and β5

strands are interrupted by loops or bulges between them with an α-helix between β3 and β4. β-strands are held side by side with hydrogen bonds. There are OB-folds of S=10 and S=8, for example the ssDNA binding protein has S=10 and the cold shock DNA binding domain has S=8. The binding face in the S=10 OB-fold is made up of β2 and β3 strands in the center and bounded by loop 1 at the bottom left, loop 4 at the top and loop 2 at the top right as shown in Figure 1.8 (1) (Arcus, 2002). OB-folds in different organisms use this binding face as a catalytic site or to bind to nucleic acids, proteins or oligosaccharides as shown in Figure 1.9.

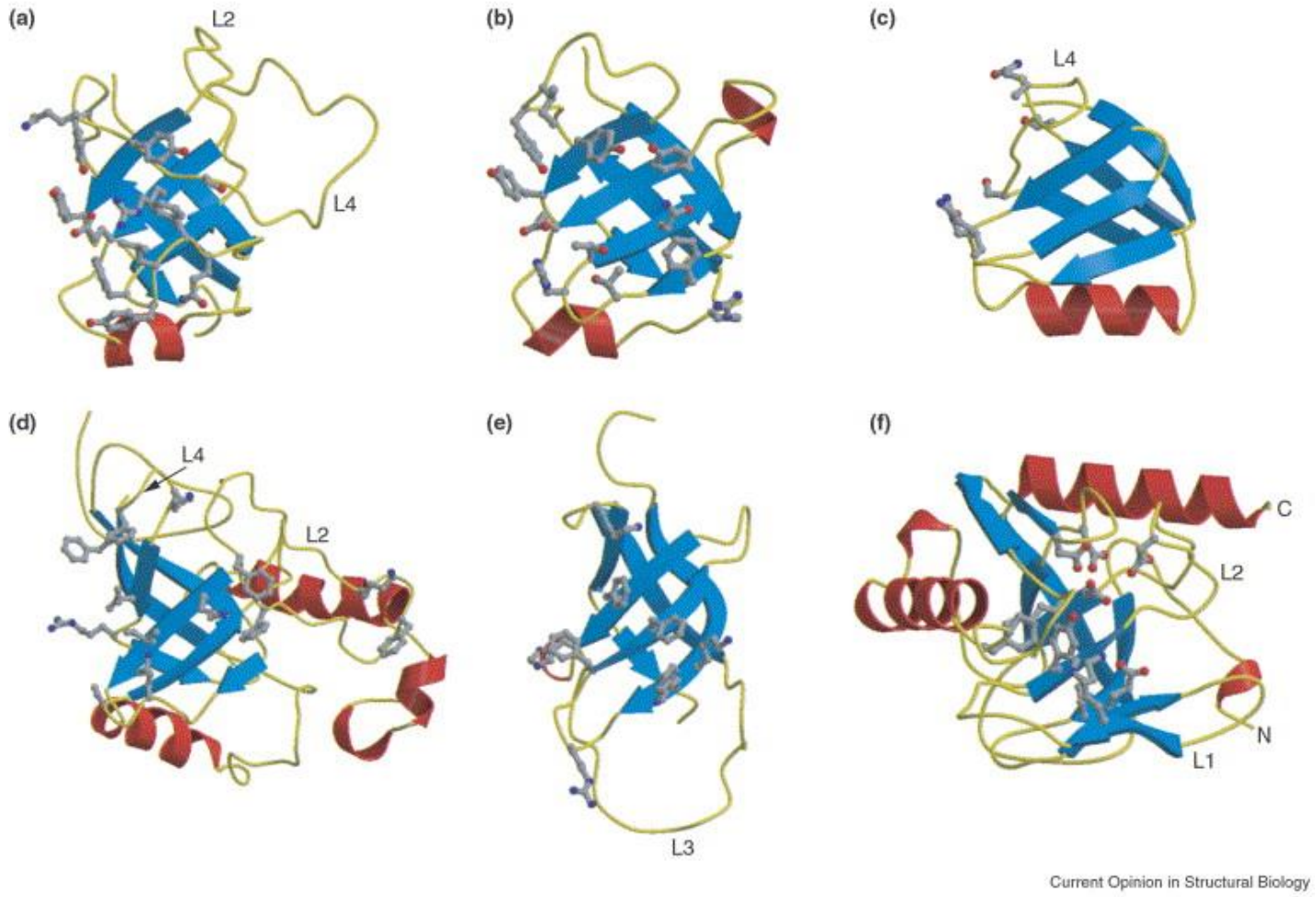


Figure 1.9 Examples of OB-fold domains

Six examples of OB-fold domains are shown. β-sheet, α-helices and loops are shown in blue, red and yellow respectively. Binding surface residues are shown as ball and stick figures; (a) N-terminal domain of Staphylococcal enterotoxin B, (b) N-terminal domain of Streptococcal superantigen SPE-c, part of dimer interface, (c) B_5 subunit of Escherichia coli (*E.coli)* Shiga like toxin, (d) structure of Cell division control protein Cdc13,(e) cold shock DNA binding domain, (f) catalytic site of *E.coli* inorganic pyrophosphatase- PPase [used with permission from (Arcus, 2002)].

OB-folds that have evolved from a common ancestor are clustered together into superfamilies. The bacterial enterotoxin superfamily has an OB-fold with the binding face specialised to bind to cell surface proteins or oligosaccharides. The nucleic acid binding superfamily has an OB-fold with the binding face specialised to bind to nucleic acids. The nucleic acid binding face uses central aromatic residues to accommodate the nucleotide base while using peripheral positively charged loops to bind to the phosphate backbone. The inorganic pyrophosphatase superfamily are groups of enzymes that uses the OB-fold binding face and its extended loops to form a cavity which is the active site for binding of metal ions and pyrophosphate (Arcus, 2002). When β-sheets of different OB-fold domains are superimposed, the helices and the ligand binding site align well showing sequence independent structural conservation (Murzin, 1993).

OB-folds have been modified in research. The anticodon recognition domain (OB-fold) of Aspartyl t-RNA synthetase from *Pyrobaculum aerophilum* was previoulsy engineered to bind hen egg-white lysozyme as a proof-of-concept (Steemson et al., 2014). Phage display technology was used to select for OB-fold clones. These engineered OB-folds have been named OBodies (Figure 1.10). This OBody scaffold has high thermal stability, is soluble, lacks disulphide bonds and tolerates mutation of up to 22% of the amino acid residues at the surface and in the loops. An OBody library was created by randomising 17 residues on its binding surface. The crystal structure of this OBody with lysozyme showed 1800 $Å^2$ buried surface area between the proteins with hydrophobic residues in the center surrounded by polar and charged residues. (Steemson et al., 2014).

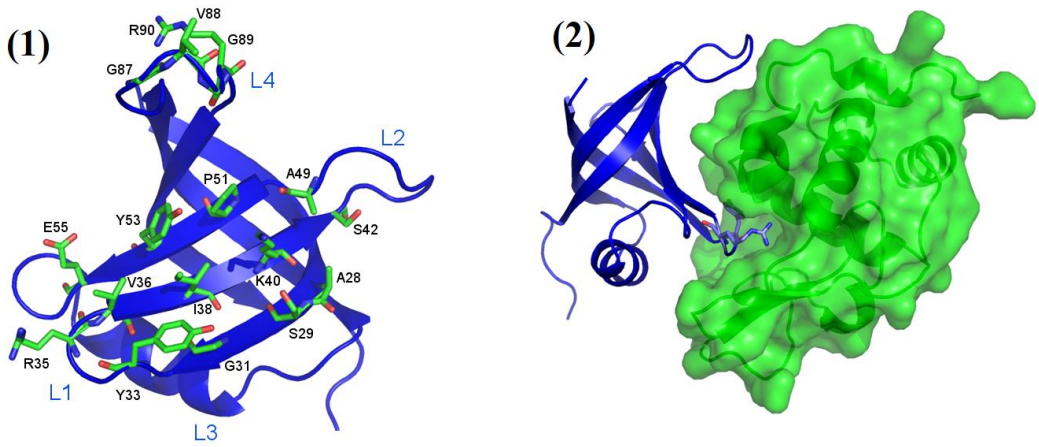
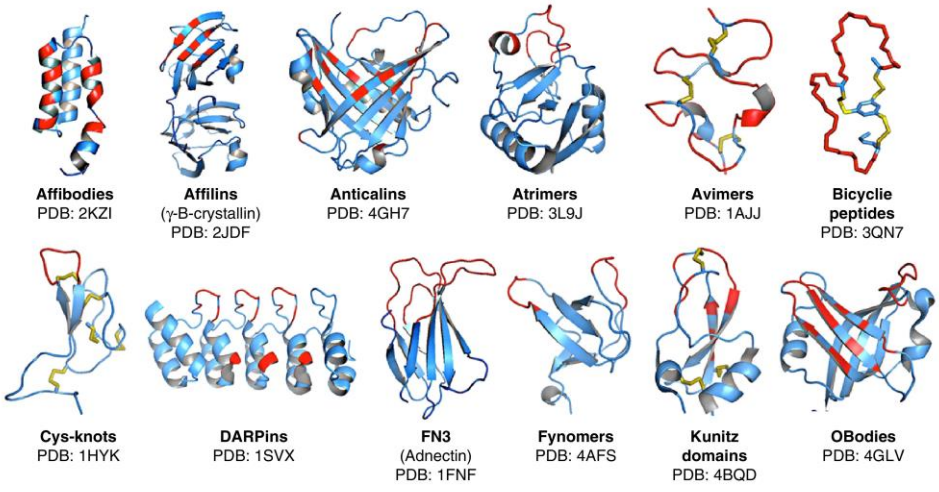


Figure 1.10 Engineered OB-fold (OBody) to lysozyme

(1) Model of OBody (blue) showing binding surface with mutated residues (numbered, green stick Figures) surrounded by numbered loops (L1-L4) (2) Model of OBody in complex with lysozyme (green) showing interaction at binding site with loop residues. [used with permission from (Steemson et al., 2014)].

These binding proteins, antibodies and non-immunoglobulin scaffolds have been engineered using various techniques ranging from the native immune system of mammals to *in vitro* selection techniques such as phage display and ribosome display. The following figure summarises key features of different non-immunoglobulin scaffolds (Fig 1.11)



Scaffold	Parental protein	Structure	Randomization	MW (kDa)
Affibodies	Z domain (protein A)	α-helical	Helix randomization	6
Affilins	γ-B-crystallin	β-sheet	Beta-strand randomization	20
	Ubiquitin	α/β	Beta-strand randomization	10
Anticalins	Lipocalin	β-sheet + α-helical terminus	Loop randomization Beta-strand randomization	20
Atrimers	C-type lectin (tetranectin)	α/β	Loop randomization	3 × 20
Avimers	A-domain	Ca^{2+} binding Disulfide constrained	Loop randomization	4
Bicyclic peptides	Peptide	Chemically constrained	Loop randomization	2
Cys-knots	Peptide	β-sheet Disulfide constrained	Loop randomization	4
DARPins	Ankyrin repeats	α-helical + β-turn	Helix randomization Beta-turn randomization	14-21
FN3 scaffolds (Adnectins, Centyrins, Pronectins, Tn3)	Fibronectin (type III)	β-sheet	Loop randomization Beta-strand randomization	10
Fynomers	SH3 domain (fyn kinase)	β-sheet	Loop randomization	7
Kunitz domains	Serine protease inhibitor	α/β Disulfide constrained	Loop randomization	7
OBodies	OB-fold	β-sheet	Loop randomization Beta-strand randomization	12

Drug Discovery Today

Figure 1.11 Comparison between non-immunoglobulin scaffolds

The features of the four most popular scaffolds (monobody = FN3 scaffold) with OBodies and few others are shown with relation to original domain, secondary structure, area of randomisation and size of scaffold (MW) along with their PDB structures. [used with permission from (Vazquez-Lombardi et al., 2015)].

1.5 Techniques in protein engineering

Traditionally the techniques for protein engineering required mammals or were labour intensive and expensive. Hybridomas were used to generate monoclonal Antibodies in mammals (Figure 1.12). This technology was based on using hybrid cells resulting from the fusion of an antibody producing lymphocyte with a tumour cell.

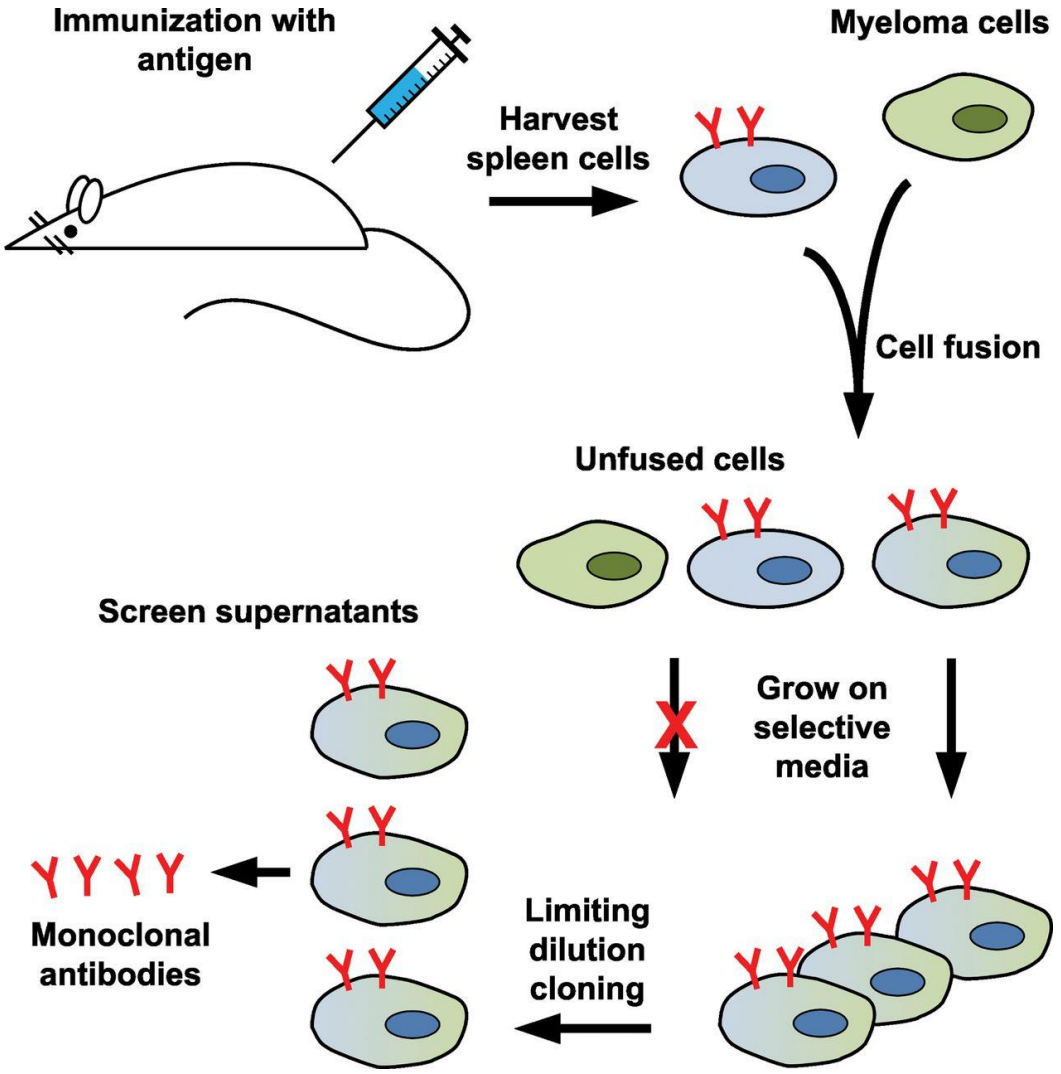


Figure 1.12 Hybridoma technology

The target antigen is injected into the mouse and antibody forming B-cells (from spleen) are isolated and fused to tumour (myeloma) cells producing hybridomas. The successful hybridomas are selected (using selective media) and screened for target specific antibodies (by limited dilution cloning). The specific monoclonal antibodies are harvested by clonal expansion. [used with permission from (Ruigrok, Levisson, Eppink, Smidt, & van der Oost, 2011)].

Hybridoma technology exploits the mammalian immune system in which a diverse set of B-cell lymphocytes each expressing a different antibody clone are exposed to a target antigen to produce target specific antibodies. The B-cells

carrying these specific antibodies are subjected to random mutations in the genes for the surface binding loops, a process called somatic hypermutation. This two-step process results in antibody clones with high affinity to the target antigen (Hoogenboom, 2005).

Later, libraries were made by cloning genes for antibody fragments into bacteria to produce antibodies. They were generated by ligating DNA fragments into lambda phage or plasmid DNA for subsequent transformation into *Escherichia coli*. These antibodies were screened by binding to radiolabelled target antigen. The screening approach limited the number of recombinants to 10^6 due to the limitation of growing a large number of agar plates for individual clones (Hoogenboom, 2005) (FitzGerald, 2000). This led to the invention of display technology which did not require mammals nor was limited by the screening approach.

1.6 Display technology

Display technology was invented when antibodies were successfully displayed on the surface of filamentous phage (G. P. Smith, 1985) and combined with PCR based gene diversity generation. This meant that each phage (or phagemid particle) containing the gene coding for the antibody variant had the equivalent antibody displayed on its surface, and phage display was born (Figure 1.13).

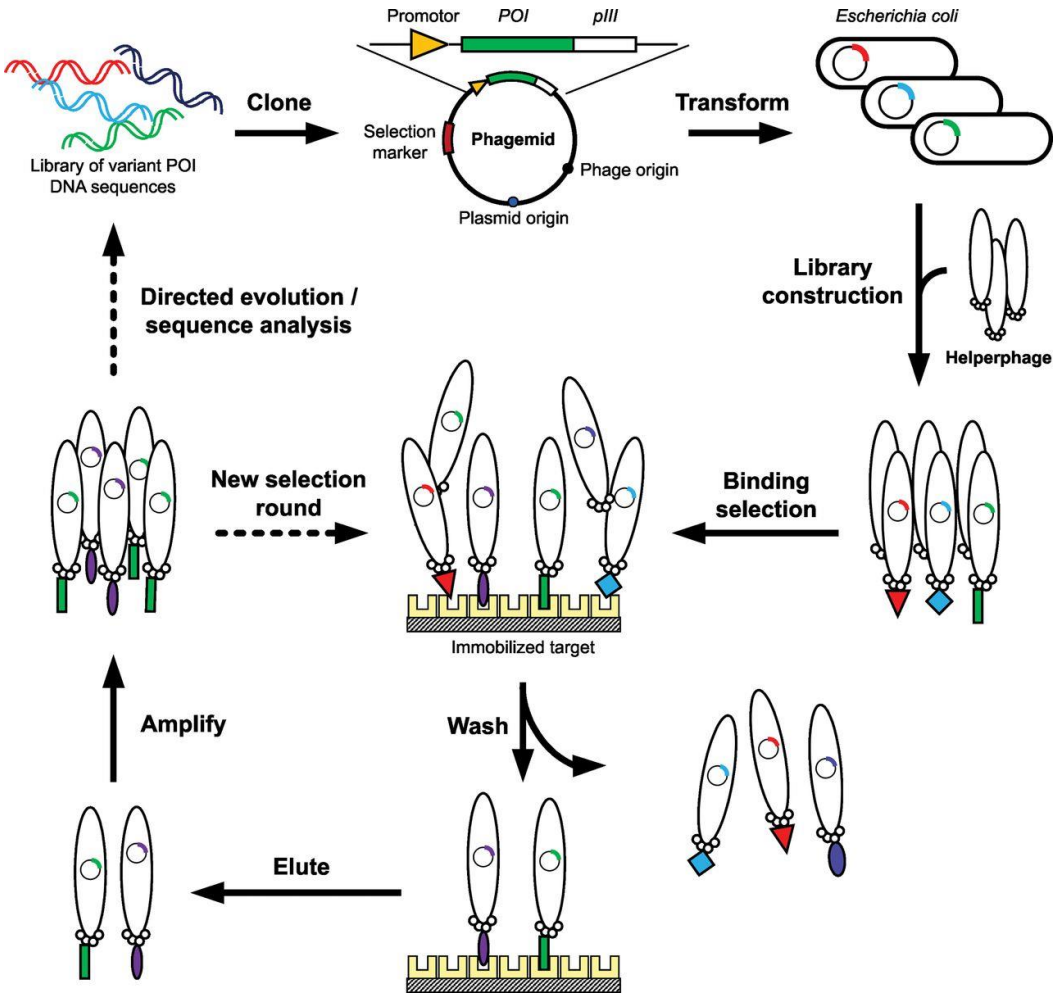


Figure 1.13 Phage display cycle

A diverse gene library is cloned into a phagemid and transformed into *E. coli*., with each cell containing an antibody variant (POI, protein of interest) which is produced on the surface of phagemid particles with a helper phage. These phagemid particles are made to bind to a target antigen and non-specific antibody variants are washed away. The specific antibody variants are eluted and amplified for sequence analysis and subsequent rounds of selection using target antigen until desired antibody variant is obtained. [used with permission from (Ruigrok et al., 2011)].

The advantage of phage display is the ability to generate antibody libraries with huge diversity from non-immunized mammals including humans as well as from artificially synthesized genes. Further, the genotype is directly linked to the phenotype by packaging of the DNA in phage and display of single clones on the exterior. Another critical step is the ability to select and then amplify chosen clones. Newer display technologies such as ribosome and yeast display have also become popular in generating monoclonal antibodies and other peptides. The display technologies share four steps with the immune system, 1) generating gene diversity similar to a diverse library of non-immunized B-cell lymphocytes, 2) coupling of phenotype to genotype similar to each B-cell coding for the surface

displayed antibody clone, 3) selective pressure with the target antigen and 4) amplification to generate target specific antibodies. It is important to note that there are only a few specific clones in a library of millions of non-specific clones. Therefore screening is challenging and display technologies must get around this problem by multiple rounds of selection with target antigen and then amplification of the selected clones (Hoogenboom, 2005).

1.6.1 Ribosome display

Ribosome and mRNA display are *in vitro* techniques that rely on a stable complex between the coding mRNA and the generated antibody fragment (Figure 1.14).

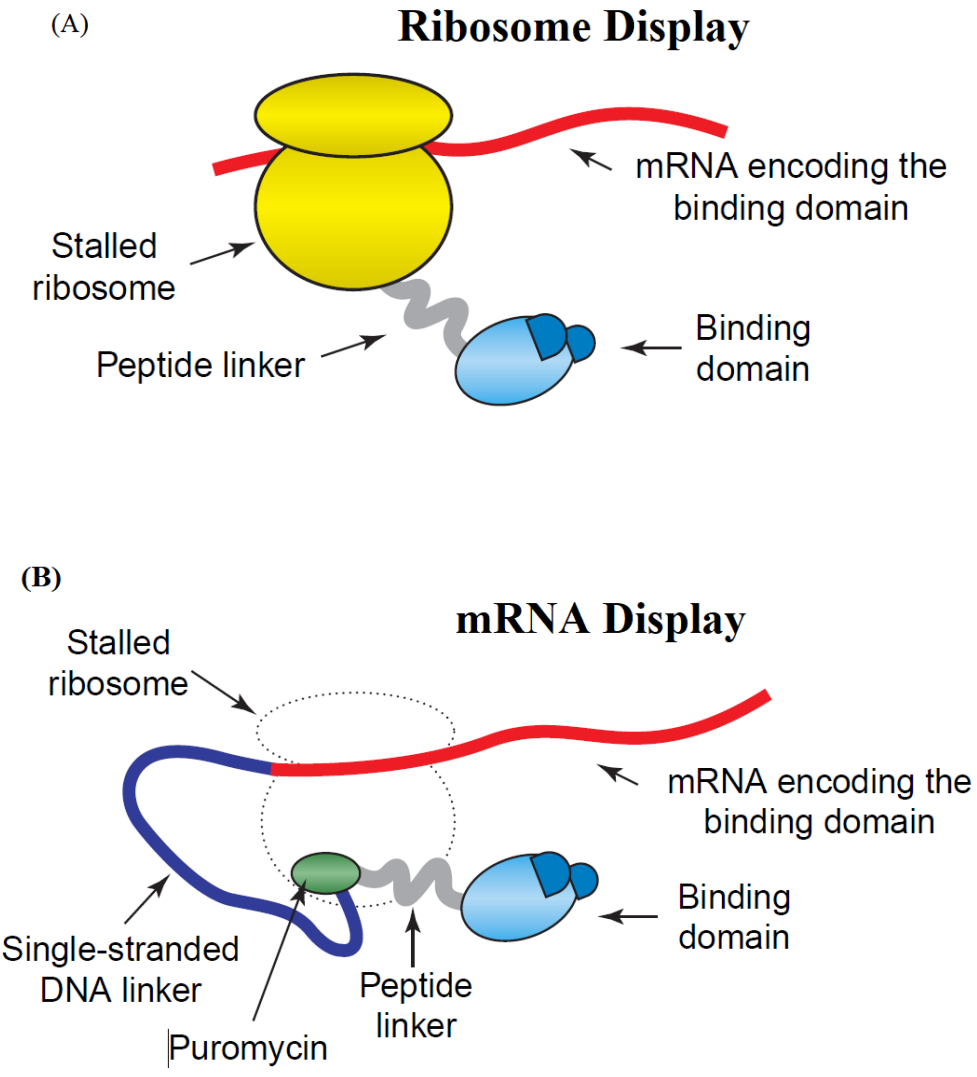


Figure 1.14 Ribosome and mRNA display

In Ribosome display (A), the coupling is between ribosome (yellow), mRNA (red strand) and antibody fragment (blue). In mRNA display (B), the coupling is between puromycin (green) at one end connecting the antibody fragment to the mRNA (red) via a DNA linker (dark blue strand). [used with permission from (FitzGerald, 2000)].

In ribosome display, direct connection of the gene with its antibody fragment is done by 'freezing' the ribosome in translation with the mRNA that encodes the antibody fragment. The ribosome-mRNA-protein (antibody) complex is stabilized in the following ways, 1) controlling magnesium ion concentration 2) addition of chloramphenicol to terminate translation and 3) cooling down the translating protein lacking stop codons. The protein of interest coupled with its mRNA is separated and DNA is regenerated by reverse transcription PCR. The generated genes that encode these antibodies can be used for subsequent rounds of selection (FitzGerald, 2000). The advantage of this display technology is its 'in built' affinity improvement process due to errors resulting from reverse transcription. This leads to generation of antibodies or peptides with very high affinity analogous to somatic hypermutation mentioned for antibodies (Hoogenboom, 2005). The other advantage of *in vitro* display technology is the large initial library diversity due to the absence of an *E.coli* transformation step (FitzGerald, 2000).

1.6.2 mRNA display

In mRNA display (Figure 1.13), a small molecule called Puromycin that resembles the aminoacyl end of transfer RNA (tRNA) is used to form a covalent complex between the mRNA and its antibody fragment. This is done by attaching puromycin to the end of a small ssDNA linker. The other end of the DNA linker is attached to the mRNA. When the ribosome reaches the junction between DNA and mRNA, it 'freezes' resulting in puromycin forming a covalent bond with the antibody fragment at the ribosomal aminoacyl site. The gene responsible for the antibody is recovered similar to ribosome display using reverse transcription (FitzGerald, 2000). Both these technologies can use *in vivo* or *in vitro* translation for producing proteins of interest on a large scale (Hoogenboom, 2005) though production of stable complexes are susceptible to ribonucleases and are cumbersome to produce (FitzGerald, 2000).

1.6.3 Yeast display

The yeast *Saccharomyces cerevisiae* has been successfully used to display antibodies on its surface. The antibodies are fused to alpha-agglutinin yeast adhesion receptor and displayed on the cell wall. Potential antibodies are 'hand-picked' via flow cytometry using fluorescent target proteins. This allows

screening for target specific antibodies down to nanomolar concentration binding to target protein. It has the added advantage of selecting for antibodies based on their kinetic properties, expression level or their binding to different sites of the target protein. Phage display using yeast for producing gene diversity has also been explored with libraries up to 10^{14} recombinants made (Hoogenboom, 2005).

1.7 Phage display

There are two types of phage based on its interaction with host cell: Lytic phage where phage assembly in the host cell cytoplasm leads to cell death and release of phage (e.g. T7 phage) and lysogenic phage where phage assembly leaves the host cell intact and the phage are exported from the cell (e.g. filamentous phage).

Filamentous phage (e.g. M13) which depend on the F-pilus of *E. coli* are most commonly used for phage display. The phage is 900 nm x 7 nm wide, the length depending on the ssDNA inside. The viral coat is made up of 2700 copies of major coat protein pVIII and capped at both ends with minor coat proteins (pIII, pVI, pVII, pIX) (Figure 1.15) (Wilson & Finlay, 1998).

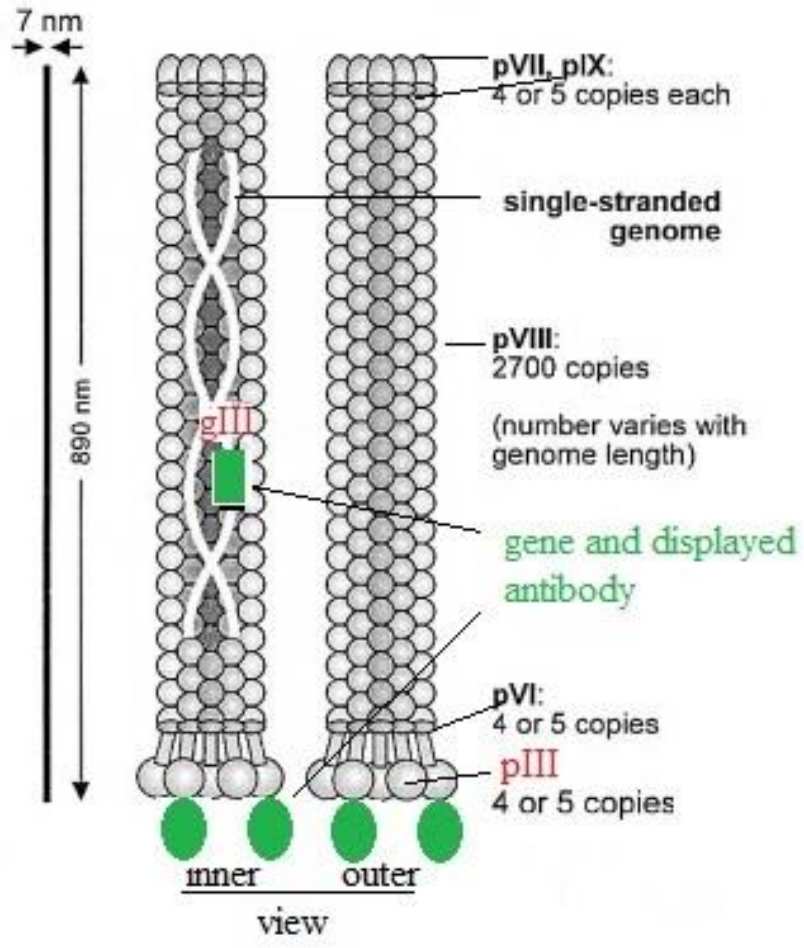


Figure 1.15 Schematic of filamentous phage

Diagram of M13 filamentous phage showing single stranded (ss)DNA in the inner cavity and major coat protein pVIII and other minor coat proteins- pIII, pVI, pVII, pIX on the outer surface. The gene of interest is fused to gene III (gIII, green rectangle) which results in the protein of interest (POI) being displayed in fusion with pIII (minor coat protein, green circle) on the surface of phage [adapted with permission from (Wilson & Finlay, 1998)].

The two coat proteins pIII and pVIII on filamentous phage have been used for the display of foreign peptides on their N-terminal end. Both these proteins have signal sequences on their N-terminal end which allow for export through the inner membrane of *E. coli* (Figure 1.15). These proteins are anchored into the inner membrane via their C-terminal hydrophobic domains (Fuh & Sidhu, 2000) followed by cleavage of the signal sequence by signal peptidase (Marvin, Symmons, & Straus, 2014; Wilson & Finlay, 1998). Foreign peptides have been inserted between these coat proteins and their signal sequences for display on the phage surface as shown in Figure 1.14 (Paschke & Hohne, 2005). The coat protein pVIII has been successful in displaying only small peptides as opposed to pIII which can display large proteins (Wilson & Finlay, 1998). Nanomolar affinity

proteins and libraries of 10^{10} individual clones have been engineered with phage display (FitzGerald, 2000).

Filamentous bacteriophage (or genus Inovirus) infect gram negative bacteria and are secreted in large amounts without killing the host cells. They can produce up to 10^{13} phage per ml of culture. The pVIII protein is made up of 50 amino acids and thousands of these proteins form the rod containing a central cavity that houses single stranded DNA (ssDNA). This hollow cylinder is closed by two pairs of minor proteins pVI-pIX and pIII-pVI at either ends as shown in Fig 1.16 (Rakonjac, Bennett, Spagnuolo, Gagic, & Russel, 2011).

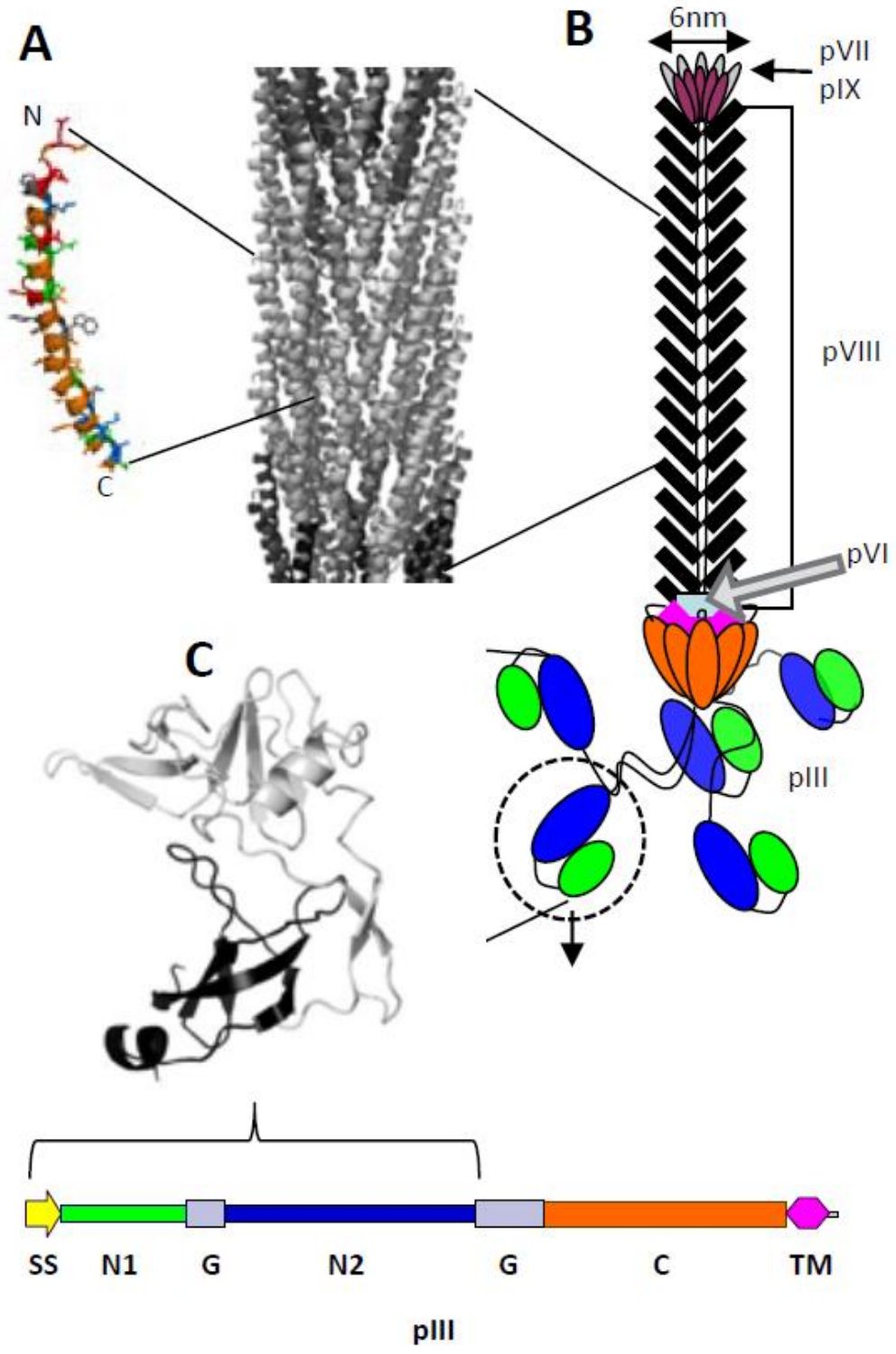


Figure 1.16 Structure of filamentous phage coat proteins pVIII and pIII

A) α-helical Structure of pVIII (PDB: 2c0w) showing packing with relation to other pVIII forming the major coat. B) Filamentous phage showing all coat proteins. C) structure of N1 and N2 domains of pIII (PDB: 1g3p) with illustration of entire pIII below it. SS: signal sequence, G: glycine linker, C: C-domain of pIII, TM: transmembrane α-helix. [adapted with permission from (Rakonjac et al., 2011)]

pVIII contains a signal sequence and is inserted into the inner membrane via YidC translocase with an N-terminal part (-) in the periplasm, a transmembrane helix in the inner membrane and a C-terminal part (+) in the cytoplasm. When the phage is assembled, the pVIII units stack against each other via hydrophobic interactions. The positively charged C-terminal part is directed inwards and interacts with negatively charged phosphate of ssDNA. The length of the phage depends on the size of the ssDNA and the contribution of four minor proteins to initiation and termination of elongation.

pVII and pIX are inner membrane proteins that initiate phage assembly. The C-terminal part of this complex interacts with packaging signal (PS) of ssDNA leading to its packaging. pIII and pVI form a terminal cap which is inserted into end of the phage and extruded. pVI is made up of three transmembrane α-helices with an N-terminal part in the periplasm and a C-terminal part in the cytoplasm as for the other coat proteins. pIII is also required for phage infection of host bacterium and is made up of three domains (N1, N2 and C) separated by glycine linkers. The C-terminal transmembrane α-helix anchors it to the inner membrane leaving five amino acids in cytoplasm as shown in Fig 1.16.

pIII initiates infection by binding to F pilus on surface of *E.coli* via N2 domain. This binding uncovers the N1 domain that binds to conserved TolA in the inner membrane. This is assisted by spontaneous pilus retraction. The change in conformation is possible via a cis-trans isomerisation of proline 213 in the N2 domain. This is followed by α-helical C-domain mediated viral uncoating, insertion of pIII in inner membrane, DNA entry into cytoplasm and insertion of pVIII in inner membrane as shown in Fig 1.17. All three domains of pIII and TolQRA complex are required for phage infection (Marvin et al., 2014).

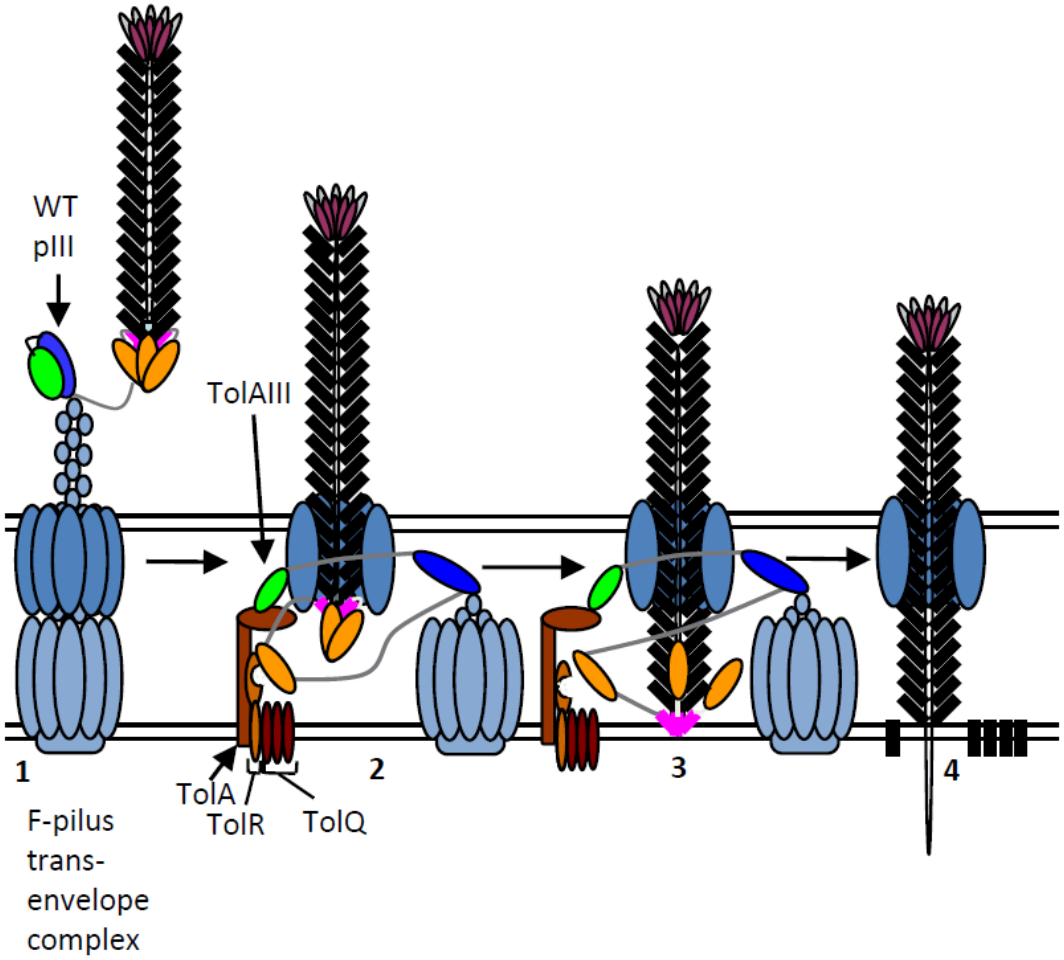


Figure 1.17 Illustration of pIII mediated infection

1) Binding of filamentous phage via N2 domain (blue oval) to F-pilus of gram negative bacteria (light blue appendage). 2) Pilus retraction and binding of N1 domain (green oval) to TolAIII (TolQRA) of inner membrane. 3) Opening of C domain and insertion of TM (pink) into inner membrane. 4) Entry of phage ssDNA into the cytoplasm with insertion of pVIII into inner membrane. TM – transmembrane α-helix of C domain (pIII). [adapted with permission from (Rakonjac et al., 2011)]

pII, pV and pX are the only proteins which remain in the cytoplasm and assist with replication. The + strand acts as a template to produce a complementary – strand at the negative origin of replication with the help of RNA polymerase. This generates an RNA primer which dissociates to complete the rest of the – strand with the help of host DNA polymerase. This dsDNA is called the replicative form (RF) and is used to produce phage proteins and more + strands via a rolling circle mechanism with the help of pII. pV binds to + strand (ssDNA) and forms the packaging substrate (dimer) and inhibits pII induced replication. pV does not bind to ssDNA at the PS, the latter binds to phage export complex (pI/pXI/pIV) spanning the inner and outer membrane, leading to initiation of phage assembly

by interaction with pVII-pIX complex. pI contains an ATP binding domain and forms the inner membrane channel with pXI, whereas pIV forms a barrel like outer membrane channel made up of three rings with the N-terminal ring facing inward and interacting with the inner membrane channel. This contiguous channel is used for phage elongation and extrusion. pIV is a homologous to type II, III secretion and type IV pilin systems collectively referred to as secretins. pV is replaced by pVIII as phage is extruded from the cell with final release of the phage contributed by the pIII-pVI complex.

The phage life cycle (Figure 1.18) involves all coat proteins being transported to the periplasm and anchored to the inner membrane. These proteins form the outer shell housing the phage DNA as the phage leaves the bacterial cell. It is important to note that different coat proteins use different channels in the inner membrane to reach the periplasm. pIII is exported through a general secretory channel (Sec translocase) in the inner membrane and pVIII is Sec independent. This has an impact on the type of proteins (e.g. antibodies) that can be displayed using pIII as Sec translocase exports only unfolded proteins (Wilson & Finlay, 1998).

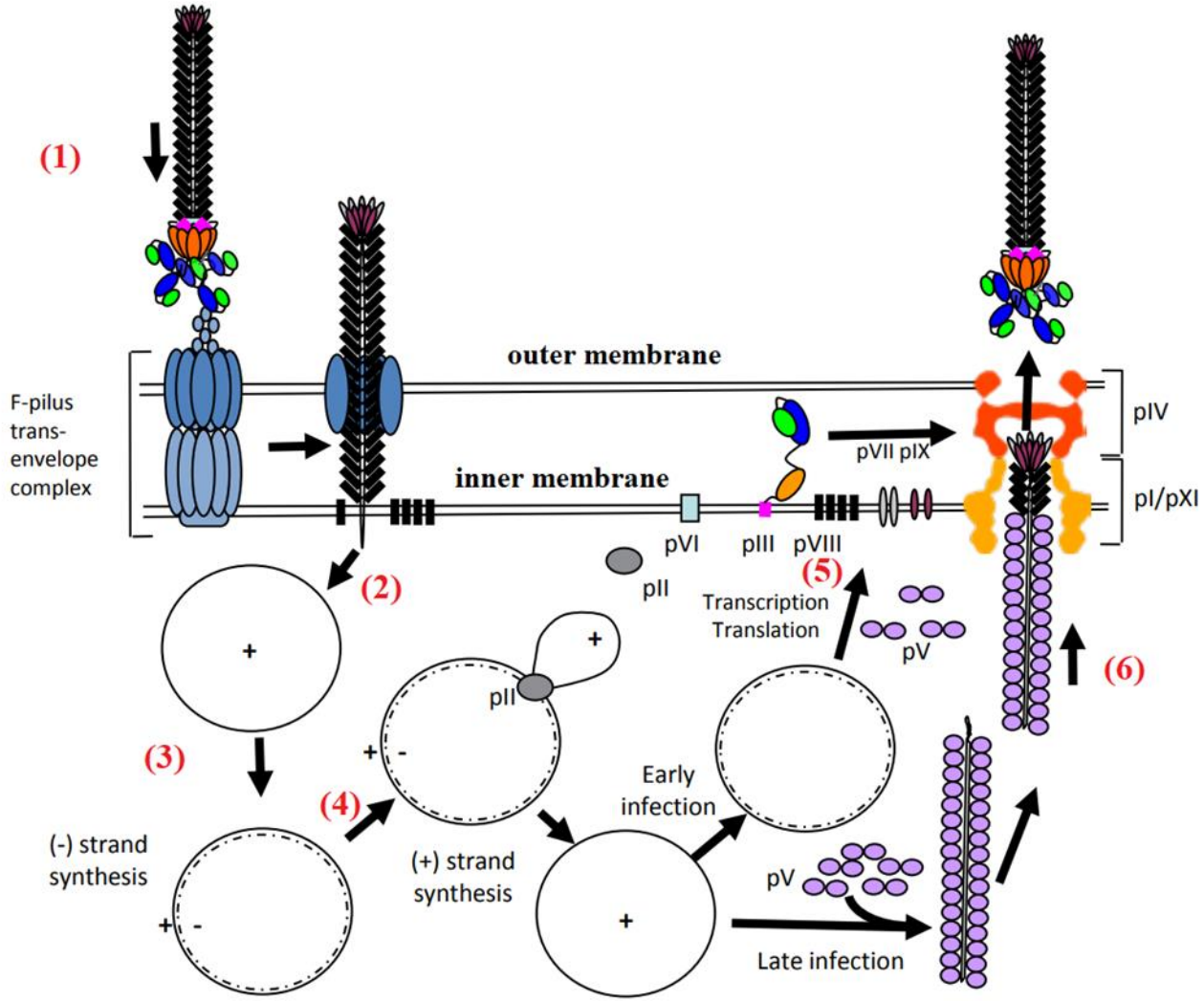


Figure 1.18 Filamentous phage life cycle

(1) Infection starts with binding of pIII of phage to the F-pilus of *E. coli* leading to (2) internalisation of its ssDNA also called (+) strand. The (+) strand is used as a template (3) to produce complementary (-) strand resulting in double stranded (ds)DNA called replicative form (RF). The dsDNA is used as template (4) to produce more (+) strands with the help of a phage protein (pII). The dsDNA is also used as a template (5) to produce phage proteins. The (+) strand is coated with another phage protein (pV) which (6) delivers it to the phage export channel (pI, pIV) to be coated with proteins before extrusion from cell. [adapted with permission from (Rakonjac et al., 2011)].

pIII is a minor coat protein where two-thirds of the N-terminal end is required for F-pilus adsorption to the host cell and infection, and one-third of the C-terminal end is required for capping and phage extrusion from the cell. The failure of capping leads to non-infectious phage whereas the failure of phage extrusion is lethal to the host cell. This is important because using pIII as a display protein may affect phage display based on the nature of the fusion protein (e.g. antibody) ranging from blocking phage infection of *E.coli* (Rakonjac et al., 2011; Wilson & Finlay, 1998) to failure of export of pIII across the inner membrane (Dammeyer & Tinnefeld, 2012).

1.7.1 pIII as display protein in phage display

The most common filamentous phage used in phage display is M13. The type of display can be monovalent (single display of POI) or polyvalent (multiple display) dependent on the vector (phage or phagemid) or helper phage used (Figure 1.19). Type 3+3, where a separate wild type (WT) gene III is provided by a separate plasmid from a helper phage is the preferred method for phage display as it selects for high affinity POIs, due to low display of number of the POI coupled to protein III on surface of phagemid particles. Type 3 where gene III is fused to gene for POI or type 33, where gene III and gene for POI are separately present in the same phage are reserved for the first round of selection to screen for the most binding clones due to increased display in number of POI on surface of phage (Hoogenboom, 2005). Type 3 and 33 favours low affinity due to combined binding from multiple low affinity POI and target interactions, also called avidity. One limitation of Type 3 display is the POI may interfere with phage infection due to absence of wt pIII required for adsorption to F-pilus in *E. coli*, which is overcome in type 33 and 3+3 due to separate wt gIII. These three formats have also been used with pVIII as a display protein (referred to as Type 8, 88 & 8+8) (Wilson & Finlay, 1998).

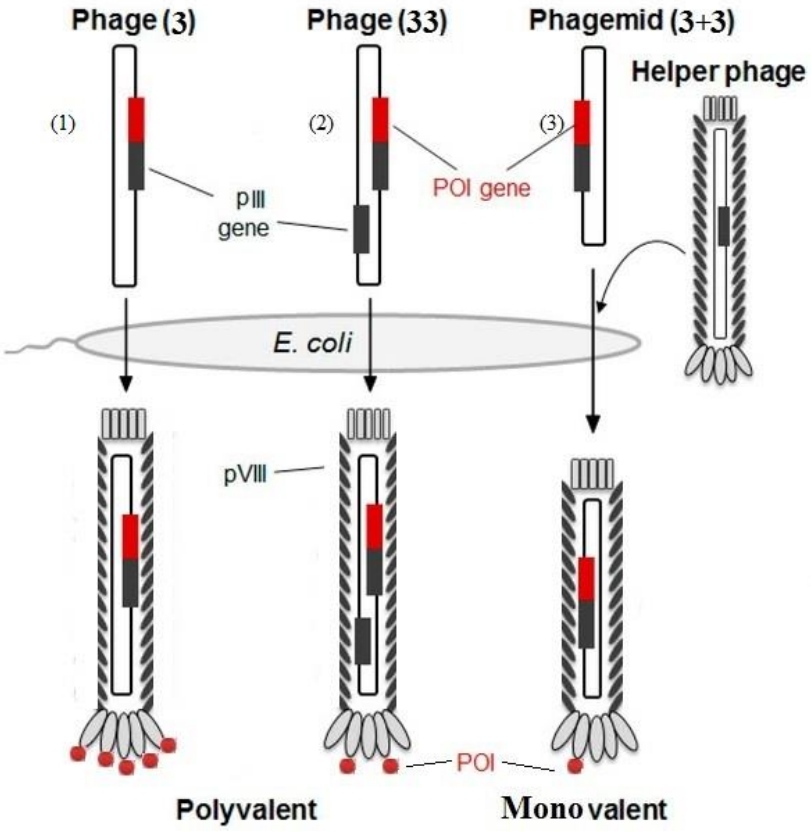


Figure 1.19 Different formats of phage display using pIII

(1) Type (3) where wild type (wt) gene III (gIII, grey) is fused to gene for foreign protein (red). This leads to display of foreign protein (red circle, POI) on all pIII (grey oval). (2) Type (33) where a separate wt gIII is present. This leads to display of both wt pIII and foreign protein. (3) Type (3+3) where a separate wt gIII is provided by a separate plasmid from helper phage. This leads to similar display as type (33). In type 3+3 the phage plasmid lacking a functional pIII is called a phagemid [adapted with permission from (Hwang, 2014)].

1.7.2 Protein translocation systems for phage display using pIII coat protein

Translocation is the process of export of proteins across a cell membrane. The POI produced in the cytoplasm of the host bacterial cell has to cross the inner membrane to be displayed on the surface of the phage. The general secretory pathway (SEC) is known to transport unfolded proteins from the cytoplasm into the periplasm (E.g. display protein pIII). The twin arginine translocation pathway (TAT) is known to transport folded and cofactor proteins from the cytoplasm into the periplasm (Figure 1.20)

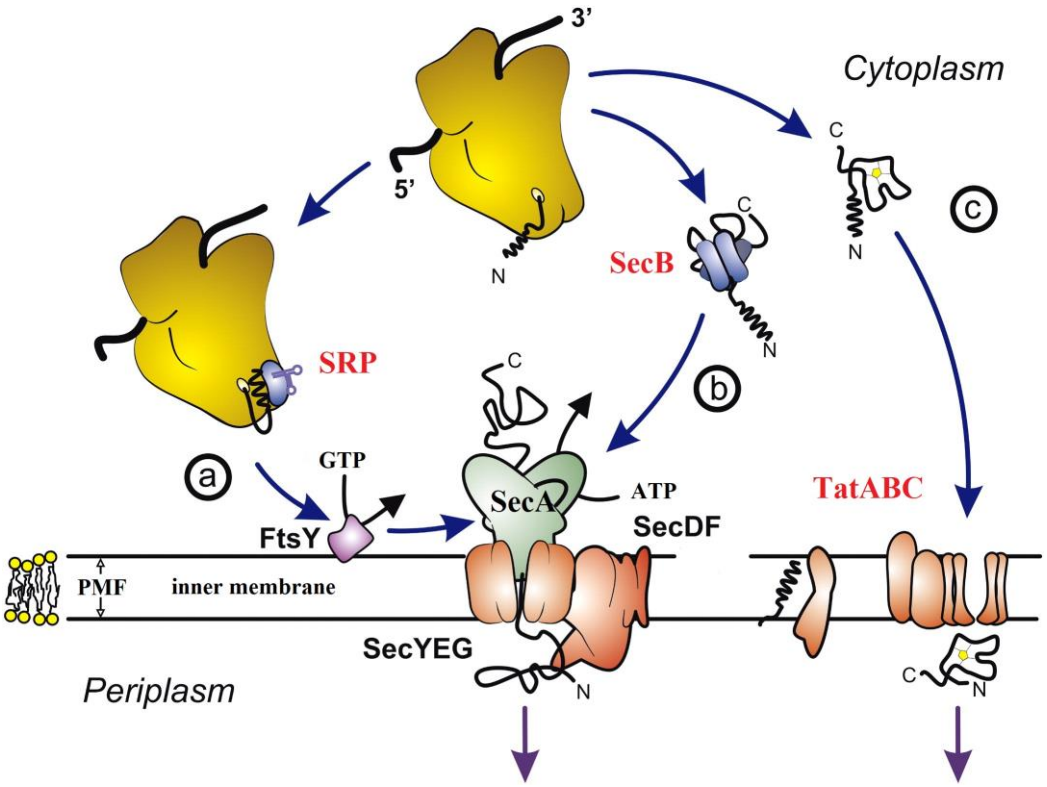


Figure 1.20 Inner membrane translocation pathways in bacteria

Sec based translocation of unfolded proteins by two sub-pathways (a) co-translational via SRP, (b) post-translational via SecB. (c) Tat based translocation of folded proteins. The translating ribosome with 5'-3'mRNA is shown as the yellow structure; (1) SecA = Sec pathway motor protein, (2) SecB = Sec-post-translational pathway specific chaperone, (3) TatABC = Tat pathway export channel, (4) SecYEG = Sec pathway export channel, (5) PMF = proton motive force and (6) ATP & GTP = energy sources [adapted with permission from (Natale, Bruser, & Driessen, 2008)].

The Sec pathway consists of ATP dependent motor protein SecA and a channel protein complex SecYEG (together called the Sec translocase) to export unfolded proteins with a N-terminal signal sequence. Unfolded proteins are recognized by either a molecular chaperone called SecB or a signal recognition particle (SRP) based on the signal sequence of the POI. The SecB chaperone binds to a fully translated protein (post-translational) and prevents it from folding to its native state and delivers it to the Sec translocase for export into the periplasm. SRP binds to the signal sequence and delivers the translating ribosome directly (co-translational) to the Sec translocase with the help of its receptor - FtsY, resulting in the protein being threaded into the periplasm. SecB is used to translocate secretory proteins whereas SRP is used for membrane proteins (Natale et al., 2008).

The Tat pathway consists of Tat A, B and C protein sub units in the inner membrane which are together responsible for protein translocation of folded proteins from cytoplasm to periplasm using proton motive force (PMF). Protein translocation in these three systems is dependent on chemical groups or proton movement. Sec translocase exports post translated unfolded proteins using ATP and in case of co-translation, GTP is preferred.

Protein signal sequences dictate the translocation pathway that a protein chooses to exit the cytoplasm into the periplasm. They are conserved across all domains of life and have a three-part structure consisting of a positively charged N-terminal segment, a hydrophobic core and a polar C-terminal segment. The signal sequence is usually found at the N-terminal end of the protein sequence and they have a cleavage site recognised by signal peptidases (Figure 1.21).

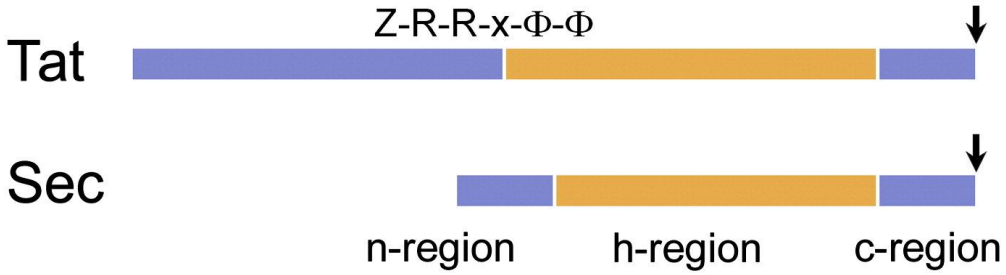


Figure 1.21 Structures of Tat and Sec signal sequences

Signal sequences that direct proteins to Tat or Sec pathway show a three-part structure; Tat has a conserved twin arginine (RR) amino acid sequence, Z denotes a polar amino acid and Φ denotes hydrophobic amino acids, n-region- N-terminal end of signal sequence, h-region- Hydrophobic core of signal sequence, c-region- C-terminal end of signal sequence. Sec has no conserved sequences. The arrow shows the cleavage site in both signal sequence by signal peptidases. [used with permission from (Natale et al., 2008)].

SecB transports proteins that have more hydrophobic amino acids. SRP transports proteins that have a positively charged N-terminal end, which make electrostatic interactions with membrane phospholipids. There is no sequence homology between signal sequences in the Sec pathway. There are two arginines conserved in the signal sequence of proteins exported by the Tat pathway. They occur at the junction of their longer N-terminal ends and the hydrophobic core of the signal sequence (Natale et al., 2008).

1.7.2.1 Phage display using post-translational Sec pathway

Phage display via the post-translational Sec pathway uses a specific chaperone called SecB which binds to the core region of the POI and stabilizes it in an unfolded state after exiting the ribosome (Figure 1.17b). SecB is required for maintaining the POI in an unfolded state and has a preference for hydrophobic and basic amino acids in the mature POI. It does not bind to the signal sequence of the POI. SecB delivers the unfolded POI to SecA of the Sec translocase system. In the absence of SecB, the signal sequence guides the protein to SecA using general chaperones. SecA is an ATPase (in the Sec translocase system) and binds to the signal sequence and core region of the POI. SecA is located in the inner membrane bound to negatively charged phospholipids and associated with the SecYEG complex (also in the inner membrane) to form the protein conducting channel (PCC). SecA couples POI export through the inner membrane with ATP binding and hydrolysis. The PCC forms a hydrophilic channel for POIs to be exported from the cytoplasm to the periplasm (Natale et al., 2008).

1.7.2.2 Phage display using co-translational Sec pathway (SRP)

SRP shares the Sec translocase system to export POIs and it is better at exporting fast folding and stable cytoplasmic proteins like DARPins, GFP (Nangola, Minard, & Tayapiwatana, 2010) or thioredoxin when compared to SecB, which may be inefficient in maintaining them in an unfolded form. Highly stable POIs have been displayed at levels up to 700-fold above conventional SecB route due to faster translocation through Sec translocase. POIs dependent on cytoplasmic cofactors for folding are poor candidates for SRP route (Steiner, Forrer, Stumpp, & Pluckthun, 2006).

1.7.2.3 Phage display using Tat pathway

The Tat pathway exports only folded POIs and has no preference for chaperones except when cofactors like molybdenum or nickel have to be added before protein folding and subsequent targeting to the Tat translocase. Tat translocase is made up of Tat A, B and C (Figure 1.17c). Tat B and C binds to the twin arginine motif of signal sequences when folded POIs are delivered by chaperones. This leads to the recruitment of Tat A leading to softening of the inner membrane and a conformational change in both Tat B and C then causes translocation of the POI

into the periplasm (Natale et al., 2008). In phage display using coat protein pIII which contains cysteines and requires an oxidising environment to form a functional protein, pIII has to be separately exported via Sec pathway and fused to the POI in the periplasm if the Tat pathway is to be employed for phage display (Nangola et al., 2010).

1.7.3 Phage display using pVIII coat protein

Phage display has been attempted on the amino terminus of the major coat protein pVIII. There are 2700 copies per phage particle (Figure 1.14) making it an ideal candidate for selection of low affinity POIs due to high avidity but a poor candidate for high affinity POIs. pVIII can display only short peptides less than 10 amino acids due to the spatial constraints from pVIII on the phage surface as well as the metabolic stress of making thousands of copies. pVIII has to be expressed in large quantities in a natural phage life cycle making it difficult if the fusion peptide is toxic to the cell. The other limitations of pVIII as a display protein is reduced diversity of phage library due to incompatibility with certain codons (Rodi & Makowski, 1999).

The minor proteins pVII and pIX lack signal sequences and spontaneously insert into the inner membrane of gram negative bacteria. However, phage display using pVII and pIX has been attempted successfully on the N terminal end and requires a signal sequence (Rakonjac et al., 2011).

1.7.4 T7 phage display

T7 is a double stranded DNA phage used for phage display. It is encapsulated in an icosahedral shell made up of 415 copies of protein 10. A variant called 10B is used to display peptides on its surface. This type of phage is released by the lysis of the cell, removing restrictions to export of proteins from cytoplasm to the periphery of the host bacteria (Rosenberg et al., 1996) (Wilson & Finlay, 1998). It is a robust structure capable of being resistant to harsh physical conditions and also multiplies rapidly. It is capable of phage display libraries up to 10^9 recombinants. However display of peptides is limited to a size of 50 amino acids due to spatial constraints of the protein shell (Rosenberg et al., 1996).

1.8 Thesis Outline

Protein folds from antibodies to non-immunoglobulin scaffolds are superior to small molecules in molecular recognition with regards to sensitivity and specificity from superior size and complexity. OB-folds are a non-immunoglobulin scaffold which shows similar potential in molecular recognition of proteins or ligands. There are various techniques to exploit the potential of these protein folds as binding agents. Phage display is one such engineering technique and there are ways to optimize phage display of non-immunoglobulin scaffolds especially OBodies.

This thesis focuses on characterisation of an OB-fold that binds to a small molecule, progesterone (P4) and using phage display to improve the binding affinity of this OBody towards P4. Chapter 2 summarise the experimental methods. Chapter 3 describes the production and characterisation of a naïve OBody called D7 that binds to P4. Chapter 4 involves the affinity maturation of OBodies including production and characterisation of the best P4 binding OBody B7. This chapter also describes the X-ray crystal structure of an OBody-P4 complex. Chapter 5 describes the characterisation of a foreign signal sequence that enhances phage display of stable proteins including OBodies. Chapter 6 provides a summary and discussion including future experiments.

2 Materials and methods

2.1 General molecular biology

2.1.1 Media and agar plates

Media (7.1.1) were prepared using standard recipes or pre-made media were made in UP H_2O and both autoclaved as per standard sterilization protocol (121°C for 20 minutes) and stored at room temperature (~ 25°C). Agar plates were made by adding 15 g bacto-agar to liquid media (1L of 2xYT, LB) and autoclaving the solution. This was stored in liquid form at 50°C or in solid form (with antibiotic) as ~25 ml (small) or ~250 ml (large) plates at 4°C before using to grow *E. coli* colonies (25 -37°C). The recipes for the media are mentioned in the appendix.

2.1.2 *E. coli* strains

The genotypes are provided in 7.1.7.

2.1.2.1 TG1

This strain was used for phage display since it has an F-pilus for phage infection and amber suppression (UAG stop codon is translated as glutamine) between display protein (pIII) and foreign protein conducive to large scale expression of only the foreign protein in non-suppression strains (Carmen & Jermutus, 2002).

2.1.2.2 DH5α

This strain was used for transformation of engineered plasmids into competent *E. coli* cells due to deletion of genes *recA* and *endA* which results in superior plasmid retention (Phue, Lee, Trinh, & Shiloach, 2008).

2.1.2.3 BL21(DE3)

This strain was used for large scale expression of OBodies as it lacks *Ion* and *OmpT* proteases and contains T7 RNA polymerase gene from λDE3 phage for high protein expression under *lac*UV5 promoter (Jeong, Kim, & Lee, 2015).

2.1.3 Induction of culture

2.1.3.1 IPTG

Isopropyl β-D-thiogalactopyranoside (IPTG) was used at 1/1000 stock concentration in culture media at OD_{600nm} between 0.4 - 0.6. The stocks (1M)

were stored at -20°C. IPTG works by inhibiting *lacI* repression of the Lac operon resulting in protein synthesis (Marbach & Bettenbrock, 2012).

2.1.3.2 Helper phage

KM13 helper phage or M13K07δIII (hyperphage) were used for phage infection of *E. coli* TG1 containing phagemids (pOB2) at OD_{600nm} between 0.4 – 0.6 to produce OBody displaying phagemid particles. Hyperphage lacks the gene for pIII thus resulting in increased display of OBody displaying pIII (instead of wild type pIII) (Rondot, Koch, Breitling, & Dubel, 2001)

2.1.4 Plasmids

The sequences of empty vectors and cloned inserts are provided in appendix 7.1.8 and 7.2 respectively.

2.1.4.1 pOB2

A second generation (*Pyrobaculum aerophilum* Aspartyl t-RNA synthetase) naïve OBody library (2GPATS) was used as the insert for this phagemid. A phagemid is plasmid which contains most of the genes in a phage but needs a helper phage for phagemid particle replication (G. P. Smith & Petrenko, 1997). This phagemid is named after the insert, pOB2 (*P. aerophilum* OBody 2nd generation) and used as OBody gene libraries (plasmid) in *E. coli* TG1 (for phage selections). This is a periplasmic expression strain.

2.1.4.2 pET28b

This is a standard protein expression plasmid from the pET system which uses a phage based T7 RNA polymerase for protein synthesis under the control of *lac* operon (Grossman, Kawasaki, Punreddy, & Osburne, 1998; Pan & Malcolm, 2000). This is a cytoplasmic expression strain.

2.1.4.3 pPROEX-HTb

This is another protein expression plasmid which is a hybrid promoter of *trp* and *lacUV5* promoters which leads to high expression of proteins and minimal leaky expression (de Boer, Comstock, & Vasser, 1983). This is a cytoplasmic expression strain.

2.1.4.4 pOPINM

This is a protein expression plasmid with a MBP fusion tag and is based on a modified pTriex2 plasmid which uses the standard T7 *lac* operon for protein expression. The pOPIN system is a suite of similar plasmids with different fusion tags which can be used for cloning of insert (gene) with a single set of primers (Berrow *et al*., 2007). This is a cytoplasmic expression strain.

2.2 Protein expression and purification in *E.coli* cytoplasm

2.2.1 Protein expression and extraction

A colony was looped using disposable sterile loops and added into 10 ml LB using appropriate antibiotic and grown overnight. This starter culture was added to a larger 1000 ml LB with appropriate antibiotic and incubated in a shaking incubator at 37°C, 200 rpm until OD_{600} 0.4 – 0.6, then induced with 1ml 1mM IPTG and incubated overnight with shaking. This culture was transferred to plastic pots and centrifuged at 6500 rpm for 20 minutes. The supernatant was discarded and the pellet was re-suspended with 10 ml binding buffer (TBS with 20mM imidazole, 7.1.3.1). A sonicator (Misonix XL-2020) was used to lyse the cells in the culture with a micro-tip at setting 5 for 15 second bursts alternating with 30 seconds rest (6 cycles) on ice. The lysed cells containing protein of interest was centrifuged again at 13000 rpm for 20 minutes. The supernatant was purified consecutively with three filters, 1.2 μM, 0.45 μM and 0.2 μM and used for Immobilized Metal-ion Affinity Chromatography (IMAC) purification on Fast protein liquid chromatography (FPLC).

2.2.2 Protein purification

2.2.2.1 Immobilized metal-ion affinity chromatography (IMAC)

IMAC is a protein purification method where imidazole ring of histidine (from polyhistidine residues of tagged protein of interest) interacts with a metal ion (Ni^{2+}) in an IMAC column. This tagged protein can be then separately eluted using high concentrations of imidazole which competes as the electron donor (to Ni^{2+}) with histidine from the tagged POI (Bornhorst & Falke, 2000).

GE Histrap HP 5 ml Nickel Sepharose column (stored in 20% ethanol) was prepared by using the following solutions* (in order):

25 ml UP H20

10 ml EDTA

25 ml 50mM NaOH

7 ml $NiCl_2$

10 ml binding buffer (TBS with 10mM imidazole)

*Note: 10 ml UP H_2O was loaded into the column between every solution

EDTA is used to chelate Ni^{2+}, NaOH as a cleaning agent and $NiCl_2$ to recharge the Ni^{2+} resin. The filtered supernatant (containing OBody) is then manually loaded into the column slowly to maximize polyhistidine based binding to the column. The column was then washed with 5 column volumes of binding buffer on FPLC (Amersham Biosciences) and then eluted with elution buffer (TBS with 1 M imidazole) over a gradient spanning 75 ml using FPLC. The fractions containing eluted (purified) OBody (indicated by 280 nm absorbance on FPLC and confirmed by 16.5% SDS-PAGE, 2.2.3.2) was concentrated using 10000 kDa (Millipore) concentrator at 3700 rpm to a volume of ~5 ml. The concentrator was equilibrated with the following solutions before loading IMAC fractions containing OBody.

5 ml 10% Tween-80 (T80)

5 ml binding buffer (TBS with 20mM imidazole)

2.2.2.2 Size exclusion chromatography (SEC)

SEC is a protein purification method which separates proteins on the basis of molecular size using porous material which traps proteins and elutes by steric exclusion leading to larger proteins being eluted first (Hagel & Janson, 1992). This is used as a secondary purification method.

A HiLoad 16/600 Superdex 75 prep grade column (S75 16/600, containing dextran with agarose as the porous medium) was loaded with the following solutions (in order):

125 ml UP H_2O

125 ml SEC buffer (TBS without 20 mM imidazole)

The concentrated protein (~5 ml) from IMAC purification was injected into a 5ml loop attached to the S75 16/600 column via the FPLC and eluted using SEC buffer. The loop was equilibrated before loading OBody as follows:

10ml UP H_2O

10ml SEC buffer (TBS without 20 mM imidazole)

The fractions containing purified OBody were collected based on the 280nm absorbance on FPLC and analyzed using 16.5% SDS-PAGE (2.2.3.2). The fractions were measured using an UV spectrophotometer (Thermo scientific nanodrop) to determine the OBody concentration. The molecular weight and extinction coefficients of the protein were determined using Prot Param web server.

2.2.3 Protein Analysis by SDS-PAGE

2.2.3.1 Gel preparation

The resolving layer (7.1.4.1) is poured into a (5) gel caster and covered with isopropan-2-ol (IPA, to prevent bubbles) for 1 hour until the gel solidifies. The IPA is poured out and stacking gel (7.1.4.2) is poured into the gel caster over the solidified resolving layer. Combs are placed in the stacking gel which carves out the wells in the gel. This is allowed to solidify for 40 minutes. Then the gel is used for protein analysis in a PAGE system or stored (wrapped in damp paper towel) at 4 °C. Note that damp paper towels are placed over the gel caster during both gel solidification steps. The recipe for the SDS-PAGE reagents are mentioned in the appendix.

2.2.3.2 Gel electrophoresis

Protein (Obody) containing fractions were separately mixed with 4x SDS loading dye (7.1.4.3) in eppendorf tubes and heated to 95 ^{o}C for 5 minutes. These heated fractions were loaded separately onto wells in 16.5% SDS–PAGE gels immersed in 1x Tris-glycine buffer (in PAGE apparatus, 7.1.4.4) and run with a protein ladder (Bio-Rad Precision Plus) at 70 V (stacking layer) for 45 minutes and 150 V (resolving layer) for 35 minutes for best resolution.

The gels were stained with Fairbanks A solution (7.1.4.5) and microwaved for 30 seconds and shaken at 120 rpm for 30 minutes. It was de-stained with 10 % acetic acid and microwaved again for 30 seconds and shaken at 120 rpm overnight. At this stage, the protein can be visualized and images recorded using a scanner.

2.2.4 Protein-ligand incubation

The fractions containing highly purified protein (OBody) analysed by SDS-PAGE were pooled and measured using nanodrop. Then the OBody solution was

incubated with P4 (Steraloids # Q2600-000, stored in 100% DMSO) at 1:0.94 molar ratio at 4 ^{o}C overnight. Graphpad molarity calculator web server was used for calculations. This mixture was highly concentrated using a 10000kDa (Millipore) concentrator (after SEC buffer equilibration).

2.3 Protein crystallography

2.3.1 Crystal trials

2.3.1.1 Sitting drop vapor diffusion method

Protein was purified as mentioned in 2.2.2. The concentrated protein-ligand mixture is loaded onto a robot liquid handler (TTPLabtech Mosquito) which mixes (100 nl:100 nl) protein-ligand solution with 96 well microplate screens (PEGRx HT - HR2-086, Crystal Screen HT - HR2-130, Index HT - HR2-134, and SaltRx HT - HR2-136) from Hampton Research. The microplates were sealed with crystal clear air tight covers and incubated at 18°C. The microplate screens were observed under the microscope for possible crystalline protein (microcrystals, needles or crystals).

2.3.1.2 Hanging drop vapor diffusion method

The wells with crystalline protein were fine-tuned by repeating the specific screen (1 µl:1 µl) in a 24 well format (hanging drop vapor diffusion method) by varying two parameters in X and Y axis (pH and concentration of precipitant) in VDX plates (Hampton Research). A 24 well grid describing the volume of individual reagents for the fine screens were designed using Hampton Research online tools. The protein solution was placed 1:1 on 22 mm silicone cover slips and sealed with grease (Glisseal – Borer Chemie) inverted on wells containing 500 µl of the crystallization condition (for vapor diffusion similar to sitting drop vapor diffusion).

2.3.1.3 Other methods

Batch seeding was a technique used for growing larger and high diffraction quality crystals. The existing crystals were diluted in fresh crystallization condition containing the precipitant (1:10 to 1:10000) and fresh purified protein (1 µl) is added to individual dilutions (1 µl) to provide a nucleation site (seed) for growing new crystals.

Streak seeding is a similar seeding technique where fresh protein was added to fresh crystallization condition containing the precipitant (1:1) and allowed to reach equilibrium at 18 °C (3 hours). Then a cats' whisker is dipped in the existing crystal condition and passed through the equilibrated mixture (1 µl) to grow new crystals.

Optimization screens like additive screen (HR2-138) and silver bullet (HR2-096) were other fine screens used in a 96-well format with the fresh crystal condition containing the precipitant (10 µl: 90 µl) and incubated (1:1) with fresh protein using hanging drop vapor diffusion method to improve the size or quality of crystals.

2.3.2 X-ray diffraction

2.3.2.1 Testing of Cryoprotectants

The fresh crystal condition (containing the precipitant) of the crystal grown (from fine screens) was mixed with different percentages of glycerol (10% to 20% v/v) and separately loaded onto a cryo-loop (Hampton Research) and irradiated with X-rays from X-ray diffractometer (Supernova X-ray source, with a gaseous nitrogen stream at 100 K) to obtain minimum ice rings during diffraction spot analysis. The crystal condition with the best cryoprotectant concentration (with minimum ice rings) was used to loop the crystal.

2.3.2.2 X-ray diffraction of crystals

The crystals were immersed in a drop of the best cryoprotectant condition and looped with a cryo-loop. The crystal in the loop was irradiated with X-rays on Supernova X-ray source to rule out salt crystals and screen for protein crystals from diffraction spot analysis. The crystals (or screen conditions) that diffracted as proteins were frozen in liquid nitrogen and sent to Australian Synchrotron (Melbourne, MX1 beam line, ADSC Quantum 210r detector) for X-ray diffraction and data collection.

2.3.2.3 Data collection

The raw diffraction data was collected in IMG and HDF5 formats for MX1 (bending-magnet beamline) and MX2 (finely focused in-vacuum undulator beamline) respectively at the Australian Synchrotron and auto-processed (in HKL

format). Crystal annealing by blocking the nitrogen stream transiently was used to improve diffraction and data collection (Heras & Martin, 2005).

2.3.3 Data Analysis and solving of structure

2.3.3.1 Indexing and integration of diffraction data

The intensities from raw diffraction data were indexed and assigned to the appropriate space group with unit cell dimensions and integrated into a single data (MTZ format) file with iMOSFLM (Battye, Kontogiannis, Johnson, Powell, & Leslie, 2011; Leslie & Powell, 2007). The MX2 beamline data was collected by Eiger detectors. Eiger2cbf program was used to convert diffraction images (HDF5 format) into miniCBF format for further indexing and integration of these images with iMOSFLM.

2.3.3.2 Data processing and reduction (scaling)

Aimless (CCP4) was used to scale the integrated data from iMOSFLM into one set of structure factors and convert auto-processed files (in HKL format) from the Australian Synchrotron into MTZ format (P. Evans, 2006; Philip Evans, 2011; P. R. Evans & Murshudov, 2013). The processed diffraction data from aimless was further analysed using two software suites : CCP-4 (Winn et al., 2011) and Phenix (Adams et al., 2010). The data contains only the amplitudes (and positions) of the diffraction spots whereas the phases have to be computationally derived to solve the protein (crystal) structure (P. Evans & McCoy, 2008). Xtriage (Phenix) was used to analyse the quality of data (sets).

2.3.3.3 Molecular replacement

The data with highest resolution (Å) and completeness (%) was used for molecular replacement (MR) to solve the phase problem (Wlodawer, Minor, Dauter, & Jaskolski, 2008). The aimless files were used by Phaser (Phenix) (McCoy et al., 2007) to find a MR solution with the protein data bank (PDB) model of the fusion protein (MBP, PDB code 1OMP).

2.3.3.4 Density fitting and model building

The MR solution containing the protein model (PDB) were used with aimless files to build the MBP-OBody fusion model using Autobuild (Phenix) (Terwilliger et al., 2008). Pymol was used to place the OBody manually in the partially built

model to complete model building (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC).

2.3.3.5 Model refinement

The MBP- B7 model was refined multiple times using Refine (Phenix) (P. V. Afonine, Grosse-Kunstleve, Adams, & Urzhumtsev, 2013; P. V. Afonine et al., 2012; Pavel V. Afonine, Grosse-Kunstleve, Urzhumtsev, & Adams, 2009; Headd et al., 2012) and Refmac5 (CCP4) (Garib N. Murshudov et al., 2011; G. N. Murshudov, Vagin, & Dodson, 1997). Wincoot was used for manual refinement of the models and ligand fitting (P4) (Emsley & Cowtan, 2004).

2.4 Functional characterisation

2.4.1 ELISA of protein-ligand binding

Enzyme-Linked Immunosorbent Assays (ELISA) are used to quantitatively measure the binding affinity of proteins to their ligands via equilibrium dissociation constants (K_d) derived from colour forming reactions (Friguet, Chaffotte, Djavadi-Ohaniance, & Goldberg, 1985) (Figure 2.1). The dissociation constant (M) is calculated as follows:

$$Kd = \frac{[A][B]}{[AB]}$$

for the equilibrium $AB \leftrightarrow A + B$

where [A] is the concentration of protein, [B] is the concentration of the ligand and [AB] is the concentration of bound protein (with ligand) (de Mol & Fischer, 2008).

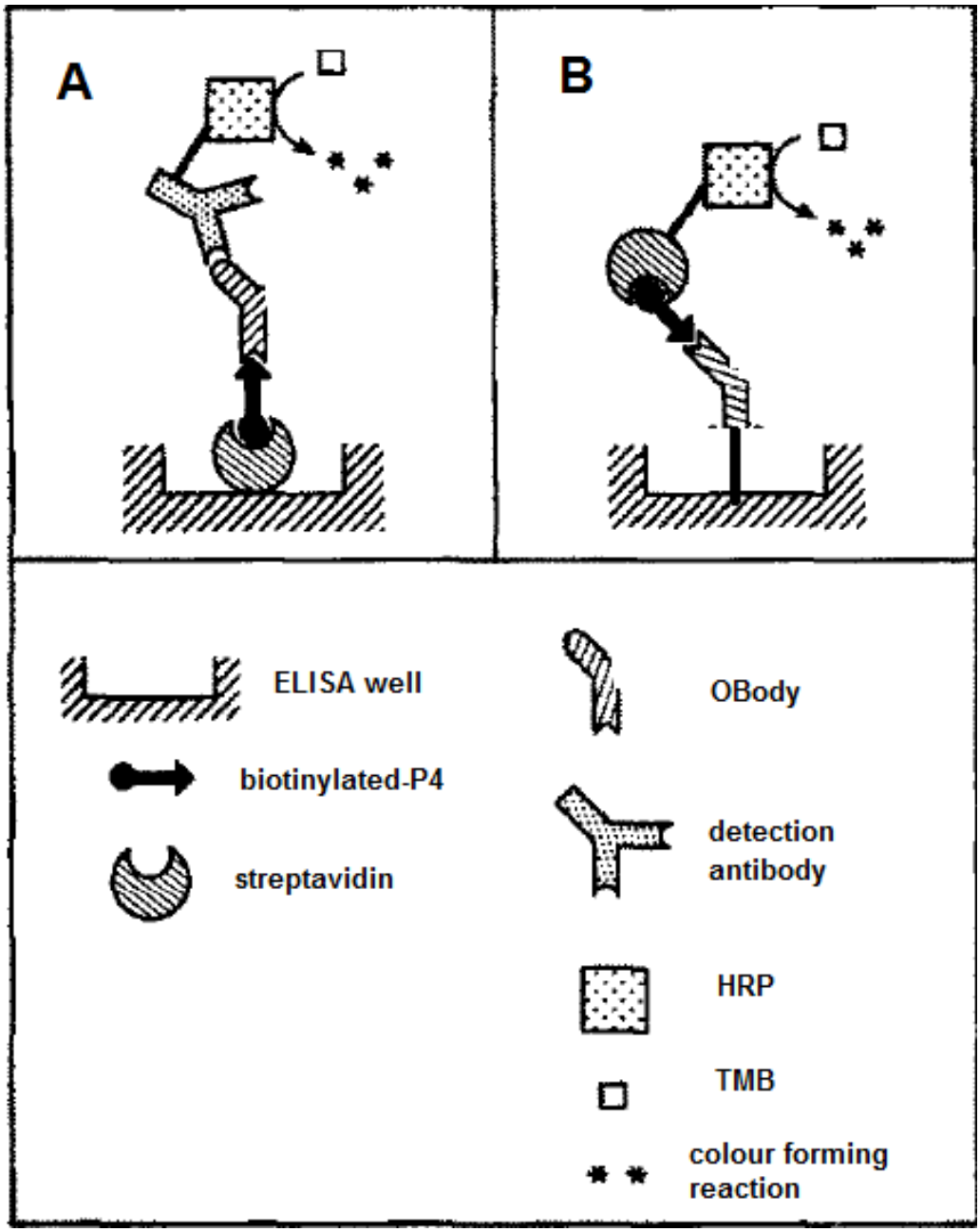


Figure 2.1 Schematic comparing different types of molecule immobilization in ELISA

The antigen immobilized ELISA (A) shows immobilized ligand (biotinylated-P4) in a streptavidin coated plate. The P4 is bound by protein (OBody/antibody) which is in turn detected by a secondary antibody conjugated with HRP. HRP is an enzyme which catalyzes TMB and produces colour indicating presence of P4 binding protein. In antibody (OBody) immobilized ELISA (B), the OBody is coated in a ELISA plate well and biotinylated-P4 is added for detection with a secondary agent (streptavidin-HRP) which uses TMB to produce colour forming reactions. [Adapted with permission from (Bodmer, Tiefenauer, & Andres, 1989)].

The following 3 ELISA systems have been used to determine protein- ligand binding (Bodmer et al., 1989; Friguet, Chaffotte, Djavadi-Ohaniance, & Goldberg, 1995).

2.4.1.1 Antigen (P4/ligand) immobilized ELISA

Progesterone 3-PEG11-biotin (P4, Cayman chemical #9000645) was immobilized at room temperature on a Nunc streptavidin (Wilchek, Bayer, & Livnah, 2006) microplate (300 µl) wells at 7.5 µg/ml (6.6 µM, 50 µl/well) in PBS (in triplicates). The positive and negative control used were biotin-P4 and PBS respectively. This was incubated for 1 hour (at room temperature) and washed with PBS using a plate washer (3x 300 µl). Skim milk (3% w/v) in PBS is used for blocking the wells to prevent non-specific binding of OBody to the plate surface. Reagent volumes (50 µl/well), repeated incubation (for 1 hour) and washing with PBS between steps are standard to these ELISA systems. The protein (Obody) solution with starting concentration of 5 µg/ml (4.4 µM) were added to the first 3 wells and serial dilutions (0.6x) were made (using PBS) in parallel (for triplicates) except for the positive and negative controls. Sigma # P1922 Anti-progesterone rat antibody (Anti-P4 Ab, at 1:1000 ratio with 3% w/v skim milk PBS) was used for detection of P4 in the positive and negative controls. PBS with 0.1% T20 was used for washing at this stage to eliminate non-specific binding. Secondary detection antibodies for the Obody and Anti-P4 Ab were anti-histag Ab (1 µg/well, Invitrogen # 37-2900) and anti-rat IgG Ab-HRP (Sigma # A9037 at 1:4000, HRP- Horse Radish Peroxidase) respectively. Horseradish peroxidase (HRP) is an enzyme which catalyzes the colour forming reaction of 3,3',5,5'-Tetramethylbenzidine (TMB). After incubation and washing, ultra-TMB (Thermoscientific # 34028) was added for detection of binding of OBody and P4. After 15 minutes, the colour forming reaction was stopped with 0.5 M H_2SO_4. The absorbance (Abs_{450}) was measured at 450nm in precision mode using Thermoscientific Multiskan Go microplate spectrophotometer. The data was processed using Graphpad Prism with non-linear regression (Böhm, 2005) using the following equation (one-site total binding model) and illustrated in Figure 2.2.

$$Y = \frac{Bmax * X}{Kd + X} + NS * X + background$$

where B_{max} is the maximum specific binding (Abs_{450}), K_d is the equilibrium dissociation constant and defines the concentration when 50% of protein is bound to ligand (half of B_{max}), Y-axis refers to the absorbance at 450nm (Abs_{450}) from the colour forming reaction which is a function of bound protein (with ligand) and X-axis refers to the variable concentrations of protein [A] or ligand [B]. NS is the

slope of nonspecific binding (y-axis/x-axis) and *background* refers to the amount of nonspecific binding with no added protein (OBody).

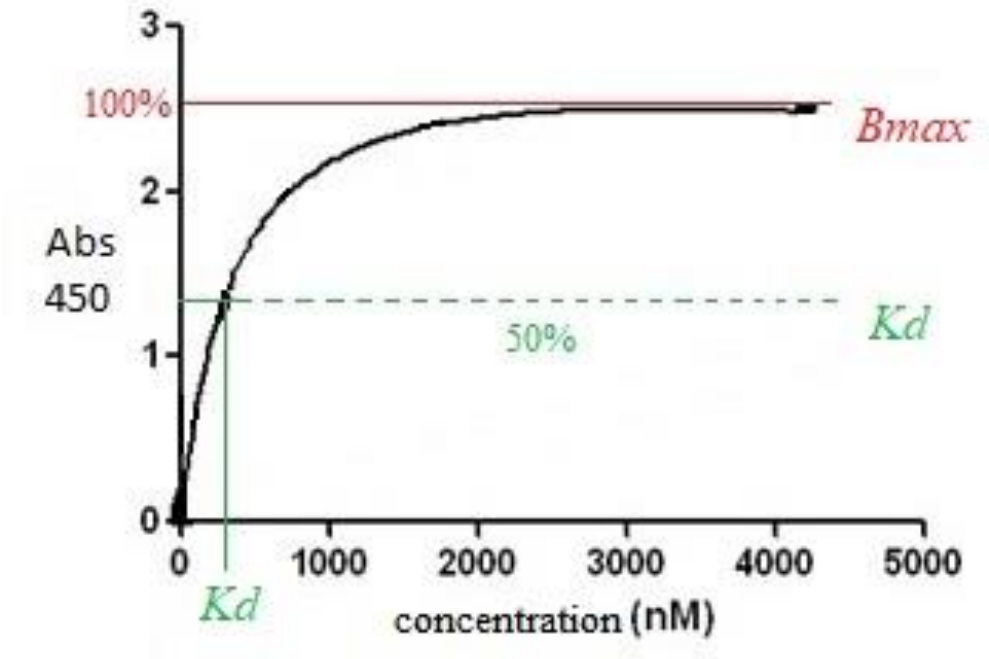


Figure 2.2 Binding curve of protein-ligand binding from ELISA

The closed circles in the graph show different concentrations of OBody (x-axis) added to constant concentration of immobilized ligand (e.g. biotinylated- P4) and the colour forming reaction from ELISA is a function of OBody-P4 binding measured as absorbance at 450nm (y-axis). The maximum absorbance (B_{max}) is a function of maximum OBody-P4 binding and the the OBody concentration (nM) producing absorbance equal to half of (B_{max}) gives the binding affinity (K_d) which is when 50% of free OBody is in equilibrium with P4 bound OBody. The binding of protein-ligand shows a characteristic hyperbole starting with protein-ligand binding at low concentrations (of protein) reflected by steady increase in Abs 450nm and saturation of binding at high protein concentrations.

2.4.1.2 Antibody (Obody / protein) immobilized ELISA

This system is more accurate because small molecule ligands diffuse faster than large protein molecules and because of reduced steric hindrance from immobilization of large protein molecules (Bodmer et al., 1989). The OBody (D7) is immobilized at constant concentration (400nM) in PBS in flat bottom 96 well microplates (Greiner-bio one #655-101). Blocking, incubation and washing steps are same as antigen-immobilized ELISA (2.4.1.1). Biotinylated-P4 is added at decreasing concentrations starting from 44 μM. In this antibody-immobilized ELISA, streptavidin-HRP (Invitrogen #434323) is used to detect OBody-P4

binding. The colour forming reaction was analysed via Abs_{450} values using Graphpad Prism, as in antigen-immobilized ELISA.

2.4.1.3 Competition ELISA

This system is highly accurate because antigen and antibody are allowed to interact freely without immobilization of a partner or partial denaturation of protein, thus reaching equilibrium. The resulting free protein is used for detection and gives a true picture of the binding between antigen and antibody (Hardy, Djavadi-Ohaniance, & Goldberg, 1997). Biotinylated-P4 was immobilized on a Nunc streptavidin plate. Another non-streptavidin 96 well microplate (Greiner-bio one #655-101) contains non-biotinylated P4 (Steraloids # Q2600-000) at different concentrations with constant concentration of OBody (D7). The OBody or positive control antibody (anti-P4 antibody, Ab) is incubated to allow binding at equilibrium between protein and ligand. This antigen-antibody (or OBody) mixture was added to the P4 (biotinylated) immobilized Nunc streptavidin plate. The unbound Obody (or anti-P4 Ab) binds to biotinylated P4 and gives an inverse binding curve. Incubation, blocking, washing, colour forming reaction and processing of Abs_{450} using non-linear regression is same as for antigen-immobilized ELISA.

2.4.2 SPR for protein-ligand binding

Surface Plasmon Resonance (SPR) is a spectroscopic technique where binding of an analyte (protein) to an immobilized ligand on a sensor chip changes the refractive index and this is the best method to determine K_d (equilibrium) and kinetic parameters (Nieba, Krebber, & Pluckthun, 1996). The change in refractive index is measured in resonance units (1000 RU = 1 ng/mm^2) (Fivash, Towler, & Fisher, 1998). The experiments are conducted with SA chip (GE Healthcare, carboxymethylated dextran pre-immobilized with streptavidin) in Biacore 3000. Two ligands are used to test for cross-specificity. Biotinylated -P4 is used to saturate the binding surface of the test flow cell and biotinylated-E2 (estradiol) is used to saturate the binding surface of the control flow cell. The OBody (in HBS-EP buffer, GE Healthcare) was flown over these flow cells at different concentrations and the maximum RU attained (after subtraction of E2 control) from saturation with protein is taken as the B_{max} (Figure 2.3).

The equilibrium analysis is done using a non-linear regression similar to ELISA (also called langmuir binding isotherm) using BIAevaluation software (ver 4.1, Biacore AB). The change in refractive index at equilibrium (R_{eq}) is given by the following equation (in RU):

$$Req = \frac{Bmax * [A]}{Kd + [A]}$$

where [A] is the concentration of the (analyte) protein, K_d is the equilibrium dissociation constant (µM) and B_{max} is the maximum binding capacity (in RU), when all binding sites on the sensor surface are occupied (de Mol & Fischer, 2008).

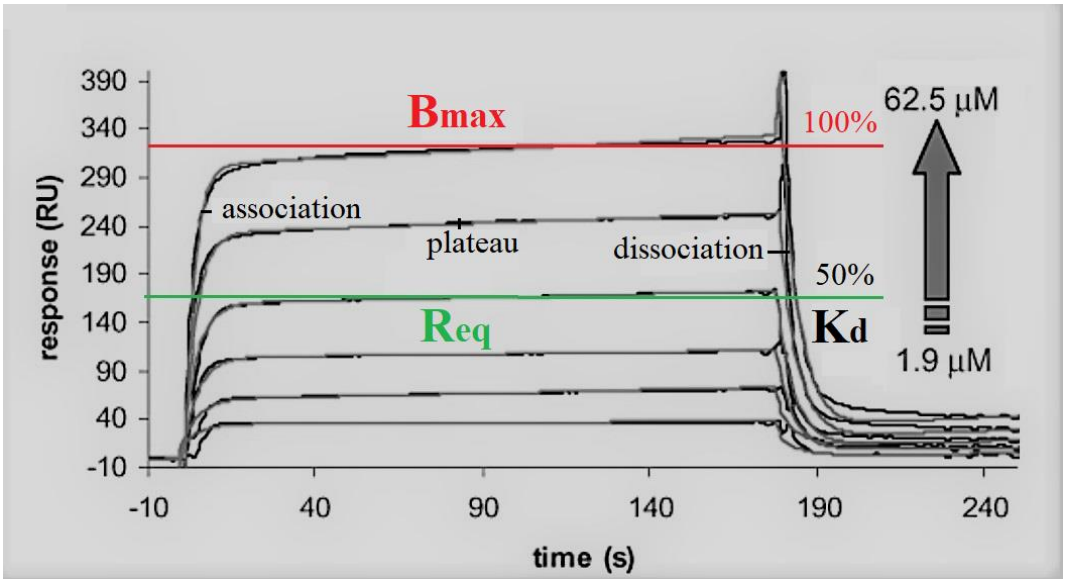


Figure 2.3 SPR sensorgram from protein-ligand binding

Different concentrations of analyte (right arrow, µM) flown over immobilized ligand (biotinylated-P4 on SA chip) creates a characteristic curve showing rapid association, plateau followed by rapid dissociation with time (x-axis). This is reflected by the change in refractive index (RU, y-axis). At the plateau, the rate of association is equal to the rate of dissociation (R_{eq}). The highest RU reflects maximum protein-ligand binding (B_{max}). The analyte (OBody) concentration (nM) producing (R_{eq}) equal to half (B_{max}) gives the binding affinity (K_d) similar to ELISA. [adapted with permission from (Munoz et al., 2010)].

2.4.3 FTSA for thermo-stability of OBody

Fluorescent Thermal Shift Assays (FTSA) or commonly known as Differential Scanning Fluorimetry (DSF) is a technique used by the pharmaceutical industry to test protein- ligand binding. It works on the principle that the fluorescent dye has a non-specific affinity to hydrophobic parts of a protein and fluoresces upon binding. This fluorescence is decreased by water in the protein buffer. Therefore,

protein unfolding due to temperature and binding of the dye to hydrophobic parts of the protein leads to fluorescence (Figure 2.4) and gives a good idea regarding the thermostability of the protein. Since ligand binding generally stabilizes the protein, this can be used as an indication of protein-ligand interaction (Boivin, Kozak, & Meijers, 2013).

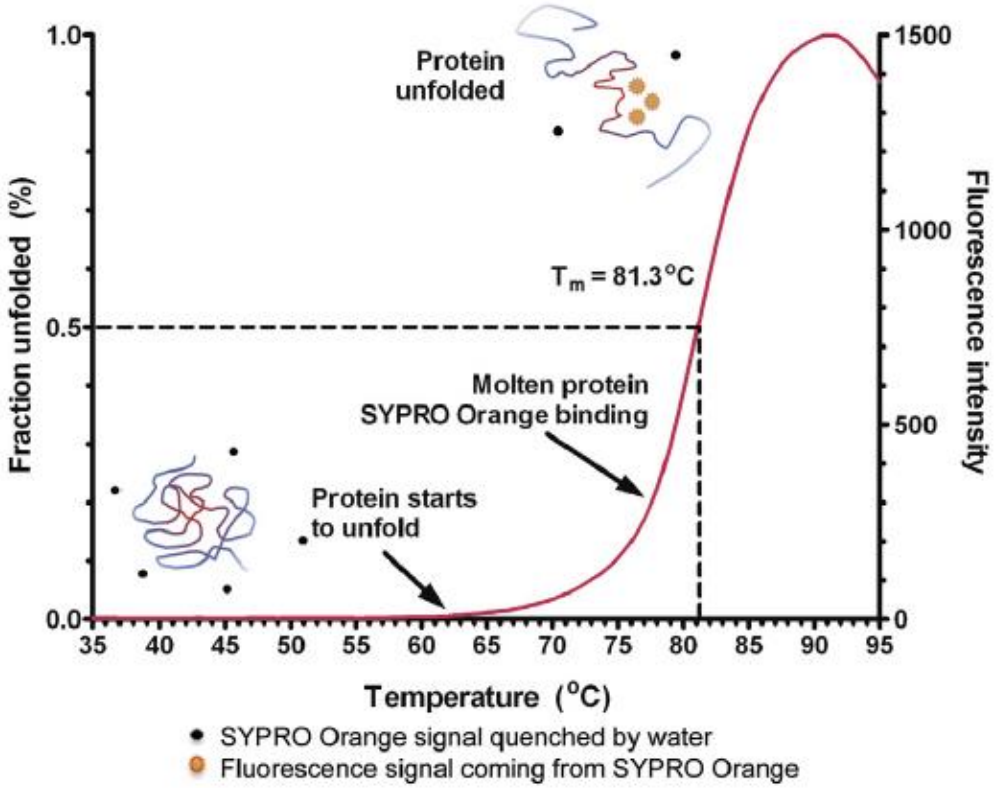


Figure 2.4 Graph of Fluorescent Thermal Shift Assay (FTSA) melt curve

The folded protein unfolds with increases in temperature (x-axis) exposing the hydrophobic core to SYPRO orange leading to increase in fluorescence (y-axis, right). The increase in fluorescence follows a sigmoidal curve and is a function of the fraction of protein unfolding (y-axis, left) and the temperature at which 50% of the protein in unfolded is called melting temperature (T_m). [used with permission from (Boivin et al., 2013)].

The melting temperature (T_m) is used to measure the thermostability and is the temperature at which 50% of protein is unfolded (Niesen, Berglund, & Vedadi, 2007). The OBody (D7 and mutants) were subjected to Thermal shift assays using SYPRO-orange (Invitrogen) dye in Rotor-gene Q-PCR (Corbett Life Science). This dye excites maximally at 470 nm and emits at 569 nm. The following recipe was used:

5x Sypro Orange dye (final concentration from stock 5000x)

3 µg OBody (in PBS pH 7)

In total 20 µl UP H_2O

(Gain setting- 4)

The data is processed as the first derivative of the melting curve (δF/δT) (Bohling, Wittwer, King, & Elenitoba-Johnson, 1999) which is equal to change in fluorescence (δF) divided by the change in temperature (δT), plotted against the temperature (T); and the temperature of maximum (δF/δT) is the T_m (Gandham, Volk, & Gorenstein, 2015).

2.5 DNA manipulation and analysis

2.5.1 Cloning of Obody DNA sequences into *E. coli*

The gene for the Obody (D7) was digested from the phagemid (pOB2) using NEB restriction enzymes NcoI, NotI or SacII and transferred into expression plasmids (pET28b, pPROEX HTb or pOPINM, 7.1.8) using custom primers (7.1.9) ordered from Integrated DNA Technologies (IDT, Singapore). The signal sequence and foreign fragment (YscC, 7.1) was not included in these cytoplasmic expression plasmids. The restriction enzyme digestion is explained in detail in 2.5.6.

The insert (donor gene) and vector (recipient plasmid) which were digested using the same restriction enzymes was ligated using NEB T4 DNA Ligase (#M0202) in a vector to insert ratio of 1:3 and is explained in detail in 2.5.7. The ligated plasmid was transformed into electro-competent (EC) *E. coli* cells using electroporation or chemical transformation and is explained in detail in 2.5.8.

2.5.2 Confirmation and storage of clones

The resulting colonies on the plate were tested for successful cloning of donor insert into recipient plasmid by colony Polymerase Chain Reaction (PCR) where individual colonies were tested for successful ligation of transformed plasmid using flanking primers and is explained in detail in 2.5.4.1. The successful transformed colonies were grown in 5 ml LB cultures (with appropriate antibiotic) at 37°C overnight at 200 rpm in a shaking incubator.

The *E. coli* cultures (grown overnight) were harvested for storage and plasmid purification. A part of the culture is stored (under sterile conditions) with 25% glycerol (v/v) at -80°C. The remaining culture was used for plasmid extraction with a commercial kit (QIAprep spin miniprep kit # 27106). The manufacturer's protocol was used and the purified plasmid concentration measured by UV

spectroscopy (Thermo-scientific nanodrop). The plasmid with an appropriate primer was sent (to Massey Genome Sequencing) for confirmation of gene sequence (at recommended concentrations). The sequencing results were analysed using Geneious R8 (Biomatters) software.

2.5.3 Primer design

In silico design of PCR primers were done using Geneious (Biomatters) sequencing analysis software and ordered online using Integrated DNA Technologies (IDT, Singapore). Where applicable, primers were designed with restriction enzyme cut sites which matched the sites in plasmids used for cloning (as discussed in section 2.5.6). All the primers were delivered in lyophilized format which had to be reconstituted to 100 µM using TE buffer as per provided instructions. The 100 µM primers were diluted using UP H_2O to 10 µM and used for PCR reactions.

Primers for phage display libraries: The primers used in construction of Affinity Maturation (AM) library used NNK codons where N= G, A, T or C and K= G or T. This allows elimination of stop codons UAA and UGA, but UAG also called 'amber' will be suppressed and replaced with glutamine in *E. coli* TG1 strain (Inokuchi *et al.* 1979).

2.5.4 Polymerase Chain Reaction (PCR)

PCR was used to magnify DNA fragments of interest using custom built primers which flanked both ends. Most of the PCR reactions were typically performed in 15 µl reactions (final volume) for analysis of insert into vector backbone during sub-cloning or 50 µl reactions for amplification of DNA fragments for cloning or construction of phage library. Two polymerases were used depending on routine PCR reactions (*Taq* polymerase) or PCR with proofreading (*Pfx* Polymerase) for high fidelity DNA amplification for designing libraries.

1) *Taq* polymerase (Invitrogen #10342) was used for most PCR reactions with the same protocol described in 2.5.4.1. scaled up 3.33x times for a 50 µl reaction.

2) Invitrogen Platinum *Pfx* Polymerase (#11708) was used for site directed mutagenesis (to randomize DNA fragments for construction of Affinity maturation phage library) using primers with NNK codons. This enzyme is high

fidelity and has proofreading 3’ to 5’ exonuclease activity. The following mixture was used:

5 µl 10x amp buffer

1.5 µl 10 mM dNTP

1 µl 50 mM MgSO4

0.4 µl *pfx* polymerase

1.5 µl 10 µM primers (forward and reverse)

1 µl 50 ng DNA template

Up to 50 µl UP H_2O

This mixture is prepared on ice and run on a PCR machine with the following settings:

1) 94 °C 3 minutes
2) 94 °C 15 seconds
3) 50 – 70 °C, 30 seconds (temperature varied with DNA fragments)
4) 68 °C, 30 seconds (go to 2, x30 times)

2.5.4.1 Colony PCR

The individual colonies on a plate are stabbed with sterile pipette tips and incubated in separate 10 µl LB (with appropriate antibiotic) for 10 minutes at room temperature. This is used in PCR reaction mixtures, where every reaction mixture contains:

1.5 µl 10x –MgCl2 (Taq) buffer

0.45 µl 50 mM MgCl2

0.3 µl 10 mM dNTP

0.75 µl 10 µM primers (forward and reverse)

0.15 µl Taq polymerase

1 µl LB (containing the colony stab)

Up to 15 µl UP H2O

This mixture is prepared on ice and run on a PCR machine with the following settings:

1) 95 °C 2 minutes
2) 95 °C 20 seconds
3) 55 °C 30 seconds
4) 72 °C 45 seconds (go to 2, x29 times)
5) 72 °C 5 minutes

The PCR products are analysed for the appropriate size of the inserts on a 1% DNA gel (section 2.5.5). The successful colonies (~9 µl of LB from colony stabs) are separately grown in 5 ml LB with 1/1000 final concentration recommended antibiotic at 37°C overnight.

2.5.4.2 Overlap extension PCR (OE-PCR)

Overlap extension PCR (Heckman & Pease, 2007) was used in gene modification, where two DNA fragments sharing identical terminal segments (from complementary primers) were fused with a second PCR reaction using only flanking primers (Figure 2.5).

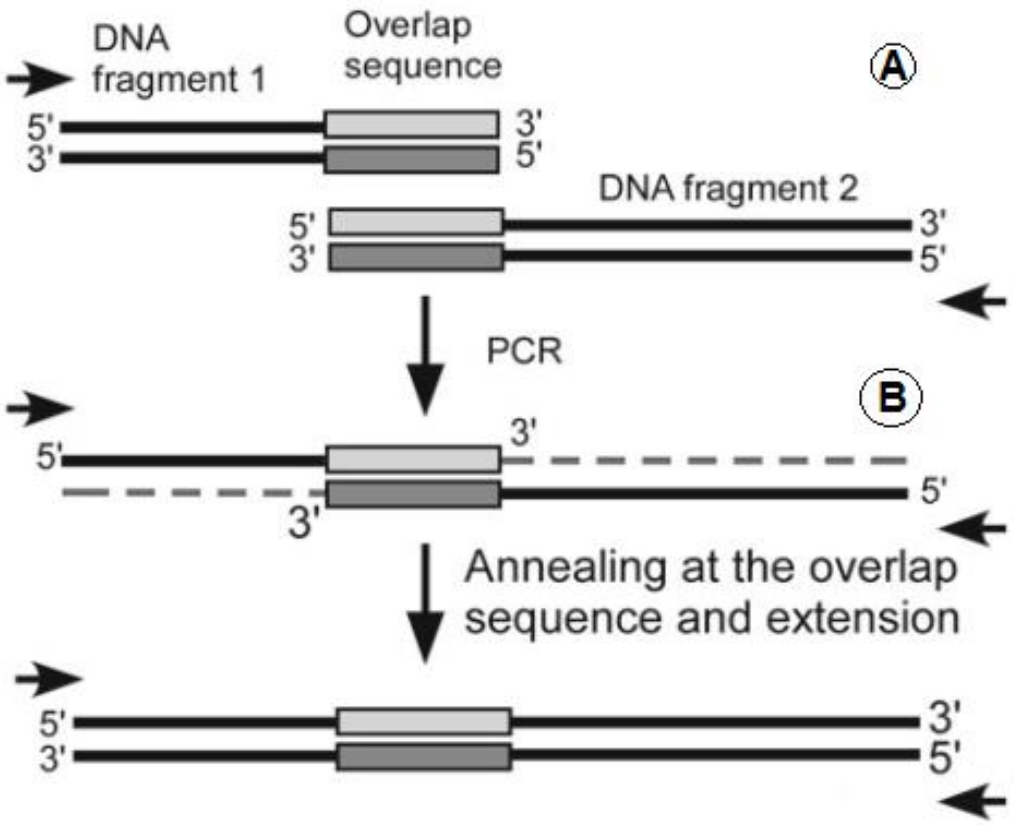


Figure 2.5 Illustration of Overlap-Extension PCR (OE-PCR)

The OE-PCR is a two-step PCR where the first-step (A, overlap) involves independent PCR to generate two fragments (DNA fragment- 1, 2 to be fused) to have overlapping ends. This is achieved by complementary (overlap) primers (at the fusion ends, in grey) which extend their nucleotide sequence into the overlapping fragment. In this illustration, DNA fragment 1 is generated by using a 5'- flanking primer (left arrow) and overlap primer (dark grey). Similarly, DNA fragment 2 is generated by 5'- flanking primer (right arrow) and complementary overlap primer (light grey). The second-step (B, extension) involves mixing these two generated fragments and generating single hybrid fragment using only flanking primers (arrows). This illustration shows annealing at complementary (fusion, in grey) ends during PCR leading to extension and generation of the hybrid fragment. [used with permission from (Cha-aim, Hoshida, Fukunaga, & Akada, 2012)].

2.5.5 DNA Agarose gel electrophoresis

The agarose gel is prepared microwaving intermittently for 1 minute until solution clears. The solution (7.1.5.1) is then poured into a DNA gel caster with combs and allowed to set for 30 minutes. The PCR product (or DNA fragment) to be visualized is mixed separately with 10x DNA loading dye (in 10:1 ratio, 7.1.5.3) and then added to the wells in a solidified agarose gel in a DNA electrophoresis system. TAE running buffer (7.1.5.2) is used and the gel is run at 90V for 30 minutes. A DNA ladder (Thermo-scientific 1Kb Plus) is run in parallel as a standard. DNA can then be visualized under blue light (due to SYBR-safe DNA gel stain). The gel with the DNA fragments of interest are excised under blue light (Invitrogen Safe Imager, wearing orange filter glasses) and DNA extracted using commercial kit (QIAquick Gel Extraction Kit # 28706) as per manufacturer's protocol.

2.5.6 Restriction enzyme digestion of DNA fragments

The purified DNA fragments which were PCR amplified were restriction enzyme digested with the appropriate enzyme dictated by their gene sequence. This was done using New England Biolabs (NEB) restriction enzymes for 37°C for 2- 5 hours with the following protocol:

10 µl 10x cutsmart NEB buffer
1 µl NEB enzyme (s)
40 µl purified (gel extracted) DNA fragment
Up to 100 µl UP H_2O

Double digestion with two appropriate enzymes or sequential digestion with one enzyme were done to obtain DNA fragments for sub-cloning into expression plasmids or gene modification for phage display. DNA was run on 1% agarose gel and required DNA fragment extracted and purified as mentioned in 2.5.5.

2.5.7 Ligation of DNA fragments

The insert and vector were ligated using the following reagents (for a 1:3 vector to insert ratio) and incubated at room temperature for 2 hours (or 18°C overnight). The reagents are:

30 fmol of digested vector
90 fmol of digested insert
1 µl NEB T4 DNA ligase (#M0202)
4 µl 5x ligase buffer
Up to 20 µl UP H_2O

The calculations for the quantity of insert and vector are as follows (bp = number of base pairs):
Insert (ng)= 90 x bp x 600/ 10^6
Vector (ng) = 30 x bp x 600/ 10^6

2.5.8 Transformation of ligated plasmid

The resulting ligation mixture was transformed into electro-competent (EC) or chemo-competent *E. Coli* DH5α/BL21/TG1 strain cells using following protocols.

2.5.8.1 Preparation of electro-competent (EC) E. coli cells

The protocol for preparing EC *E. coli* cells for transforming plasmids (molecular cloning or construction of phage library) is as follows (all steps are done at 0 °C on ice unless mentioned otherwise):

- Streak a LB plate from *E. coli* glycerol stock (- 80 °C, without plasmid) and grow 2 (500 ml) large LB cultures at 37°C.
- At OD_{600} between 0.5 to 0.7, harvest cells by centrifugation at 3000g for 15 minutes at 4°C.
- The supernatant is discarded, and pellet resuspended in (1 L) 0.1 M hepes (pH 7.5) and centrifugation repeated.
- The re-suspension of the cells (after discarding supernatant) and centrifugation are repeated with lower volumes successively.
  500 ml 0.1 M hepes
  20 ml 0.1 M hepes (with 10% glycerol)
  2.5 ml 10 % glycerol
- The final suspension of EC *E. coli* cells (in 2.5ml 10% glycerol) is freshly used at required amounts for transformation of ligated phagemids (see 2.6.1.3. for phage library construction) or made into (50 µl) aliquots and snap frozen in liquid nitrogen and stored at - 80 °C (for molecular cloning purposes).

2.5.8.2 Transformation of competent E. coli cells

Electroporation:

The EC *E. coli* cells stored at -80°C (as 50 µl aliquots) is thawed on ice for five minutes and 5 µl of ligation mixture added to it and then this mixture is transferred to a sterile electroporation cuvette (chilled on ice). The cuvette is transferred to Gene pulsar (Bio-Rad) electroporator and the EC *E. coli* cells shocked with the following settings:

Resistance= 200 Ω

Capacitance= 25 µFD

Voltage= 2.5 kV

The competent cells are recovered with 1ml SOC or LB and incubated at 37°C for 1 hour.

Chemical transformation:

Chemically competent cells were made using a similar protocol as 2.5.8.1. but resuspended with TSB media (7.1.1.5). Chemically-competent cells stored at - 80 °C as 100 µl aliquots were thawed on ice for 5 minutes and 5 µl of ligation mixture added and incubated for 30 minutes. At the end of 30 minutes, the competent cells were subjected to 42 °C for 45 seconds and then recovered for 2 minutes on ice. LB or SOC (1ml) is added for further recovery and the competent cells are incubated at 37°C for 1 hour.

2.5.9 Bioinformatic tools

DNA sequencing analysis, sequence alignments and modifications (and *in silico* protein translations) were done with Geneious (Biomatters) sequencing analysis software. The web-based BLAST (NCBI) and SWISS-MODEL server were used for *in silico* protein analysis and modelling respectively. Graphpad PRISM (ver. 6) was used for statistical analysis of ELISA data. BIAevaluation software was used for analysis of Surface plasmon resonance (SPR) data. The FTSA data and phage ELISA graphs were processed using Microsoft Excel.

2.6 Phage library and selection

This section refers to methods specific to construction of a phage library and phage selections (Chapter 4).

2.6.1 Construction of Affinity Maturation (AM) phage library

2.6.1.1 DNA Template and stop codon insertion

To design an AM library, steps were taken to prevent re-selection of the original (D7) OBody. This was done by inserting stop codons to DNA sequences of D7 mutants, C35S and double mutant (DM, C35S:C48S) which were used as templates (2.6.1.2). The expression strain for phage display is *E. coli* TG1. It does not use the stop codon UAG (amber) which is replaced with glutamine (Inokuchi *et al.* 1979). Taking this into consideration, a stop codon (TAA) was inserted into the DNA sequence of D7 mutants (positions 48 and 83 of C35S and DM respectively) to prevent re-selection of these mutants (chapter 5). Re-selection of these D7 mutants would result in pre-mature termination of expression and elimination of phage containing them. The stop codon was inserted via OE-PCR using complementary primers (containing TAA) which is explained in detail in 2.5.4.2.

2.6.1.2 Randomization of inserts

The inserts (~5 μg) were generated with OE-PCR using D7 mutants (with stop codons) and another set of complementary primers containing NNK codons where N=G, A, T or C and K= G or T. This arrangement allows elimination of stop codons UAA and UGA (U and T are interchangeable) in an amber (UAG) suppressed *E. coli* TG1 leading to possibility of all 20 amino acids (Mena & Daugherty, 2005) in the randomized positions as shown for L4_DM library (Figure 2.6).

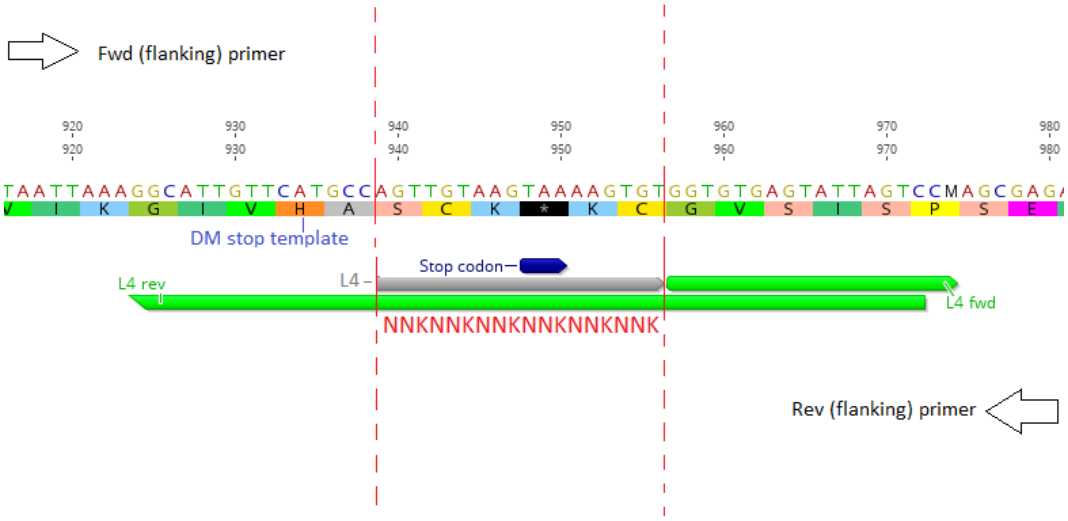


Figure 2.6 Sequence of randomisation scheme for loop (L4) on DM (stop) template

The sequence of loop (L4) is shown (inside dashed red lines) as part of the bigger OBody (DM) sequence. The complementary primers (green, for OE-PCR) show randomized areas (grey, NNK for 6 amino acids). The overlapping fragments are generated separately with forward primer (left arrow, sequence not shown) and L4 reverse (green) for one fragment, and reverse primer (right arrow, sequence not shown) and L4 forward (green) for the second fragment. The extension step (OE-PCR) follows using only flanking primers (arrows, sequence not shown) to generate randomized (single, L4) DM fragments. The stop codon in DM template is shown as TAA (*).

2.6.1.3 Ligation and transformation of library members

The vector (phagemid, pOB2) backbone from D7 was generated in large amounts (~25 µg) and the D7 template removed using standard DNA restriction digestion with NEB enzymes SacII (#R0157) and NotIHF (#R3189) as described in 2.5.6. The randomized inserts from OE-PCR were digested using the same enzymes. The vector (~12.5 µg) and inserts (~2.5 µg) were ligated as described in 2.5.7. The ligation mixture was made in 100 µl aliquots in a 96 well microplate format and incubated overnight at 16°C. This was purified using Roche High pure PCR product purification kit (#11 732 676 001).

The purified phagemids (~12 µg) were transformed into freshly made EC *E. coli* TG1 cells (as described in 2.5.8) using a standard protocol for making EC *E. coli* TG1 cells in a cold room. The ratio used for electroporation was 3 µg plasmid (~50 µl) in 300 µl EC *E. coli* per electroporation cuvette.

Two controls, vector backbone only and pUC vector (Chan, Dreolini, Flintoff, Lloyd, & Mattenley, 2002) were transformed as negative and positive controls respectively. The transformed library cells were recovered by plating dilutions on small (for counting) and large (for harvesting) 2xYT (with carbenicillin, C100) plates grown overnight at 30°C. The transformation efficiency of the AM library (~10^7 cfu/µg) and the two controls were calculated using the formula

$$transformation\ efficiency\ (in\ cfu/\mu g) = colonies/\mu g\ DNA/dilution\ factor$$

The transformed library cells were scraped from large plates with 2xYT (with 17% glycerol) and stored in aliquots at -80°C.

2.6.1.4 Phage rescue of library members

The transformed (phage) library cells equivalent to OD_{600}= 0.1 was grown in 2xYT (with C100 and 1% glucose) to an OD_{600}= 0.5 and super-infected with helper phage (KM13) using the formula

$$total\ amount\ of\ helper\ phage = OD600 * (5\ x\ 10^8) * volume\ of\ culture * MOI$$

where MOI is Multiplicity of Infection. It is the amount of phage that infects one *E. coli* cell (KM13 = 10). The amount of phage required to infect 1 ml of *E. coli* TG1 at OD_{600} = 1 is experimentally determined to be 5 x 10^9 cfu for KM13.
The super-infected culture was incubated at 37 °C for an hour with gentle shaking (150 rpm). The phagemid (pOB2) contains an ampicillin resistance gene (AmpR) whereas KM13 has a kanamycin resistance gene (KanR). The culture was centrifuged at 3300g for 10 mins and supernatant removed and the resulting pellet re-suspended with 2xYT (with C100 and kanamycin, K50) and incubated at 25°C overnight (200 rpm) for 18 hours, to facilitate survival and reproduction of superinfected *E. Coli* TG1 containing both antibiotic markers and production of OBody displaying phages.

2.6.1.5 PEG precipitation of library members

The protocol for PEG precipitation of rescued phage (library) was as follows:

- The culture was centrifuged at 3300g for 30 minutes (4 °C) to recover supernatant (containing phage). The supernatant was incubated with 100

ml PEG/NaCl (20% Polyethylene glycol 6000, 2.5 M NaCl) for an hour on ice (in sterile conditions). All steps were done on ice unless mentioned elsewhere

- The PEG/NaCl solution precipitates phage. This mixture was re-centrifuged and supernatant (containing PEG/NaCl) was discarded this time. The PEG precipitated phage was resuspended with 8 ml PBS with 2 ml PEG/NaCl for 20 minutes and re-centrifuged again.
- The supernatant was discarded and phage pellet re-suspended with 5 ml PBS. This was re-centrifuged (at max G to remove cell debris) and the supernatant (containing phage) was filtered using 0.45 µl filter and stored in aliquots (with 15% glycerol) at -80°C.
- Phage titre determination was done by diluting PEG precipitated phage 10^6 times in 2xYT and infecting 1ml freshly grown *E. coli* TG1 (at OD_{600} =0.5, 37°C, 1 hour, 150rpm as in 2.6.1.4.). A fraction of the infected *E. coli* TG1 cells were diluted and plated (C100, 2% glucose) at 30°C overnight and colonies counted. The formula for phage titre determination is:

$$phage\ titer\ (in\ cfu/ml) = colonies * dilution\ factor * 10$$

where 10 is the dilution related to volume of infected *E. coli* TG1 cells (0.1ml) used for phage titre determination (AM library = 10^{13} cfu/ml).

2.6.2 Phage selection of Affinity maturation library

2.6.2.1 Blocking phage and beads

A calculated volume of PEG precipitated phage (~10^{12}) was blocked with equal volume 3% (w/v) skim milk in PBS (MPBS) to prevent non-specific adhesion and to ensure phage-ligand binding (during phage selection) is solely through the OBody displayed on the surface of minor coat protein III (pIII) of phage. A manufacturer recommended amount (~50 µl) of streptavidin/neutravidin beads (Themoscientific Dynabeads M280 #11206D or sera-mag speed beads) to be used for binding biotinylated-P4 (bio-P4) was blocked separately with 1ml 3 % MPBS and both were incubated on a rotating platform for an hour (at room temperature). A magnetic strip was used to isolate the blocked beads from the supernatant. Note

that streptavidin/neutravidin binds to biotinylated-P4 which binds to specific OBody displaying phage.

2.6.2.2 Phage de-selection and selection

A small volume (~50 µl) of blocked streptavidin/neutravidin beads were incubated with PEG precipitated phage to encourage non-specific binding of phage to the beads and these beads were discarded (with non-specific binding phage) using the magnetic strip.

The OBody ligand (bio-P4, with starting concentrations of 1 µM) was added in blocked phage and incubated for 2 hours (at room temperature) to select for strong P4 binding OBodies. The (OBody displaying) phage binding to bio-P4 was removed using blocked streptavidin/neutravidin beads and washed with PBS (0.1 % Tween 20) 10 times using the magnetic strip (to remove non-specific phage).

2.6.2.3 Elution of selected phage and infection of E. coli TG1

The beads were separated from the bound phage (with trypsin cleavable site) using trypsin (1 mg/ml) and vigorously shaken at 37°C for 30 minutes. The selected phage (~500 µl, after discarding separated beads using the magnetic strip) were used to infect freshly made *E. Coli* TG1 cells (~4.5 ml) as mentioned in 2.6.1.4. The resulting infected cells were designated as the *output*. The PEG precipitated phage before selection was also separately used to infect *E. Coli* TG1 cells as in 2.6.1.5. These unselected infected cells were designated as the *input*. These *E. coli* TG1 cells were plated on small (for counting) and large (for harvesting) 2xYT (C100, 2% glucose) agar plates as described in 2.6.1.3. The total output was calculated using the formula:

$$Total\ Output\ (in\ cfu)\ =\ colonies\ x\ dilution\ factor\ x\ 50$$

where 50 is the dilution related to volume of infected *E. coli* TG1 cells (0.1ml) used for output titre determination same as described in 2.6.1.5. These *E. coli* TG1 cells were used as inputs for subsequent rounds of selections or stored in -80°C.

2.6.2.4 Subsequent rounds of phage selections

The *E. coli* TG1 (*output*) cells was used as *input* for subsequent phage selections. The phage was rescued, and PEG precipitated as described in 2.6.1.4 and 2.6.1.5. respectively. Successive selections used lower concentrations of ligand, shorter

incubation times and stringent washing conditions (up to 20 times) to select for high affinity OBody (phage). After 5 to 7 rounds of selection, the last two rounds were used to isolate P4 binding OBodies (by colony picking from small output plates) and testing their binding using phage ELISA (with immobilized bio- P4).

2.6.3 Phage ELISA

2.6.3.1 Phage rescue of selected phage

The colonies from (small) output plates from last two rounds of phage selections were individually picked with sterile pipette tips and incubated independently in 96 well plates containing 2xYT (C100 and 2% glucose) at 37°C, 150 rpm overnight. A plate replicator (in 70% ethanol and flame sterilized) was used to inoculate these micro-cultures into 96 deep well plates (containing 2xYT, C100 and 2% glucose) ensuring no cross-contamination. The micro-cultures are stored at -80°C (in 17% glycerol) and the deep well plates are incubated at 37°C, 200 rpm until solution turns turbid (~5 hours). The turbidity indicates OD_{600}~ 0.5 followed by super-infection (37°C, 150 rpm, 1 hour) with KM13 as described in 2.6.1.4 (~10^{10} cfu/ml). Subsequent centrifugation (3300g), re-suspension (2xYT, C100, K50) and incubation (25°C, overnight) was as described in 2.6.1.4.

2.6.3.2 Phage ELISA of selected phage

Biotinylated-P4 was immobilized in a 96 well streptavidin coated microplate and blocked with 3% MPBS as described in 2.4.1.1. The deep well plates (from overnight 25°C incubation) are centrifuged at 3300g (10 minutes) and the supernatant (~50 µl, containing phage with OBody) is used to detect immobilized P4 on the streptavidin microplate. The independently bound phage were detected by the antibody (anti-M13 Ab-HRP, at 1:4000 ratio in 3% MPBS, Sigma Aldrich #27-9421-01) which detects a coat protein (of M13 phage). The HRP was used for the colour forming reaction as described in 2.4.1.1 (Figure 2.7)

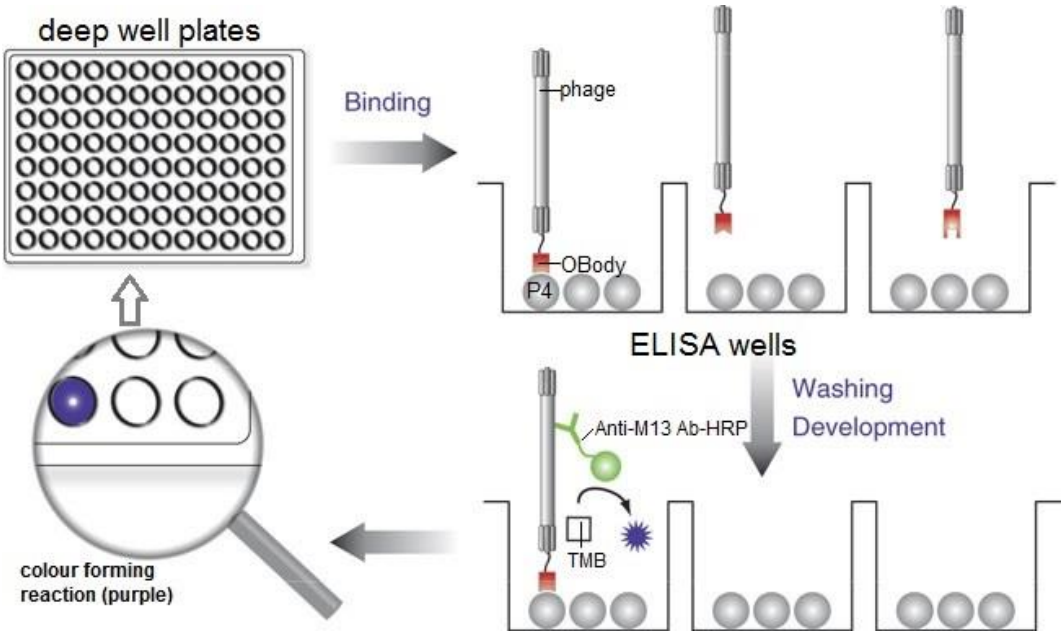


Figure 2.7 Cartoon of Phage ELISA system

The individual phage clones (with OBody) are grown in deep (96) well plates (upper left). These phage were independently added to an ELISA (96-well) plate with immobilized P4 (grey circles, upper right). The selected OBody binds to P4 and unbound phage are washed away. The bound phage is detected by anti-M13 Ab-HRP (green) which produces a colour forming reaction (violet) with TMB (lower right). The well containing a P4 binding OBody is animated (lower left) and the corresponding deep well (upper left, or glycerol stock) is used for large scale expression and characterisation of the specific OBody clones. [adapted with permission from (Lee et al., 2007)].

The phage (with OBody) with Abs_{450} > 0.25 were sub-cloned into expression vector pPROEX-HTb as described in 2.5.1 and confirmed for OBody-ligand (P4) binding (2.4.1.1). The confirmed OBodies were expressed in large scale (2.2) and binding affinity determined using SPR (2.4.2).

3 Characterisation of anti-progesterone OBody (D7)

3.1 Introduction

An OBody designated D7 binds to the small molecule progesterone (P4) and it was previously generated from a naïve library using phage display. An OBody that can detect P4 has many applications in the dairy industry. D7 is an OBody from the second generation *Pyrobaculum aerophilum* Aspartyl t-RNA synthetase (2GPATS) library selected against the small molecule target P4. D7 is soluble and expressed in large scale using pET28b in *E. coli* BL21 strain. A simple ELISA using immobilized (biotinylated) P4 confirmed D7 binding to progesterone. Biophysical characterization of D7 provides key insights into the structure and function of an anti-progesterone OBody, with the further goal of improving binding affinity. X-ray crystallography is one such technique which can provide a detailed understanding of the D7 structure and its specific interactions with P4 bound. An analysis similar to that reported for the OBody bound to HEL from the first generation naïve library (1GPATS) (Steemson et al., 2014).

3.1.1 Structure of Progesterone (P4)

P4 is a steroid, is secreted by the corpus luteum and placenta of animals, and acts on the female reproductive system. It is secreted in milk and is a good indicator of embryo survival after artificial insemination (Stronge et al., 2005). It is a small molecule by nature like the female hormone estradiol and both have natural binding partners such as their specific receptors outlined in Figure 3.1 (Tanenbaum, Wang, Williams, & Sigler, 1998).

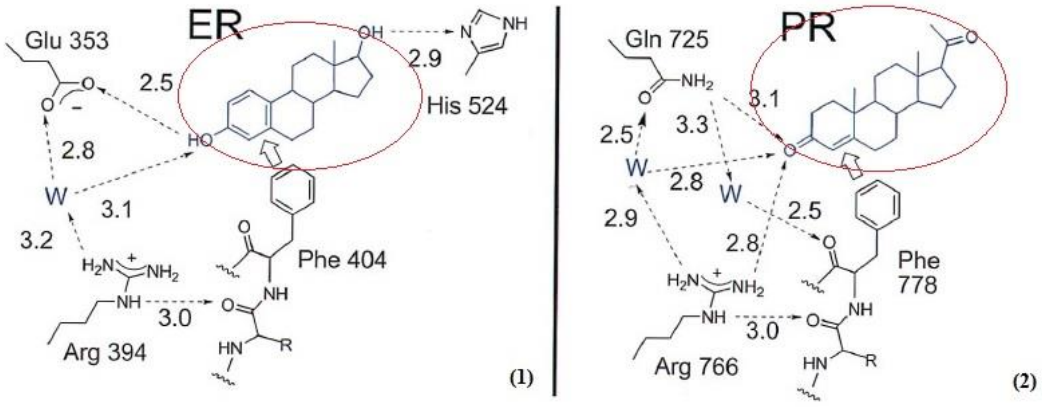


Figure 3.1 Comparison of Estradiol (E2) and Progesterone(P4) bound by their receptors ER and PR

Atomic details of (1) the estradiol receptor (ER) binding to estradiol (red oval) and (2) the progesterone receptor (PR) binding to progesterone (red oval). Both show similarities in the steroid rings (pentagonal or hexagonal shapes) but differences in terminal polar groups (O and OH groups). In both instances, the steroid ring of phenylalanine (Phe) in the receptors stacks against the steroid ring of their respective small molecules (thick arrows) whereas the terminal polar interactions are depicted by hydrogen bonds (dashed arrows) including that made with water molecules (W). [used with permission from (Tanenbaum et al., 1998)].

Progesterone (P4) and estradiol (E2) are very similar molecules, the main difference being carbonyl groups (C=O) at each end of P4, whereas hydroxyl groups (OH) occupy similar positions in E2 (Figure 3.1, red circles). Their respective receptors (ER and PR) provide hydrogen-bond acceptors from E2 and donate hydrogen bonds to P4. This is possible because the OH groups at either ends of E2 are hydrogen donors and the carbonyl (O) groups at either ends of P4 are hydrogen acceptors. Both P4 and E2 share steroid rings which hydrophobically interact with the aromatic rings of the receptor amino acids. Water mediated hydrogen bonds play a role at the ends of both P4 and E2 in the context of their receptors. In both ER and PR, the small molecules (E2 and P4) form the structural core of their respective receptors from cooperative hydrophobic and hydrogen bonding between the molecules and their receptors. (Tanenbaum et al., 1998). This gives an understanding of the context of the binding pocket that might be required for an artificially synthesized binding partner (D7) to P4.

3.1.2 Differences between periplasmic and cytoplasmic expression

Protein folding in the cytoplasm is assisted by chaperone systems like trigger factor (TF), DnaK-DnaJ-GrpE (Hsp 70), GroEL (Hsp 60), IbpAB (sHsps) and ClpB as shown in Fig 3.2. TF is the first chaperone the protein encounters when it exists the ribosome. It assists in protein folding along with downstream DnaK. The latter is part of Hsp 70 which binds to hydrophobic regions of proteins and along with partners DnaJ and GrpE assist folding and release of the protein in an ATP dependent manner. This cycle can repeat, the protein may be transferred to GroEL, subjected to further folding steps or end up aggregating. Hsp 70 increases solubility of proteins (Kolaj, Spada, Robin, & Wall, 2009).

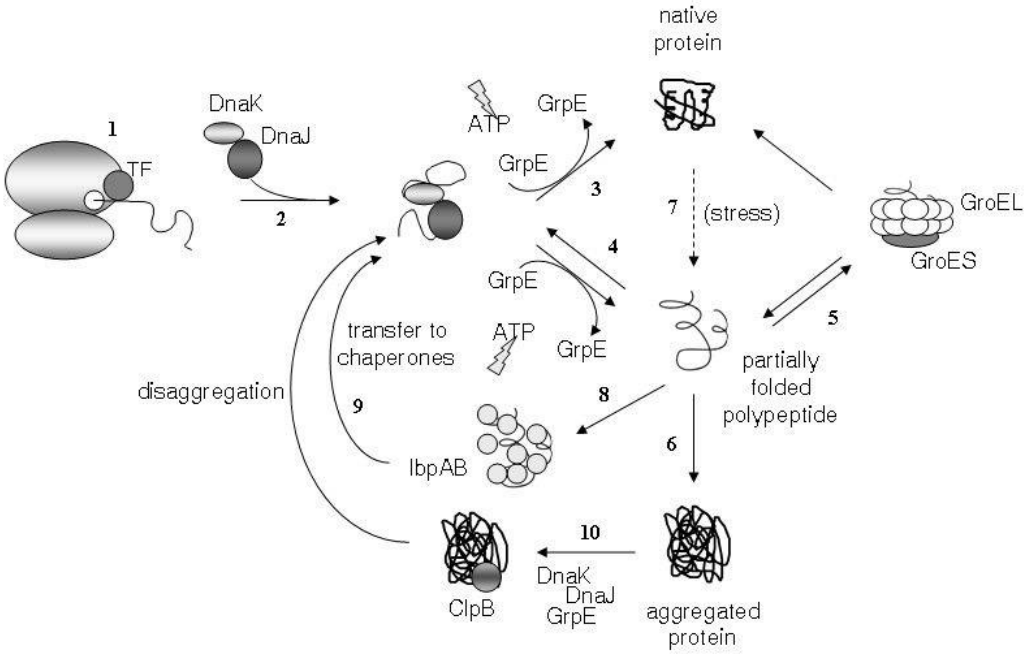


Figure 3.2 Protein folding in reducing cytoplasmic environment

1) The trigger factor (TF) binds to the protein when it exits the ribosome. 2) and 3) refers to protein being folded by Hsp 70 (DnaK/ DnaJ/GrpE). When the protein does not fold to its native form, it goes through repeated cycles with Hsp 70 (4) or Hsp 60 (GroEL, 5) and sHsps (IbpAB). Aggregated proteins tend to use the aid of ClpB to fold to its native form (9, 10). [adapted with permission from (Kolaj et al., 2009)]

GroEL is a double heptameric ring structure with a central cavity, which binds to hydrophobic parts of a protein and internalises it with a lid made of a single heptameric ring called GroES (Hsp 10) providing an environment for the protein to fold until there is no exposed hydrophobic protein parts, in a ATP dependent manner. Small Hsps like IbpA and IbpB act as reservoirs of unfolded protein under stress and help fold proteins with Hsp 70. ClpB is a large star shaped, hexameric chaperone with a central channel which assists with disaggregating proteins with Hsp 70.

Some proteins are secreted to the periplasm for folding. The general secretory system (Sec) consisting of SecB chaperone, SecA motor protein and SecYEG pore is commonly used pathway. SecB maintains protein in ATP independent unfolded form, SecA translocates protein through SecYEG into the periplasm in an ATP dependent manner. SecDF prevent reverse translocation of the protein. Signal recognition particle (SRP) based FtsY receptor translocates proteins via Sec or YidC pore. Tat pathway consisting of Tat ABC translocates folded proteins to the periplasm. The translocation is dependent on signal sequences and assisted by chaperones in the cytoplasm as shown in Fig 3.3.

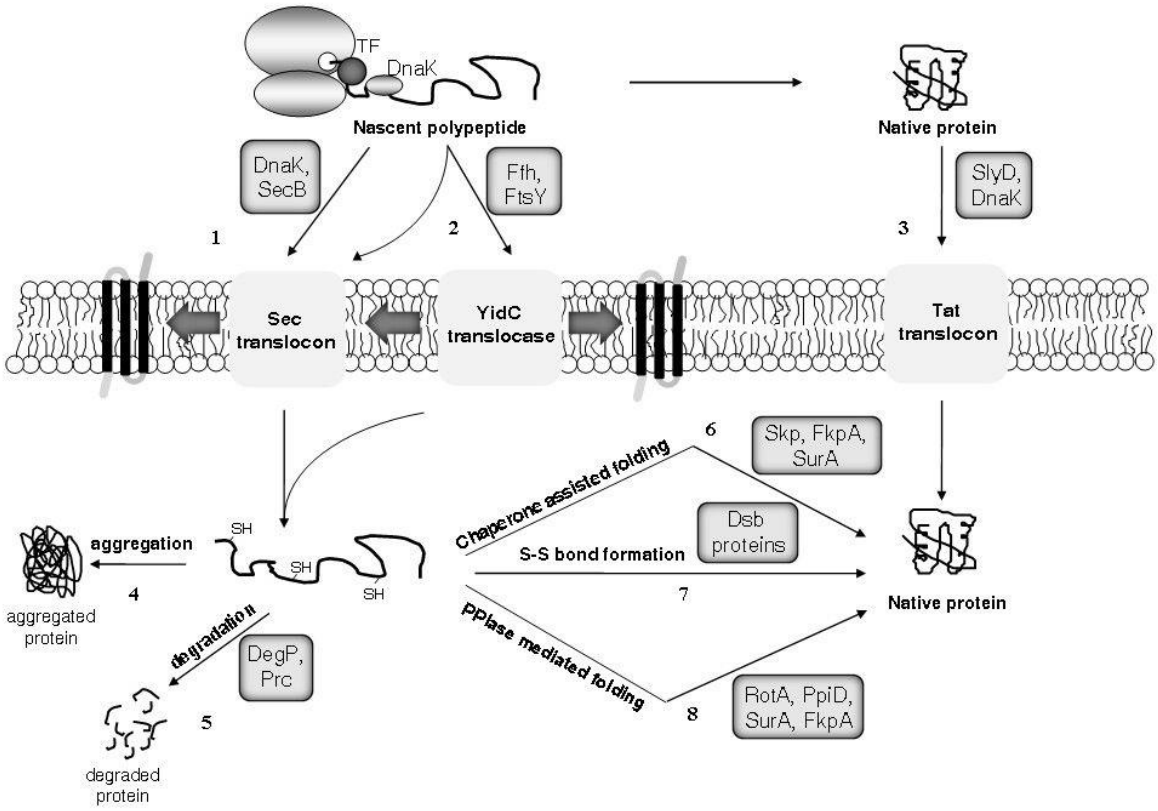


Figure 3.3 Protein folding in oxidising periplasmic environment

1) Hsp 70 (DnaK) or SecB maintain the protein exiting the ribosome in its unfolded form while delivering it to Sec translocon for translocation into periplasm. 2) Signal sequence dependent co-translational translocation occurs through Sec or YidC translocase via SRP (FtsY) receptor. 3) Folded proteins are translocated through Tat translocon as shown in Fig 1.20. 4) Proteins in periplasm can aggregate due to formation of non-native disulfides, 5) get degraded by proteases, 6) undergo chaperone assisted folding, 7) repair its disulfides by oxidoreductases (Dsb) or 8) undergo cis- trans isomerisation (PPiases) to fold into its native form. [adapted with permission from (Kolaj et al., 2009)]

The proteins in the periplasm after translocation and signal sequence cleavage, can be folded with chaperones (Skp, SurA, FkpA), disulfide catalysts (Dsb), isomerases (PPiase), degraded with proteases (DegP) or end up aggregating. DsbA and DsbC catalyse disulfide formation and isomerisation respectively. They are maintained in their oxidised state via DsbB and DsbD respectively. DsbEG are other catalysts. SurA and FkpA are key Peptidyl Prolyl cis/trans isomerases. Many of the chaperones have more than one function.

Disulfide bonds are the result of oxidation of thiol groups between two cysteines. Disulfide bonds can undergo one of the three states – oxidation, reduction and isomerisation. Different compartments of a prokaryotic *E.coli* cell contribute to the above states with the cytoplasm and periplasm providing a reducing and

oxidising environment respectively. These reactions are assisted by oxido-reductases from the thioredoxin and glutathione/glutaredoxin pathways contributing to the lack of oxidation in the cytoplasm and Disulfide bond (Dsb) system contributing to oxidation and isomerisation of thiol groups in the periplasm respectively. This is illustrated in Fig 3.4 (Berkmen, 2012).

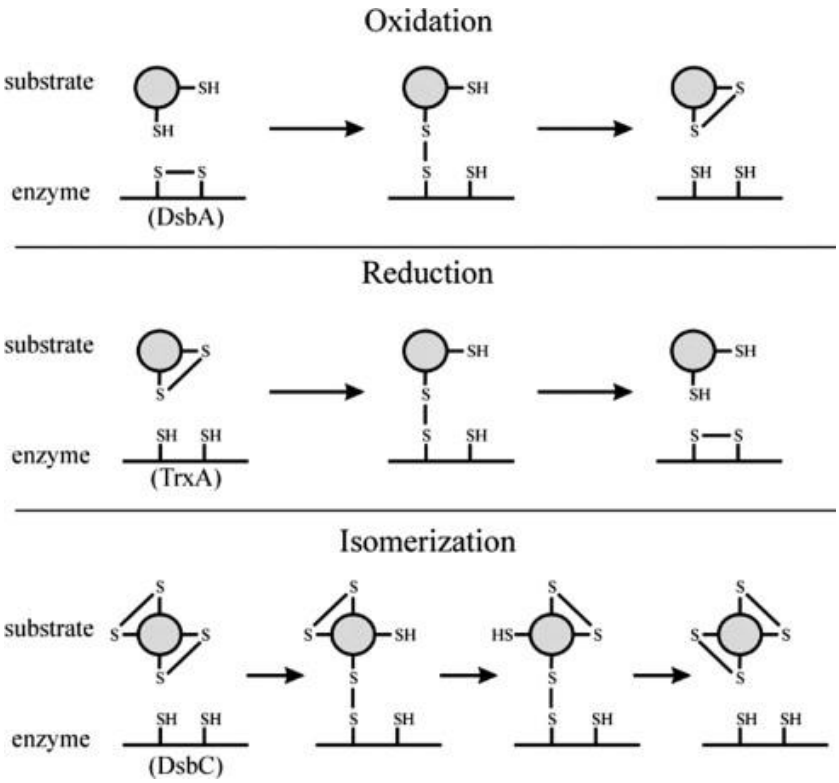


Figure 3.4 Illustration of redox reactions

1) Oxidation involves the enzyme in oxidised state (DsbA) to form a transient disulfide bridge prior to transferring its oxidised state to the reduced substrate (-SH). 2) Similarly, an enzyme (TrxA) in reduced state transfers its reduced state to the oxidised substrate (S-S). 3) Isomerisation involves similar formation of transient oxidised state between enzyme (DsbC) and substrate while the disulfides on the substrate are shuffled resulting in formation of native disulfides. [adapted with permission from (Berkmen, 2012)]

These reactions are possible due to the presence of a catalytic site on the oxido-reductases which has a Cys-X-X-Cys motif which transfers its oxidative state of thiol groups to the reduced state of thiol groups of the protein substrate or vice versa, (where X is any amino acid). This involves an intermediate short lived disulfide formation between the Dsb enzyme and substrate.

Disulfide bonds exist primarily to serve one of the three functions of structure, signaling or catalysis of reactions. Structurally speaking, disulphide bonds can either produce native folding of the protein or add to the stability of protein.

Different compartments of *E.coli* are required for expressing proteins which depend structurally on disulphide bonds. Most proteins that require disulfide bonds for native folding are extra-cytoplasmic where proteins may be exposed to environmental damage. Disulfide bonds are commonly found in periplasmic proteins of prokaryotes like *E.coli.* The cytoplasm is not capable of oxidising thiol groups to produce disulfides for proteins. Signaling disulfide bonds are part of the oxidative stress response which responds to reactive oxygen species (ROS). Catalytic disulfide bonds are caused by oxido-reductases where the electron transfer events of oxidation or reduction are dependent on the amino acids at the catalytic site as well as the environment in which the enzyme is expressed.

The OBody D7 was selected against ligand progesterone (P4) using phage display with protein assembly occurring in the periplasm of *E.coli* TG1. This hints at the possibility of a structural contribution of disulfides to this OBody. DsbA is a strong oxidase that has been shown to oxidise non-native disulfides in proteins. The gene coding for the D7 OBody was sub-cloned into an expression vector pET28b and produced in the cytoplasm in *E.coli* BL21. The reducing environment of cytoplasm contributed by the two pathways is shown in Fig 3.5.

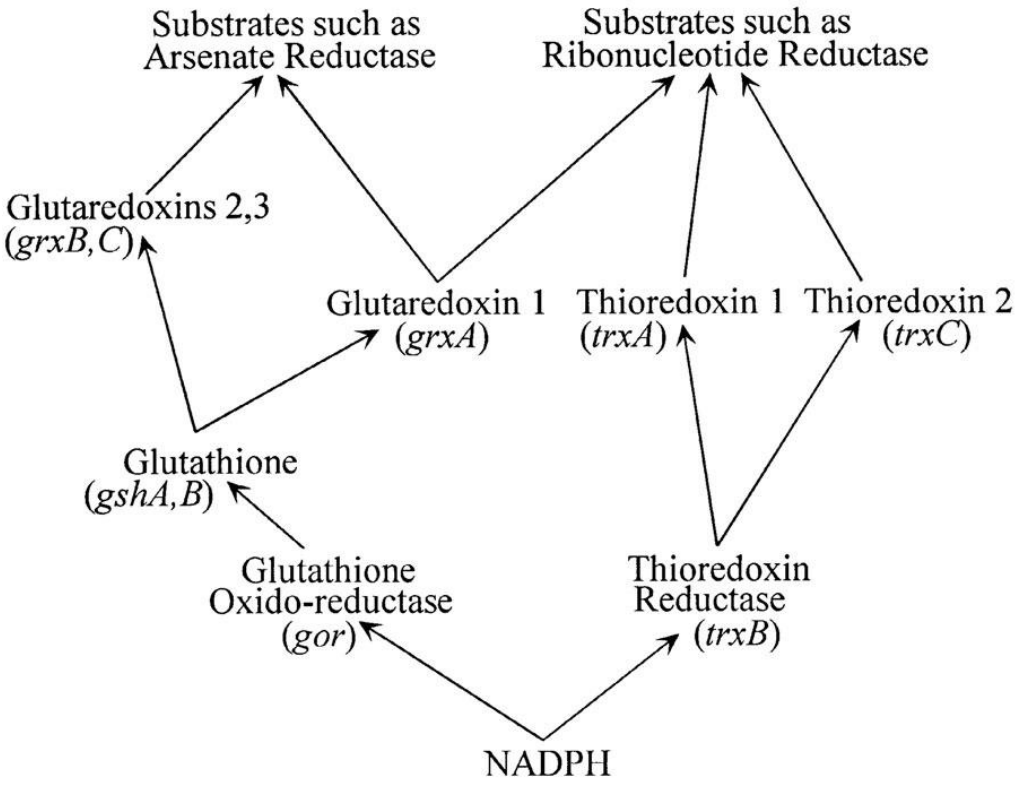


Figure 3.5 Reducing environment in cytoplasm

The reducing environment of the cytoplasm is maintained primarily by two pathways: 1) Glutathione/glutaredoxin pathway with the glutaredoxins (*grxABC*) maintaining substrates in a reduced state. The glutaredoxins maintain their reduced state via glutathione which is synthesized by synthetases (*gshAB*). Glutathione is maintained in its reduced state via glutathione oxido-reductase (*gor*). 2) Thioredoxin pathway with the thioredoxins (*trxAC*) maintaining substrates in a reduced state. The thioredoxins maintain their reduced state via thioredoxin reductase (*trxB*). Both the pathways receive their reducing potential from NADPH. [adapted with permission from (Stewart, Aslund, & Beckwith, 1998)]

The reductases, thioredoxins (*trxA,C*) and glutaredoxins (*grxA,B,C*) are maintained in a reduced state by thioredoxin reductase (*trxB*) and glutathione respectively, which ultimately receive their reducing potential from NADPH. (Stewart et al., 1998).

3.1.3 Structure of an OB-fold

OB-folds are single stranded nucleic acid binding proteins using similar binding surfaces for protein-ligand interactions. OB-folds vary from 70 to 150 amino acids in length with variability determined by loops with a conserved β-barrel (1.4). The binding interface in most OB-folds is adjacent to β2 and β3. The loops involved in this interaction contribute to this binding via hydrophobic aromatic interactions, hydrogen bonds and electrostatic interactions (Theobald, Mitton-Fry, & Wuttke, 2003).

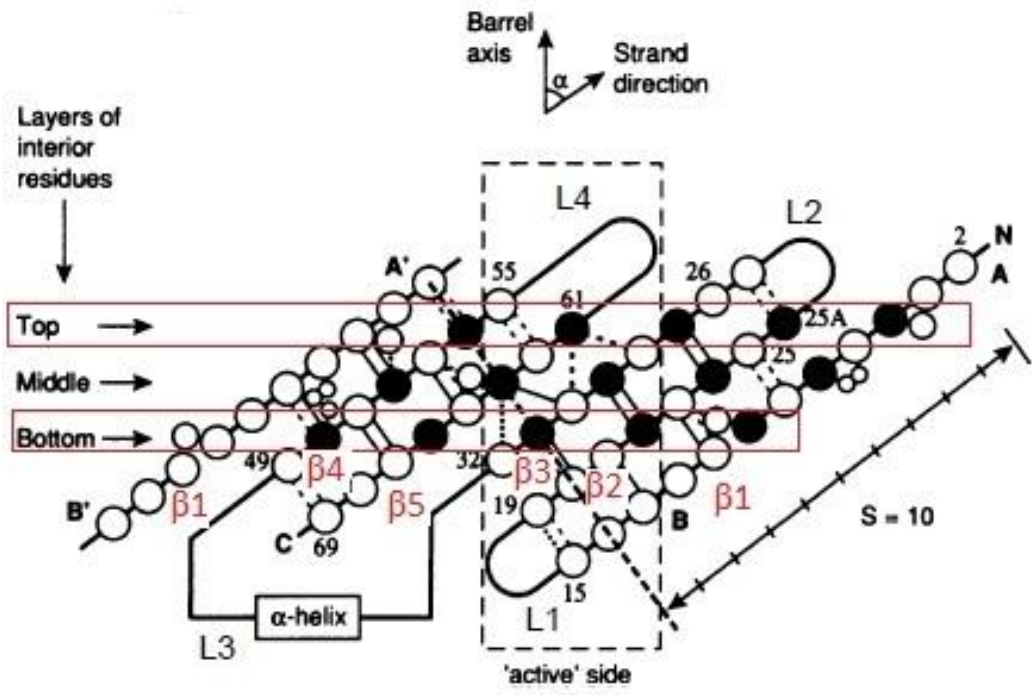


Figure 3.6 Model of OB-fold showing active side

The model of an 'open' β-barrel with the β-sheets (numbered) laid out side to side shows three layers stabilizing it (top, middle and bottom) with each layer contributed by one amino acid per β-sheet (solid black circles). The active side is formed by one side of the β-barrel (dashed box) bound by loop-1 (L1) at the bottom and loop-4 (L4) at the top [adapted with permission from from (Murzin, 1993)].

The binding (active) side of an OB-fold is shown in figure 3.6. The β-sheets arrange themselves in three layers to stabilise the scaffold internally, contributing one amino acid per sheet per layer (Murzin, 1993). The top layer varies with respect to residues, but the middle and bottom layer are dominated by hydrophobic residues that point into the interior of the barrel. The α-helices are usually perpendicular to the barrel axis. The twisting of the β-sheets brings loops 1 and 4 onto the same side of the barrel, and the shallow groove created between them acts as a binding site for elongated ligands (Murzin, 1993). Modifying this (binding) surface is the strategy for designing OBodies to P4.

3.1.4 Preliminary OBodies work

The OB-fold (anticodon recognition domain) of Aspartyl t-RNA synthetase from *Pyrobaculum aerophilum* was engineered using phage display to bind a different ligand, hen egg-white lysozyme (HEL) with a binding affinity of 3 nM (K_d) and this work was done by a previous PhD student in the lab (Steemson et al., 2014). This OBody has high thermal stability, is soluble and lacks disulfide bonds.

The binding affinity (K_d) tells us a lot about the interaction between a protein and its ligand and higher affinity translates into superior results and applications. The K_d was calculated for the Anti-HEL OBody named native_library_8 (NL8) to be 35 μM using surface plasmon resonance (SPR). The crystal structure showed a 1600Å^2 buried surface area between the OBody NL8 and the protein ligand HEL with hydrophobic residues in the center surrounded by polar and charged residues. NL8 inserts its L1 into the substrate binding groove (of HEL) and a residue R35 of NL8 binds to the HEL active site similar to residue H60 of the HEL inhibitor YkfE (Figure 3.7). Acidic residues in the active side of NL8 are bound to complementary basic residues on HEL (Steemson et al., 2014).

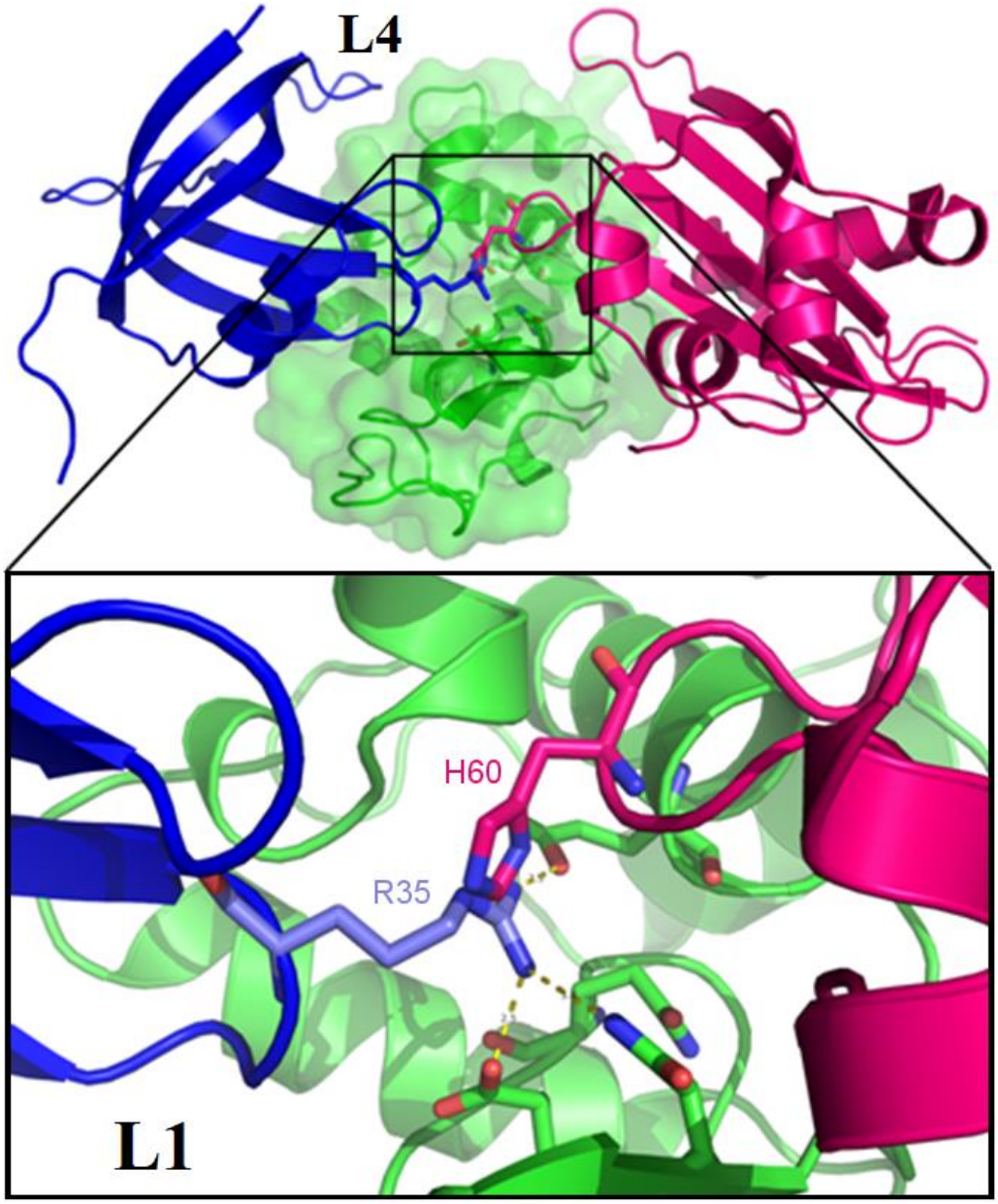


Figure 3.7 Model showing OBody NL8 interacting with HEL in comparison to YkfE (HEL inhibitor)

A cartoon model from crystal structure of NL8 (blue structure) bound to HEL (green structure) shown in comparison to similar interaction of HEL inhibitor (YkfE, pink structure) with HEL. The inset shows a magnified image of L1 residue R35 of NL8 interacting with active site of HEL. Similarly, residue H60 of YkfE interacts with active site of HEL [used with permission from (Steemson et al., 2014)].

The OBody NL8 was further randomised by Steemson at 10 residues to facilitate better contact with HEL (including improving L4 contact of NL8 with HEL). The resulting stronger binding clones that were selected using phage display had a consensus in β-sheet residues indicating the versatility of the binding surface. The affinity calculated from SPR for OBody named Affinity Maturation 1st Library 10 (AM1L10) was K_d = 4.6 µM.

To improve affinity further, a wider area of the OBody was randomised using Error prone PCR (EP-PCR), which produces random mutations over a large area of the protein surface. This resulted in OBody named Affinity Maturation 2nd Error Prone Library 6 (AM2EP06) with a K_d = 250 nM. The crystal structure of AM2EP06 with HEL showed a similar binding surface as NL8 and AM1L10 but with the L4 making contact with HEL and having improved buried surface area by ~100Å.

A third affinity maturation library was made where mutation was focused on L4 making contact with HEL. A stop codon was inserted on L4 to prevent re-selection of existing OBodies and to ensure selection of L4 randomised clones (using NNS codons where S= C or G). The resulting strong affinity clones showed a common selective pressure towards hydrophobic or small hydrophilic residues in L4. The highest affinity recorded by SPR was for OBody named Affinity Maturation 3rd Library 9 (AM3L09) (K_d = 3nM). The buried surface area is 1800 Å^2. Electrostatic (charged) interactions were present in two complementary (loops) areas of the OBody-HEL complex, one on L1 (R35) and the other on L4 (D91) (Figure 3.8). A similar pattern was observed with regards to intermolecular water hydrogen bonding networks (not shown here). To summarize, the common binding site of the OB-fold is used in binding to HEL in all the anti-HEL OBodies, with a central hydrophobic patch surrounded by complementary charged interactions with HEL. OBody AM3L09 (K_d = 3nM) differs from OBody NL8 (K_d = 35uM) as it has more polar and charged interactions with HEL.

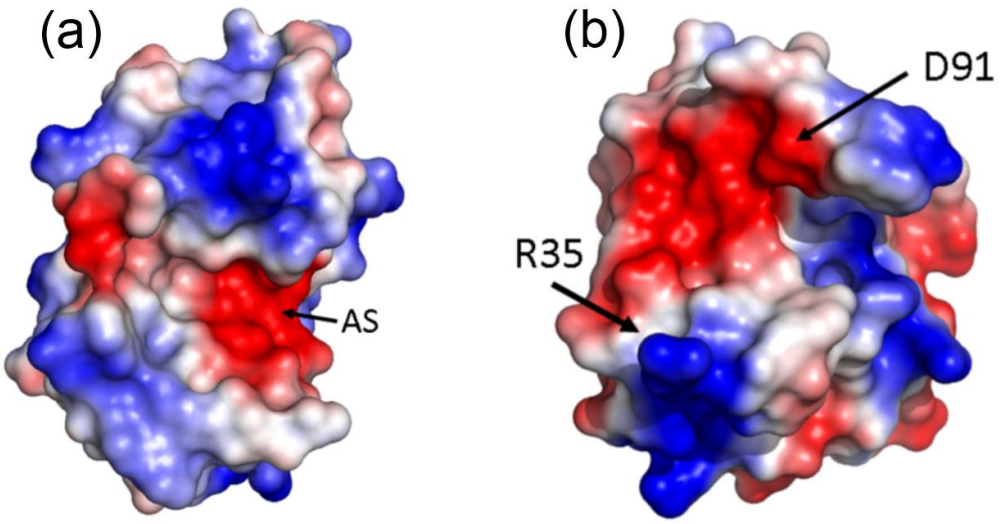


Figure 3.8 Animated electrostatic models of HEL and AM3L09

Complementary electrostatic interactions between (a) active site (AS) of HEL which interacts with (b) loop (L1) residue R35 of OBody AM3L09. Similarly, loop (L4) residue D91 of NL8 interacts in a complementary fashion with HEL (red area – negatively charged, blue area – positively charged) [used with permission from (Steemson et al., 2014)].

3.1.5 Re-design of OBody library for P4

A second generation naïve (2GPATS) OBody library was created by Dr. Mark Liddament to be able to select for OBodies with better binding qualities, with randomisation (of charged or hydrophobic residues not involved in binding) at the periphery of active side of wild type OB-fold (Figure 3.9) to try to eliminate steric hindrance. The naïve library was created by designing a library with mutations targeting 15 residues coding for the binding surface of 1GPATS library. The library was produced such that $2x10^{10}$ individually diverse clones were generated compared to 10^8 for 1GPATS (Figure 3.10).

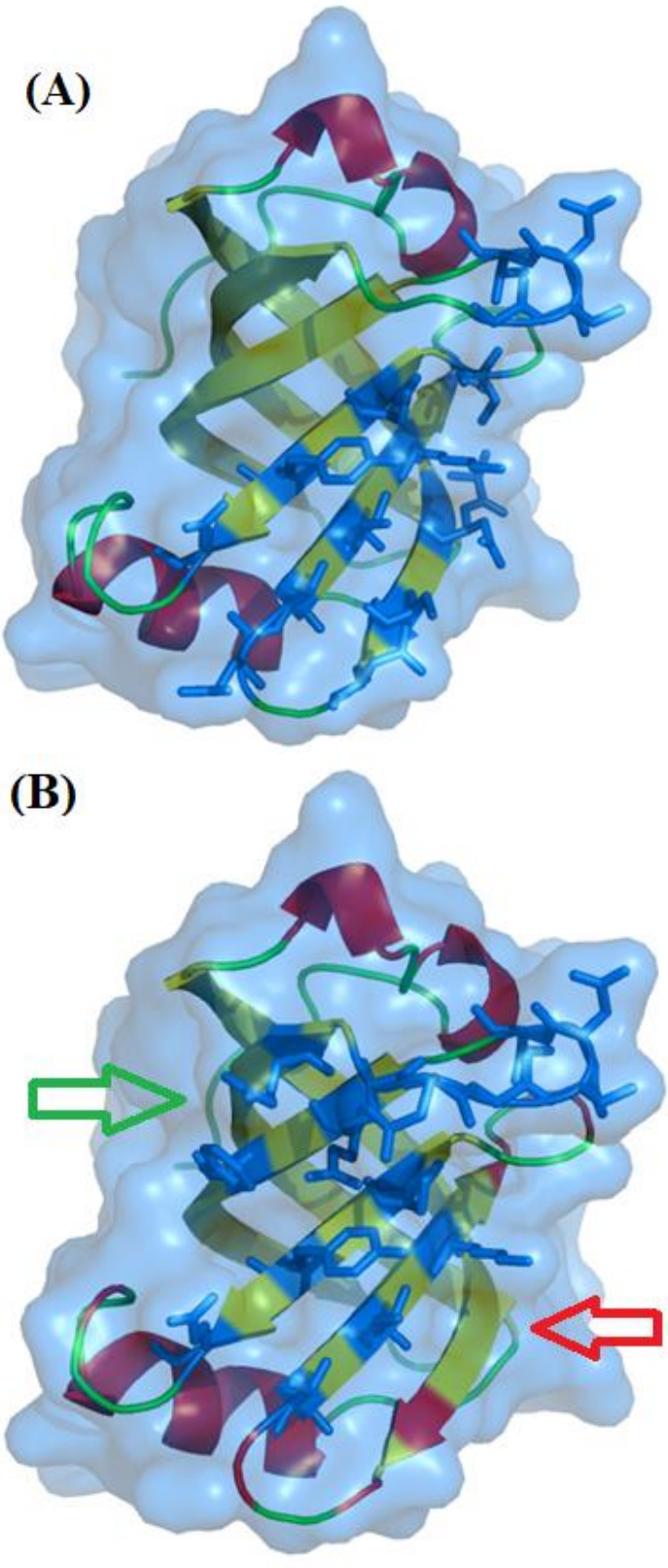


Figure 3.9 Comparison of OBodies from 1GPATS with 2GPATS model

Cartoon model of an OBody from the first generation OBody library (1GPATS, A) in comparison to 2GPATS (B) showing randomized residues as blue stick figures with new residues added for randomization at the periphery of the active side (green arrow) and residues which have been omitted from randomisation (red arrow) in 2GPATS. The β-sheets are shown in yellow, coils in green, α-helices and retained mutations in red.

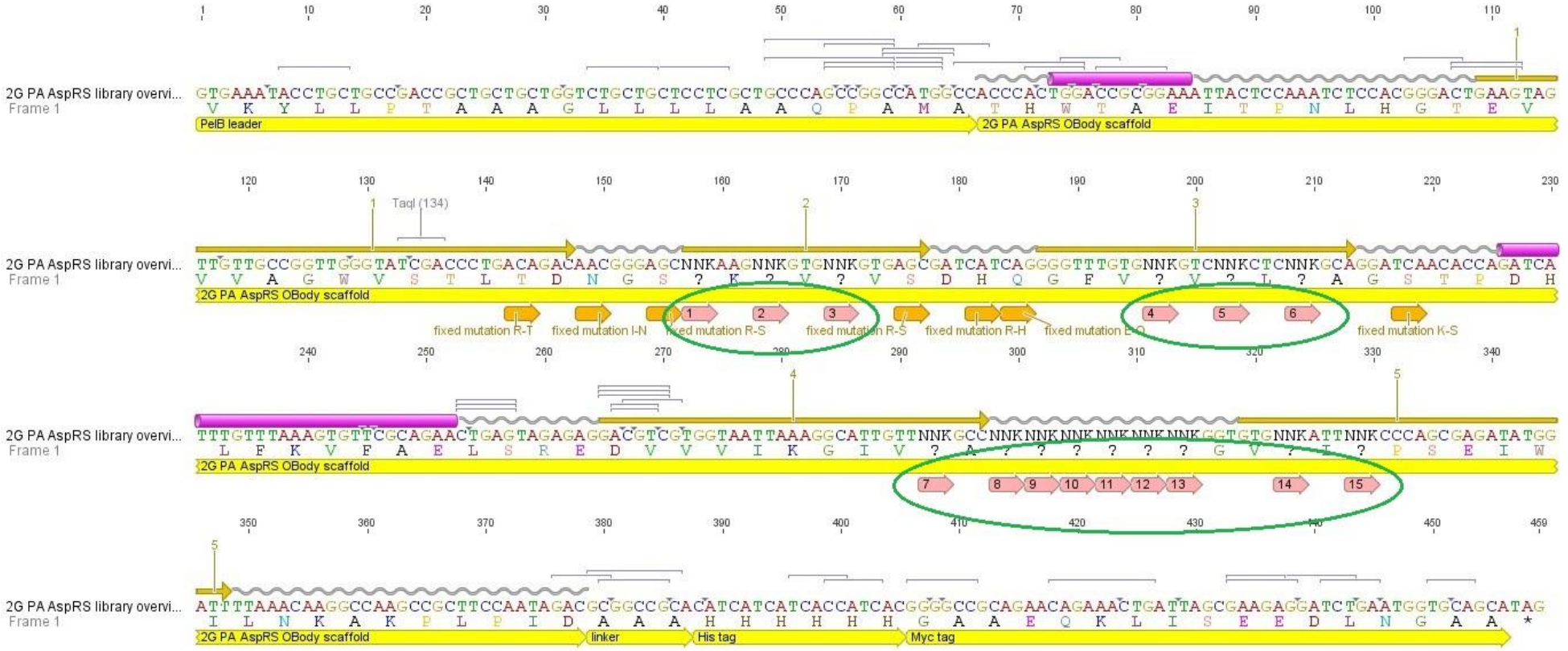


Figure 3.10 Gene sequence of 2GPATS naïve library

Insilico construction of naïve library (using geneious sequencing software) showing protein sequence of a typical OBody (with DNA sequence on top of it) with green circles indicating randomised areas specific to 2GPATS (referring to light pink arrows). On top of the protein sequence, bright pink cylinders refer to α-helices, wavy lines refer to coils, thin arrows refer to β-strands and orange arrows adjacent to light pink arrows below the protein sequence refers to retained (fixed) mutations from 1GPATS. PelB signal (leader) sequence and OBody (scaffold) sequence are annotated (yellow) and randomised mutations are accompanied by NNK codons with (?) referring to random amino acids.

A sequence comparison between P4 binding OBody (D7, 2GPATS, 7.2.1) and the HEL binding OBody (AM3L15, 1GPATS) shows the conservation of core structural features such as β-strands forming the barrel and high variability at the loop region (L4) (Figure 3.11). There is overall 75% sequence similarity between D7 and AM3L15.

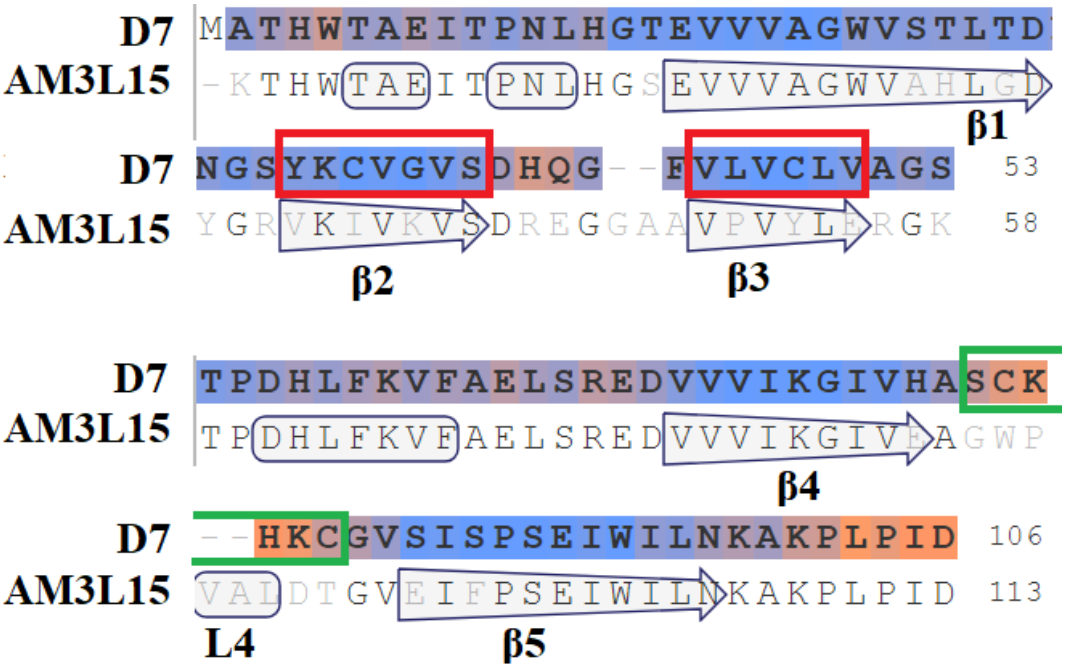


Figure 3.11 Comparison of D7 (2GPATS) with AM3L15 (1GPATS)

Protein sequence comparison (generated using SWISS-MODEL) of D7 (2GPATS) with AM3L15 (1GPATS) showing conserved residues as part of the core β-barrel with modified residues in β2 and β3 (red boxes) and modified residues in L4 (green box).

My work started with the confirmation of OBody D7 in cytoplasmic expression plasmid pET28b in the absence of foreign fragment (YscC) by DNA sequencing. OBody D7 was then expressed and purified in large scale using protein purification techniques and characterised functionally to assess P4 binding.

3.2 Results

3.2.1 Large scale expression and purification of D7

D7 was sub-cloned into expression vector pET28B (7.2.2) and expressed in *E. coli* BL21 strain using standard bacterial media (2xYT, LB) and purified by Immobilized metal affinity Chromatography (IMAC, Figure 3.8) using standard buffers (TBS, PBS) containing imidazole. Initial purification showed that there were two distinct bands, one between 10 -15 kDa and another between 20 - 25 kDa on a 16.5% SDS-PAGE gel. The OBody (D7) has a theoretical molecular

weight of 12.7 kDa possibly indicating that the band at around 25 kDa corresponds to an SDS resistant dimer. The SDS-PAGE image also showed the possibility of higher order oligomers (Figure 3.12).

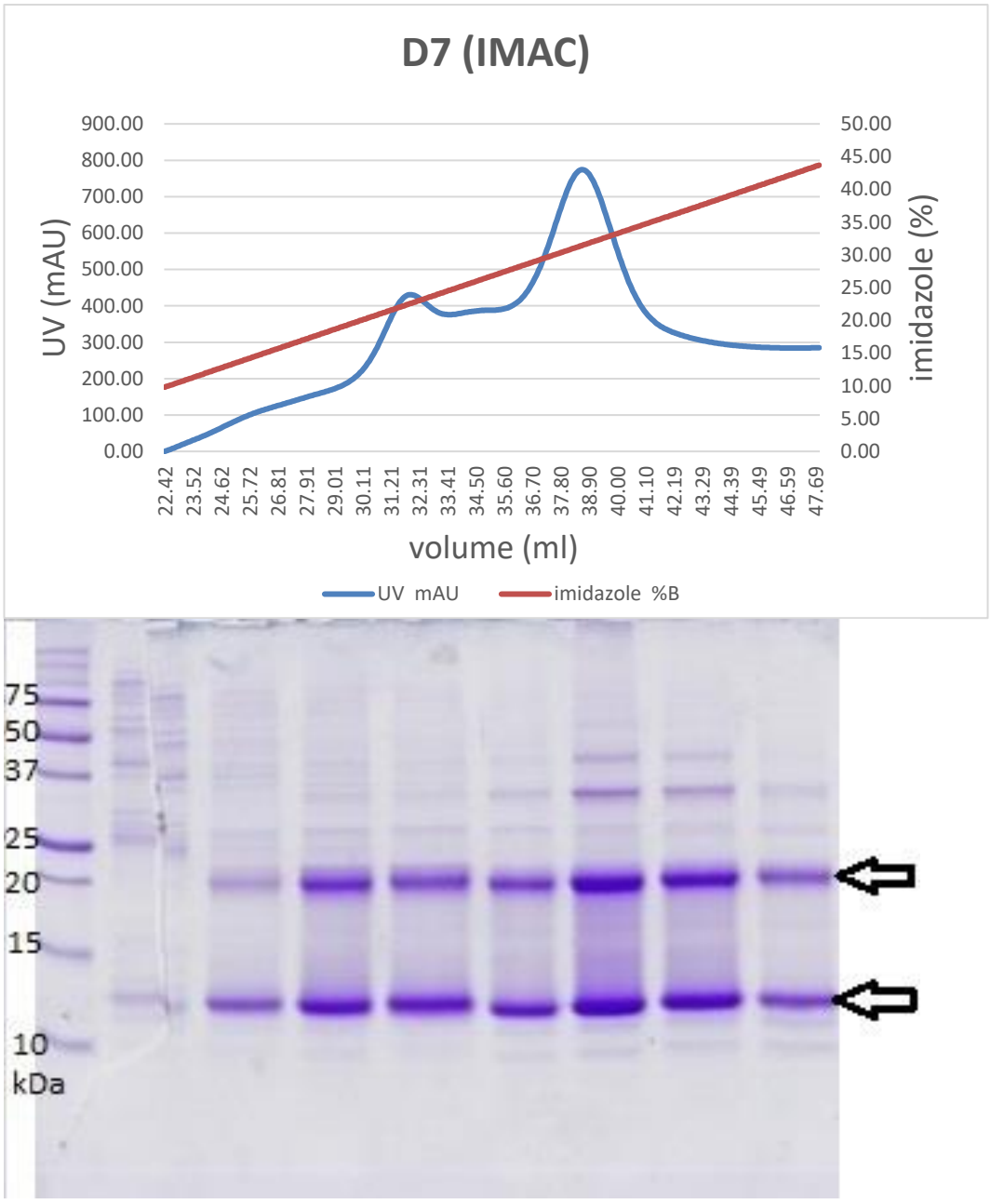


Figure 3.12 UV elution profile and SDS-PAGE image of IMAC purified D7

The top image shows the UV absorbance curve (blue, 280nm) in mAU (left y-axis) showing the elution of D7 over an imidazole gradient (red) in percentage (right y-axis) during Histrap (IMAC) purification. The blue peak indicates maximum D7 elution. The x-axis refers to the flow of IMAC buffers (7.1.3.2) containing imidazole. IMAC is explained in detail in 2.2.2.1. The bottom image is 16.5% SDS-PAGE showing IMAC fractions containing purified D7 with reference to a protein ladder (left) showing possible sizes of proteins. D7 elutes as a monomer (12.7kDa) and dimer (25kDA) both indicated by arrows. There is also trace oligomerisation at ~37 kDa. Note that each lane in SDS-PAGE corresponds to the elution fractions of two UV peaks in IMAC.

The fractions containing D7 were pooled and concentrated to further purify using an S200 10/300 GL or an S75 16/600 column for size exclusion chromatography (SEC). D7 eluted as a single peak corresponding approximately to the monomer (Figure 3.13). However, SDS-PAGE showed no separation of monomer and dimer species (Figure 3.14).

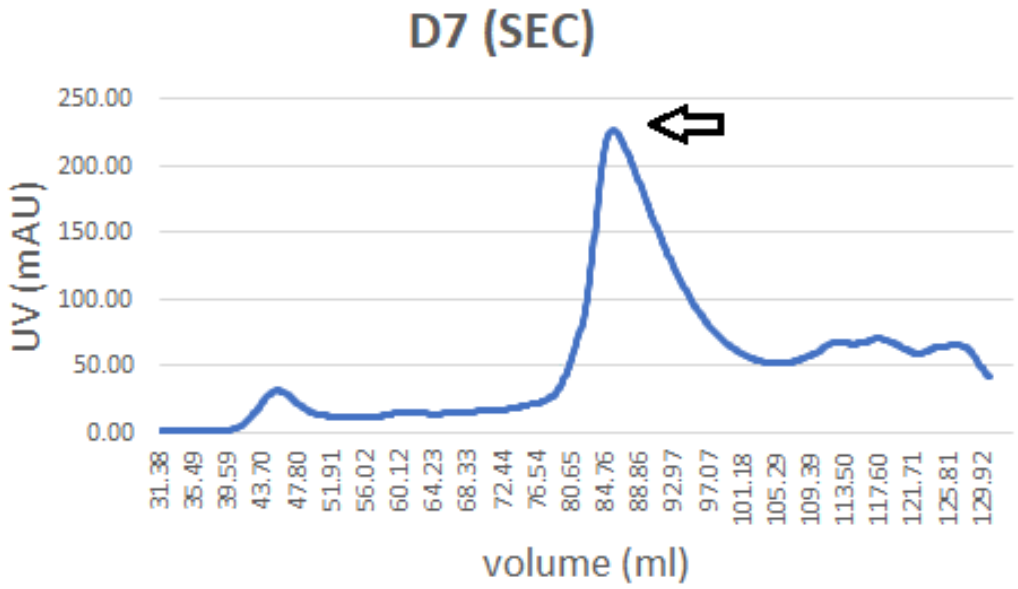


Figure 3.13 Purification of D7 on S75 16/600 column

Elution profile of D7 on S75 16/60 shows a single peak (arrow) corresponding to the monomer (12.7kDa). Y-axis is the intensity of absorbance by aromatic rings of protein at 280nm in milli-Absorbance Units (mAU). X-axis is the volume (ml) of buffer (7.1.3.2) the protein elutes at and is inversely proportional to the size (MW) of the protein in solution (Hong, Koza, & Bouvier, 2012).

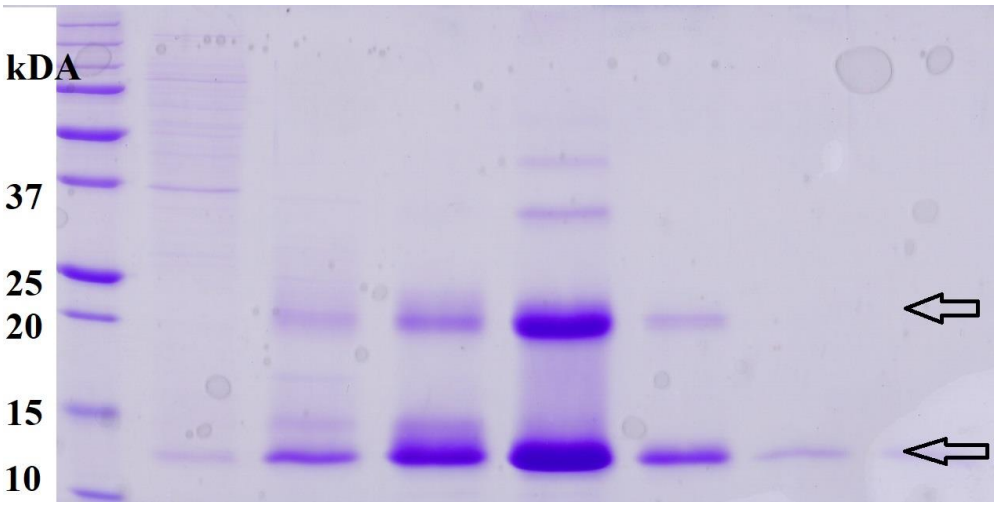
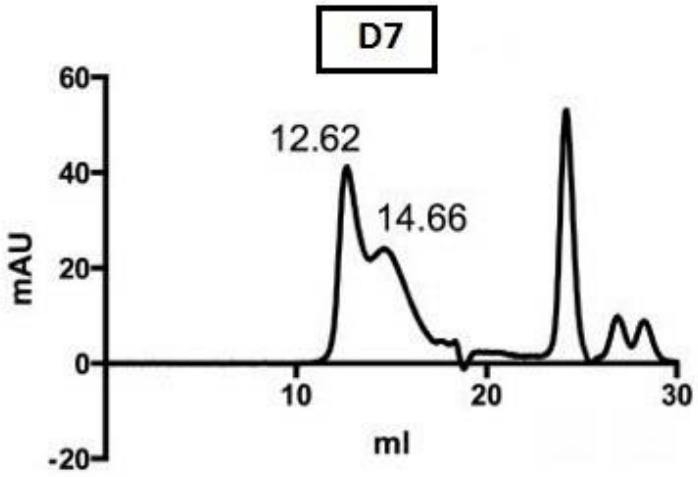


Figure 3.14 SDS-PAGE image of S75 16/600 purified D7

16.5% SDS-PAGE image showing S75 16/60 (SEC) fractions containing purified D7 with reference to a protein ladder (left) showing possible sizes of proteins. D7 still elutes as a

monomer (12.7kDa) and dimer (25kDA) both indicated by arrows. There is also trace oligomerisation at ~37 kDa. Note that the lane with strong dimer in SDS-PAGE corresponds to the UV peak (arrow) in SEC.

Samples were sent to the University of Otago for analysis. An interesting observation is that SEC done at University of Otago was sensitive enough to separate the monomer from the dimer (Figure 3.15). The other interesting observation is the dimer to monomer concentration is low at low concentrations of D7 as shown in SDS-PAGE (Figure 3.16).



	MW (kDa)	Retention volume (ml)
Thyroglobulin	670	8.14
gamma-globulin	158	8.95
Ovalbumin	44	10.88
Myoglobin	17	13.04
Vitamin B12	1.35	18.49

Figure 3.15 SEC analysis of D7

Elution profile of D7 on SEC (top) shows two peaks (numbered) corresponding to the dimer (12.62ml) and monomer (14.66ml) with relation to the standards (bottom) provided by Biorad. Y-axis is the intensity of absorbance by aromatic rings of protein at 280nm in milli-Absorbance Units (mAU). X-axis is the volume (ml) of buffer the protein elutes at and is inversely proportional to the size (MW) of the protein in solution. This was carried out in collaboration with University of Otago.

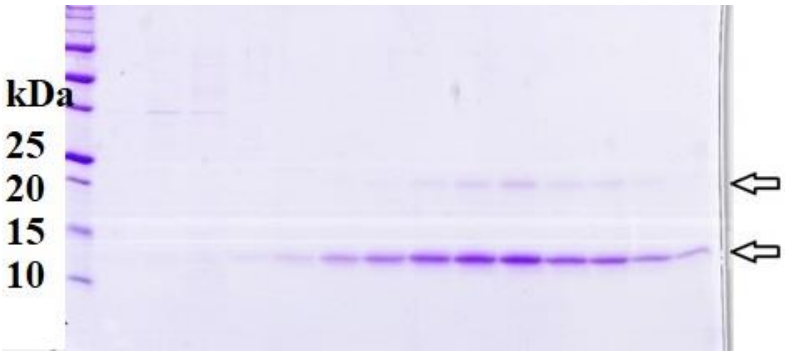


Figure 3.16 SDS-PAGE image of low yield D7

D7 fractions from S200 10/300 GL (SEC) showing low dimerization at low concentration. Upper arrow shows the position of (low quantity) dimer. Protein ladder is shown in left and monomer (D7) band is shown by lower arrow. Note that each lane in SDS-PAGE corresponds to different elution fractions.

Pure D7 was concentrated up to 10 mg/ml and crystal trials were laid out using four custom screens (Hampton Research). There was no crystal growth in any of the screens and the protein precipitated beyond this concentration in most wells. It was hypothesized that a lack of homogeneity (Dale, Oefner, & D'Arcy, 2003) or decreased protein stability (leading to protein aggregation) (Gosavi, Mueser, & Schall, 2008) may be a factor in lack of crystals. The possible contribution from intermolecular disulphide bonds was also considered a possible reason for inhomogeneity.

3.2.2 Strategies to improve protein stability

The factors contributing to protein stability can be divided into protein (intrinsic) or buffer (external) causes (Trivedi, Laurence, & Siahaan, 2009). Basic strategies to improve protein stability and homogeneity with different buffer conditions were tested first (Gosavi et al., 2008). Different types of buffers (TBS, PBS, Hepes), pH (4-9) conditions and increasing salt (Gotte & Libonati, 2014) did not have any positive effect on the stability of D7. The stability was assessed by concentrating the purified protein by centrifugation (2.2.2) and higher stability was reflected by achievement of higher concentrations without protein precipitation. Addition of 5% glycerol did improve the protein stability as D7 could be concentrated further than 10 mg/ml. D7 stability was confirmed by protein concentration estimation (using nanodrop) and lack of visible protein precipitation at high concentration. However, this did not lead to formation of crystals and most crystal screens showed precipitation (aggregation).

My hypothesis is that oxidation of one or more of the four cysteines is responsible for the formation of intermolecular disulphide based dimers (Benvenuti & Mangani, 2007). Treatment with reducing agents like dithiothreitol (upto 10 mM DTT) or β-mercaptoethanol (upto 20 mM BME) can eliminate intermolecular disulfides (Bondos & Bicknell, 2003). Treatment with DTT (Figure 3.17) in gel purification buffer resulted in the partial elimination of the dimer band on SDS-PAGE and crystal trials were laid down using 5 mM DTT or 10 mM BME.

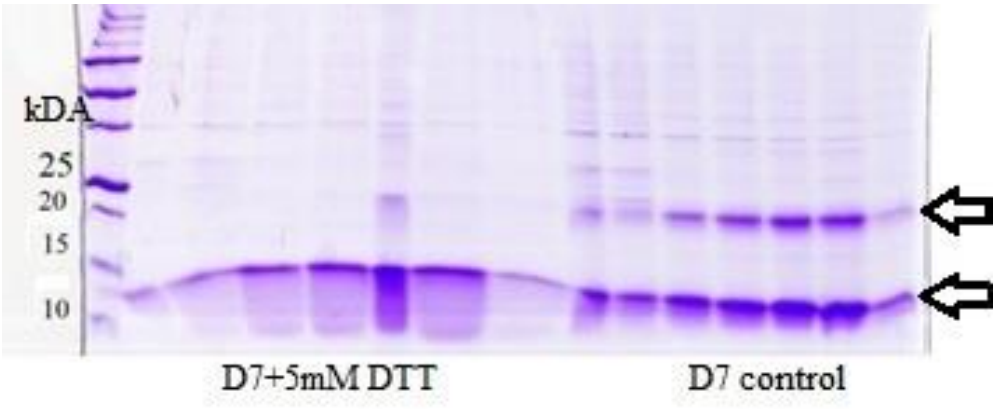


Figure 3.17 SDS-PAGE image of D7 with DTT

D7 was incubated with 5mM DTT and run on SDS-PAGE with untreated D7 control. The dimer band (upper arrow) has disappeared in D7 fractions treated with DTT. The protein ladder is on the left and the monomer (D7) band is shown by the lower arrow. Note that each lane in SDS-PAGE corresponds to different SEC elution fractions treated with and without DTT.

Using DTT or BME was not as effective on a large scale reflected by protein aggregation at high concentrations possibly from oligomerisation (Wang, 2005). OBodies have been previously concentrated up to 40 mg/ml and crystallized with equimolar ligand (Steemson et al., 2014).

D7 was selected against progesterone (P4) by phage display and P4 was added as a stabilizing ligand (which decreases the free energy of the complex and theoretically improves the chances of crystal lattice formation). The (P4) progesterone (1mM, from steraloids) stored in ethanol was incubated with concentrated D7 at room temperature for 30 minutes and centrifuged to recover the supernatant containing the D7:P4 complex and crystals trials were laid out. There were still no protein crystals.

3.2.3 Construction of D7 mutants

The next step to try to eliminate dimerization (and the contributions to oligomerisation and inhomogeneity) by replacing cysteines with serines with the goal to prevent disulphide bonds. The four cysteines are at amino acid positions 35, 48, 81 and 85. Cysteines 35 (c35) and 48 (c48) are on β2 and β3 respectively. Cysteines 81 (c81) and 85 (c85) are both on the loop L4 (Figure 3.18).

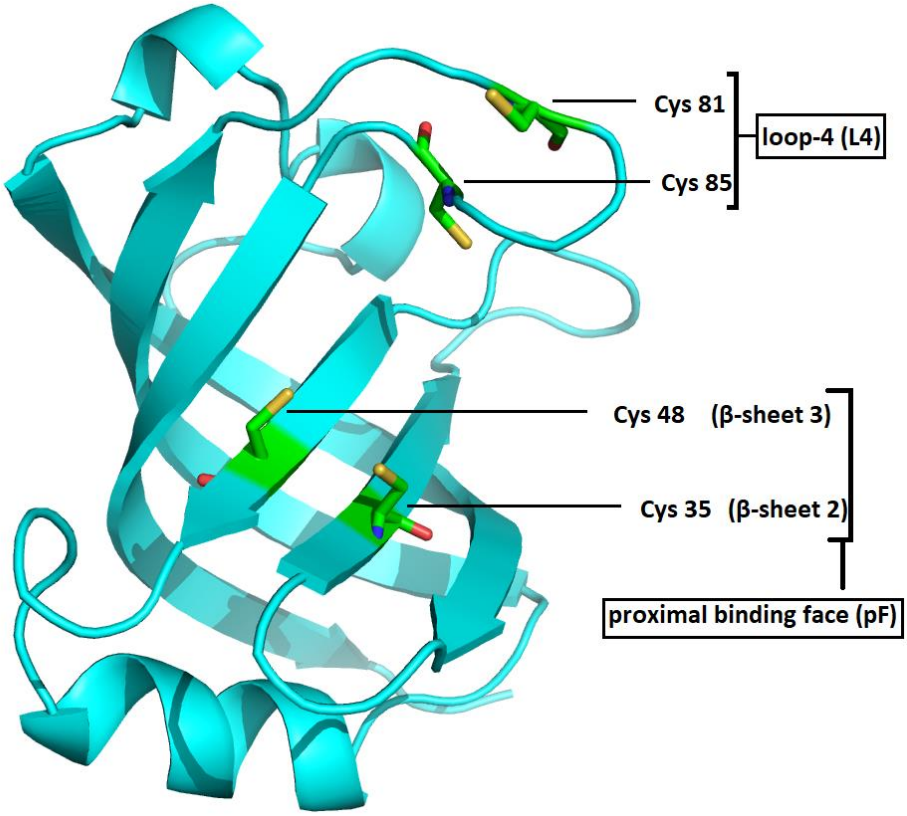


Figure 3.18 D7 model showing position of cysteines

A cartoon representation (created with Pymol) of OBody D7 (cyan) indicating the location of four cysteines (green) is shown. The proximal binding face (pF) consists of β-sheets 2 and 3 (including C35 and C48 as labelled) and loop-4 (L4) includes C81 and C85 as shown.

Four mutants were created with single point mutations (cysteine to serine) at these positions to produce c35s, c48s, c81s and c85s. This was achieved using overlap-extension PCR (OE-PCR) and this technique has been described in 2.5.4.2. Complementary overlap primers were designed for each mutant with the nucleotide sequence that replaces cysteine with serine and overlap fragments were generated with flanking primers for pOB2 (or YscC). The overlap fragments for C35S are 606 bp and 378 bp and for C48S are 640 bp and 334 bp. The overlap fragments for C81S are 743 bp and 240 bp and for C85S are 758 bp and 231 bp.

The flanking primers were used in the extension step of OE-PCR to fuse the overlap fragments to generate D7 mutants which are of the same length (944 bp) and was confirmed by DNA sequencing (Figure 3.19). The double mutant (DM, C35S:C48S) was similarly designed by using C35S as a template for OE-PCR using overlap primers for C48S. The fused overlap fragments were sub-cloned without the YscC fragment (483 bp) into the expression vector pPROEX HTb and confirmed using appropriate primers.

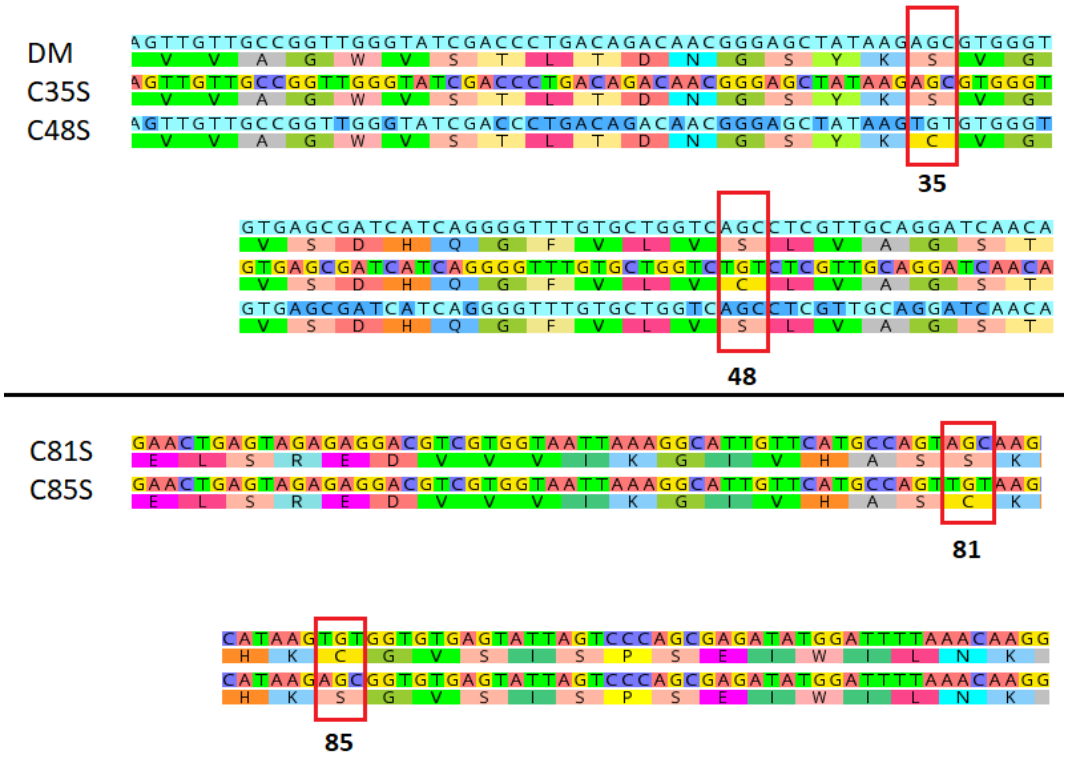


Figure 3.19 DNA sequencing of D7 mutants

The DNA sequencing results (courtesy of Massey Genome Sequencing) of fused overlap fragments for D7 mutants sub-cloned into expression vector pPROEX HTb are shown. In the image (Geneious), the labelled D7 mutants (left) show the relevant mutations (red boxes) with amino acids.

3.2.4 Purification of D7 mutants

Replacement of cysteines with serines on the β2 and β3 showed decreased dimerization in reducing (SDS-PAGE) gel image (Figure 3.20). Replacement of L4 cysteines did not affect dimerization.

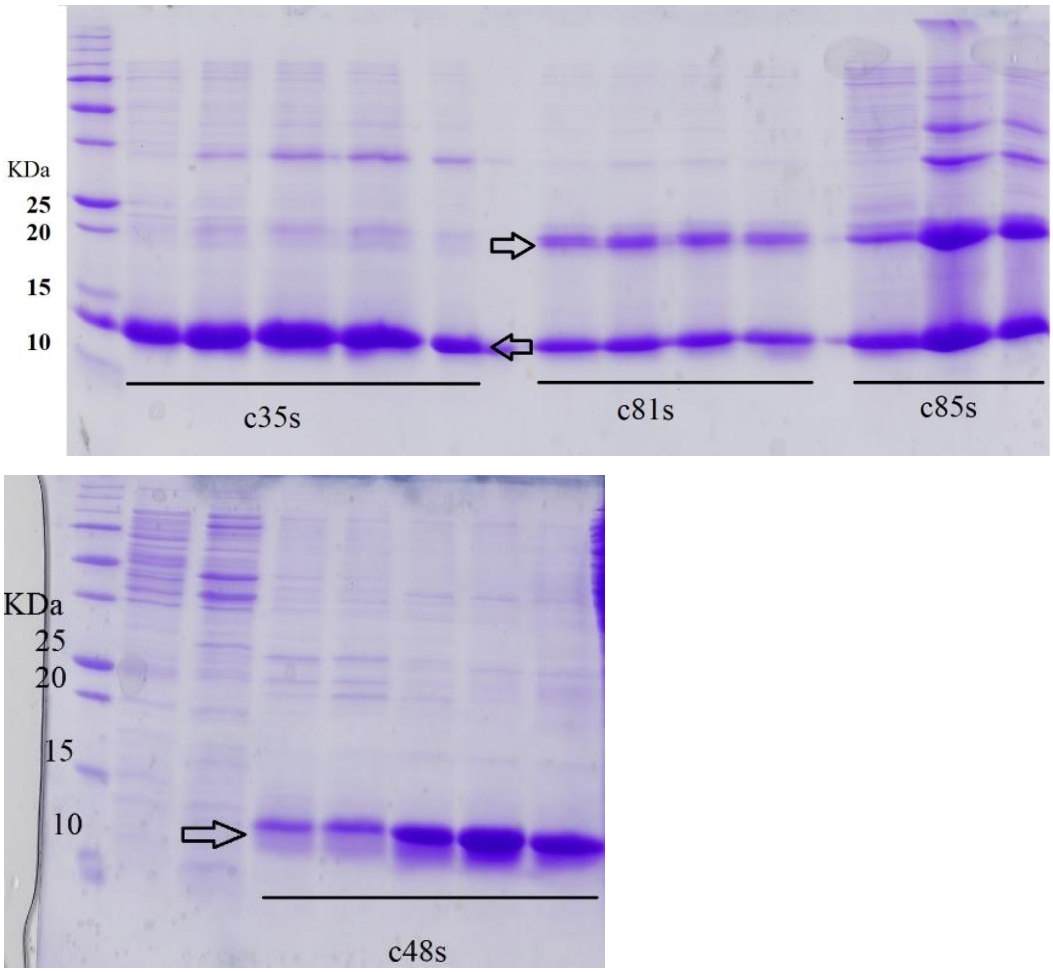


Figure 3.20 SDS-PAGE images of D7 mutants

16.5% SDS-PAGE image showing IMAC fractions containing purified D7 (cysteine) mutants with reference to a protein ladder (left) showing possible sizes of proteins. D7 mutants elute both as a monomer (12.7kDa) and dimer (25kDA) in both L4 mutants c81s and c85s (top) indicated by appropriate arrows. β-sheet mutants c35s (top) and c48s (bottom) are visible only as monomers. Note that upper arrow refers to dimer and lower arrow refers to monomer of D7 mutant. Each lane in SDS-PAGE corresponds to different elution fractions per D7 mutant (7.1.3.1, 10 mM BME).

c35s and c48s point mutants were purified using SEC and concentrated with P4 and crystal trials were laid down. There were no crystals, probably from the persistence of oligomerisation at high concentrations (Figure 3.21).

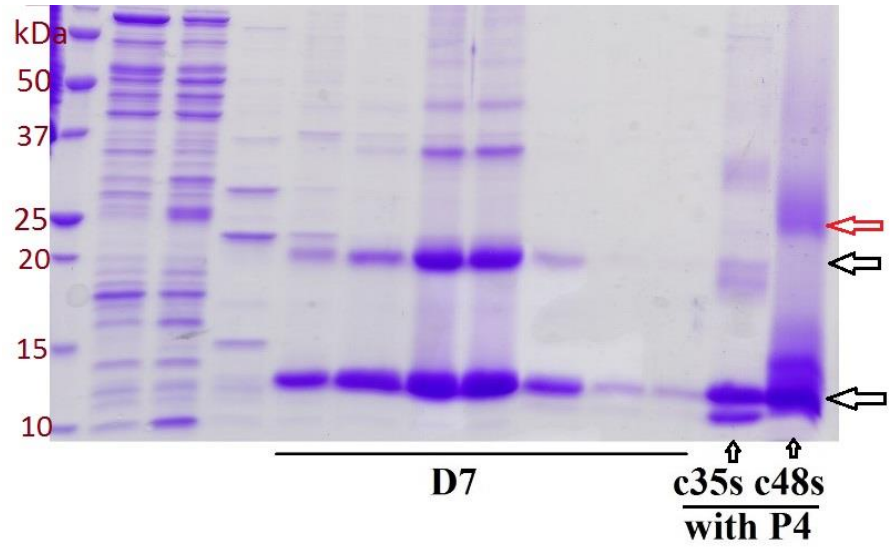


Figure 3.21 SDS-PAGE image of β-sheet mutants

16.5% SDS-PAGE image showing (concentrated) D7 mutants c35s and c48s (right) treated with P4 in comparison to untreated D7 used as a control (center) with protein ladder (left) showing possible sizes of proteins. The black arrows refer to established D7 monomer and dimer sizes. c48s lacked the dimer band but showed some kind of oligomerisation (red arrow).

It was decided to make a double mutant (DM) with both cysteines at positions 35 and 48 replaced with serines to eliminate the dimerization seen at high concentrations. DM was produced using a new vector pPROEX-HTb as part of the optimisation for expression in *E. coli*. Creating a DM however did not eliminate the dimer (Figure 3.22). An alternative strategy of modifying β-sheet mutants from serines to alanines (c35a, c48a) did not have satisfactory results at higher concentrations either (Figure 3.23).

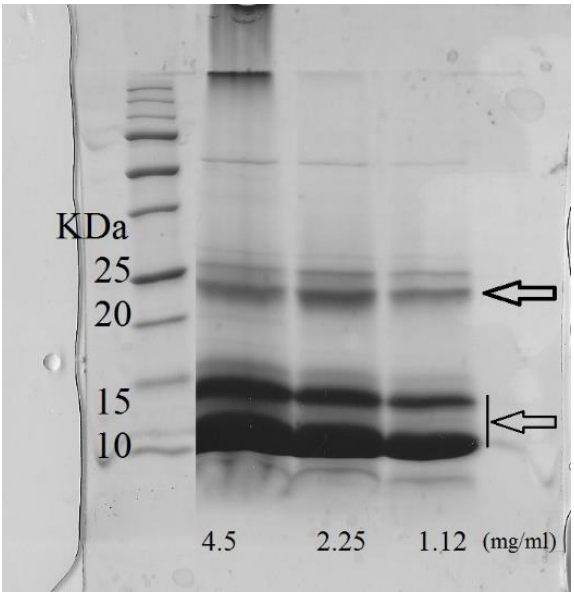


Figure 3.22 SDS-PAGE image of DM at high concentrations

16.5% SDS-PAGE image showing (concentrated) DM at different concentrations with protein ladder (left) showing possible sizes of proteins. The black arrows refer to established D7 monomer and dimer sizes. Note that there are two bands in the monomer and dimers bands which is the result of DM with and without a cleaved histag (from pPROEX, 7.1.3.1, 10 mM BME).

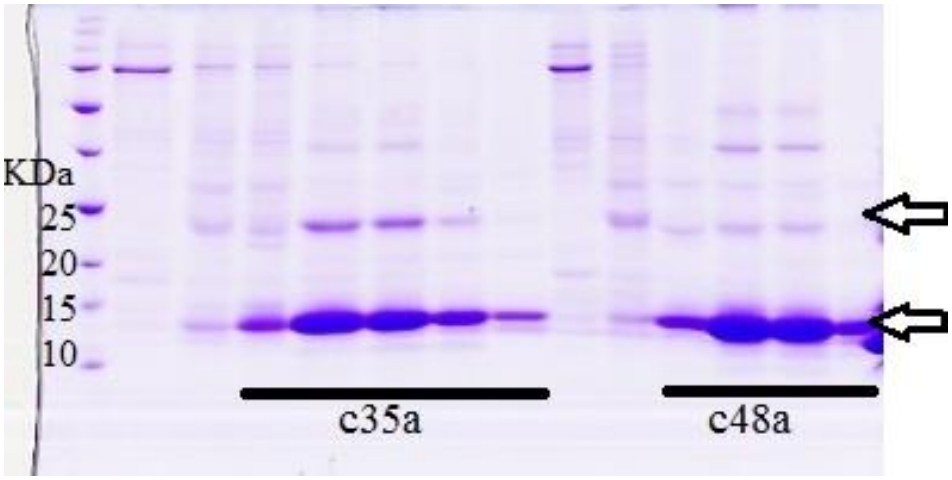


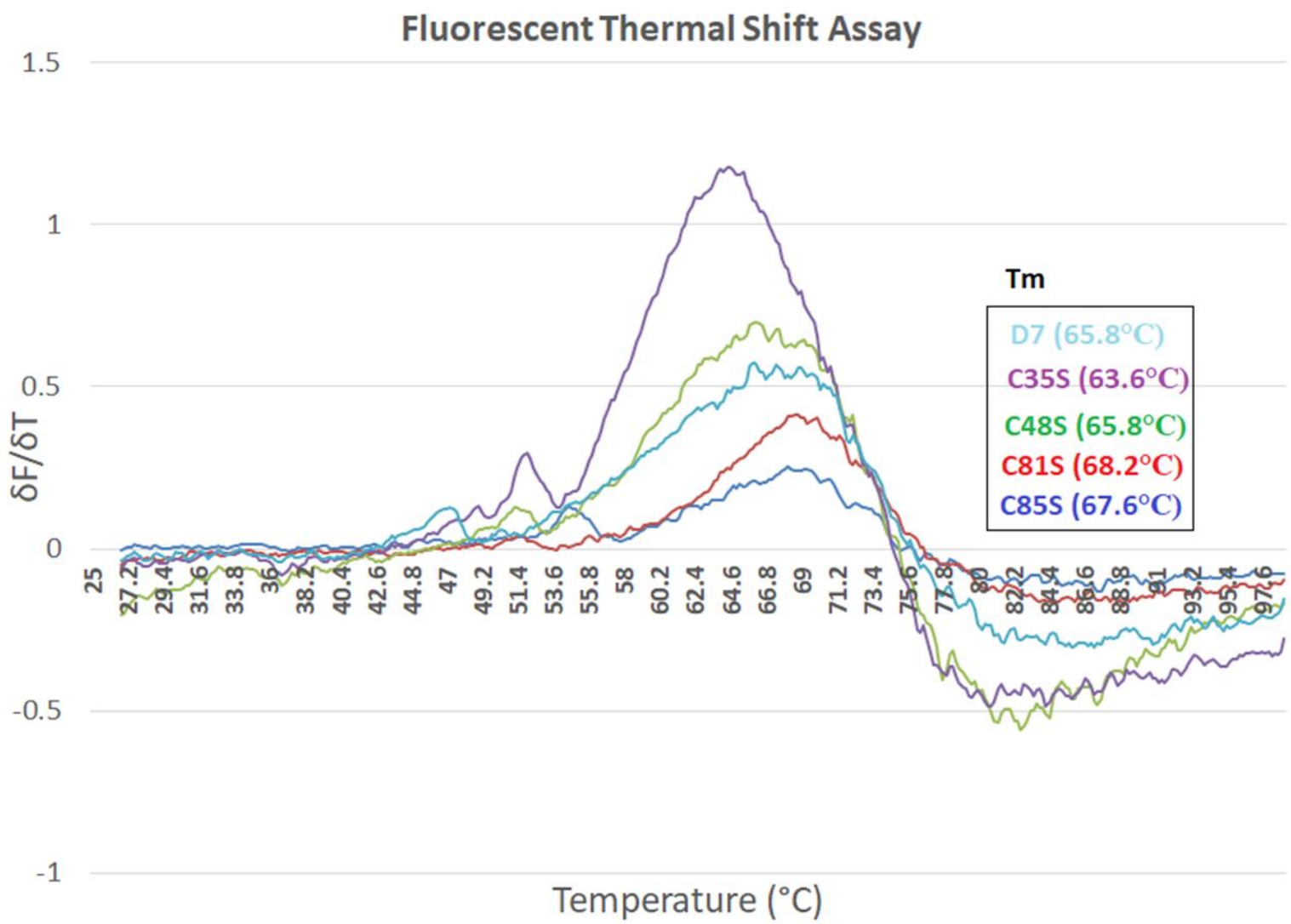
Figure 3.23 SDS-PAGE image of D7 (β-sheet) alanine mutants

16.5% SDS-PAGE image showing (SEC purified) D7 mutants c35a and c48a with protein ladder (left) showing possible sizes of proteins. The black arrows refer to established D7 monomer and dimer sizes. Each lane in SDS-PAGE corresponds to different elution fractions per D7 mutant (7.1.3.1, 10 mM BME).

Replacement of remaining cysteines (L4) with serines was pursued as a last attempt to eliminate oligomerisation but the binding affinity for P4 was compromised as assessed by ELISA (3.2.6).

3.2.5 Functional characterisation of D7 mutants: *FTSA*

Protein stability of D7 and the cysteine mutants was assessed using Fluorescent Thermal Shift Assays (FTSA) (Boivin et al., 2013). D7 and the mutants were subjected to thermal shift assays using SYPRO-orange dye in Rotor-gene Q-PCR machine. The principle of a thermal shift assay is that when the protein is heated to high temperatures, it unfolds and the dye binds to the hydrophobic residues and fluoresces (Niesen et al., 2007). The melting temperature (T_m) which is the temperature at which half of the protein completely unfolds (reflected by midpoint of maximum fluorescence) is used to gauge the stability of a protein. This is easily visualized by δF/δT which is the change in fluorescence (δF) with change in temperature (δT) (Bohling et al., 1999). Thermal shifts showed that all D7 mutants were stable, reflected by T_m values similar to that of D7 (65.8°C, Figure 3.24) (Gandham et al., 2015).



The δF/δT (Y-axis) during a thermal shift assay at increasing temperatures (X-axis) has been shown for D7 and mutants (c35s, c48s, c81s and c85s). The T_m for all four mutants are similar (63°- 68°C) to that of OBody D7 reflecting their stability. Replicates were not used. However, this experiment was repeated independently at least two times (7.3).

Figure 3.24 Fluorescent Thermal Shift Assays of D7 and mutants

3.2.6 Functional characterisation of D7 mutants: *ELISA*

Binding affinity of D7 and mutants were assessed using Enzyme Linked Immunosorbent Assay (ELISA) (Friguet et al., 1985, 1995; Hardy et al., 1997; Isobe & Nakao, 2003). To establish the sensitivity and specificity of D7 binding to P4, an ELISA (Bodmer et al., 1989) was performed with immobilized biotinylated P4 (on a streptavidin 96 well microplate) using D7 in different pH buffers (pH 4-7), and the specificity of D7 (to P4) was tested with the same buffer range but with immobilized biotinylated E2.

3.2.6.1 Binding of D7 to P4

The binding of D7 to P4 varied along pH 4-7 in a linear fashion assessed by equilibrium binding curves for different pH values (Figure 3.25). The binding curve (for a certain pH value) was generated by incubating varying concentrations of protein (D7) with constant concentrations of the ligand (P4) and the concentration of D7 in complex with P4 is directly reflected by the absorbance at 450nm (Abs450) from the ELISA colorimetric reaction (Pollard, 2010). The equilibrium binding curve (for a certain pH) is then used to calculate the equilibrium dissociation constant (K_d) which is reflective of the affinity of protein for the ligand (at that certain pH). The lower the K_d (nM) the higher the binding affinity between protein and ligand (Heinrich, Tissot, Hartmann, & Cohen, 2010). This is seen in increasing affinity of D7 to P4 with decrease in pH; K_d = 133 nM (pH 4), 145 nM (pH 5), 312 nM (pH 6) and 405 nM (pH 7). There was complete loss of D7 binding to P4 at pH 9. The calculated parameters for non-linear regression model of the experimental data is summarized in Table 1.

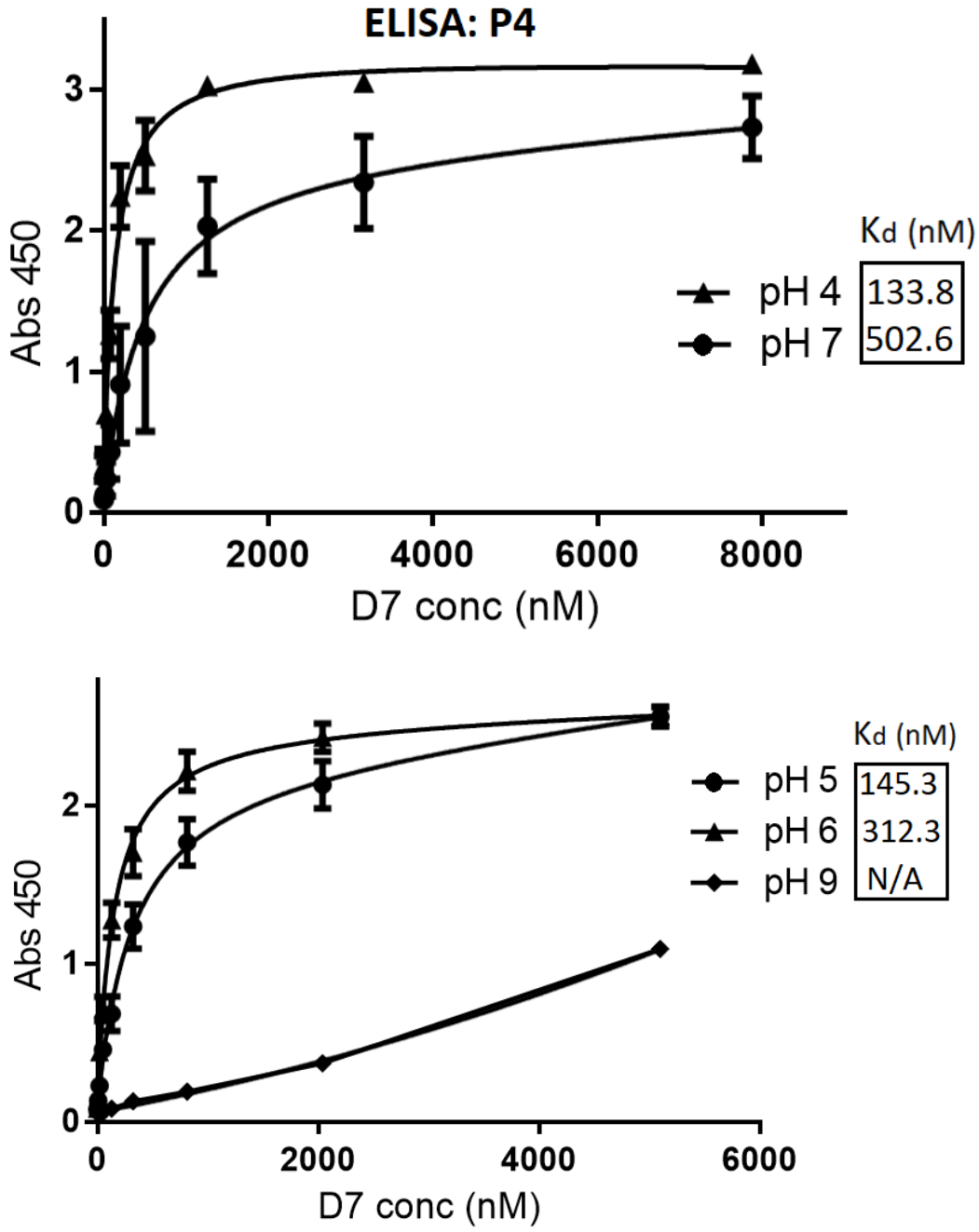


Figure 3.25 Equilibrium binding curves of D7 to P4 from pH 4 to 9

The binding curves at equilibrium from ELISA of D7 with immobilized P4 is shown for different pH with their calculated binding affinity (K_d) using Graphpad Prism for non-linear regression (right). Absorbance at 450nm (Y-axis) and D7 concentration in nM (X-axis) for the binding curves from ELISA colorimetric reaction is shown. Raw ELISA data is provided in appendix (7.4.1).

Table 3.1 ELISA of OBody D7 at different pH

The table shows Graphpad prism generated values for non- linear regression from ELISA data for OBody D7 binding to P4 at different pH. The standard error and 95% confidence intervals (nM) refer to the binding affinity (K_d). R square is an important parameter indicating goodness of fit of the binding curve to experimental data. The values closer to 1 indicate a good fit.

Non-linear fit	**pH 4**	**pH 5**	**pH 6**	**pH 7**	**pH 9**
Bmax (Abs450)	3.11	2.48	2.21	2.54	Very low
***Kd* (nM)**	133.8	145.3	312.3	502.6	No binding
Std. Error (nM)	27.39	24.36	45.95	214.1	N/A
95% conf. int (nM)	77.52 – 190.1	95.21 – 195.4	217.8 – 406.7	62.44 – 942.7	N/A
R square	0.97	0.98	0.99	0.93	N/A

3.2.6.2 Specificity testing of D7 to P4 and E2

The specificity of D7 to P4 was tested by replacing P4 with a similar small molecule E2 under the same pH conditions. At lower pH values (pH 4) there was a loss of specificity as D7 bound to both immobilized P4 and E2 but at a higher pH (pH 7) specificity was restored to P4 (Figure 3.25) and binding to E2 was lost (Figure 3.26). Binding of D7 irrespective of the ligands was strongest at lower pH values.

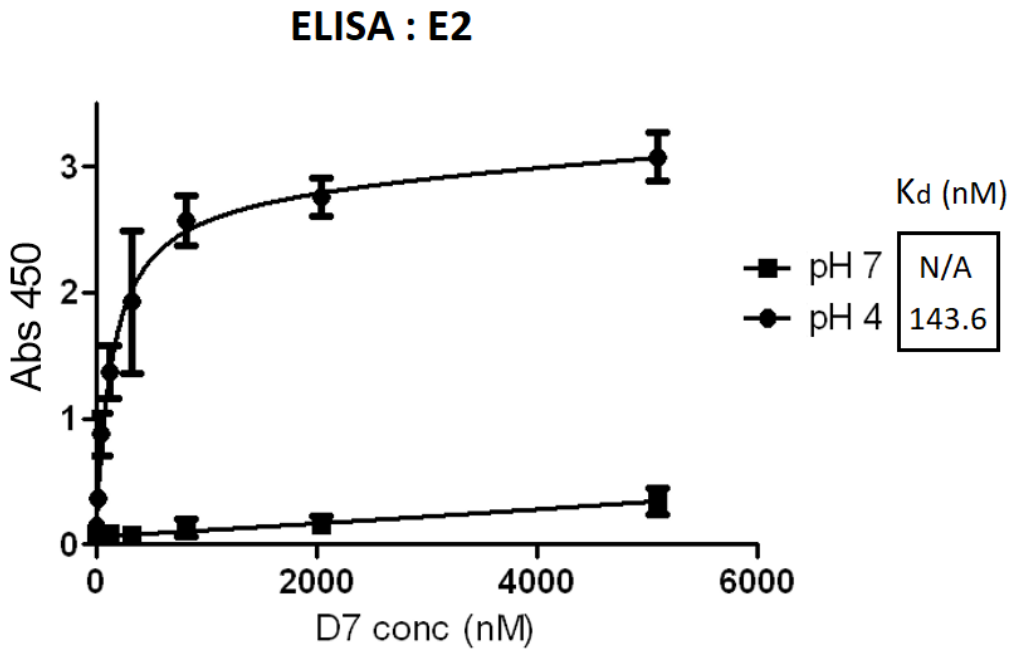


Figure 3.26 Equilibrium binding curves of D7 to E2

The binding curves (per pH) at equilibrium from ELISA of D7 with immobilized E2 is shown at pH 4 and pH 7 with the calculated binding affinity (K_d) using Graphpad Prism for non-linear regression (right). Absorbance at 450nm (Y-axis) and D7 concentration in nM (X-axis) for the binding curves from ELISA colorimetric reaction is shown (left) and K_d

values (right) are 143.6 nM (pH 4) for E2. Note loss* of D7 binding affinity to E2 at pH 7. Raw ELISA data is provided in appendix (7.4.2). * The binding curve for D7 binding to E2 shows no saturation.

P4 and E2 are predominantly hydrophobic molecules and it is highly likely that OBody D7 has predominantly hydrophobic interactions with P4 and E2. This is observed by strong binding of D7 to both P4 and E2 at pH 4. However, at pH 7 binding of P4 to D7 is maintained but binding of E2 to D7 is lost. This is possibly due to the loss of a favourable electrostatic interaction with increase in pH of buffer, possibly due to the titration of an acidic or basic amino acid. The possible acidic amino acids for the range pH 4 to pH 7 are aspartic acid and glutamic acid and one basic amino acid histidine. The only non- hydrophobic parts of P4 and E2 are the different polar groups at their end positions (3.1.1). The binding of P4 and E2 to D7 was tested with salt (500 mM NaCl) at pH 4 and pH 7 to shield charged interactions between these small molecules and OBody D7 (Figure 3.27). The binding of P4 to D7 from pH 4 to pH 7 is minimally disrupted by salt (500 mM NaCl, Figure 3.27) as expected (3.2.6.1) unlike loss of binding of E2 to D7 at pH 4 (E2 does not bind to D7 at pH 7) confirming that there is a significant charged interaction involved in the binding of E2 to D7. It is important to note that at pH 4, the acidic amino acids are not charged and histidine is positively charged (pKa of histidine side chain = 6.04). This suggests that the involved amino acid is histidine due to salt disruption of D7 binding to E2 at pH 4 (Jifeng Zhang 2012).

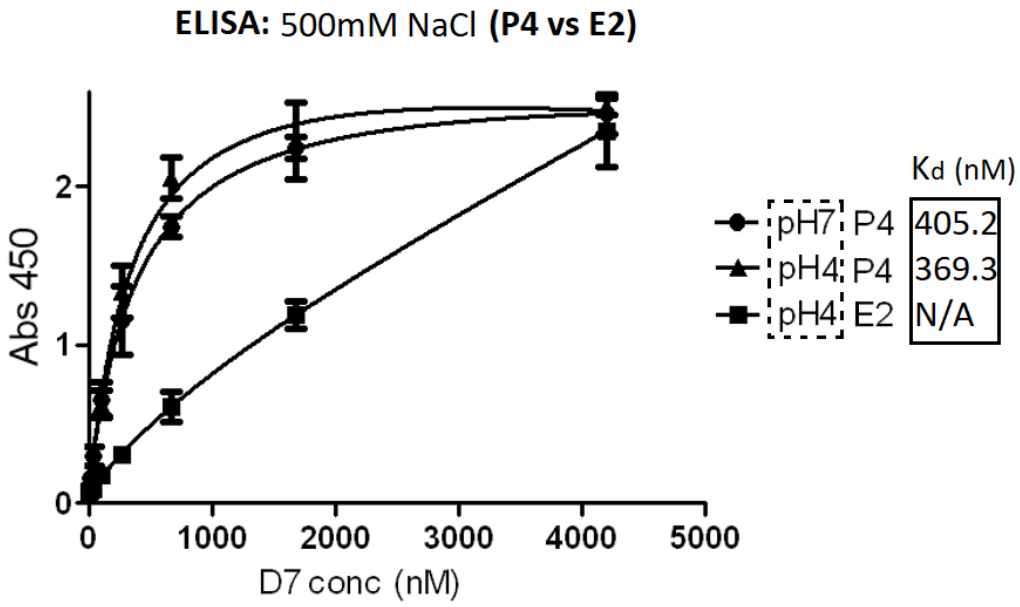


Figure 3.27 Equilibrium binding curves of D7 to P4 and E2 with high salt

The binding curve (per pH) at equilibrium from ELISA of D7 with immobilized P4 and E2 are shown at pH 7 + 500 mM NaCl and pH 4 + 500 mM NaCl (left) and calculated binding affinity (K_d) using Graphpad Prism for non-linear regression are shown (right). Absorbance at 450nm (Y-axis) and D7 concentration in nM (X-axis) for the binding curves from ELISA colorimetric reaction is shown (left) and K_d values (right) show 369.3 nM (pH4 + 500 mM NaCl) and 405.2 nM (pH 7 + 500 mM NaCl) for P4 and no* binding (pH 4 + 500 mM NaCl) for E2. Raw ELISA data is provided in appendix (7.4.2). * The binding curve for D7 binding to E2 shows no saturation.

My hypothesis was confirmed by D7 binding to P4 being relatively unaffected (pH 4 + 500 mM NaCl, pH 7 + 500 mM NaCl) (Figure 3.27) whereas loss of D7 affinity to E2 was observed at pH with highest binding affinity (pH 4 + 500 mM NaCl). No ELISA was done for D7 binding to E2 at pH 7 since there is no affinity between them. The calculated parameters for non-linear regression model of the experimental data is summarized in Table 2.

Table 3.2 ELISA of OBody D7 with different targets

The table shows Graphpad prism generated values for non- linear regression from ELISA data for OBody D7 binding to P4 and E2 at different pH and salt (500 mM NaCl). The standard error and 95% confidence intervals (nM) refer to the binding affinity (K_d). R square is an important parameter indicating goodness of fit of the binding curve to experimental data. The values closer to 1 indicate a good fit.

Non-linear fit	**Progesterone (P4)**				**Estradiol (E2)**		
	pH 4	**pH 4 + 500 mM NaCl**	**pH 7**	**pH 7 + 500 mM NaCl**	**pH 4**	**pH 7**	**pH 4 + 500 mM NaCl**
Bmax (Abs450)	3.11	3.09	2.54	2.76	2.83	Very low	0.78
***Kd* (nM)**	133.8	369.3	502.6	405.2	143.6	Not binding	Low binding
Std. Error (nM)	27.39	65.98	214.1	57.02	31.77	N/A	N/A
95% conf. int (nM)	77.52 – 190.1	233.7 – 505	62.44 – 942.7	288 – 522.5	78.28 – 208.9	N/A	N/A
R square	0.97	0.99	0.93	0.99	0.97	0.82	0.99

3.2.6.3 ELISA with D7 mutants

Therefore pH 7 was established as the buffer standard for determining the K_d of D7 mutants since it had a good balance of sensitivity and specificity to P4. β sheet mutations (c35s and c48s) retained affinity to P4 whereas L4 mutant (c81s) lost affinity to P4, which was shown by ELISA (Figure 3.28) with K_d = 394 nM (c35s) and 1 μM (c48s).

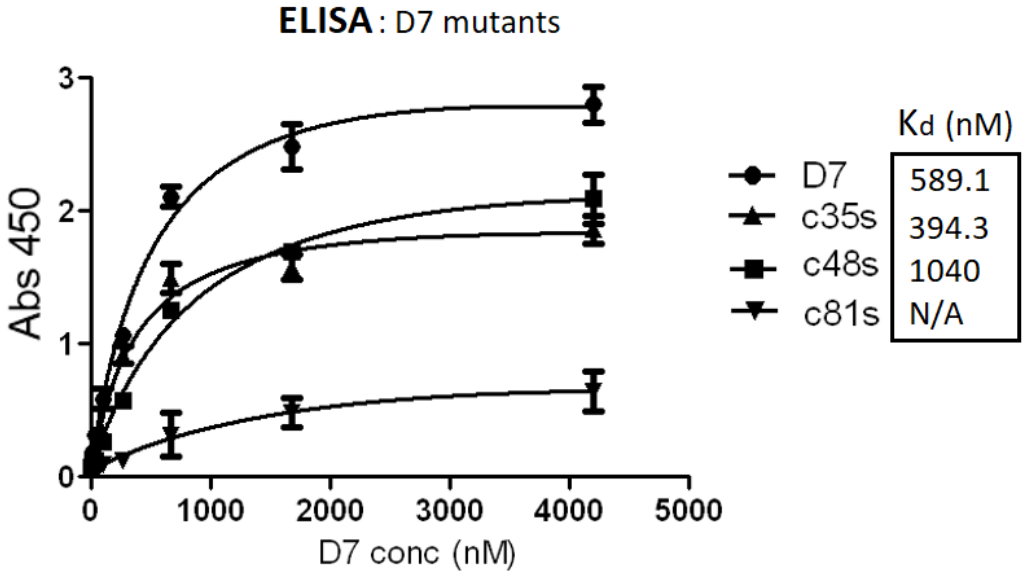


Figure 3.28 Equilibrium binding curves of D7 mutants to P4

The binding curves at equilibrium from ELISA of D7 and mutants (C35S, C48S and C81S) with immobilized P4 at pH 7 is shown (left) and calculated binding affinity *(K_d)* using Graphpad Prism for non-linear regression are shown (right). Absorbance at 450nm (Y-axis) and D7 or mutant concentration in nM (X-axis) for the binding curves from ELISA colorimetric reaction is shown (left) and K_d values (right) show no* affinity with loop (L4) mutant (c81s). Raw ELISA data is provided in appendix (7.4.3). * The binding curve for c81s binding to P4 shows no saturation.

Interestingly, these experiments were conducted with alanine mutants (C35A and C48A) in a 'reverse' system where constant concentration of D7 mutants were immobilized on the 96 well microplate and varying concentrations of (biotinylated) P4 was used to produce colorimetric reaction (Figure 3.29) (Bodmer et al., 1989). A double mutant of the same cysteines (DM, C35S: C48S) was also compared to D7 mutants with the same system. Both ELISA systems confirmed the finding that the loop (L4, C81S) was critical for binding of D7 to P4 and extensive modification of the β- sheet cysteines (DM, serine and alanine mutants) retained binding to P4.

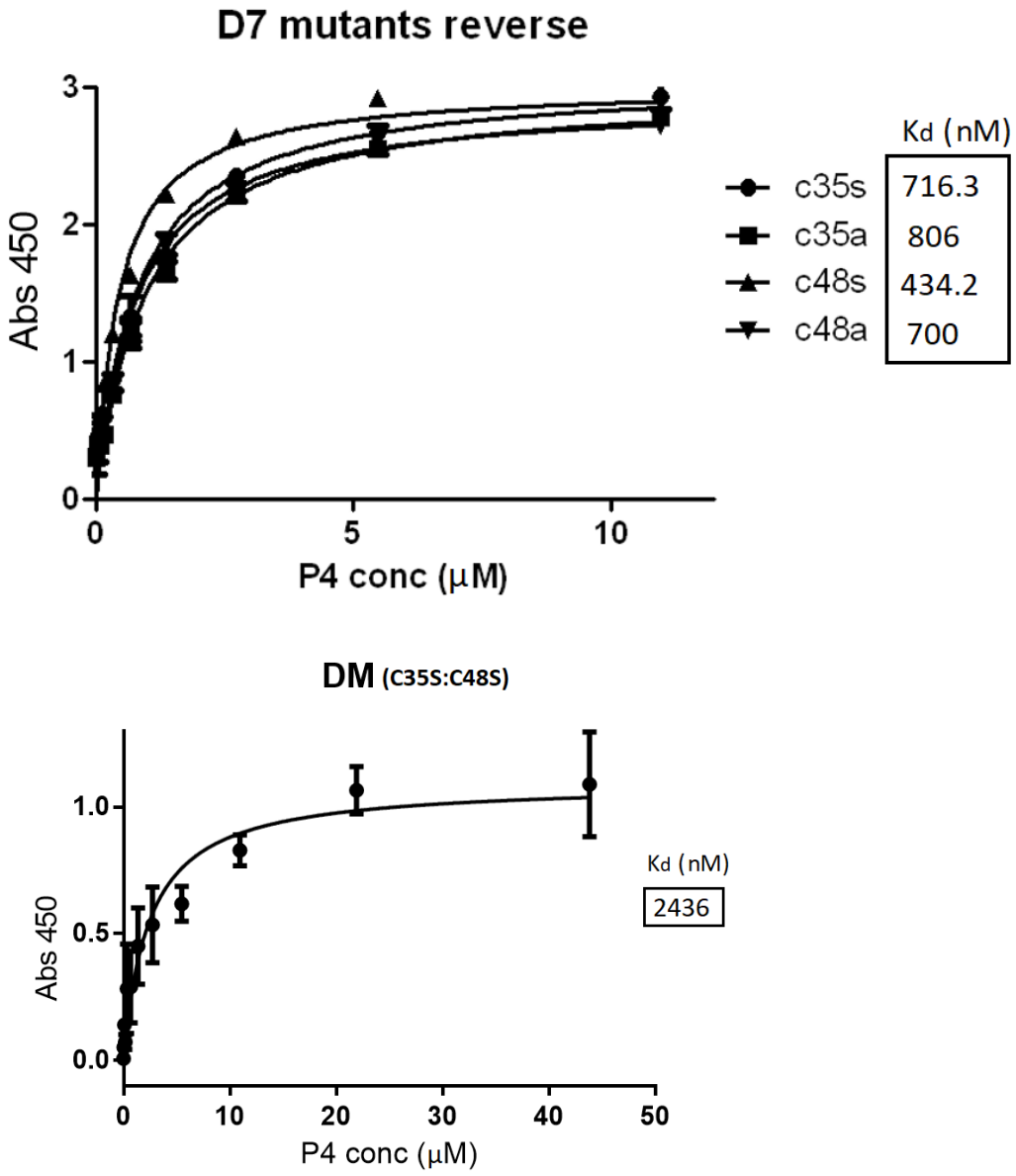


Figure 3.29 Equilibrium binding curves of D7 mutants to P4 (reverse ELISA)

The binding curves at equilibrium from ELISA of immobilized D7 mutants (C35A, C35S, C48A, C48S and DM) with increasing concentrations of P4 at pH 7 is shown with legend (top land bottom) and calculated binding affinity (K_d) using Graphpad Prism for non-linear regression are shown (box). Absorbance at 450nm (Y-axis) and P4 concentration in µM (X-axis) for the binding curves from ELISA colorimetric reaction are shown (top and bottom). Raw ELISA data is provided in appendix (7.4.3). The calculated parameters for non-linear regression model of the experimental data is summarized in Table 3.

Table 3.3 ELISA of OBody D7 mutants to P4

The table shows Graphpad prism generated values for non- linear regression from ELISA data for OBody D7 or mutant binding to P4. The standard error and 95% confidence intervals (nM) refer to the binding affinity (K_d). R square is an important parameter indicating goodness of fit of the binding curve to experimental data. The values closer to 1 indicate a good fit. The values in parenthesis refer to the reverse ELISA statistics.

Non-linear fit	**D7**	**C35S**	**C48S**	**C81S**	**DM**	**C35A**	**C48A**
Bmax (Abs450)	3.66	2.15 (2.92)	2.96 (2.95)	1.52	(1.09)	(2.78)	(2.79)
***Kd* (nM)**	589.1	394.3 (716.3)	1040 (434.2)	low binding	(2436)	(806)	(700)
Std. Error (nM)	98.04	85.76 (81.72)	240.8 (34.62)	N/A	(499)	(91.34)	(64.92)
95% conf. int (nM)	387.5 – 790.6	218 – 570.7 (549.6 – 882.9)	544.9 – 1535 (363.6 – 504.8)	N/A	(1422 – 3450)	(619.7 – 992.2)	(567.7 – 832.6)
R square	0.99	0.98 (0.9584)	0.99 (0.98)	0.90	(0.9)	(0.96)	(0.97)

3.2.6.4 Comparison with anti-P4 antibody (Ab)

An ELISA was designed where varying concentrations of a commercial anti-P4 antibody (Ab) was used instead of an OBody (D7), with a constant concentration of immobilized P4 to validate the ELISA system and compare an anti-P4 antibody (Figure 3.30) with the anti-P4 OBody (D7) at a favourable pH for the antibody (pH 7) (Reverberi & Reverberi, 2007). The Ab had a K_d (=59 nM) two-fold stronger than the best value obtained for D7 at low pH (K_d=133.8 nM at pH 4, Figure 3.25).

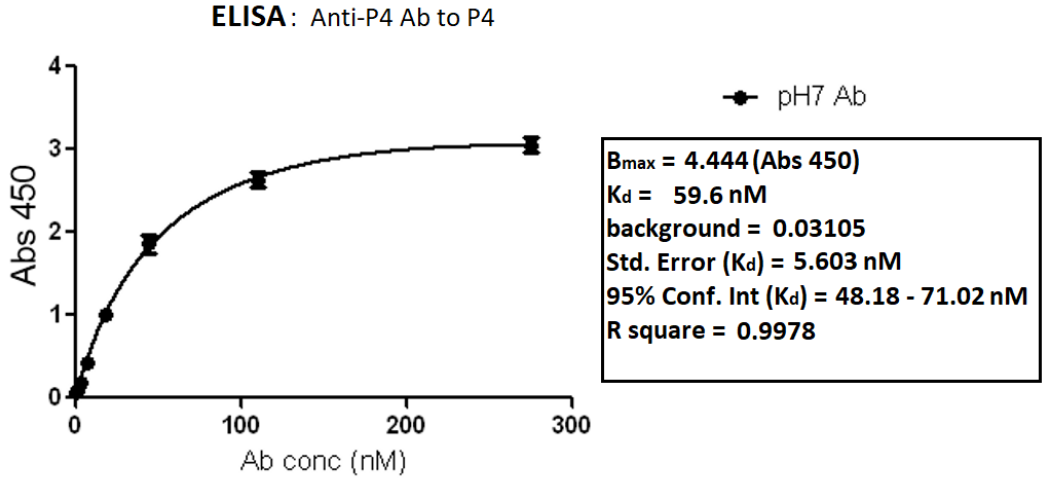


Figure 3.30 Equilibrium binding curve of anti-P4 antibody (Ab) to P4

The binding curve at equilibrium from ELISA of Ab with immobilized P4 at pH 7 is shown with legend (left) and calculated parameters using Graphpad Prism for non-linear regression are shown (right). Absorbance at 450nm (Y-axis) and Ab concentration in nM (X-axis) for the binding curve from ELISA colorimetric reaction is shown (left) and K_d is 59.6 nM (right). Raw ELISA data is provided in appendix (7.4.4).

3.2.7 MBP fusion to D7 for crystallization

The next strategy for crystallization was fusion of D7 to affinity tags such as MBP which is a 40 kDa protein (Smyth, Mrozkiewicz, McGrath, Listwan, & Kobe, 2003) and has been successful in crystallizing difficult proteins in the past (Costa, Almeida, Castro, & Domingues, 2014; Hayhurst, 2000; Kapust & Waugh, 1999; Moon, Mueller, Zhong, & Pedersen, 2010; Nallamsetty & Waugh, 2006; Waugh, 2016). Therefore the pOPIN system was employed for fusing MBP to D7 (7.2.3) with the goal of increasing solubility further and improving crystallization properties (Berrow et al., 2007). The molecular weight (MW) of the hybrid

protein was calculated to be ~54 kDa (MBP-D7, including non-cleavable histag). However, the SDS-PAGE gel image showed the protein band at a lower level at 50 kDa (Figure 3.31, B) but a second band was seen at double the size (~100 kDa) indicating the (intermolecular disulfide based) dimeric nature of D7. MBP without fusion OBody showed a similar lower band midway between 37 and 50 kDa (Figure 3.31, C) confirming that the success of fusion of MBP and Obody D7. MBP-D7 was purified first using IMAC (Histrap, Figure 3.31) and then by SEC (S75 16/600, Figure 3.32) similar to previously employed protein purification protocol, the latter indicating the oligomerisation nature of D7 (third band at ~ 200 kDa). Crystallization trials did not yield any success and MBP-D7 precipitated at high concentrations similar to its non-MBP fusion counterpart.

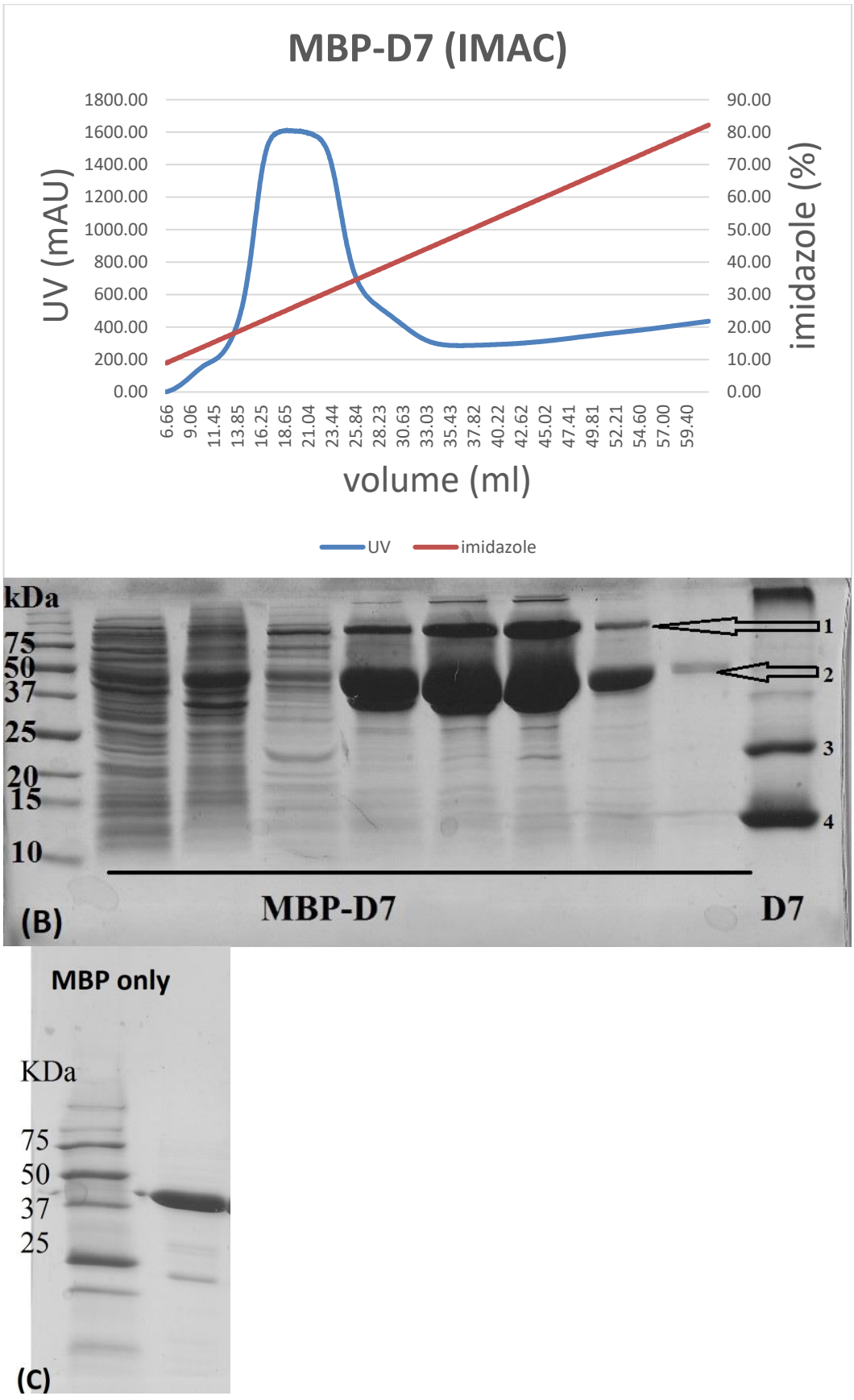


Figure 3.31 UV elution profile and SDS-PAGE: IMAC purified MBP-D7

The top image shows the UV absorbance curve (blue, 280nm) in mAU (left y-axis) showing the elution of MBP-D7 over an imidazole gradient (red) in percentage (right y-axis) during Histrap (IMAC) purification. The blue peak indicates maximum MBP-D7

elution. The x-axis refers to the flow of IMAC buffers (7.1.3.1, 10mM BME) containing imidazole. IMAC is explained in detail in 2.2.2.1. The middle (B) image is 16.5% SDS-PAGE showing IMAC fractions containing purified MBP-D7 with reference to a protein ladder (left) showing possible sizes of proteins and D7 as a second reference (right) with numbered protein sizes of interest. MBP-D7 (arrows) elutes as a monomer (54kDa, 2) and dimer (110kDA, 1) similar to its non-MBP counterpart D7 (monomer,12.7kDa, 4 and dimer, 25kDa, 3). The bottom image (C) shows only purified MBP similar to (B). Note that the lanes of MBP-D7 with strong dimer in SDS-PAGE (B) correspond to the elution fractions of UV peak in IMAC.

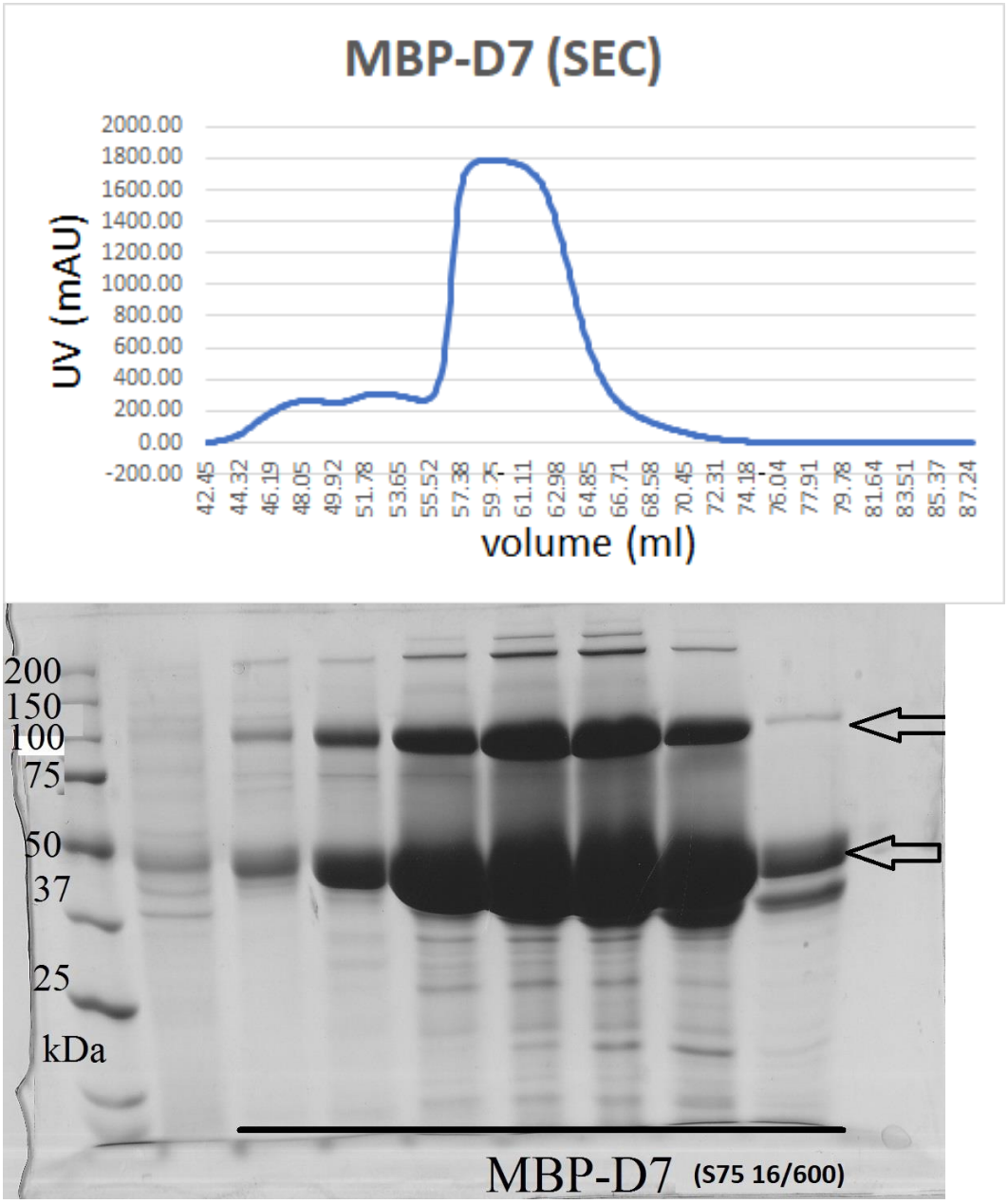


Figure 3.32 UV elution profile and SDS-PAGE: SEC purified MBP-D7

The elution profile of MBP-D7 (top image) on S75 16/600 column shows a single peak corresponding to the monomer (54 kDa). Y-axis is the intensity of absorbance by aromatic rings of protein at 280nm in milli-Absorbance Units (mAU). X-axis is the volume (ml) of buffer (7.1.3.1, 10 mM BME) the protein elutes at and is inversely proportional to the size (MW) of the protein in solution (Hong et al., 2012). A 16.5% SDS-PAGE (bottom image) is showing S75 16/600 (SEC) fractions containing purified MBP-D7 with reference

to a protein ladder (left) showing possible sizes of proteins. MBP-D7 still elutes as a monomer (54kDa) and dimer (110kDA) both indicated by arrows similar to its non-MBP fusion counterpart D7 (not shown here). Note a possible oligomer band at ~ 200 kDa. Note that the lanes in SDS-PAGE correspond to the elution fractions of UV peak in SEC.

3.3 Discussion

Structural characterisation of OBody (D7) with ligand progesterone (P4) involves growing protein crystals. One of the key requirements for crystallization is protein solubility (Gosavi et al., 2008) and monodispersity (Dale et al., 2003). Crystallization trials showed that D7 aggregated at high concentrations and was a mixture of monomers, dimers and possibly higher oligomers (as seen in SDS-PAGE). This is most likely due to mixed intermolecular disulphide bonds. This was interesting since the OBody D7 which was assembled in the oxidising periplasmic environment of *E.coli* TG1 (which favours formation of disulfides), expressed as a dimer in the reducing cytoplasmic expression strain of *E.coli* BL21. However, reduction of OBody D7 with reducing agents did not affect P4 binding.

Intra-molecular disulfide bonds have been known to be thermodynamically favourable for protein folding as it minimizes the search for lowest free energy conformation state (Chuang, Chen, Yang, Lyu, & Hwang, 2003). Disulfide bonds are only formed in the oxidizing environment of the periplasm in *E.coli* and their correct formation is catalyzed by two enzymes DsbA and DsbC oxidoreductases (Figure 3.33)(Trivedi et al., 2009). Proteins with disulfides are possible only if it is beneficial for proper folding, and are formed with the help of chaperone DsbA (Baneyx & Mujacic, 2004). Non-native disulfides are eliminated or repaired by chaperone DsbC (Qin, Wang, & Thirumalai, 2015).

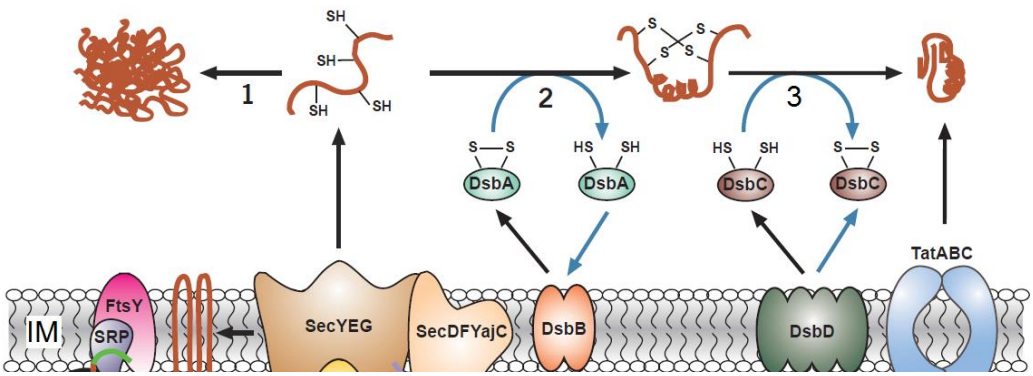


Figure 3.33 Protein folding in the periplasm

An illustration of proteins (brown thread like structure) translocating through the inner membrane (IM) into the periplasm via SecYEG (unfolded protein) and TatABC (folded protein) pathways (explained in detail in Figure 1.20) is shown. The unfolded protein (containing cysteines) enters the periplasm via SecYEG (Sec Pathway) and chooses the following routes: 1) form abnormal disulfides and aggregate 2) form normal disulfides with chaperone DsbA and 3) repair of abnormal disulfides with DsbC. DsbB and DsbD recharging the chaperones DsbA and DsbC are also shown (S- disulphide, SH- free cysteine) [adapted from (Baneyx & Mujacic, 2004)].

Since OB-folds are highly stable (Zhao, Schmitt, & Fisk, 2016) it is likely that disulfides in D7 (Obody) are inter-molecular (dimeric) in nature (Szilágyi, Kardos, Osváth, Barna, & Závodszky, 2007). This phenomenon is seen in GFP which is a stable protein domain but forms intermolecular (non-native and disulphide based) oligomers in the periplasm of *E. coli* (Aronson, Costantini, & Snapp, 2011; Dammeyer & Tinnefeld, 2012).

A conservative approach of *in vitro* reduction was tried first to improve solubility and monodispersity. This included reducing inter-molecular disulfides with DTT and BME (reducing agents). Addition of stabilizing ligand (P4) was also tried since the binding of a ligand may promote crystallization (Böhm, 2005; Celej, Montich, & Fidelio, 2003; Vodnik, Zager, Strukelj, & Lunder, 2011).

ELISA of OBody D7 at different pH values showed a decrease in binding affinity for P4 at higher pH and loss of specificity (E2) at lower pH possibly due to loss of positively charged interaction. This may be unfavourable with regards to binding affinity but favourable with regards to specificity to P4. There are at least six histidines in OBody D7 excluding the hexahistidine tag that may be responsible for this phenomenon.

When *in vitro* reduction did not yield satisfactory results, replacement of cysteines (mutations) and functional assessment (FTSA, ELISA) of their role in binding to P4 were chosen as the next strategy with the goal to improve crystallizability. OE-PCR has been successful in mutating cysteines (35, 48, 81 and 85) to serines in the OBody D7. This led to the assessment of contribution of cysteines to dimerization of D7 and its binding to P4.

Removal of all cysteines was not possible as this also deleted P4 binding. A protein tag which has been used as a crystallization chaperone – MBP was fused to D7 and concentrated. This did improve solubility dramatically as D7 was

concentrated up to 70mg/ml. The mono-dispersity was unaffected and did not produce any crystals.

The final strategy was to completely eliminate the cysteines but retaining the binding affinity to P4. An improvement in D7 affinity to P4 was also desired from its comparison with anti-P4 Ab affinity to P4. This can be achieved by a D7 affinity maturation (phage) library to randomize those regions which have cysteines i.e. β2 (c35), β3 (c48) and L4 (c81, c85) to select for cysteine deficient OBodies that bind to P4, by phage display (Clementi et al., 2012; Dufner, Jermutus, & Minter, 2006). The possible outcome from phage selections with this new library will be isolation of clones that are stronger (P4) binding and the OBodies that lack cysteines will improve chances of crystallization due to improvement in protein solubility (from lack of aberrant intermolecular disulfides) (Bondos & Bicknell, 2003) and monodispersity (due to elimination of oligomerisation) (Benvenuti & Mangani, 2007).

4 Structural and functional characterisation of improved anti-progesterone (P4) OBodies

4.1 Introduction

There are three techniques to engineer a high affinity binding protein: 1) rational mutagenesis based on design 2) computational protein design and 3) directed evolution (Tobin, Richards, Callender, & Wilson, 2014). Rational mutagenesis refers to using prior biochemical knowledge to improve the structure or function of a protein of interest (POI). Computational protein design uses molecular simulations based on existing knowledge of related proteins to design POI with improved structure or function. There is a lack of prior knowledge of binding between an OBody (D7) and the small molecule (P4) and no existing model to simulate a stronger P4 binding OBody. This leaves us with directed evolution as the optimum strategy to improve on the structure and binding function of this P4-binding OBody. Directed evolution is a random mutational approach with limited knowledge of the (binding surface of) POI to artificially select for high affinity binding proteins to specific targets (similar to natural selection). This can be achieved by random mutagenesis, focused mutagenesis or DNA recombination (Packer & Liu, 2015). The crystal structure of an OBody with its ligand P4 would enable detailed analysis of the binding interface and enhance optimisation of P4 binding OBodies.

4.1.1 Directed evolution with phage display

Phage display is a protein engineering technique which relies on making gene libraries of mutants of binding proteins and creating a phenotype-genotype link via packaging the gene in phage particles and displaying the translated protein on the surface of the phage (Binz et al., 2005; Carmen & Jermutus, 2002; Dufner et al., 2006; Hoogenboom, 2005; Packer & Liu, 2015; Rothe, Hosse, & Power, 2006; Sidhu & Koide, 2007). An artificial selective pressure is introduced to isolate mutants that bind to a specific target. These mutants can be recovered, amplified and subjected to multiple rounds of selection (or phage selection) to select for high affinity binding proteins to the target. When multiple rounds of phage selections no longer produce any further stronger binding proteins to the specific target (due to isolation of best possible mutants for the particular library gene design), the best phage clone representing the highest affinity-binding to the

target is sequenced and analysed for further improvement of protein – ligand binding.

4.1.2 Affinity maturation with focused mutagenesis and phage display

Affinity maturation (AM) is a two-step process which consists of somatic hypermutation of B-cell receptors (BCR, antibodies) and selection of these B-cells with target antigen during the immune response in animals (Weiser et al., 2011). This can be artificially replicated by focused mutagenesis via PCR for hypermutation and phage display for selection. The best mutant is used as a template for the construction of new gene libraries by various strategies of randomization based on previous understanding of the binding surface of the protein, ranging from random mutagenesis using error-prone PCR (McCullum, Williams, Zhang, & Chaput, 2010) to combinatorial mutagenesis focused on specific stretches of gene sequences (Hawkins, Russell, & Winter, 1992; Parker, Griswold, & Bailey-Kellogg, 2011; Steemson et al., 2014; Verma, Grigoryan, & Bailey-Kellogg, 2015) and the gene libraries are used for new rounds of phage selection to obtain stronger binding mutants to the target.

OBodies have been characterized to use a specific binding surface for binding (Arcus, 2002; Murzin, 1993; Steemson et al., 2014). The P4 binding OBody D7 has been found to have cysteines on this surface (β-sheets 2 and 3 and loop 4) and ELISA of cysteine mutants further indicate their involvement in OBody function (Table 3.3). Affinity maturation was planned for optimization of OBody D7 to improve its binding affinity to progesterone (P4) and eliminate cysteines to improve crystallizability.

4.1.3 Construction of pF and L4 based (AM) libraries

The OBody D7 was selected from a naïve library using directed evolution via phage display. D7 was mutated at the cysteines (C35, C48, C81 and C85) based on the prior knowledge of cysteines being responsible for intra or inter-molecular disulphide bonds as discussed in chapter 3. Functional characterization of D7 indicated that cysteines in loop 4 (L4) were critical in P4 binding but did not yield crystals possibly due to presence of cysteines and resulting intermolecular disulfides. The presence of cysteines in loop four (L4) may be a limiting factor in crystallization (due to heterogeneity from dimerisation or instability at high

concentrations). Alternatively, the cysteines on L4 may be crucial for P4 binding. Thus, construction of AM libraries by focused mutagenesis of loop (L4) and another sparing the loop but involving the β-sheets i.e. proximal binding face (pF), and selection via phage display were chosen as strategies to improve the binding affinity of D7 (Figure 3.14) to P4 and make it more amenable for crystallization.

The best mutant was a double cysteine to serine mutant (DM, C35S:C48S) where the altered cysteines occurred at positions on the β2 and β3 strands. This OBody (DM) was randomized at the loop (L4) for the construction of an AM library (L4_DM) with the goal of selecting for higher affinity P4-binding OBodies without any cysteines via phage display and to improve crystallizability. Since the cysteines in this loop may be critical for binding function, the parallel conservative library (pF_35) was designed by eliminating only the cysteines on the β- sheets by randomisation of β3 (pF, centered around C48) on another high affinity D7 mutant (C35S, 3.2.6.3) leaving the loop intact.

4.2 Results

4.2.1 Construction of AM library

4.2.1.1 Point mutations for construction of AM libraries

Affinity maturation (AM) libraries (pF_35 and L4_DM) were designed to improve the binding affinity and crystallizability of D7 (with P4) and is described in 2.6. The AM libraries were designed in steps. To prevent 'carry-over' of the original OBody (D7) in the AM libraries, stop codons were strategically inserted into positions 48 and 83 of D7 by OE-PCR to generate pF_stop and L4_stop respectively (Figure 4.1).

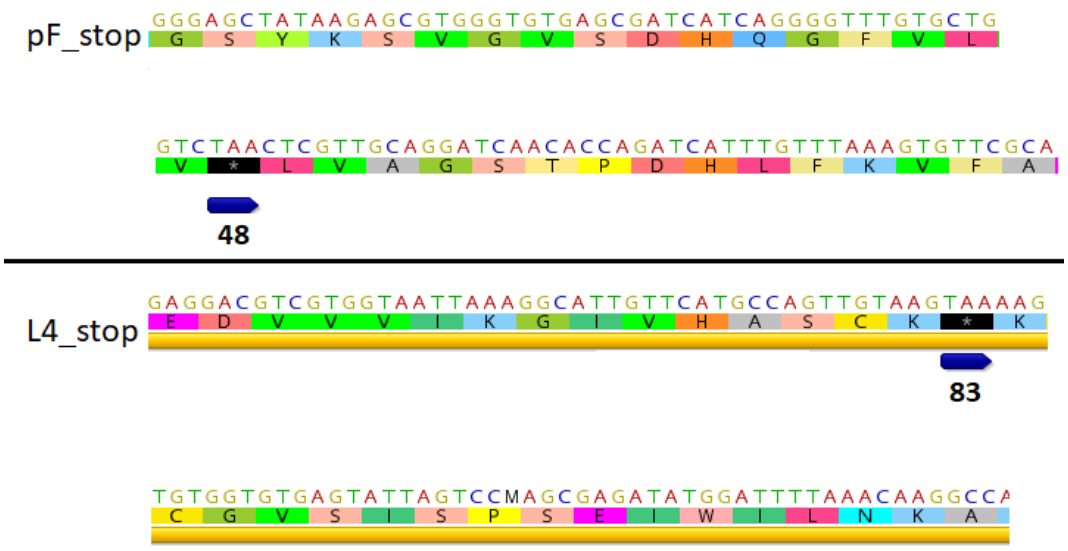


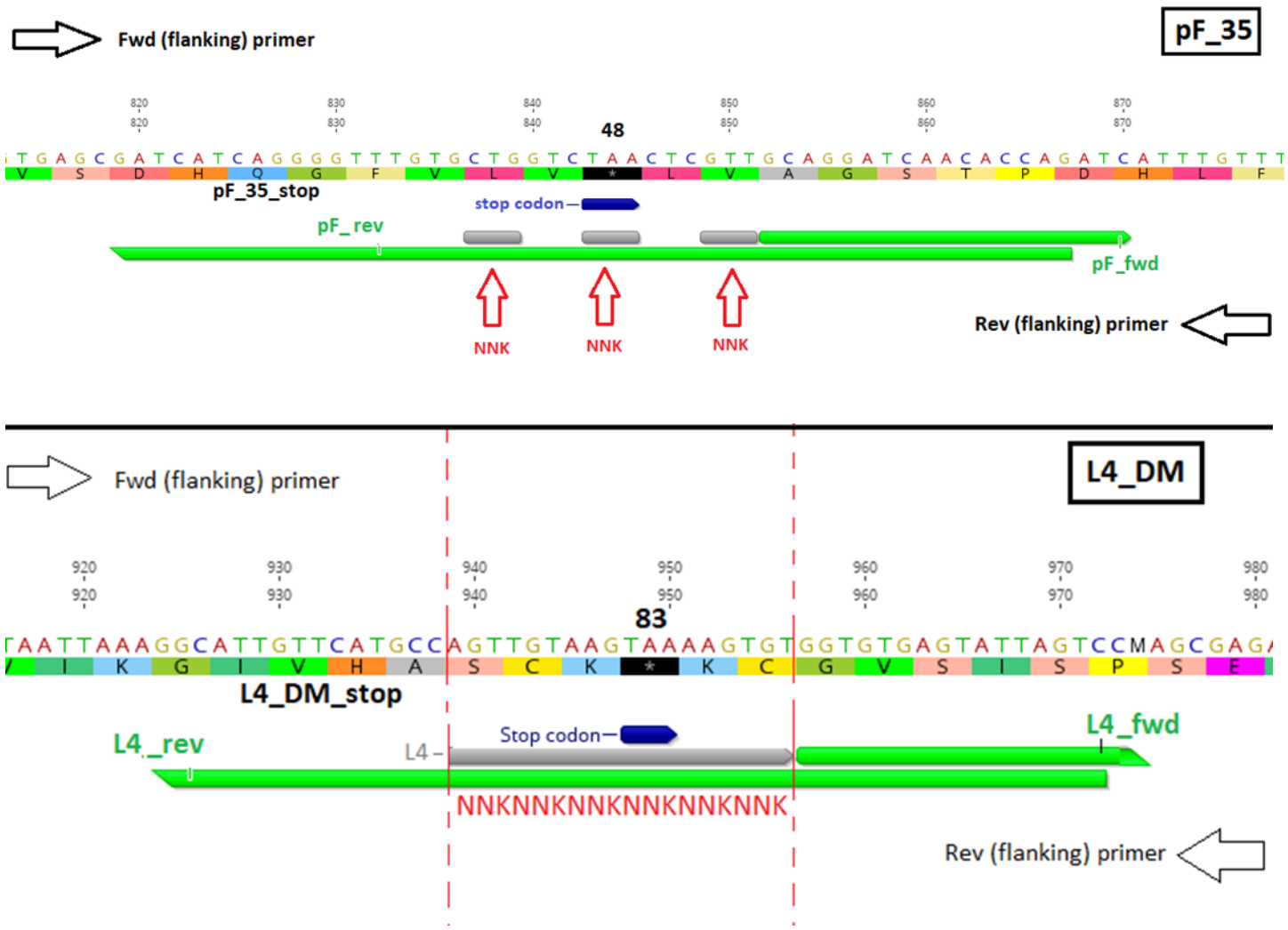
Figure 4.1 DNA sequencing of stop codon insertion into D7 for AM library construction

The DNA sequencing of fused overlap fragments of D7 containing stop codons (pF_stop and L4_stop) at position 48 and 83 respectively are shown. The labelled D7 stop mutants (left) show the stop codon (TAA) indicated by blue tab.

Point mutations were made in these mutants (pF_stop and L4_stop) using overlap primers for C35S thus producing pF_35_stop and L4_35_stop mutants respectively using OE-PCR. A point mutation was made in L4_35_stop using overlap primers for C48S thus producing L4_DM_stop mutant using OE-PCR.

4.2.1.2 Randomization of AM libraries

pF_35_stop and L4_DM_stop DNA sequences were used as templates to generate overlapping fragments using randomization primers for construction of the respective libraries. The primer sets used are pF_forward (or reverse) and L4_forward (or reverse) for the appropriate libraries (Figure 4.2). Only the reverse (rev) primers had NNK codons where N=G, A, T or C and K= G or T. This arrangement allows elimination of stop codons TAA and TGA in an amber (TAG) suppressed *E.coli* TG1 leading to the possibility of all 20 amino acids (Mena & Daugherty, 2005) in the randomized positions. The PCR trials consisted of temperature optimisation (50 °C to 57 °C) to generate randomized overlap fragments (pF_35_rand_A and L4_DM_rand_A). The annealing temperatures of the PCR which gave the best results was 54.6 °C to 56.4 °C for pF_35_rand_A and 50.1 °C to 50.8 °C for L4_DM_rand_A. These randomized overlap fragments were generated in large quantities to create diversity of DNA sequences.



The sequence of β-sheet 3 (top) and loop 4 (bottom) from pF_35 and L4_DM libraries respectively are shown as part of larger sequence of OBody D7 mutants (labelled). The complementary primers (green, for OE-PCR) show randomized areas (grey, NNK) also highlighted in red. The overlapping fragments are generated separately with Forward primers (left arrows, sequence not shown) and pF/L4 reverse (green) for forward fragments, and Reverse primers (right arrows, sequence not shown) and pF/L4 forward (green) for reverse fragments. The numbers 48 and 83 refer to the stop codon residues for pF_35 and L4_DM respectively shown as TAA (*). The extension step (OE-PCR) follows using only flanking primers (arrows, sequence not shown) to generate single randomized pF_35 and L4_DM fragments.

Figure 4.2 Sequence of randomisation scheme for β-sheet 3 (pF_35) and loop (L4_DM) on D7 stop mutants

The non-randomized overlap fragments (pF_35_B and L4_DM_B) were generated and fused with the randomized overlap fragments with OE-PCR and confirmed (Figure 4.3).

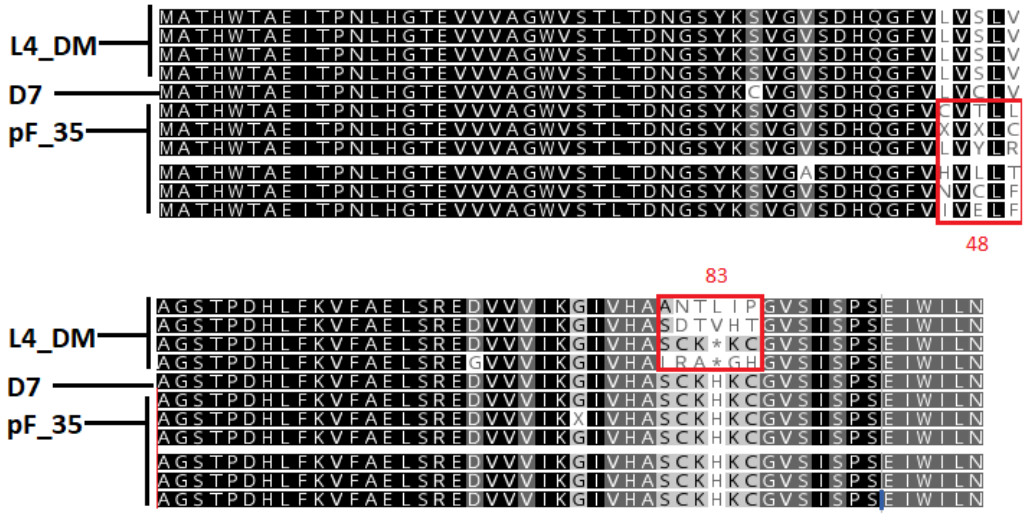


Figure 4.3 DNA sequencing of pF_35 and L4_DM

The DNA sequencing of fusion of the appropriate overlap fragments with OE-PCR are shown with randomized residues shown in red boxes with relation to numbered amino acid (E.g. 48) using naïve library OBody D7 as a reference.

These randomized fusion products (insert) were restriction enzyme digested with Sac2 and Not1-HF (NEB) as described in 2.5.6. The YscC phagemid (vector, from D7) was restriction enzyme digested with the same set of enzymes. The AM inserts and YscC phagemid were ligated as described in 2.5.7 and transformed into electro-competent *E. coli* TG1 cells as described in 2.5.8.

4.2.2 Phage selections of AM libraries

The transformed AM libraries (pF_35 and L4_DM) were pooled and grown as explained in 2.6.1.4. The purified rescued phage was precipitated using PEG (2.6.1.5) and used for bio-panning against target P4 as explained in 2.6.2. The library panning are summarized in Table 4. Hyperphage (M13K07δIII, which lacks wild type gIII) was used to rescue phage in early rounds of selections to enhance phage display of moderate affinity P4 binding OBodies (Rondot et al., 2001). The phagemid particles isolated from successive library panning rounds

were used for successive phage selections. Streptavidin and neutravidin beads were used in alternative selections (to isolate phage via biotinylated P4) to prevent selection of non-specific streptavidin binding OBodies (Vodnik et al., 2011). The antigen concentration and duration of selection was reduced from round 2 onwards to facilitate stringent selection conditions for high affinity P4 binding OBodies. There is low phage yield after round 4 which possibly indicates loss of P4 binding OBodies in subsequent library panning rounds (illustrated in Figure 4.4).

Table 4.1 summary of AM library affinity-screening

The table summarizes the conditions of phage selection rounds (R1 to R7). Phage from selections are used in successive rounds as shown under input library. The phage library volume after blocking with 3%MPBS (ml) and target antigen (biotinylated-P4, uM) is mentioned. The selection beads (streptavidin or neutravidin) to isolate phage via biotinylated-P4 and helper phage used for phage rescue are shown. The total input titre (cfu) and output titre (cfu/ml) shown are used to calculate the yield (%).

Selection round (R)	Input library	Library volume	Antigen biotin P4	Selection beads	Selection duration (hours)	Helper phage type	Total input titer(cfu)	Output titre (cfu/ml)	yield
R1	AM	0.3 ml	1 uM	Streptavidin	2	KM13	1.0E+12	1.9E+04	<0.01 %
R2	R1	1 ml	4 uM	Neutravidin	3.5	M13K07δIII	1.5E+11	5.4E+07	3.6%
R3	R2	1 ml	2.5 uM	Streptavidin	3	M13K07δIII	8.5E+11	1.2E+08	1.4%
R4	R3	1 ml	2.5 uM	Neutravidin	2	KM13	1.8E+13	2.4E+08	0.13%
R5	R4	1 ml	1.5 uM	Streptavidin	1.5	KM13	2.5E+13	1.5E+07	<0.01%
R6	R5	1 ml	1 uM	Neutravidin	1	KM13	4.1E+14	1.8E+07	<0.01%
R7	R6	1 ml	0.5 uM	Neutravidin	0.45	KM13	9.5E+14	1.8E+07	<0.01%

Figure 4.4 summarizes the output from phage selections in colony forming units (cfu/ml) and shows that there was enrichment of possible P4-binding OBodies until round 4 (R4) followed by a significant drop in phage output. This is in agreement with the selection pressure where the antigen concentration (µM) and duration of incubation (hrs) have been decreased in subsequent selection rounds and is reflected by most P4 binding OBodies isolated from R4 and R5 (Table 4.1) beyond which there was low recovery of P4 binding OBodies.

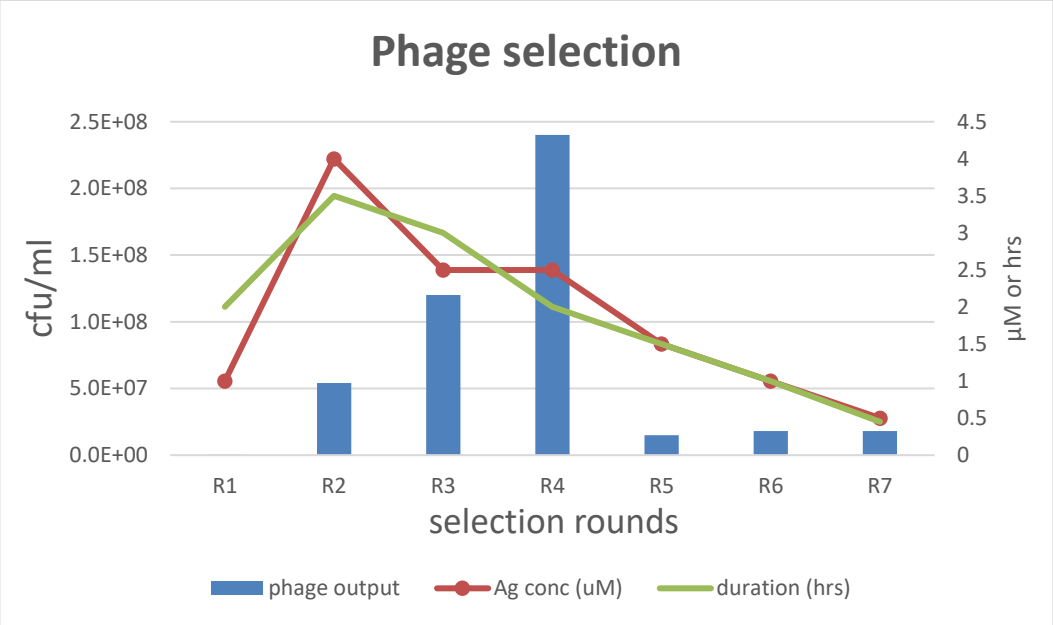


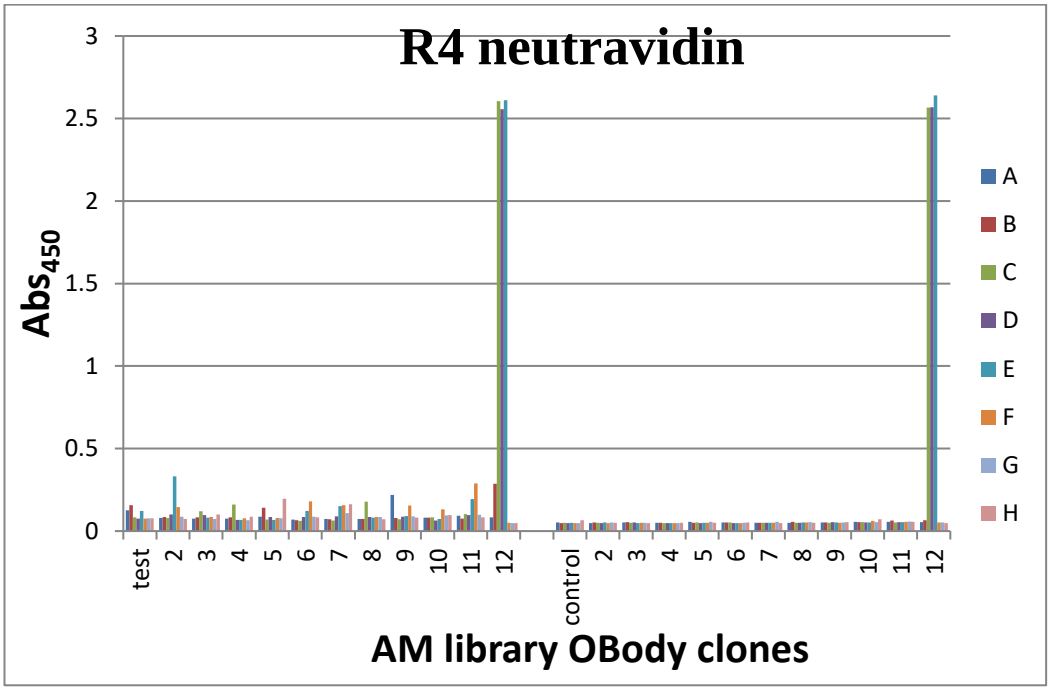
Figure 4.4 Bar graph showing phage selection of AM library

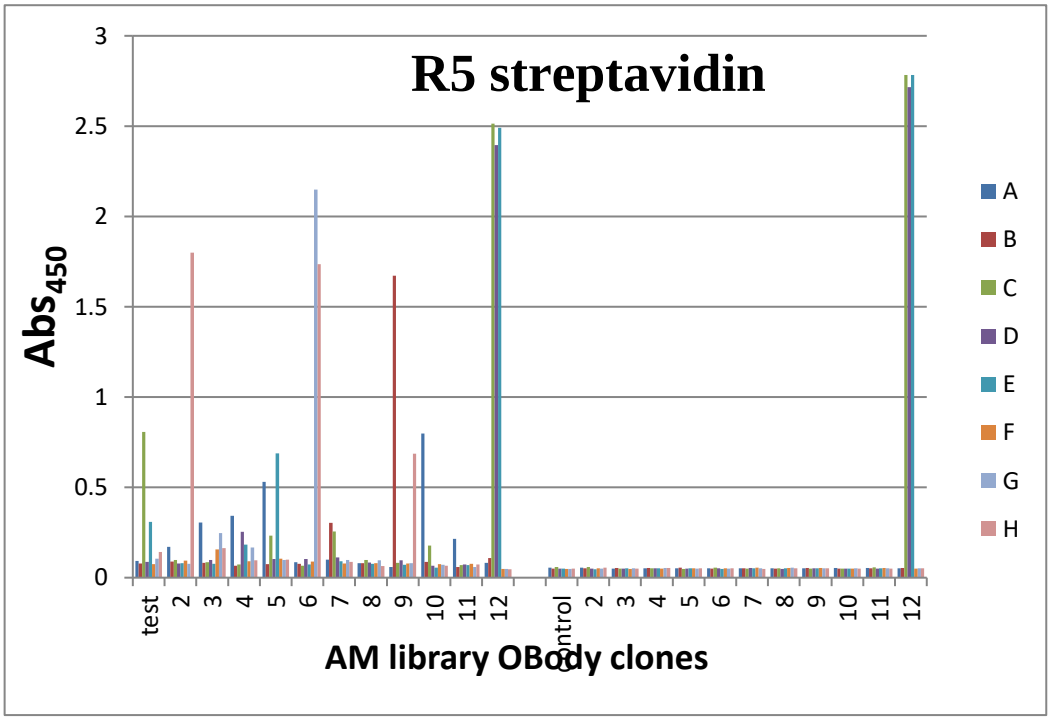
The phage selection rounds (R1-R7, x-axis) of OBodies to P4 are shown as bar charts with relation to phage output (cfu/ml, left x-axis). The right x- axis shows both antigen concentration (µM, red line) and duration of incubation (hrs, green line) with reference to selection rounds. The increase in phage output up to R4 indicates enrichment of possible P4 binding OBodies followed by a decrease in phage output indicating loss of possible P4 binding OBodies.

4.2.2.1 Phage ELISA and isolation of P4 binding OBodies

The phagemid particles from library panning rounds were clonally isolated in agar plates (C100, 2.6.2.4) and assayed by phage ELISA for binding to P4 as described in 2.6.3. The P4 binding OBodies were isolated from three rounds (R4, R5 and R7) as illustrated in Figure 4.5. However, R5 showed the highest number of Individual OBody clones with high signal in ELISA (~10 times higher than background) which indicates selection of specific P4 binding OBodies compared to round 4 with the highest phage output (Figure 4.4). It is important to note that most of the isolated OBody clones were non-specific in binding and the

magnitude of absorbance has no correlation with binding affinity (Bodmer et al., 1989).





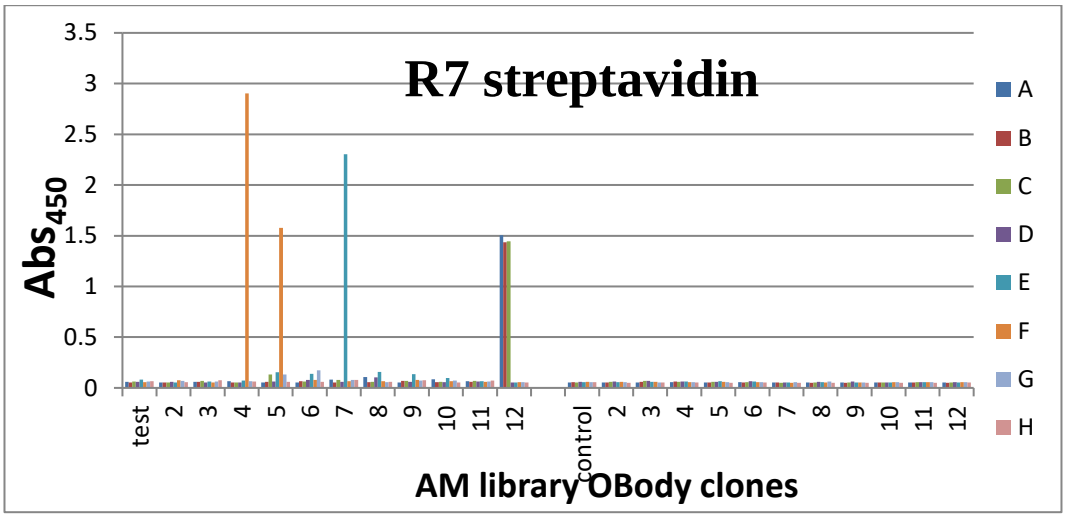


Figure 4.5 Phage ELISA of selections

An illustration of phage ELISA of labelled selection rounds showing Abs450nm (y-axis) and individual OBody clones (x-axis and legend). The individual OBody clones are tested for p4 binding in duplicates (one with P4 immobilized as test and the other with PBS as control). The individual OBody clones are in 96-well format (12x8). The first three wells in column 12 is the positive control (anti-P4 antibody). The positive P4 binding OBodies and positive control show high Abs450nm values.

The distinct P4 binding OBodies were confirmed by sequencing using flanking primers by PCR as shown in Figure 4.6 (7.2.7). These OBodies have been functionally characterised in 4.2.4.

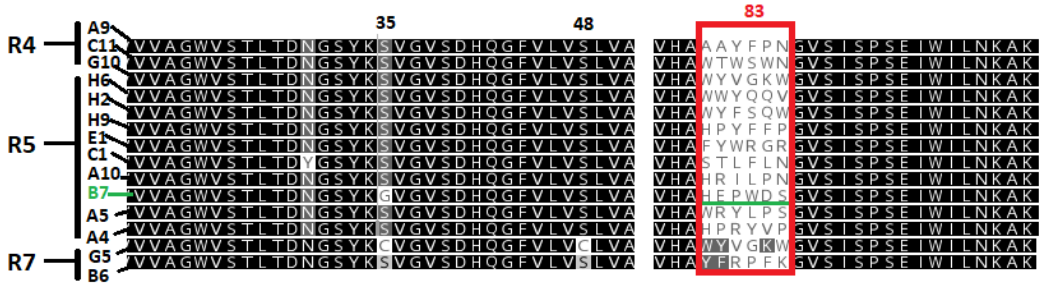


Figure 4.6 DNA sequencing of P4 binding OBodies from AM library

The DNA sequences of P4 binding OBody clones from round 4 (R4) to round 7 (R7) are shown (left label) with randomized loops (L4, red box) in relation to numbered amino acids (E.g. 83). The best P4 binding OBody B7 is highlighted in green.

4.2.3 Analysis of P4 binding OBodies

The strong P4-binding OBodies were screened by phage ELISA and confirmed by large scale expression and surface plasmon resonance (SPR) analysis. The P4 binding OBodies screened from phage ELISA mostly consisted of L4 randomized OBodies. The amino acid sequence alignment (Figure 4.7) of the selected OBodies from the AM libraries shows an L4 with both polar and hydrophobic

amino acids with preference towards hydrophobic amino acids especially tyrosine (Y) and tryptophan (W) at the periphery, and polar (or charged) amino acids especially glutamine (Q) at the center of L4. The OBody B7 is an exception as it has charged amino acids in the periphery i.e. histidine (H), glutamic acid (E) and aspartic acid (D), and hydrophobic amino acids i.e. tryptophan (W) and proline (P) in the center of L4. Overall, there is no consensus in amino acid sequence in L4 of selected OBodies from AM library.

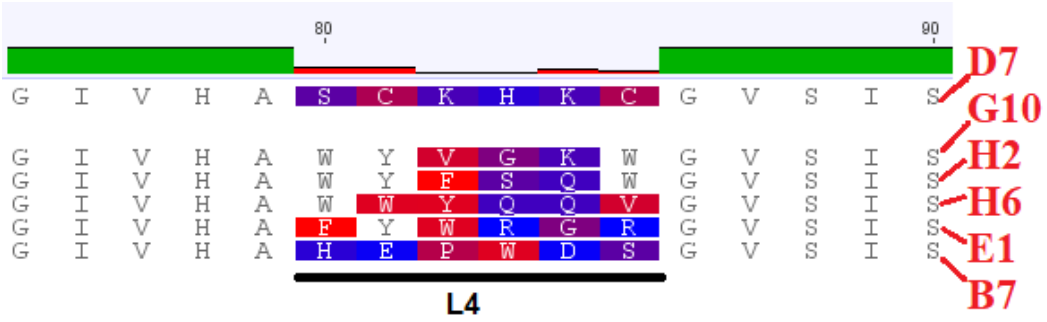


Figure 4.7 Amino acid alignment of P4 binding OBodies (AM library)

The amino acid sequence of D7 centered around L4 is shown in comparison to the P4 binding OBodies selected from AM library. The OBodies are labelled (right) in red. This alignment is generated using geneious software showing common amino acid sequences (green bar, non-coloured) around L4 (black line). The sequences in L4 are coloured based on lack of consensus (blue=polar or charged, red= hydrophobic).

The OBodies selected from AM library were sub-cloned into pPROEX-HTb and expressed in *E. coli* DH5α strain and purified as described in 3.2.1. These P4 binding OBodies (without cysteines) were purified with IMAC which confirmed the correlation between the presence of cysteines and dimerization (4.2.5). They purified only as a monomer unlike D7 (which contained cysteines and shows dimer and multimer bands on the SDS-PAGE gel).

4.2.4 Surface Plasmon Resonance (SPR) analysis of P4-binding OBodies

These OBodies were tested for binding affinity (K_d) using SPR due to the ease of use and high accuracy at the same time (Van Der Merwe, 2001). D7 and its mutant (DM) were analyzed using SPR to establish a standard (Figure 4.8) and their K_d values were within 5-fold of magnitude of ELISA results in Chapter 3 (Friguet et al., 1995; Hardy et al., 1997; Heinrich et al., 2010).

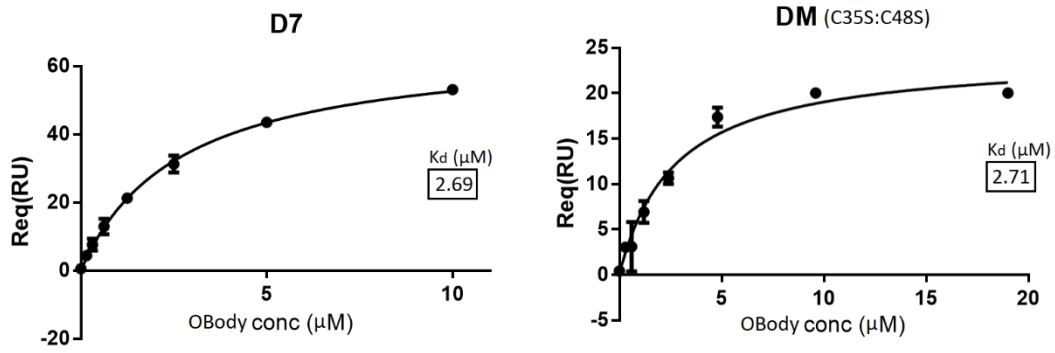


Figure 4.8 Equilibrium binding curves of naïve library OBodies (SPR)

The binding curves at equilibrium from SPR generated sensorgrams are shown for different concentrations of D7 or mutant (DM) binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (R_{eq}, in RU) and the x-axis shows the concentration of analyte (OBody, in μM). The relevant statistics are summarized in Table 4. SPR is explained in detail in 2.4.2 and raw data provided in appendix (7.5).

It is important to note that it is the difference in magnitude of R_{eq} (RU) at varying concentrations of analyte that is indicative of binding at equilibrium (not the absolute magnitude) similar to Abs 450nm of ELISA (Bodmer et al., 1989). The SPR analysis (Figure 4.9) showed that most of the OBodies from AM library had a K_d in the same range of D7 (in low μM).

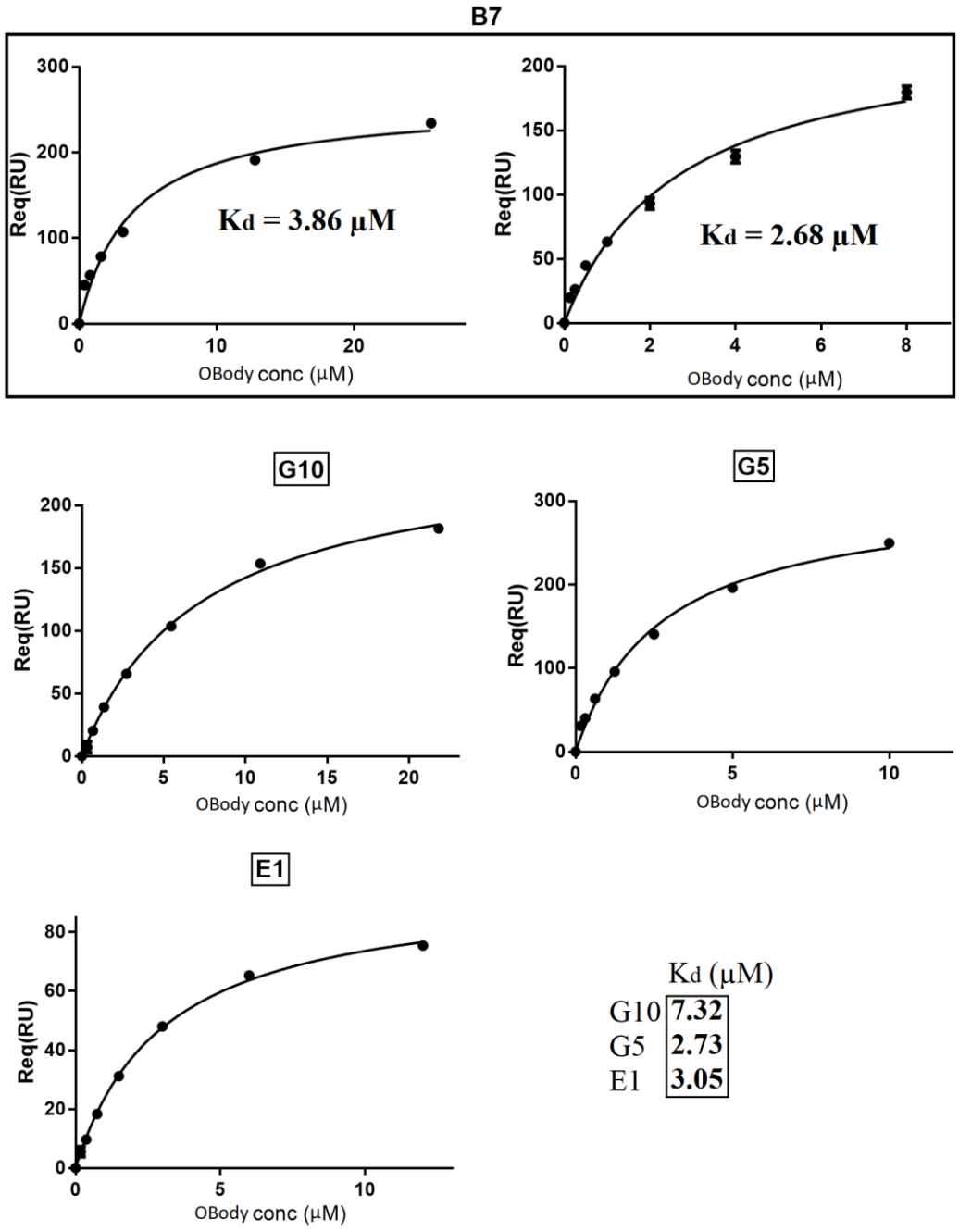


Figure 4.9 Equilibrium binding curves of AM library OBodies (SPR)

The binding curves at equilibrium from SPR generated sensorgrams are shown for different concentrations of affinity maturation (AM) library OBodies binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (R_{eq}, in RU) and the x-axis shows the concentration of analyte (OBody, in µM). The binding affinity (K_d) of the OBodies is indicated. The relevant statistics are summarized in Table 4. SPR is explained in detail in 2.4.2. and raw data provided in appendix (7.5).

The analyzed OBodies from the AM library have been summarized with relation to D7 and two of its mutants (C35S, DM) in Table 5. All selected OBodies lack cysteines except clone G5 from round 7 (R7) which has the same L4 sequence as G10 and shows a re-selection of cysteines on β2 and β3. However, there was an

increase in binding affinity in G5 (K_d ~ 2.7 μM) as compared to G10 (K_d ~ 7.3 μM) thus indicating a contributory role of cysteines (in β-sheets) in binding of OBodies to P4.

Table 4.2 SPR of P4 binding OBodies

The table shows Graphpad prism generated values for non- linear regression from SPR data for AM OBody binding to P4. The standard error and 95% confidence intervals (μM) refer to the binding affinity (*Kd*). R square is an important parameter indicating goodness of fit of the binding curve to experimental data. The values closer to 1 indicate a good fit. The values of another independent SPR analysis of OBody B7 is provided in parenthesis.

Non-linear fit	**D7**	**DM**	**B7**	**G10**	**G5**	**E1**
Bmax (Abs450)	66.8	24.19	259.4 (230.6)	246.6	310.5	125.4
***Kd* (μM)**	2.69	2.71	3.86 (2.68)	7.32	2.73	3.05
Std. Error (μM)	0.21	0.46	0.54 (0.35)	0.40	0.25	0.4
95% conf. int (μM)	2.24 – 3.14	1.73 – 3.7	2.67 – 5.05 (1.91 – 3.44)	6.45 – 8.2	2.19 – 3.28	3.31 – 5.05
R square	0.99	0.97	0.98 (0.9831)	1	0.99	1

4.2.5 Large scale expression and purification of B7

The OBody B7 (7.2.4) from L4 based AM library expressed in high yield and lacked cysteines with binding affinity comparable to D7 (K_d = 2.7 μM, SPR). Therefore, B7 was chosen for crystallization trials and expressed in *E. coli* BL21 (DE3, pPROEX, 7.2.5) strain using standard bacterial media (2xYT, LB) and purified by Immobilized metal affinity Chromatography (IMAC, Figure 4.10 and 4.11) and size exclusion chromatography (SEC, Figure 4.12 and 4.13) using standard buffers (TBS, PBS) similar to 3.2.1.

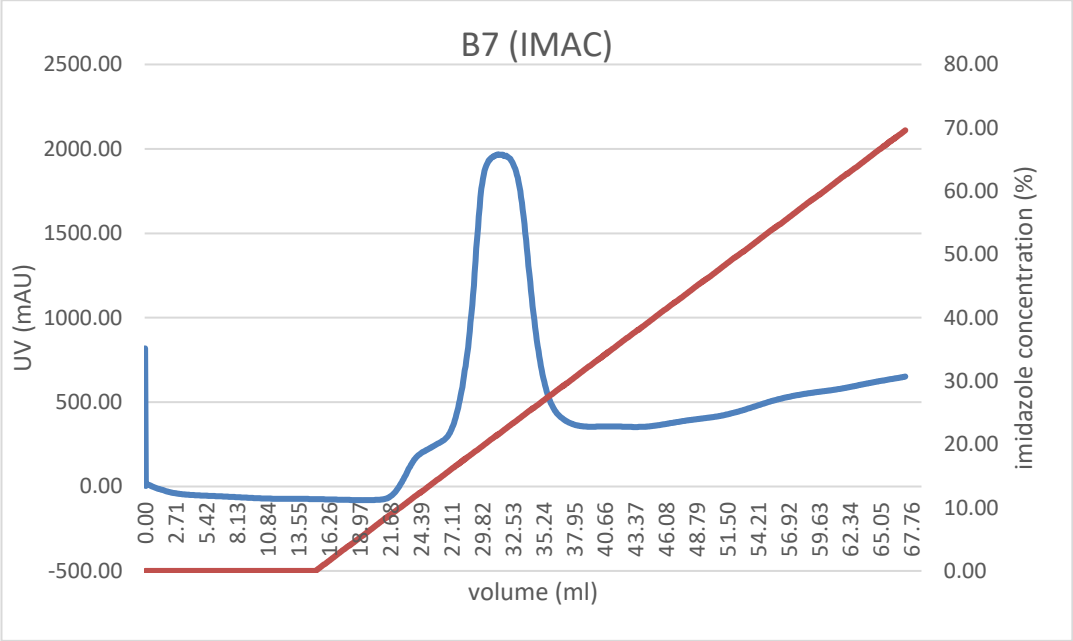


Figure 4.10 Purification of B7 on Histrap column

The UV absorbance curve (blue, 280nm) in mAU (left y-axis) shows the elution of B7 over an imidazole gradient (red) in percentage (right y-axis) during Histrap (IMAC) purification. The blue peak indicates maximum B7 elution. The x-axis refers to the flow of IMAC buffers (in ml, first 15ml of binding buffer followed by elution buffer containing imidazole, 7.1.3.1). IMAC is explained in detail in 2.2.2.1.

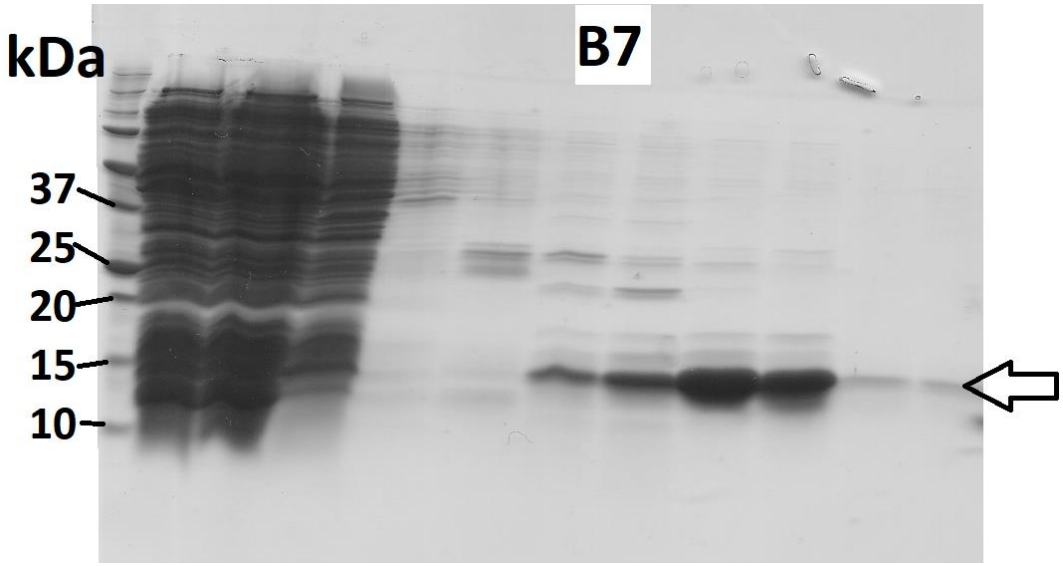


Figure 4.11 SDS-PAGE image of Histrap purified B7

16.5% SDS-PAGE image showing Histrap (IMAC) fractions containing purified B7 with reference to a protein ladder (left) showing possible sizes of proteins. B7 elutes as a monomer (14.5 kDa, arrow). Note that each lane in SDS-PAGE containing OBody B7 corresponds to elution fractions of UV peak in IMAC.

The OBody B7 was expressed in large amounts and was highly stable compared to its predecessor OBody D7 presumably due to lack of cysteines. Purification

confirmed loss of dimerization (as compared to D7) with 16.5% SDS-PAGE (Figure 4.11 and 4.13) and confirmed the size of the monomer (B7 = 14.5kDa).

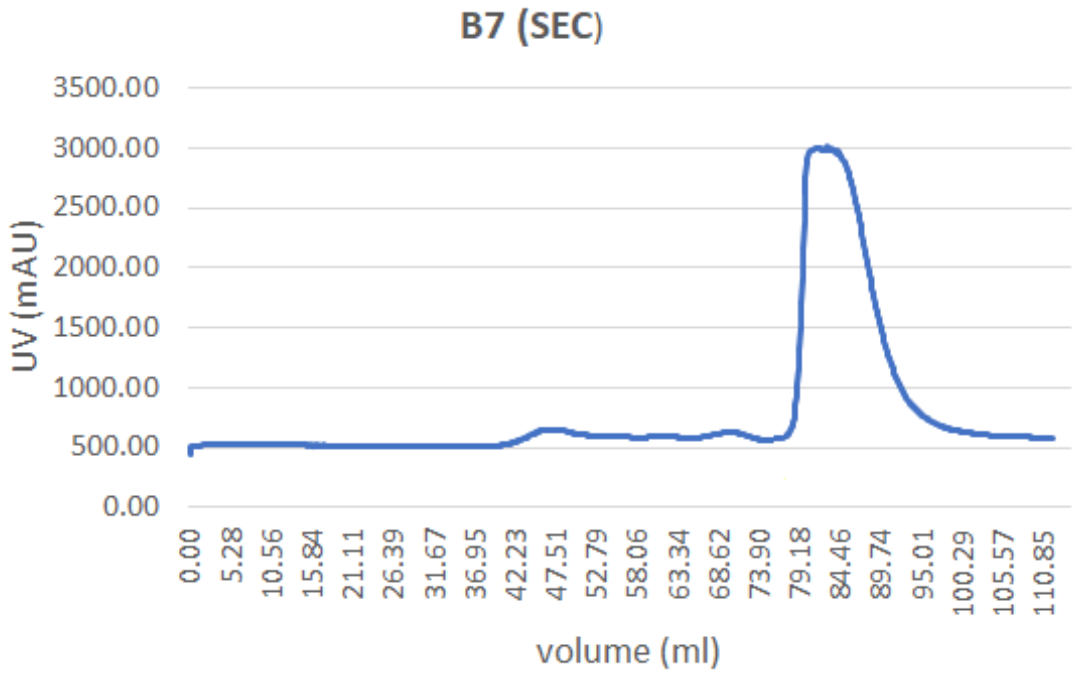


Figure 4.12 Purification of B7 on S75 16/60 column

The UV absorbance curve (blue, 280nm) in mAU (y-axis) shows the elution of B7 during S75 16/600 (SEC) purification. The x-axis refers to the flow of SEC buffer (in ml, 7.1.3.1). The blue peak indicates maximum B7 elution and corresponds to the size of monomer (14.5 kDA). SEC is explained in detail in 2.2.2.2.

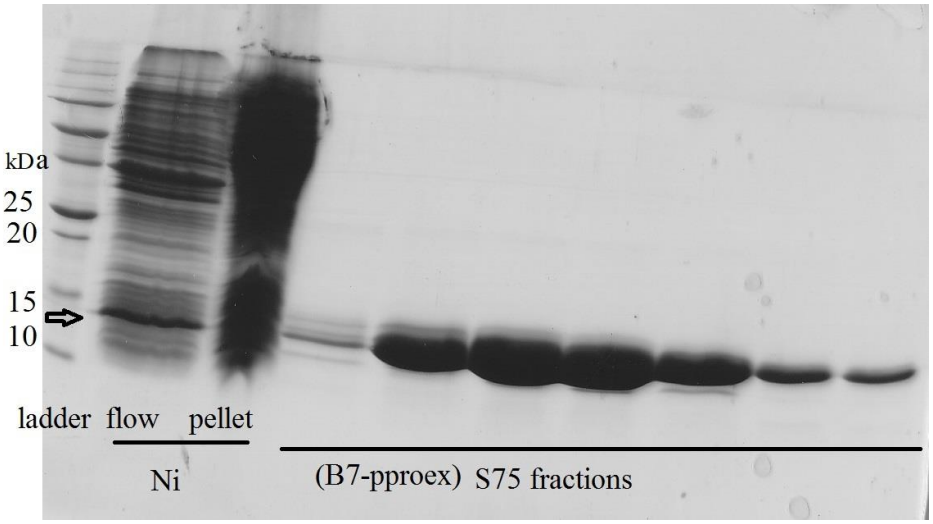


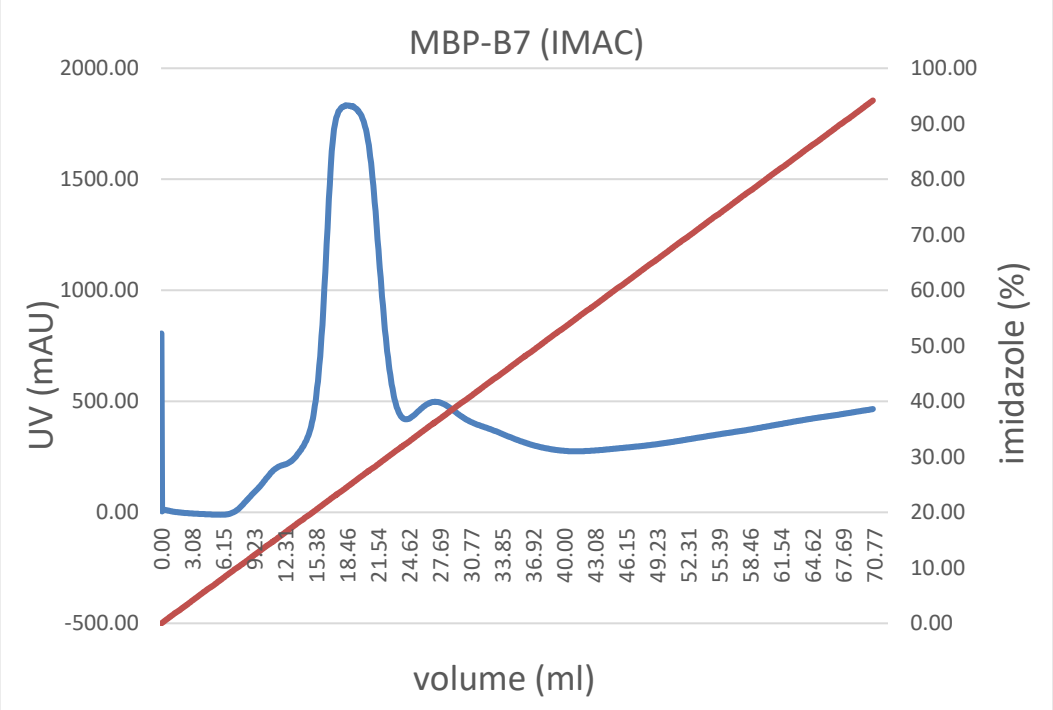
Figure 4.13 SDS-PAGE image of S75 16/600 purified B7

16.5% SDS-PAGE image showing S75 16/600 (SEC) fractions containing purified B7 with reference to a protein ladder (left) showing possible sizes of proteins. B7 elutes as a monomer (14.5 kDa) indicated by an arrow between 10 and 15 kDa. Note that each lane in SDS-PAGE containing OBody B7 corresponds to elution fractions of UV peak in SEC with the exception of IMAC (Ni) fractions.

The OBody B7 was concentrated and incubated with P4 (as described in 2.2.4) and crystal trials were put down with PEGRx HT - HR2-086, Crystal Screen HT - HR2-130, Index HT - HR2-134, and SaltRx HT - HR2-136 as described in 2.3.1. The crystals from these trials were grown in larger volumes by varying the pH and concentration of precipitant of the mother liquor. The larger crystals tested on X-ray diffractometer did not diffract or were salt crystals (2.3.2). Crystallization attempts included removal of N-terminal histag using recombinant tobacco etch virus protease (rTEV) and expression with short C-terminal histag using pET28b plasmid. This did not lead to growth of protein crystals and fusion to crystallization chaperone (MBP) was undertaken with the hopes of crystallization of this P4 binding OBody.

4.2.6 Fusion of crystallization chaperone MBP to OBody B7 (MBP-B7)

The OBody B7 was fused to MBP (7.2.6) successfully using the pOPIN system as described in 3.2.7. It was purified with IMAC (Figure 4.14) and SEC (Figure 4.15) similar to 4.2.5.



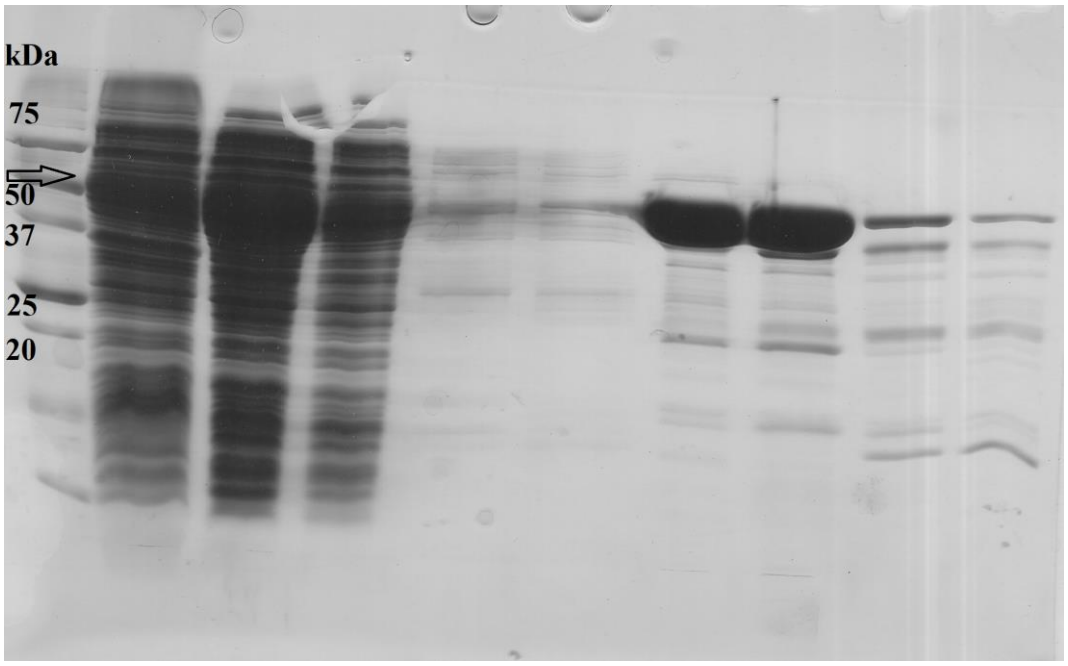


Figure 4.14 Purification of MBP-B7 on Histrap column

The top image shows the UV absorbance curve (blue, 280nm) in mAU (left y-axis) showing the elution of MBP-B7 over an imidazole gradient (red) in percentage (right y-axis) during Histrap (IMAC) purification. The blue peak indicates maximum MBP-B7 elution. The x-axis refers to the flow of IMAC buffers containing imidazole (7.1.3.1). IMAC is explained in detail in 2.2.2.1. The bottom image is 16.5% SDS-PAGE showing Histrap (IMAC) fractions containing purified MBP-B7 with reference to a protein ladder (left) showing possible sizes of proteins. The arrow indicates the size of this hybrid protein (~54 kDa). Note that two lanes in SDS-PAGE containing MBP B7 correspond to elution fractions of UV peak in IMAC.

The SEC elution of MBP-B7 showed two peaks and as there are no cysteines (Figure 4.15) this is possibly due to non-covalent association/oligomerisation (Gotte & Libonati, 2014; Hashimoto & Panchenko, 2010; Nishi, Hashimoto, Madej, & Panchenko, 2013). The SDS-PAGE analysis of the OBody fused to MBP showed an extra band (high MW peak, red arrow) above the 75 kDa level but most of the MBP-B7 eluted as monomer (black arrow, 54 kDa). Both peaks were separately subjected to crystallization trials (as described in 2.3).

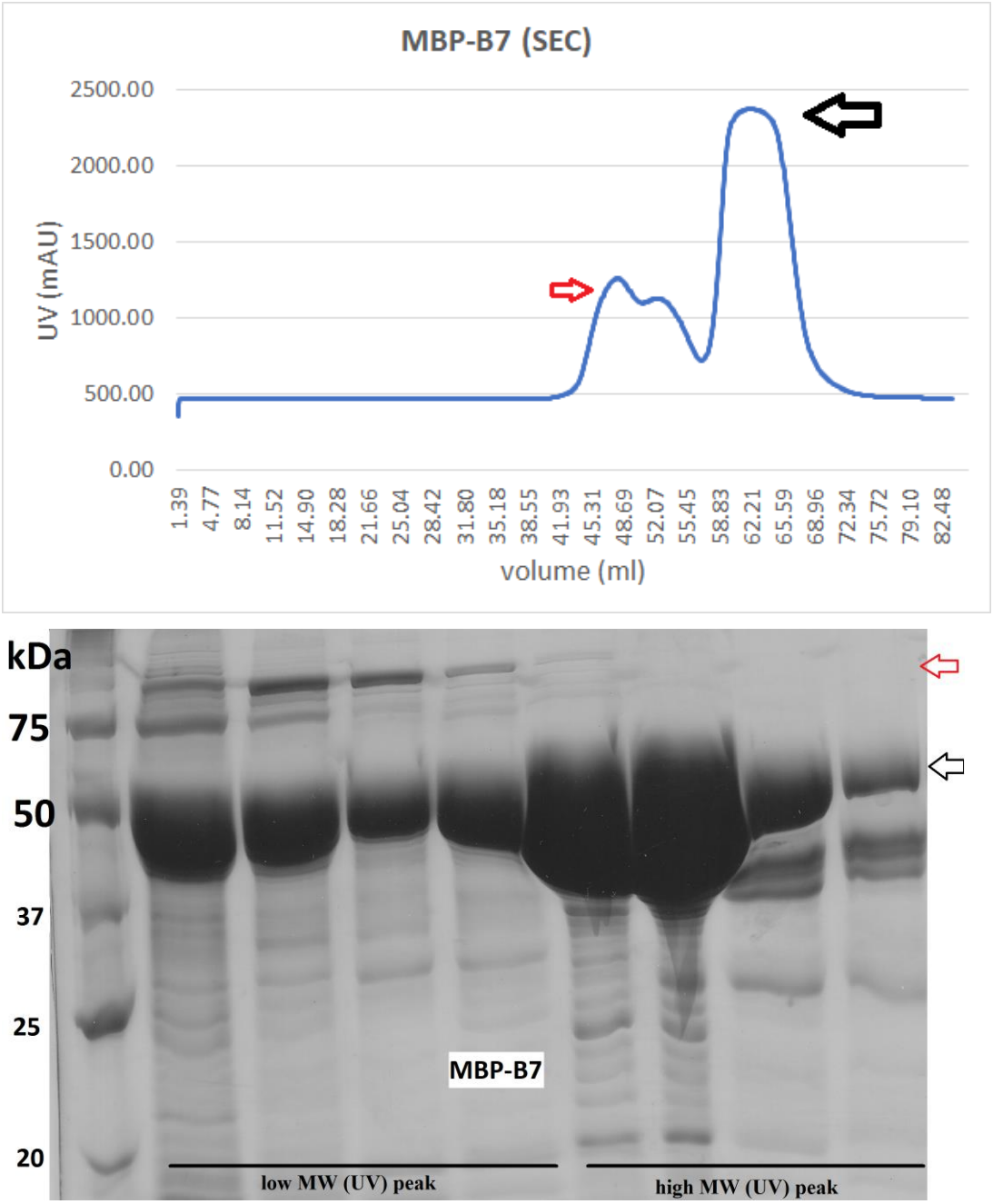


Figure 4.15 Purification of MBP-B7 on S75 16/60 column

The top image shows UV absorbance curve (blue, 280nm) in mAU (y-axis) showing the elution of MBP-B7 during S75 16/600 (SEC) purification. The low MW peak indicates maximum MBP-B7 elution (black arrow). The x-axis refers to the flow of SEC buffer (in ml, 7.1.3.1). The UV curve did produce two peaks with trace oligomerisation (non-covalent) possibly producing the high MW peak (red arrow). SEC is explained in detail in 2.2.2.2. The bottom image is 15% SDS-PAGE showing S75 16/600 (SEC) fractions containing purified MBP-B7 (black arrow) with reference to a protein ladder (left) showing possible sizes of proteins. The fractions represent MBP-B7 with relation to SEC elution (high and low MW peaks). MBP-B7 elutes as a monomer (54 kDa) indicated by a black arrow between 50 and 75 kDa. Note extra band above 75 kDa (high MW peak, red arrow) which is possibly due to non- covalent oligomerisation of MBP-B7.

The MBP-B7 from the low MW peak (Figure 4.15) produced protein crystals for the first time (after incubation with P4) that diffracted to a resolution of ~8 Å

using the sealed source X-ray diffractometer (XRD) at the University of Waikato. These crystals were then sent to the Australian Synchrotron to collect diffraction data to solve the crystal structure of MBP-B7 with P4. The MBP-B7 crystals from the low MW peak (~54 kDa) were tested for retention of binding (via B7) to P4 using SPR. Interestingly, SPR showed the K_d of MBP- B7 (688 nM, Figure 4.16) to be 4.3 times higher than B7 alone. (K_d = 2.7 μM).

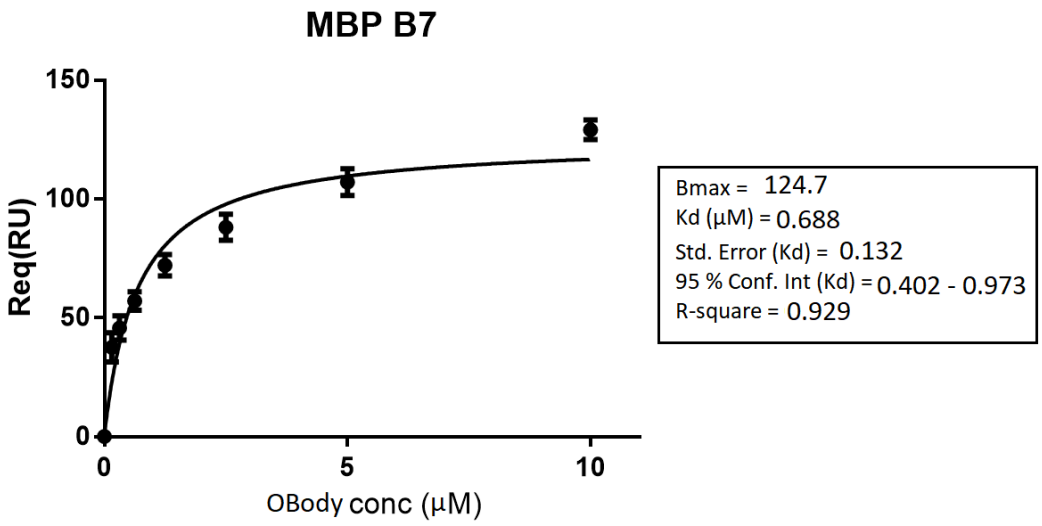


Figure 4.16 Equilibrium binding curve of MBP-B7 using SPR

The binding curve at equilibrium from SPR generated sensorgram is shown for different concentrations of MBP-B7 binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (R_{eq}, in RU) and the x-axis shows the concentration of analyte (OBody, in μM). The relevant statistics are summarized in the right. SPR is explained in detail in 2.4.2.

4.2.7 Crystal structure of fusion protein (MBP-B7) with progesterone (P4)

4.2.7.1 Crystal conditions

MBP-B7 was incubated with P4 (in a ratio of 1:1 as described in 2.2.4) and concentrated to ~65 mg/ml and crystallization screens were put down in drops of a 1:1 ratio as described in 2.3.1. PEG 24 (30% w/v PEG monomethyl ether 2000, 0.1M Tris pH 8, Figure 4.17B) and PEG 33 (28% w/v PEG 4000, 0.1M Tris pH 8, Figure 4.17A) are conditions from PEGRx HT (HR2- 086) screen that produced diffracting protein crystals (Smyth et al., 2003). These conditions were refined using a grid format in 6x4 large well plates with the precipitant (PEG 4000 or PEG monomethyl ether 2000) ranging from 24 to 34% w/v and pH values ranging from 7.9 to 8.2.

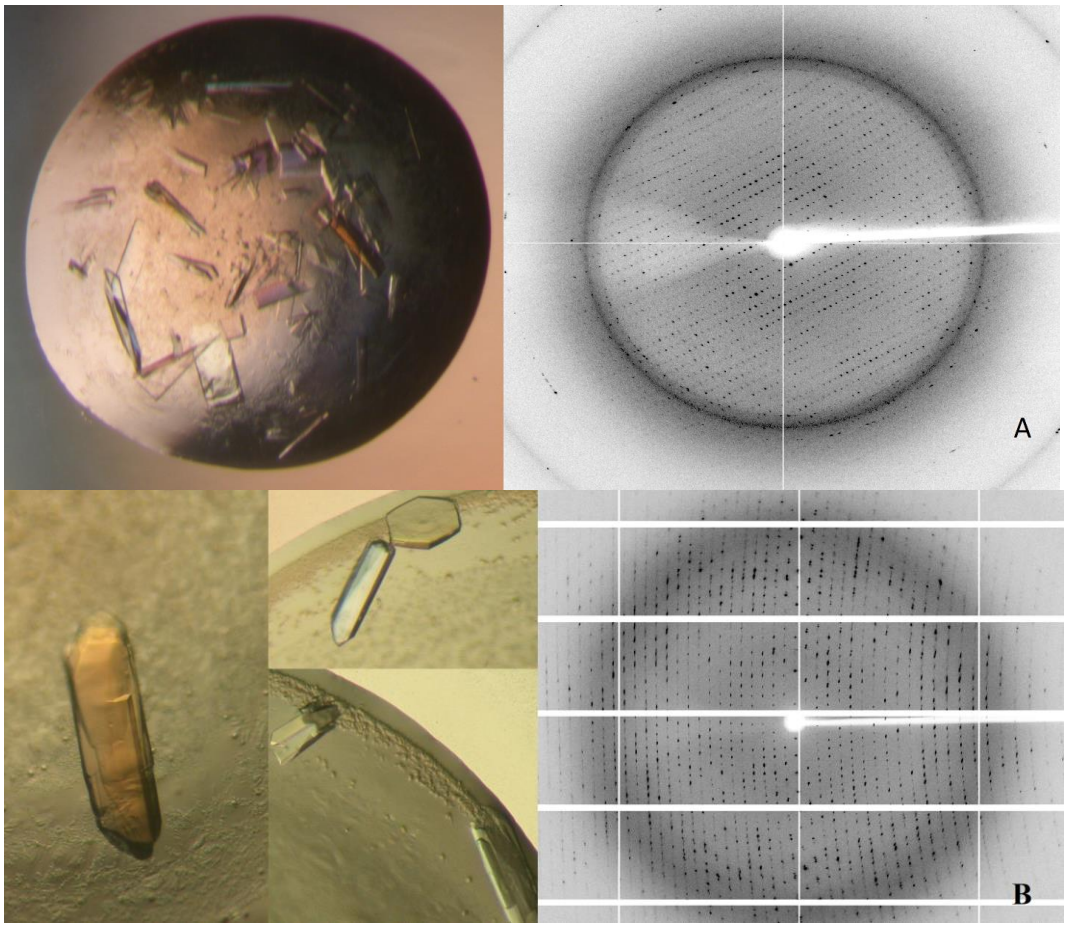


Figure 4.17 Crystals of MBP-B7 (with P4) under X-ray diffraction

An image of MBP-B7 (left, with P4) crystals under an optical microscope (Nikon SMZ800) shows crystals of different sizes in the 2 µl crystal drop (1:1) containing protein with PEG 33 (A) or PEG 24 (B)(# HR2-086). A X-ray diffraction image from the best diffracting crystal for each condition are also shown (right) i.e. 2.5 Å dataset (A) and 1.9 Å dataset (B) collected at the Australian Synchrotron.

4.2.7.2 Data processing

The raw diffraction data (collected at the Australian synchrotron, 7.6) were indexed using iMOSFLM (2.3.3.1) and integrated using aimless software (2.3.3.2). The data collection statistics (Table 6) showed that both crystal conditions PEG 24 (30% w/v PEG monomethyl ether 2000, 0.1M Tris pH 8) and PEG 33 (28% w/v PEG 4000, 0.1M Tris pH 8) from PEGRx HT (HR2- 086) screen produced crystals that belonged to $P2_1$ space group (2.3.3.1). The best condition to solve the crystal structure of B7 with P4 was PEG 24 (30 % w/v PEG MME 2000, 0.1M Tris pH 8) due to two reasons i.e. the dataset was of higher resolution (1.9 Å) and does not have significant crystal pathology like twinning (PEG 33). This 1.9 Å dataset (PEG 24) was processed to a resolution of 2.5 Å and showed data completeness of 94.3 % in its highest resolution shell (2.3.3.2). The

R_{merge} which reflects data quality and the signal to noise (I/sigI) ratio was overall within acceptable range ~0.2 (less than 0.8) and 3.9 (greater than 2) respectively (Karplus & Diederichs, 2012). The CC ½ was 0.844 at 2.5 Å (i.e. greater than 0.4)(P. R. Evans & Murshudov, 2013). The phases were estimated using molecular replacement with MBP as a model (PDB: 1OMP) (Sharff, Rodseth, Spurlino, & Quiocho, 1992) using Phaser (Phenix suite, 2.3.3.3) (P. Evans & McCoy, 2008). Similarly, other datasets (belonging to PEG 24 and PEG 33) were solved with molecular replacement ranging in resolution from 1.9 Å to 3.5 Å. This MR solution from Phaser (which contains the experimentally derived intensities and computationally derived phases with relation to the MBP model) was used to build an electron density map and model of B7 with model building software Auto-build from Phenix suite (2.3.3.4). The solution contained 4 molecules of MBP-B7 in the asymmetric unit (ASU) consistent with the calculated Matthews co-efficient (Matthews, 1968).

Table 4.3 Data collection statistics of OBody B7

The values in brackets are indicative of highest resolution shell.

Crystal condition	*PEG 33 (PEG 4000)*	*PEG 24 (PEG MME 2000)*
Crystal pathology	twinning	translational NCS
Space group	P 1 2_1 1	P 1 2_1 1
Unit cell (Å)	61.8, 142.8, 109.5	65.9, 193.9, 82.4
Unit cell (°)	90, 90.37, 90	90, 97.7, 90
Resolution (Å)	47.59– 2.70	97.11–2.50
R_{merge}	0.153 (1.177)	0.204 (0.518)
R_{pim}	0.059 (0.455)	0.100 (0.258)
Completeness (%)	92.8 (93.8)	98.6 (94.3)
No. of observations	368459 (33369)	350581 (21103)
Unique reflections	48648 (4485)	70181 (4303)
Multiplicity	7.6 (7.4)	5.0 (4.9)
Average mosaicity	0.36	1.29
Mean I/sigI	13.3 (2.0)	3.9 (1.9)
Mean (I) CC (1/2)	0.997 (0.613)	0.978 (0.844)
No. of molecules in asymmetric unit (ASU)	4	4
Wavelength (Å)	0.9537	0.95373

4.2.7.3 Crystal pathology

The PEG 24 data had translational NCS (tNCS) which occurs when two or more molecules have a symmetry function locally but not as a crystal in the translational direction (Dauter & Jaskólski, 2016). Phaser (Phenix) uses maximum likelihood approaches for molecular replacement which takes into consideration deviations of translation and orientation and assigns it to the same asymmetric unit to produce a successful structure solution (Sliwiak, Jaskolski, Dauter, McCoy, & Read, 2014). Refmac5 (CCP4) uses local and global based NCS restraints which compensates for local differences in NCS related chains during refinement. In this way, the tNCS data was used for final data refinement.

The PEG 33 data had twinning which is a pathology where the diffraction spots result from the crystals being in two different orientations (Parsons, 2003; Zwart, Grosse-Kunstleve, Lebedev, Murshudov, & Adams, 2008). The intensities of diffraction spots in twinned crystals are the result of overlap of diffraction from the two orientations and was also resolved using Phaser and refined with intensity based twin refinement Refmac5. This data was used only for initial model building and was not included in structure analysis.

4.2.7.4 Data refinement

The data from PEG 24 was used for solving the structure of OBody B7 bound to P4. The refinement of this data was challenging due to tNCS and thus model building was initially done with PEG 33 (twinned) data using Auto-build (Phenix suite, 2.3.3.4) (Read, Adams, & McCoy, 2013). Auto-build generated only parts of OBody B7 with trace skeletal outline of the β-sheets of the β-barrel and parts of the α-helix. A previous OBody structure (PDB: 4GLV) was mutated *in silico* to resemble B7 with Wincoot and positioned (and merged) with the model of B7 (from Auto-build) using Pymol. This hybrid model was used to build the rest of B7 by rounds of model building and refinement with Refine from the Phenix suite. Small parts of the B7 model could not be built (e.g. flexible loops (L4) and the linker adjoining MBP and B7) and were deleted from the hybrid model. The partially refined MBP-B7 structure revealed the electron density for the position of P4 bound to OBody. Wincoot was used to fit this ligand and refined.

The model from this twinned data (PEG 33) was used to construct a final model for tNCS (PEG 24) data using refinement with Refmac5 (CCP4 suite, 2.3.3.5). The MBP part of the model refined perfectly with tNCS data whereas residues of the OBody did not match the electron density. These poorly fitting residues were removed and refined in cycles with Refmac5 (CCP4) and Refine (Phenix). This led to improvement in overall model and poorly fitting residues were added to match the electron density. This resulted in building of the OBody and R_{free} was 26% (lower than 30%) for the final model (Kleywegt & Brünger, 1996; Wlodawer et al., 2008). The refinement statistics from the PEG 24 data used to solve the MBP-B7 structure with progesterone (P4) has been summarized in Table 7.

Table 4.4 Data refinement statistics of OBody B7

The table lists out important parameters indicating refinement and summary of best dataset.

Crystal condition	*PEG 24 (PEG MME 2000) MBP-B7*	*MBP only*
R_{work}	0.22	0.25
R_{free}	0.26	0.29
Bond length (rms)	0.01	0.01
Bond angle (rms)	1.66	1.38
Total no. of atoms	14261	12093
Total no. of protein atoms	13791	11818
No. of waters	378	275
Average B-factors (protein)	32.50	30.8
Average B-factors (water)	33.59	36.08
Ramachandran plot		
amino acids in favored regions (%)	89 %	95.6 %
amino acids in allowed regions (%)	10.7 %	4.27 %
amino acids in disallowed regions (%)	0.3%	0.13 %

4.2.8 Structure of OBody B7 and interaction with P4

4.2.8.1 Comparison of AM library OBody (B7) to naïve library OBody (D7)

The affinity maturation (AM) library OBody B7 (4.2.5) is similar in structure to naïve library OBody (D7, Figure 3.14) and differs only in six amino acids in the loop (L4, Figure 4.18) and lack of cysteines (including proximal binding face, pF). The naïve library OBody D7 has a different loop (L4, with cysteines at positions 81 and 85) and cysteines at two positions in pF (35 and 48). The progesterone (P4) packs against tryptophan (W472) on the opposite side of the proximal binding face on the β-barrel of the OBody B7. This OBody-P4 binding is expected to be the same for OBody D7 due to same sequence of amino acids at the progesterone (P4) binding site.

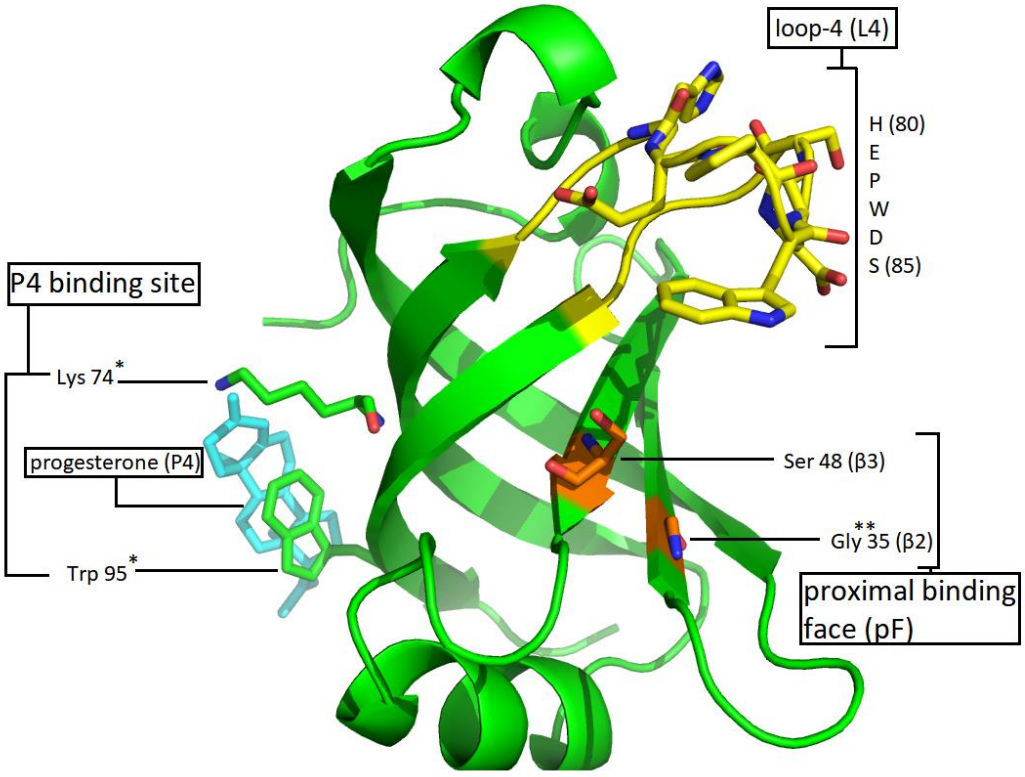


Figure 4.18 Model of AM OBody (B7) with progesterone (P4)

A cartoon representation of OBody B7 (green) indicating the mutated loop (L4) with residues (yellow), location of eliminated cysteines in β-sheets 2 and 3 (orange) and P4 binding site on the other side of β- barrel with the residues binding P4 (cyan) constructed using Pymol is shown. The residue numbers are indicated and * refers to numbering without MBP (Lys 451 and Trp 472). ** refers to a random mutation during phage selection of AM library (S35G).

The 2GPATS OBodies are also similar to 1GPATS OBodies engineered to bind HEL by phage display. The models generated from solved structures of OBodies

from both generations (B7 and AM3L15) have been superimposed in figure 4.19 to show the similarities and differences. The r.m.s.d of the backbone atoms between the two OBodies is 0.61 Å. The β-barrel and α-helix overlap well without any change in overall structure. The loops show major differences with 2GPATS OBody possessing a smaller loop 2 (L2) and a larger loop 1 (L1) and loop 4 (L4) when compared to 1GPATS OBody. The design of 2GPATS library involved shortening of L2 by two amino acids (removal of GA spacer, figure 3.7) and elongation of L4 by addition of two amino acids. This was done due to non-involvement of L2 in the binding interface whereas L4 lies in the periphery of the binding interface and elongation was hypothesized to contribute in ligand binding. However, elongation of L1 was not planned during library design and is the result of different amino acids at the intersection of L1 with the β1 and β2 strands.

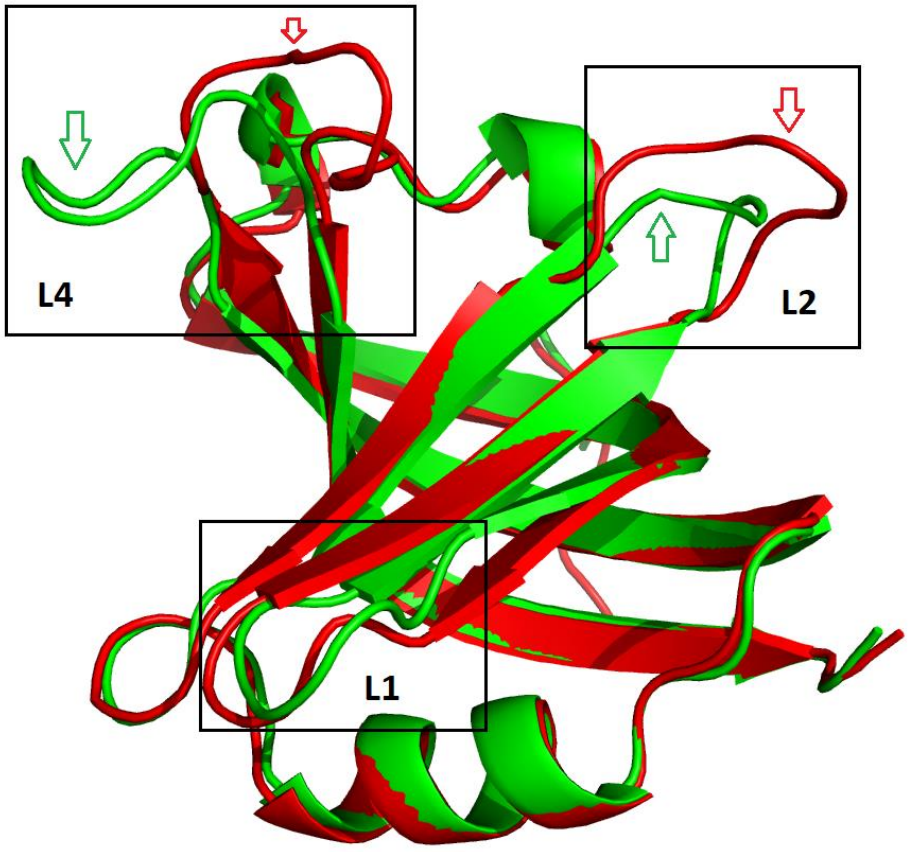


Figure 4.19 Overlap of 2GPATS OBody (B7) and 1GPATS Obody (AM3L15)

A cartoon representation of OBody B7 (green) overlapping OBody AM3L15 (red) showing the agreement with the overall β-barrel structure and the α-helix. The differences between these two models lie in the labelled loops indicated by boxes. The coloured arrows indicate the loops of the appropriate OBodies.

4.2.8.2 Crystal packing of MBP-B7

MBP-B7 packs as four copies in the asymmetric unit (ASU) of the crystal. The ASU is the smallest indivisible part of the crystal which can be used to form a unit cell of the crystal by rotation or translation under crystal symmetry (Krissinel & Henrick, 2007). The unit cell is required to construct the entire crystal by translation in all three dimensions. There is a long flexible linker between MBP and OBody B7 allowing MBP molecules to pack together with the B7 occupying the space in the cleft between MBP molecules as shown in Figure 4.20. This also allows two B7 molecules to form a non-covalent dimer as hypothesized from SEC purification of MBP-B7 (Figure 4.15). This is possible via a hydrophobic pocket between the B7 monomers which allows for P4 binding and is discussed in 4.2.8.4.

The arrangement of MBP-OBody fusion proteins leads to formation of intermolecular polar bond interactions in two interesting ways. Firstly, there is a complementary interaction involving alanine (A379) and glutamine (Q375) of the two opposing MBP-B7 linkers as shown in the Figure 4.20. Amino acid 378 is the first amino acid of the OBody with methionine (M378) interacting with glycine (G376) of the opposing MBP molecule in the MBP-OBody linkers. Secondly the alpha helices of the MBP molecules form intermolecular polar bond interactions with adjacent opposing OBody molecules (threonine T345-histidine H418 and serine S367-tryptophan W400). Most of these interactions are of moderate affinity based on the intermolecular distances (~ 2.5 Å) and result in robust crystal packing of these molecules. The amino acids that are completely missing from any of the four MBP-B7 in the ASU are Y422, S437, P471 and D473.

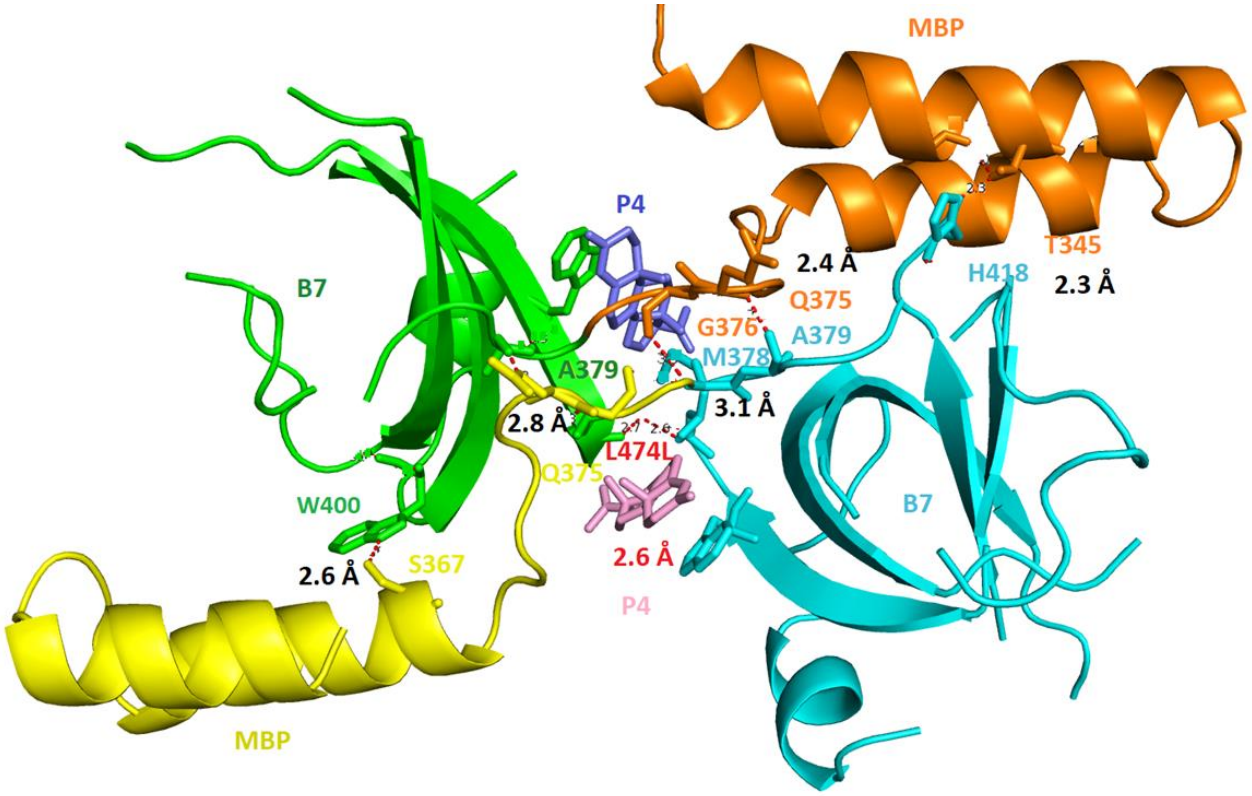


Figure 4.20 Crystal packing of MBP and OBody B7

A cartoon representation of the crystal packing between two MBP- OBody fusion proteins created with pymol is shown. The MBP OBody pairs (yellow – cyan and brown - green) and the bound P4 molecules (pink and blue) with colour appropriate labels (amino acid and residue number) and intermolecular polar bond distances are indicated. The interaction of L474 of two OBodies via a water molecule is highlighted in red. This is an aerial view of the two OBodies (looking down the β-barrels, green and cyan) stacked against each other supported by two α-helices of adjacent MBP molecules (C- terminal, yellow and orange) with the adjoining MBP -B7 linkers between the two OBodies.

Other amino acid residues potentially contributing to the intermolecular interaction were assessed by using the PDBePISA tool (Table 4.5). It confirms the amino acids already established to interact with P4 (hydrogen bonds) but also indicates the binding surface area (BSA) to be dominated by hydrophobic amino acids (Phe 374, Trp 382 and Trp 400). Buried surface area is a stronger indicator of intermolecular interaction than hydrophobicity (Chen, Sawyer, & Regan, 2013).

Table 4.5 Interactions between MBP and Obody B7

The amino acid residues contributing to the intermolecular interaction between two MBP-OBody molecules (A and B) and relevant values are shown. ASA and BSA refer to accessible surface area and buried surface area respectively (in Å). ΔiG refers to solvation energy effect (in kcal/mol). The OBody residues are highlighted in **bold**.

Interfacing residues	**H-bond**	**ASA**	**BSA**	**ΔiG**
A: Asp 363		72.56	27.72	0.07
A: Thr 366		101.20	48.94	0.73
A: Ser 367		63.82	22.03	-0.15
A: Leu 370		109.32	58.48	0.94
A: Val 372		132.50	17.25	0.28
A: Leu 373		113.27	84.44	1.13
A: Phe 374		163.01	112.59	1.74
A: Gln 375	H	155.26	32.78	0.21
A: Gly 376		61.26	37.56	-0.23
A: Thr 377		95.51	22.93	0.37
A: Met 378	H	127.89	71.43	1.63
A: Ala 379	H	43.81	42.13	0.06
A: Thr 380		97.86	22.09	- 0.12
A: Trp 382		142.49	116.34	1.27
A: Ala 398		23.84	21.50	0.22
A: Trp 400	H	127.58	104.12	1.20
A: Asp 417		41.99	20.42	-0.09
A: His 418	H	118.14	89.45	0.01
A: Gly 420		61.44	24.33	-0.14
A: Val 447		74.18	39.42	0.63
A: Ile 473		59.92	10.95	-0.02
A: Leu 474		129.21	60.70	0.44

Interfacing residues	**H-bond**	**ASA**	**BSA**	**ΔiG**
B: Tyr 341		124.64	54.42	0.76
B: Thr 345	H	80.15	50.09	0.49
B: Asn 349		37.05	10.05	-0.11
B: Asp 363		75.25	31.96	0.15
B: Thr 366		100.33	52.65	0.78
B: Ser 367	H	57.25	48.12	0.01
B: Leu 370		92.93	60.00	0.96
B: Leu 373	H	123.64	93.53	1.21
B: Phe 374		155.55	107.98	1.63
B: Gln 375	H	152.31	45.27	0.32
B: Gly 376	H	54.05	33.01	-0.16
B: Thr 377		92.60	24.58	0.03
B: Met 378		107.99	57.91	1.31
B: Ala 379	H	42.88	39.86	0.08
B: Thr 380		70.86	21.99	-0.11
B: Trp 382		198.84	72.30	0.73
B: Ala 398		43.72	17.70	0.22
B: Gly 399		32.31	11.92	0.18
B: Trp 400	H	177.09	114.26	1.17
B: Glu 445		74.93	10.21	0.16
B: Val 447		62.69	33.58	0.54
B: Leu 474		103.30	53.59	0.37
B: Asn 475		144.59	23.23	-0.17

4.2.8.3 Interactions between MBP molecules

The MBP molecules interact with each other with at least three molecules involved in intermolecular polar bonds of moderate affinity (~ 2.5 Å, figure 4.21). There is a complementary interaction between glutamic acid (E309) and lysine (K305) between two MBP molecules. The other notable interactions are between asparagine (N124) and lysine (K127) between the same molecules and between glutamic acid (E172) and lysine (K219) with the third MBP molecule. This combined with strong interactions between MBP and OBody (B7, 4.2.8.2) result in strong crystal packing of these molecules.

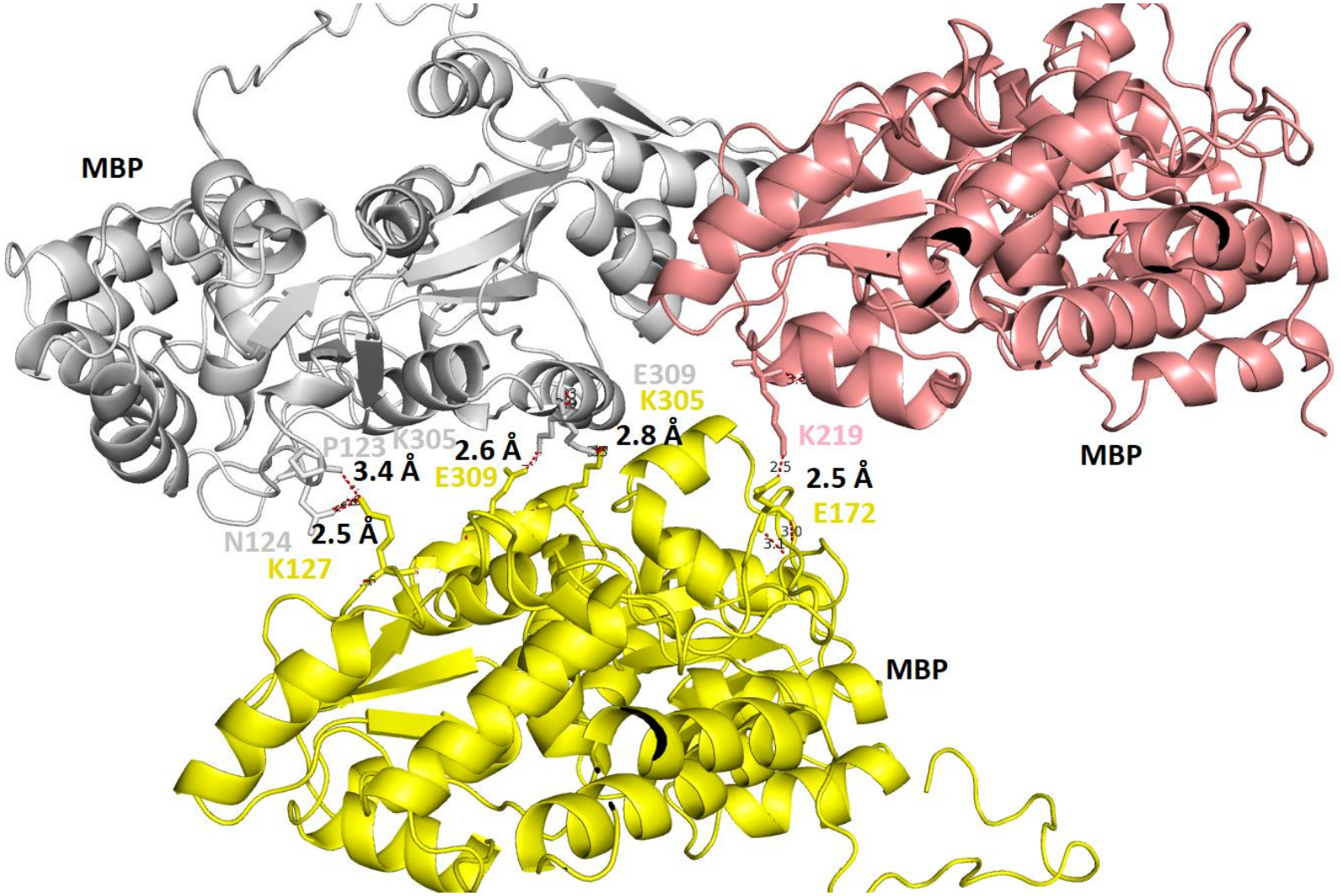


Figure 4.21 Crystal packing of MBP molecules

A cartoon representation of the crystal packing between MBP molecules created with pymol is shown. The three interacting MBP molecules (grey, pink, yellow) with colour appropriate labels (amino acid and residue number) and intermolecular polar bond distances are indicated.

4.2.8.4 Binding between OBody B7 and progesterone (P4)

Surprisingly, the P4 molecule fits in a hydrophobic pocket created between two OBodies with the steroid rings of P4 packed against the aromatic rings of tryptophan W472 of one OBody (Figure 4.22). This hydrophobic pocket is fortified by valine (V449) and methionine (M378) of the adjacent OBody molecule, the latter amino acid forming a part of the MBP-OBody linker which forms the majority of the wall of the hydrophobic pocket with charged glutamic acid (E371). The rest of the hydrophobic pocket is supported in one direction with valine (V396) and a charged lysine (K451) of the adjacent OBody (with the carbonyl end of P4 forming a hydrogen bond with a water molecule in the vicinity) and in another direction by asparagine (N475) and leucine (L474) of the opposite OBody. Similarly, another P4 molecule is found in the hydrophobic pocket supported by tryptophan W472 of the opposite OBody with two-fold symmetry as shown in figure 4.20.

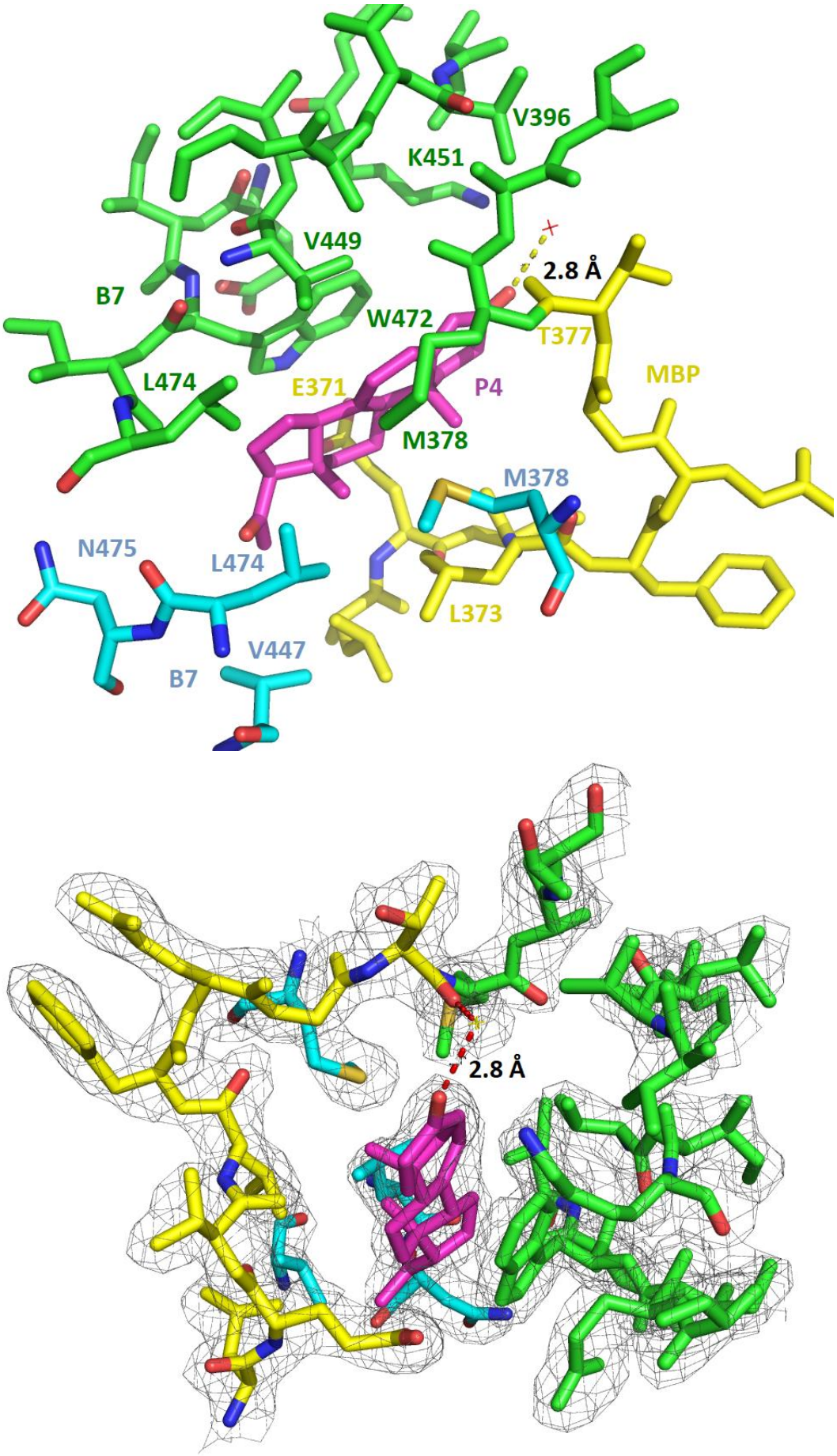


Figure 4.22 Interaction surface between OBody B7 and P4

The atomic detail image of P4 (A, magenta) in the hydrophobic pocket between two OBodies (B7) is shown as a model (a) and with electron density (b). This Pymol generated image also shows P4 packed against tryptophan 472 of B7 (green) surrounded by adjoining MBP linker (yellow) and opposite OBody (cyan). The colour appropriate labels (charge, amino acid and residue number) and hydrogen bond with water is indicated. Note the change in orientation of electron density image (b) from the direction of the hydrogen bond.

The amino acid residues potentially contributing to the intermolecular interaction were assessed by using the PDBePISA tool (Table 4.6). The majority of the buried surface area is formed by leucines (amino acids 370, 373 and 474 of both OBodies). The hydrophobic amino acids W472 (involved in packing against P4 molecule), M378 (first amino acid of OBody) and a charged amino acid E371 (MBP) contribute to the rest of the buried surface area. The overall buried solvent accessible surface area between the two Obodies is 1320 Å.

Table 4.6 Interactions at binding interface of Obody B7

The amino acid residues contributing to the binding interface between two MBP-OBody molecules (A and B) and P4 with relevant values are shown. ASA and BSA refer to accessible surface area and buried surface area respectively (in Å). ΔiG refers to solvation energy effect (in kcal/mol). The OBody residues are highlighted in **bold**.

Interfacing residues	**H-bond**	**ASA**	**BSA**	**ΔiG**
A: Leu 370		109.32	23.93	0.38
A: Glu 371		112.98	26.91	-0.02
A: Val 372		132.50	15.58	-0.18
A: Leu 373		113.27	28.83	0.46
A: Gly 376		61.26	17.85	0.29
A: Thr 377		95.51	15.99	-0.14
A: Met 378		127.89	16.20	0.23
A: Val 449		16.05	14.36	0.23
A: Lys 451		81.54	17.84	-0.55
A: Trp 472		82.36	54.37	0.85
A: Leu 474		129.21	16.91	0.27
P4: A		487.98	341.37	-4.46
P4: B			105.56	-1.10
B: Met 378		107.99	28.96	1.04
B: Val 447		62.69	17.57	0.28
B: Leu 474		103.30	33.49	0.54

The surface representation of the model from Figure 4.20 and 4.22 is shown with P4 (Figure 4.23) to better visualise the P4 molecule with relation to the two MBP fused OBodies (B7). The binding interface is in the form of a hollow tube lined

with hydrophobic amino acids with either ends lined by charged amino acids (positively charged area towards K451 of adjacent OBody and negatively charged area towards N475 of the opposite Obody). The P4 molecule fits in this pocket with its hydrophobic steroid rings via aromatic stacking, and the polar ends of the P4 molecule are exposed towards the charged ends. Most of the hydrophobic pocket is supported by the two opposing OBodies and adjacent MBP-B7 linker with two charged ends contributed by the two Obodies (centered around K451 of adjacent OBody and N475 of the opposite OBody). Similarly, another P4 molecule shows a two-fold rotation with the two B7 molecules.

In chapter 3, it was established that there was a pH dependent change in binding affinity *(K_d)* and specificity to P4 (3.2.6.1 and 3.2.6.2). Histidine was hypothesized as the amino acid (pKa ~ 6) to contribute to this phenomenon. However, since the binding interface is the same for OBody D7 and B7, there are no histidines in the vicinity of the binding interface (Figure 4.20 and 4.22) to explain this pH dependency or specificity and may be the result of indirect effects on protein stability and disruption of OBody dimers required for P4 binding.

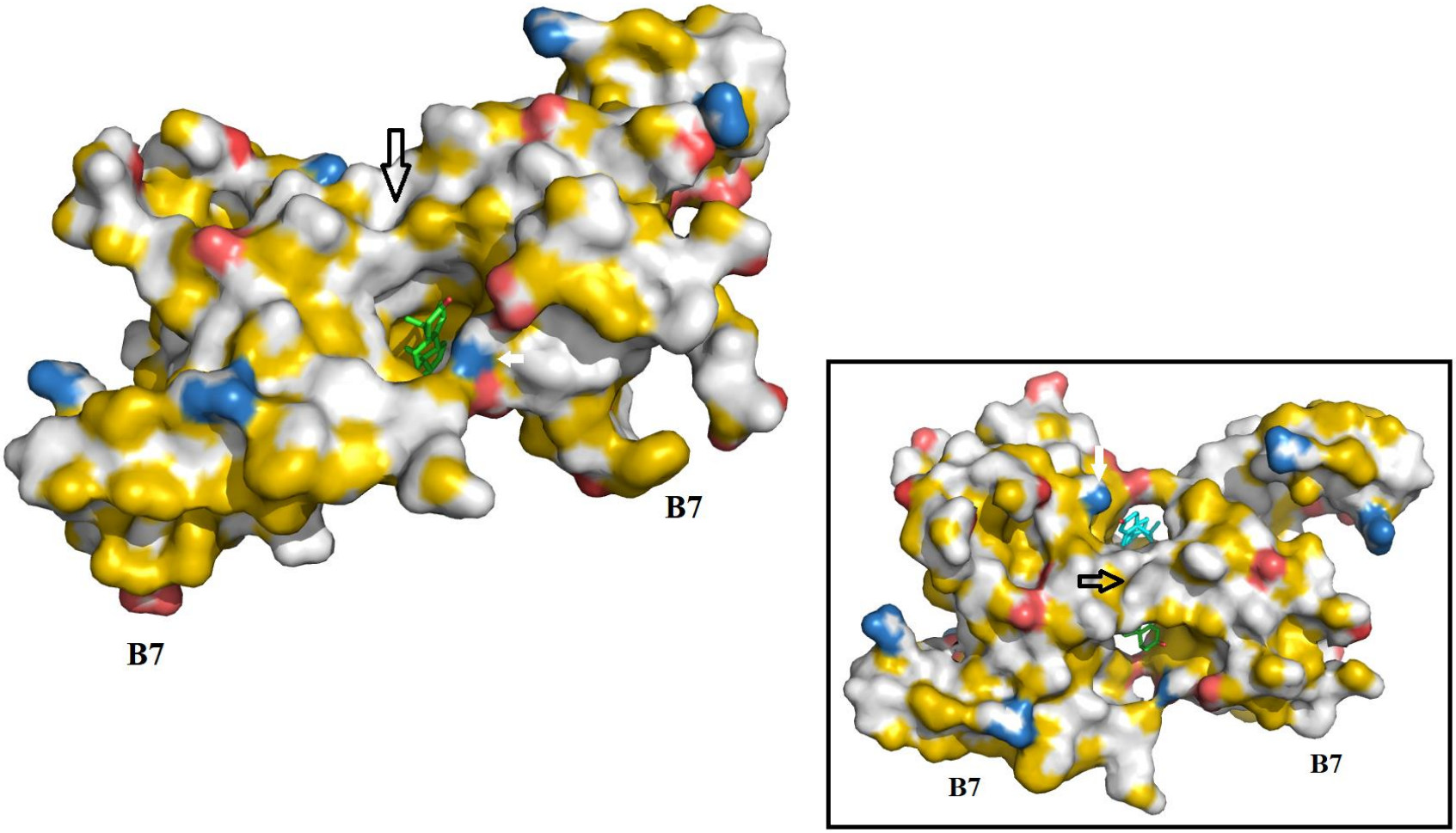


Figure 4.23 Surface representation of B7 with P4

The cartoon model of two OBodies forming the binding interface with P4 with reference to surface is shown from the side (blue – positive charged, red – negative charged, yellow - hydrophobic) with the top view in inset. The P4 molecules (green and cyan) can be seen in the hydrophobic channels created by the two OBodies and their MBP linkers (black arrow) with two-fold symmetry. The positively charged area centered around K451 (white arrow) is also indicated.

4.3 Discussion

4.3.1 Affinity maturation of OBodies

It was hypothesized that the cysteines in OBody D7 may be an impediment to crystallization due to formation of intermolecular disulfides. Therefore, affinity maturation (AM) libraries were designed from the functional mutants of D7. DM (C35S:C48S) was chosen as the OBody template for a loop 4 (L4) based AM library which lacked cysteines in the β-sheets (β2 and β3) and the cysteines in the loop were randomized to eliminate them. C35S was chosen as the OBody template for proximal binding face (pF) based AM library where only β3 (centered around C48) was randomized as a conservative approach to improve binding function if the cysteines on L4 proved to be critical for P4 binding.

OE-PCR was used to insert stop codons into D7 mutants to prevent re-selection during phage display and to create PCR fragments for randomization of nucleotide sequences coding for the binding surfaces. This PCR based randomization of the nucleotide sequences for the binding surfaces using randomization primers and temperature optimization resulted in the construction of two AM libraries (pF_35 and L4_DM) based on D7 mutants C35S and DM respectively.

The P4 binding OBodies were phage selected from the L4 based AM library which was designed based on the rationale that L4 is critical in binding between D7 and P4 as described in 3.2.6. The conservative randomization of pF based AM library possibly did not offer enough selective advantage for P4 binding compared to L4 based AM library thus explaining lack of OBodies from pF based AM library.

4.3.2 Randomization of loop (L4) and P4 binding

The OBodies from the L4 based AM library were confirmed by sequencing to vary only in L4 residues (randomized) in most P4 binding OBodies. These OBodies did not have a consensus in loop (L4) sequence thus indicating poor involvement in P4 binding. These OBodies were tested for binding affinity using SPR which showed no improvement in binding affinity with K_d in low micromolar values as naïve OBody D7 (4.2.4). Most OBodies lacked cysteines

and the best candidate B7 was subjected to crystallization trials as described in 4.2.5 and 4.2.6.

The isolation of these OBodies however reflected the selection pressure of decreasing amounts of target (P4) and duration of incubation leading to OBody output drastically dropping after round 4 (R4) with most P4 binding OBodies isolated from R4 and R5. The solved structure of B7 with P4 explains the lack of improvement in OBody affinity to P4 possibly due to recruitment of a novel binding interface not seen in OB-folds found in nature (Arcus, 2002; Murzin, 1993) instead of randomized L4 as shown in 4.2.3. This novel binding interface was common to all OBodies from the second generation (*Pyrobaculum aerophilum* Aspartyl t-RNA synthetase) naïve OBody library (2GPATS) library and the AM library. The crystal condition (containing PEG MME 2000) used to solve this structure (of B7) showed retention of the novel P4 binding interface and (non-covalent) dimerization of the OBody required for P4 binding.

Previous studies of OB-fold (OBody) binding to ligands show binding at the 'Oligomer binding site' (Figure 1.7) (Arcus, 2002; Murzin, 1993; Steemson et al., 2014; Theobald et al., 2003) whereas this P4 binding OBody represents a novel dimeric arrangement and binding interface for small molecule ligands. However, the sequence of loops has a strong influence on protein stability (Hashimoto & Panchenko, 2010). Therefore, variations in binding affinity of OBodies to P4 from both libraries probably reflected the stability provided by the cysteines in 2GPATS library OBodies (in the form of D7) or stability provided by L4 residues (in the form of OBodies selected from AM library) with regards to dimeric arrangement required for formation of hydrophobic pocket for P4 binding. OBody B7 fused to MBP may have improved affinity to P4 over OBody D7 indirectly due to same stabilizing effect and not due to stronger interaction with P4 as the binding interface is theoretically the same (Trivedi et al., 2009). However, the AM library did contribute to elimination of cysteines from OBodies resulting in crystallization with P4.

4.3.3 Structure of OBody bound to P4

Crystallization was challenging with pure B7 probably due to inherent tendency to dimerize thus affecting homogeneity (similar to D7). Crystallization was

successful with fusion to MBP and the crystals diffracted to 1.9 Å resolution with PEG based conditions though they had pathologies like twinning or translational non-crystallographic symmetry (tNCS). The highest resolution, non-twinned data was used for structure solution using Phaser and refinement with Refmac5. The diffraction data used to solve the MBP-B7 structure with P4 does indicate that there is a propensity for the OBodies to dimerize (forming a hydrophobic pocket to contain P4) though selected P4 binding OBodies were randomized on the conventional binding surface not used for P4 binding.

The binding between B7 and P4 is an interesting pattern where two OBodies (B7) come together via a novel binding interface, forming hydrophobic pockets (Figure 4.23) to house two P4 molecules (supported on adjacent tryptophans W472) with two-fold symmetry. This is supported by their MBP counterparts via interactions with α-helices and linkers (4.2.8.2). The end result is the formation of two hydrophobic channels able to accommodate a P4 molecule each with the polar ends free to make interactions with possibly charged amino acids exterior to the hydrophobic channels. The P4 binding to OBody is similar to P4 binding with P4 receptor by aromatic stacking as mentioned in 3.1.1 (Tanenbaum et al., 1998) except that this results in two OBodies forming a non-covalent dimer to bind P4 molecule in a manner not documented before.

Comparison of the solved structure (PEG MME 2000) and the one used for initial model building (PEG 4000) showed different positions of OBody B7 with respect to MBP due to the flexible MBP-B7 linker. However, the position of P4 and packing of the adjacent OBodies was similar and both belong to $p2_1$ space group.

4.3.4 Future experiments

The binding affinity of P4 binding OBodies has plenty of room for improvement as the hydrophobic pockets between the two OBodies can be made more compact by modifying residues on the novel binding interface such that hydrophobic amino acids can be added and charged amino acids removed in such a way as to improve hydrophobicity between the OBodies and for P4. However, charged residues can be placed outside the hydrophobic pocket or existing charged (or polar) amino acids can be brought closer (figure of 8, Figure 4.24) to the carbonyl

groups of the P4 molecule to form strong hydrogen bonds to contribute to overall binding.

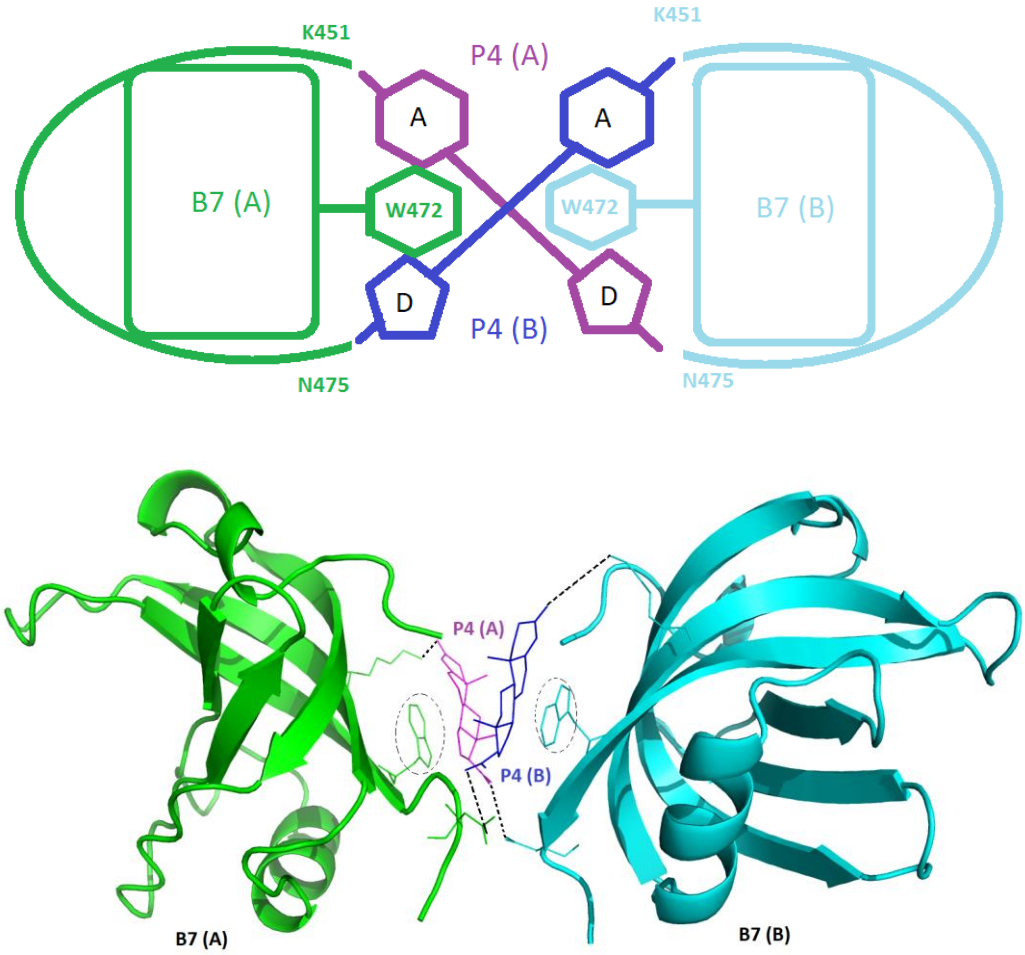


Figure 4.24 Improved binding model of B7 with P4

The schematic 'figure of 8' binding pattern (top) of B7 to P4 is shown with coloured labelling (with relevance to the Pymol generated cartoon model of B7 binding to P4 from solved crystal structure of MBP-B7 with P4, bottom). The schematic shows a side view of the OBody β-barrels (B7 A, green and B7 B, cyan from MBP-B7 A and MBP-B7 B respectively) and the arrangement of P4 molecules (P4 A, purple and P4 B, blue with labelled rings) between them. The steroid rings of the P4 molecules are packed against aromatic rings of tryptophan 472 of MBP-B7 from nearest B7. The orientation of P4 molecules is also shown with the end containing pentagonal ring (ring D) towards the bottom and the end containing hexagonal ring (ring A) towards the top. The model (bottom) follows the schematic where both P4 pack against W472 (dashed circles) of adjacent B7. P4 (A) forms hydrogen bonds with K451 of B7 (A) and N475 of B7 (B) denoted by dashed lines. P4 (B) makes similar interactions with B7 (A) and B7 (B).

W472 involved in aromatic stacking of P4 can be mutated to alanine to analyse its contribution to P4 binding. The amino acids around W472 can be replaced with larger hydrophobic amino acids and tested for P4 binding by SPR. Similarly, the

amino acids near the charged ends of the hydrophobic channel (centered around K451 and N475) can be mutated into serines and analysed for specificity of P4 binding with pH dependence since the difference between P4 and E2 is with respect to their polar groups. Rational design and randomization of this novel binding interface can be used to optimize P4 binding. A radical approach is to add cysteines peripheral to the novel binding interface to form intermolecular disulfides to covalently strengthen the dimerization and resulting trapping of P4 molecules.

5 Characterisation of YscC for phage display

5.1 Introduction

5.1.1 The YscC fragment and phage display

The signal sequence of PelB (Pectate Lyase B) (Bergquist & Gibbs, 2007; Hayhurst, 2000; Singh et al., 2013; Speck, Arndt, & Muller, 2011; Thie, Schirrmann, Paschke, Dubel, & Hust, 2008; Velappan et al., 2010; Wilson & Finlay, 1998; Zhao et al., 2016) is the preferred signal sequence used to translocate proteins using the conventional SecB route and is produced as a fusion with N-terminal end of POI (and pIII) from the encoding phagemid. Similarly, the second generation *Pyrobaculum aerophilum* Aspartyl t-RNA synthetase (2GPATS) naïve library was designed with PelB as a signal sequence at the N-terminal end of the OBody. D7 which was selected against the target progesterone (P4) had a foreign insertion in its open reading frame (ORF) of the encoding phagemid (at the Nco1 cut site) as shown in Figure 5.1 (7.2.1).

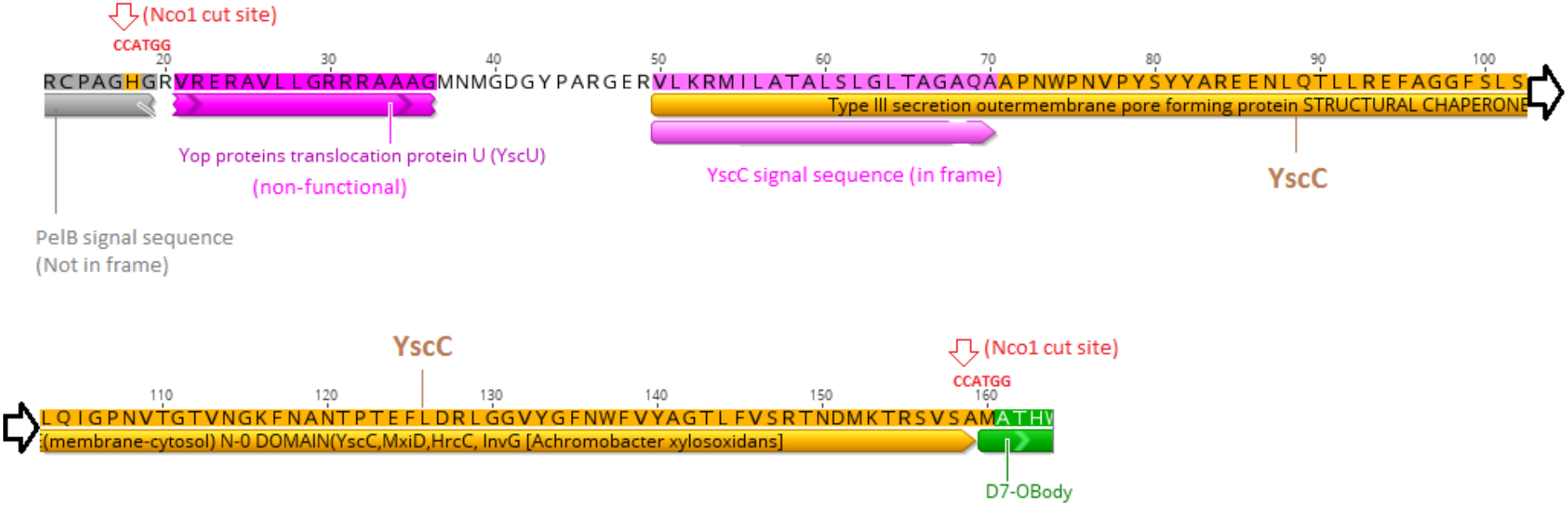


Figure 5.1 Insertion of foreign YscC fragment in (D7) phagemid

Illustration of insertion of foreign YscC fragment between signal sequence PelB (exteme top left, grey) and OBody (D7, extreme bottom right, green) generated by Geneious software. The YscC fragment (pink and orange) shows the short end sequence of YscU domain (non-functional, left pink) and N0 domain of YscC (orange) with signal sequence (right pink) in frame with the OBody (D7). The restriction enzyme cut site (nco1, C^CATGG) where the foreign insertion occurs is also shown (red, top left and bottom right). Note that signal sequence PelB is not in frame with D7 (due to insertion of YscC fragment).

The foreign genomic fragment is from *Achromobacter xylosoxidans* (UniprotKb: A0A179TE40) and was inserted in the phagemid (pOB2) Nco1 restriction site between the 3' end of the PelB signal sequence and the 5' end of the OBody (D7). This foreign fragment has a 3'end partial remnant of the sequence coding for YscU (Yops protein translocation protein U) and a 5'end sequence coding for N0 domain of a type 3 secretin family protein called YscC (InvG/EscC/MxiD)(Zheng et al., 2007). The PelB signal sequence is non-functional since it is out of frame with the sequence coding for D7 (OBody) whereas YscC signal sequence and N0 domain is in frame with D7. YscU is highly unlikely to be functional as it is short and truncated.

D7 was the only OBody from 2GPATS library which was soluble and well expressed in *E. coli*. It was hypothesized that this foreign signal sequence (refererred as ssYscC in this chapter) or the N0 domain of YscC had a positive effect in phage display of D7 due to an alternative route for translocation of stable OBodies (like D7) or efficient translocation of the OBody (Hueck, 1998; Korotkov, Gonen, & Hol, 2011; Kostakioti, Newman, Thanassi, & Stathopoulos, 2005).

5.1.2 The phagemid (YscC) for AM library construction

The phagemid used for construction of a second generation *Pyrobaculum aerophilum* Aspartyl t-RNA synthetase naïve OBody library (2GPATS) is called pOB2. When OBody D7 was selected with P4 as the target antigen, D7 was over-represented in phage selections (screened by phage ELISA and confirmed by DNA sequencing). DNA sequencing of the phagemid revealed an insertion of a foreign segment between the signal sequence (PelB) and the the OBody (D7), consisting of a signal sequence and the N0 domain from secretins, a close homologue of YscC) from *Yersinia.* The secretin (YscC) is a component of the type 3 secretory system (T3SS) and has sequence homology with other gram-negative bacteria, EscC of Entero Pathogenic *E. coli* (EPEC) and InvG of *Salmonella* (Burkinshaw & Strynadka, 2014; Hueck, 1998; Zheng et al., 2007). It was hypothesized that the foreign segment may have enhanced the display of D7 on the surface of phage due to improved transport via the inner membrane due to the signal sequence (Steiner et al., 2006) or the protein domain (Zheng et al., 2007). This may be due to either maintenance of the unfolded form of OBody in

the cytoplasm during delivery to the inner membrane (Baneyx & Mujacic, 2004) or enhancing OBody folding in the oxidising environment of the periplasm (Trivedi et al., 2009). Thus, the AM libraries (pF_35 and L4_DM) were designed with the foreign segment retained in phagemid pOB2 and the phagemid was renamed YscC (based on the T3SS component of *Yersinia*).

The (N0) periplasmic domain of outer membrane (OM) secretin YscC (EscC/InvG) that belongs to type III secretion system (T3SS) of gram-negative bacteria is also called an injectisome and injects virulent proteins into eukaryotic host cells. The injectisome consists of ring structures and a protruding needle. Secretins form one ring and are common to bacterial export pathways like T2SS, T3SS, T4 pilus and filamentous phage extrusion. They consist of 12 – 14 subunits forming a ring structure used to translocate proteins across the outer membrane. The secretin consists of a N-terminal periplasmic domain and conserved protease resistant C-terminal region (secretin homology region) in the outer membrane (Fig 5.2) (Spreter et al., 2009).

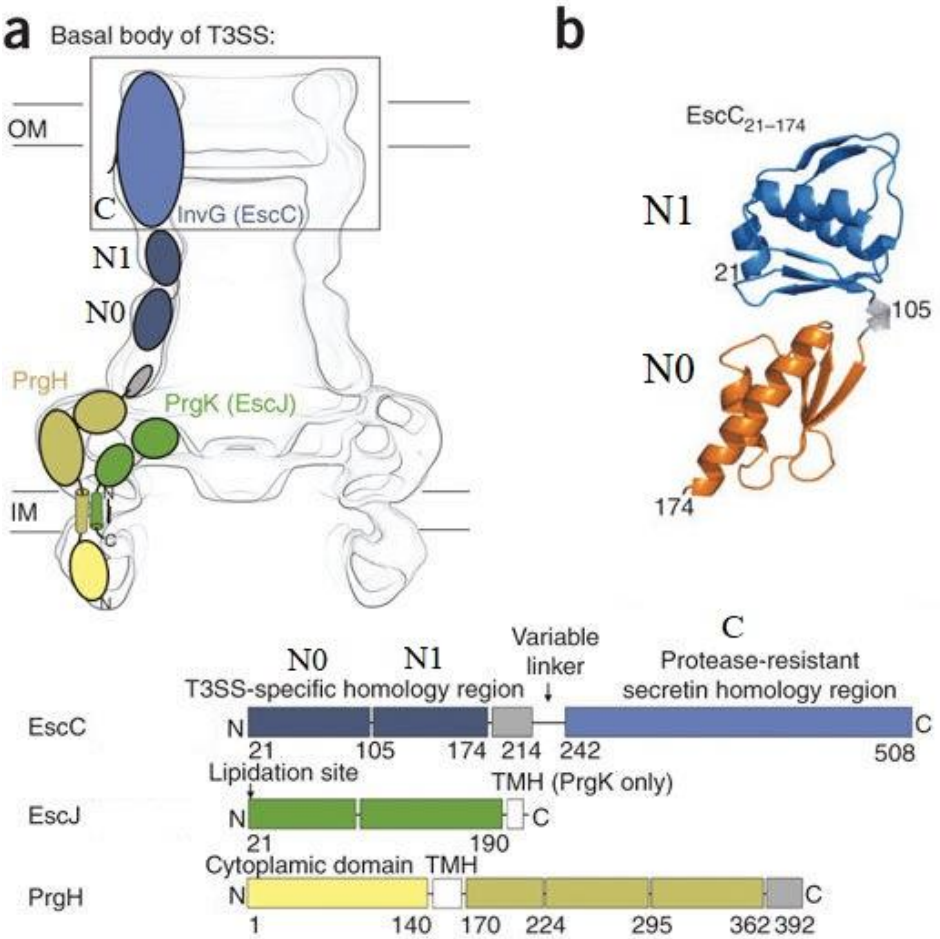


Figure 5.2 Structure of foreign YscC fragment in (D7) phagemid

a) The YscC (InvG/EscC) is shown with its three domains (N0, N1 and C) as part of T3SS. IM components (EscJ and PrgH) are also shown. b) structure of N-terminal domains (21-174) of YscC (EscC) is shown with flexible linker at position 105. Note that EscC and its domains are shown in comparison with the IM components of T3SS.

The crystal structures of N-terminal part of YscC (InvG/EscC) show two domains connected by a flexible linker (105) forming a hydrophobic interface between two domains which makes it possible for the subunits to form a ring structure. The residues in the core of the domains, the linker and the interface are conserved across T3SS species and the modular architecture of InvG overlaps with inner membrane (IM) ring subunits PrgH and PrgK (EscJ).

The N-terminal part of YscC (InvG) has a wedge-shaped structure made up of two α-helices folding against a β-sheet. This conserved feature along with the linker allows subunits of three membrane proteins (InvG, PrgH and PrgK) to form the ring structures of T3SS. The filamentous phage pIV protein is homologous to InvG.

This N-terminal fragment of YscC (InvG) was hypothesized to improve phage display of OBody D7 due to the isolation of single P4 binding OBody. This fragment also contains a unique signal sequence in frame with periplasmic domain (InvG) and OBody D7. Extra-cytoplasmic proteins like InvG contain signal sequences which facilitate translocation across IM and maybe useful for phage display of stable proteins. Phage assembly involves incorporation of minor coat proteins (e.g. pIII) in the inner membrane followed by incorporation into phage. Extra-cytoplasmic proteins in gram-negative bacteria can have either of the following fates after translocation: stay attached to OM, assemble to flagella/pili, injected to host cells or released extra-cellularly (Gagic, Ciric, Wen, Ng, & Rakonjac, 2016).

Phage display involves translocation of pIII fused protein in unfolded form across the IM to successfully display protein due to the requirement of functional disulfide formation in the periplasm. Most pIII based phage display is via its Type I signal sequence that translocates sandwiched proteins using the general secretory (Sec) pathway. Transmembrane α-helices have also shown successful translocation of phage display proteins across the IM. A few proteins have been displayed with (Type 2 and 4) and without signal sequences (Ciric et al., 2014). This was possible due to extra-cytoplasmic proteins with their native targeting

factors cloned in frame with pIII lacking a signal sequence. Most of these targeting factors have a conserved structure but are sequence independent and target proteins to the appropriate chaperones of the translocation systems. Most secreted proteins have signal sequences and transmembrane proteins have hydrophobic transmembrane α-helices. Most of the signal sequences are cleaved by corresponding membrane associated signal peptidases (I, II and IV) after translocation. pVIII has a Type 1 signal sequence whereas pVI, pVII and pIX have only transmembrane α-helices.

5.1.3 Phage display and inner membrane translocation pathways

M13 bacteriophage reproduce in *E. coli* and are released without lysis of the host bacterial cell. This requires that the proteins translated on the ribosome (in the cytoplasm) are translocated to the periplasm through the inner membrane through special export channels as described in 1.7.2. Phage display requires that the protein of interest (POI) to be displayed on the surface of phage are exported through the same export channels. The most important channels are SecYEG translocase belonging to the general secretory (Sec) pathway (which specializes in translocation of unfolded proteins and co-translational translocation via SRP) and Tat translocase belonging to twin arginine translocation pathway (which specializes in export of folded proteins) as shown in Figure 1.20 (Holland, 2004; Natale et al., 2008; Paschke & Hohne, 2005; Saier, 2006; Thie et al., 2008; Valent et al., 1998).

The Sec pathway is generally used for protein translocation in phage display as the display protein pIII has to be maintained in unfolded form in the cytoplasm for translocation into the periplasm for proper folding (due to cysteines in pIII that require native formation of disulfides in the oxidizing environment of periplasm catalyzed by Dsb oxidoreductases)(Baneyx & Mujacic, 2004; Trivedi et al., 2009) and several attempts to translocate pIII using the Tat pathway have been unsuccessful due to pIII misfolding (Nangola et al., 2010). However, the Sec pathway is not a good candidate for translocation of highly stable or fast folding cytoplasmic proteins (e.g. GFP, DARPin) since these proteins will fold after exiting the ribosome (post translation) and prior to binding of SecB chaperone and delivery to Sec translocase (Aronson et al., 2011; Dammeyer & Tinnefeld, 2012; Dinh & Bernhardt, 2011; Fisher & DeLisa, 2008; Huber & Beckwith, 2006). The

Tat pathway intuitively is the best option for these proteins but display protein (pIII) is functional only if translocated in an unfolded form via Sec pathway.

5.1.4 Strategies for highly stable protein translocation

Several research groups have found strategies (Figure 5.3) to enhance export and folding in the periplasm, one being SRP phage display which uses the co-translational branch of Sec pathway (Steiner, Forrer, & Pluckthun, 2008; Steiner et al., 2006) and a second being Tat-mediated phage display (TDP) (Paschke, 2006; Paschke & Hohne, 2005). In these scenarios, SRP bypasses the SecB chaperone and fusion protein is threaded directly into Sec translocase whereas with TDP, c-terminal pIII and the POI are separately translocated through Sec and Tat translocases respectively (via independent expression and appropriate signal sequences, Fig 5.3B) and fuse in the periplasm via respective cysteine-based leucine zippers (Jun and Fos).

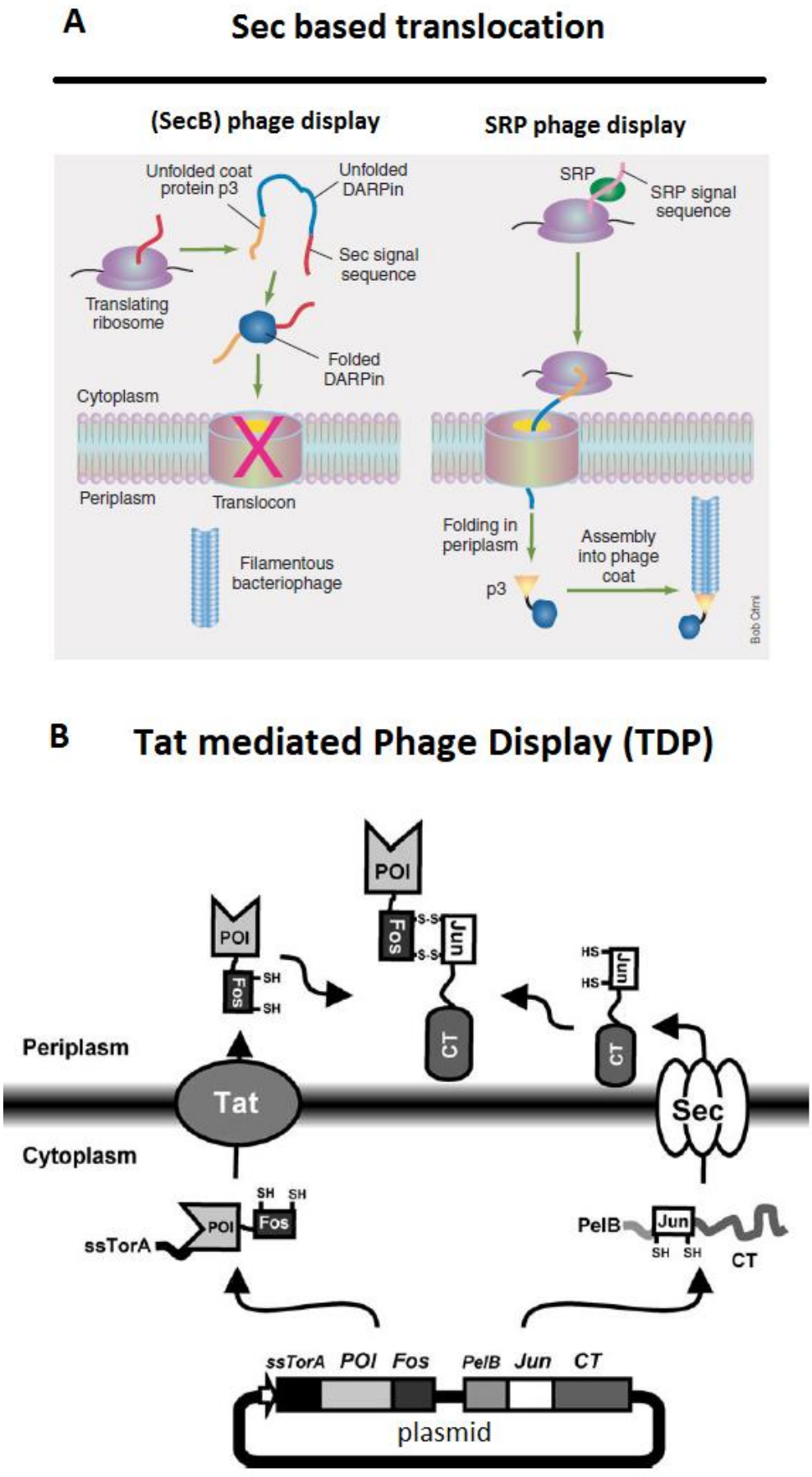


Figure 5.3 Alternative strategies for stable protein translocation

Illustrations for protein translocation strategies are shown. (A) Sec based translocation compares conventional (SecB, post-translational, left) phage display with SRP phage display (co-translational, right) where SecB (not shown here) is incapable of translocating fast folding (DARPin) proteins whereas SRP manages to thread the stable protein directly into Sec translocase. (B) Tat-mediated phage display (TDP) shows separate genes (on plasmid) for POI and c- terminal domain of pIII (CT) which are independently translated (with POI fused to Fos linker and pIII fused to Jun linker) and translocated through Tat and Sec pathways respectively followed by fusion at the periplasm via cysteine based (Fos- Jun) leucine zippers. (Huber & Beckwith, 2006; Paschke, 2006)

In this scenario N0 domain of YscC which has homologs in other species (YscC/InvG/MxiD/EscC; UniprotKb: A0A179TE40) (Burkinshaw & Strynadka, 2014; Diepold & Wagner, 2014; Gaytan, Martinez-Santos, Soto, & Gonzalez-Pedrajo, 2016; Kimbrough & Miller, 2002; Schuch & Maurelli, 2001) is translocated through Sec translocase and is capable of exporting a highly stable protein like the OB-fold (OBody) with pIII circumventing the need to use cysteine based leucine zippers (as in Tat phage display). This also provides a niche where SRP phage display is problematic when highly stable proteins require cofactors to bind them in cytoplasm before translocation into periplasm (Steiner et al., 2006).

This led us to test the hypothesis that the phage selection had fortuitously provided us with a signal sequence that was very efficient at delivering fast folding and stable POIs to the periplasm along with pIII and without the need for parallel export through SEC and TAT pathways.

5.2 Results

5.2.1 Characterisation of YscC fragment

5.2.1.1 Testing YscC for better phage display

This phagemid (named YscC) was compared with the phagemid (pOB2) from 2GPATS library by comparing phage display of the best P4 binding OBody B7 in both phagemid constructs. pOB2-B7 (Figure 5.4) was constructed by removing the YscC fragment (425 bp) from the original B7 phagemid using Nco1 restriction enzyme. Both the constructs were confirmed by DNA sequencing using flanking primers.

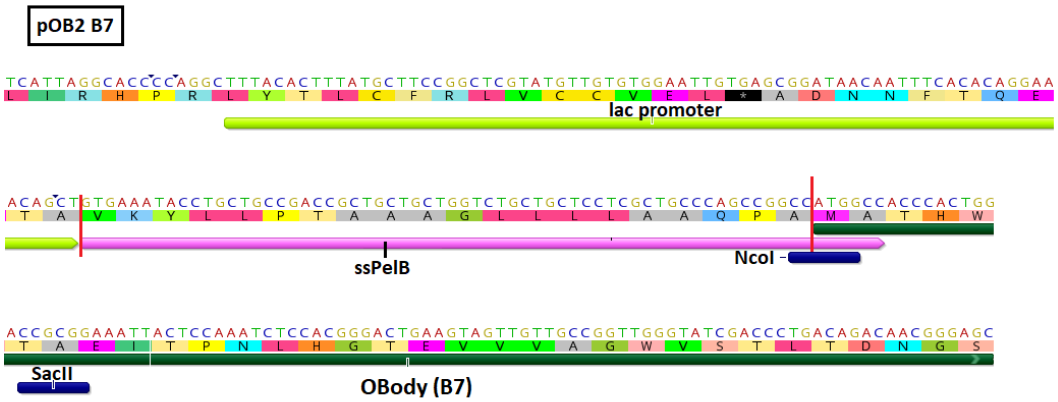


Figure 5.4 DNA sequencing of B7 gene without YscC fragment

The DNA sequence of pOB2 B7 with signal sequence (pink with red borders) of PelB fused to 5'- end of gene for OBody (B7, green) is shown. The restriction enzyme cut sites (blue) are also indicated.

The resulting constructs named YscC-B7 and pOB2-B7 were tested by phage ELISA to compare the performance of phage display from these two constructs. The phage ELISA was performed as described in 2.6.3 and confirmed that the YscC did provide a significant advantage in display of B7 as shown by increased absorbance at 450nm (Figure 5.5)

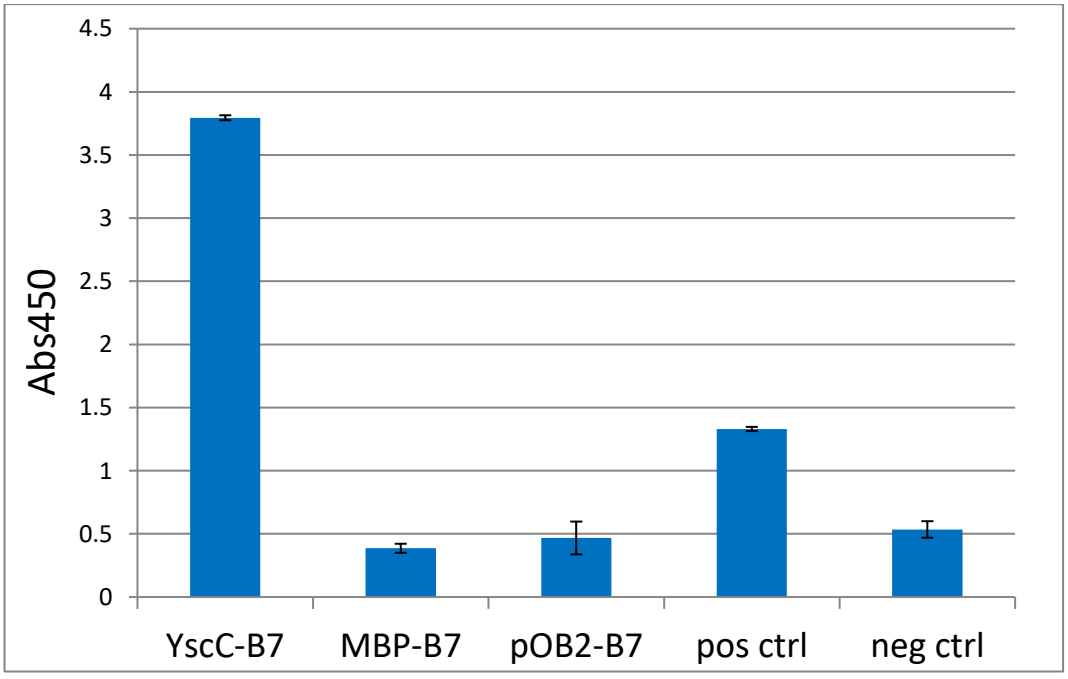


Figure 5.5 Comparison of YscC and pOB2 for phage display of B7

The bar graph shows the phage ELISA of B7 expressed from three different phagemid constructs: pOB2, MBP and YscC. Biotinylated-P4 was immobilized as the target. Anti-P4 antibody was used as the positive control. YscC-B7 was used in an empty well as a negative control. The y-axis shows the absorbance (Abs450) at 450nm which is a function of the display level of B7 on the surface of phage. Note that pOB2 and MBP

share the same signal sequence (PelB) and YscC has its own signal sequence (ssYscC). The experiment was performed in triplicates.

The phage ELISA of B7 with the two constructs shows that YscC produces significantly higher absorbance (Abs450) than pOB2 and positive control confirming that the YscC produces a significant improvement in phage display of OBodies. B7 is known to be well expressed and free of cysteines, thus indicating a possible translocation difference between the two phagemids, with and without YscC. This experiment was performed using KM13 helper phage for infection which has low display of foreign protein fusion to pIII (Carmen & Jermutus, 2002; Paschke, 2006; Rondot et al., 2001) and the results have been replicated many times independently. The next step was to determine whether the signal sequence (ssYscC) or mature domain (of YscC) was responsible for the enhanced phage display of OBody B7 (5.1.1).

Phage display with longer fusion proteins have been shown to be better translocated via the Sec translocase probably due to efficient chaperoning by SecB (Dinh & Bernhardt, 2011). MBP is one such protein which is a naturally occurring periplasmic protein and does not contain cysteines. To test the theory whether YscC was providing the same benefit of a longer fusion protein, a large protein MBP was fused to B7 (in a pOB2 phagemid with PelB as the signal sequence) and compared to YscC-B7. The (pOB2) MBP-B7 construct was confirmed by DNA sequencing. Having MBP at the N-terminus of B7 does not improve display of this protein on the surface of phage (as detected by phage ELISA). The lack of access of B7 for P4 binding in MBP B7 has been ruled out previously with SPR studies in 4.2.4.

5.2.1.2 Testing of signal sequence (ssYscC) for better phage display

A B7 construct ssYscC B7 was made using YscC-B7 to test if the signal sequence (ssYscC) or mature domain of YscC was responsible for better phage display (reflected by strong absorbance at 450nm as in Figure 5.4). The YscC gene coding for mature domain (N0, Figure 5.1) was removed from YscC-B7 using custom order primers leaving only the short signal sequence (ssYscC B7, 21 amino acids) with a non-functional intergenic fragment of 19 amino acids (Figure 5.6).

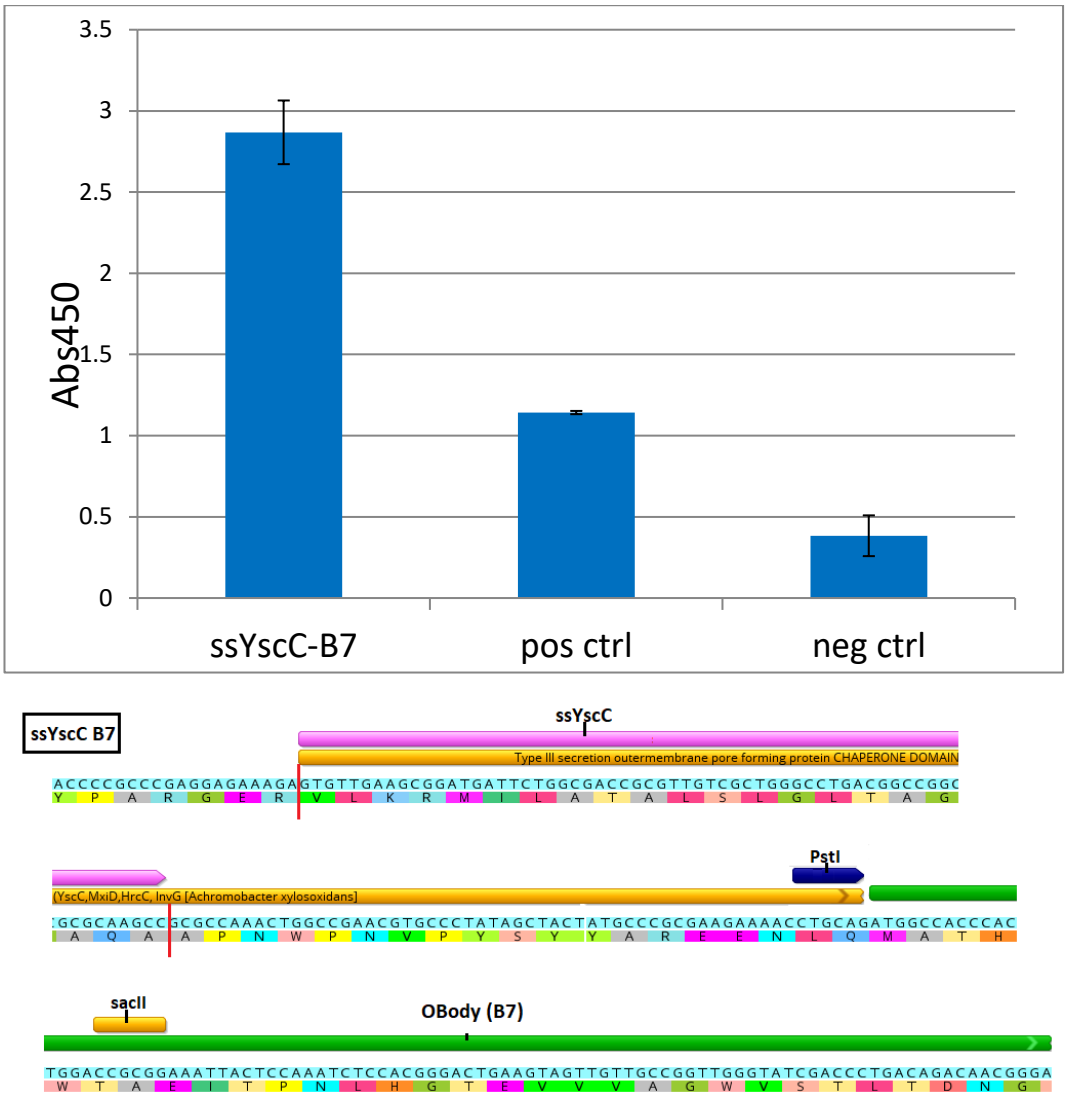


Figure 5.6 Phage display of B7 with signal sequence- ssYscC

The bar graph shows the phage ELISA (top) of B7 expressed on ssYscC B7 which contains only the signal sequence of foreign insertion YscC (ssYscC) fused to B7. Anti-P4 antibody was used as the positive control to detect P4. ssYscC B7 was used in an empty well as a negative control. The y-axis shows the absorbance (Abs450) at 450nm which is a function of the display level of B7 on the surface of phage. The DNA sequence (bottom) of ssYscC B7 showing signal sequence (pink with red borders) of YscC fused to 5'- end of gene for OBody (B7, green) separated by non-functional intergenic fragment (19 amino acids). The restriction enzyme cut site used (Pst1) for removing majority of YscC (yellow) is also indicated. The experiment was performed in triplicates.

The phage ELISA results show that the signal sequence (ssYscC) is responsible for better phage display of OBody B7 reflected by high Abs450nm levels as that of YscC B7 (Figure 5.4) thus ruling out the involvement of N0 domain (YscC) in better phage display.

5.2.1.3 The structure of signal sequence ssYscC

The cleavage site of the signal sequence was predicted using two signal sequence web servers Phobius and Signal-blast as shown in Figure 5.7. The tripartite structure (Figure 1.21) of this signal sequence from Phobius shows no large hydrophobic amino acids in the H-region which would have been indicative of a SRP dependent signal sequence. The presence of two positively charged amino acids in the N-region is highly suggestive of post-translational Sec dependent signal sequence. The lack of twin arginines in the N-region rules out a Tat dependent signal sequence. A few signal sequences have been shown for comparison (Natale et al., 2008)

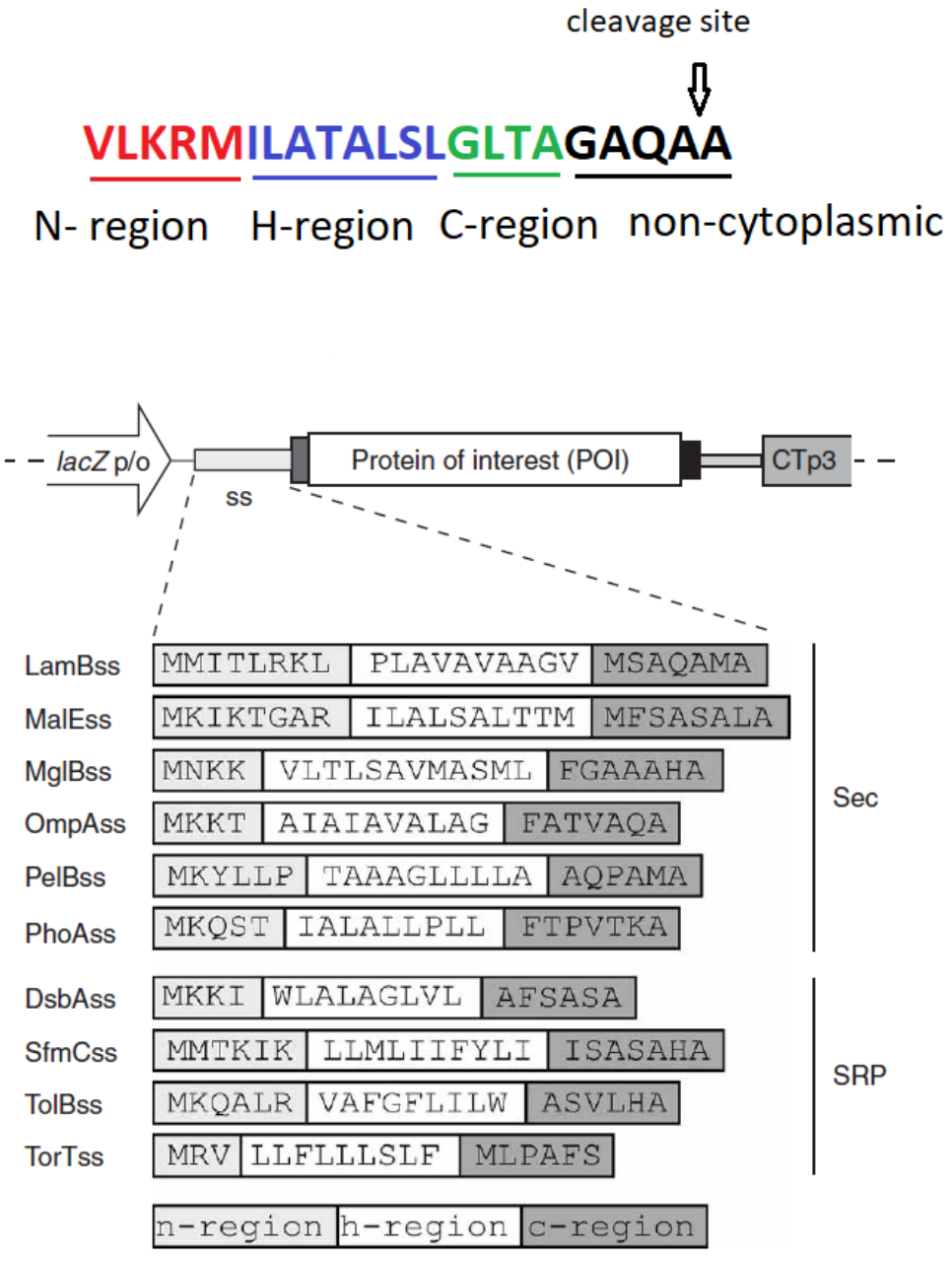


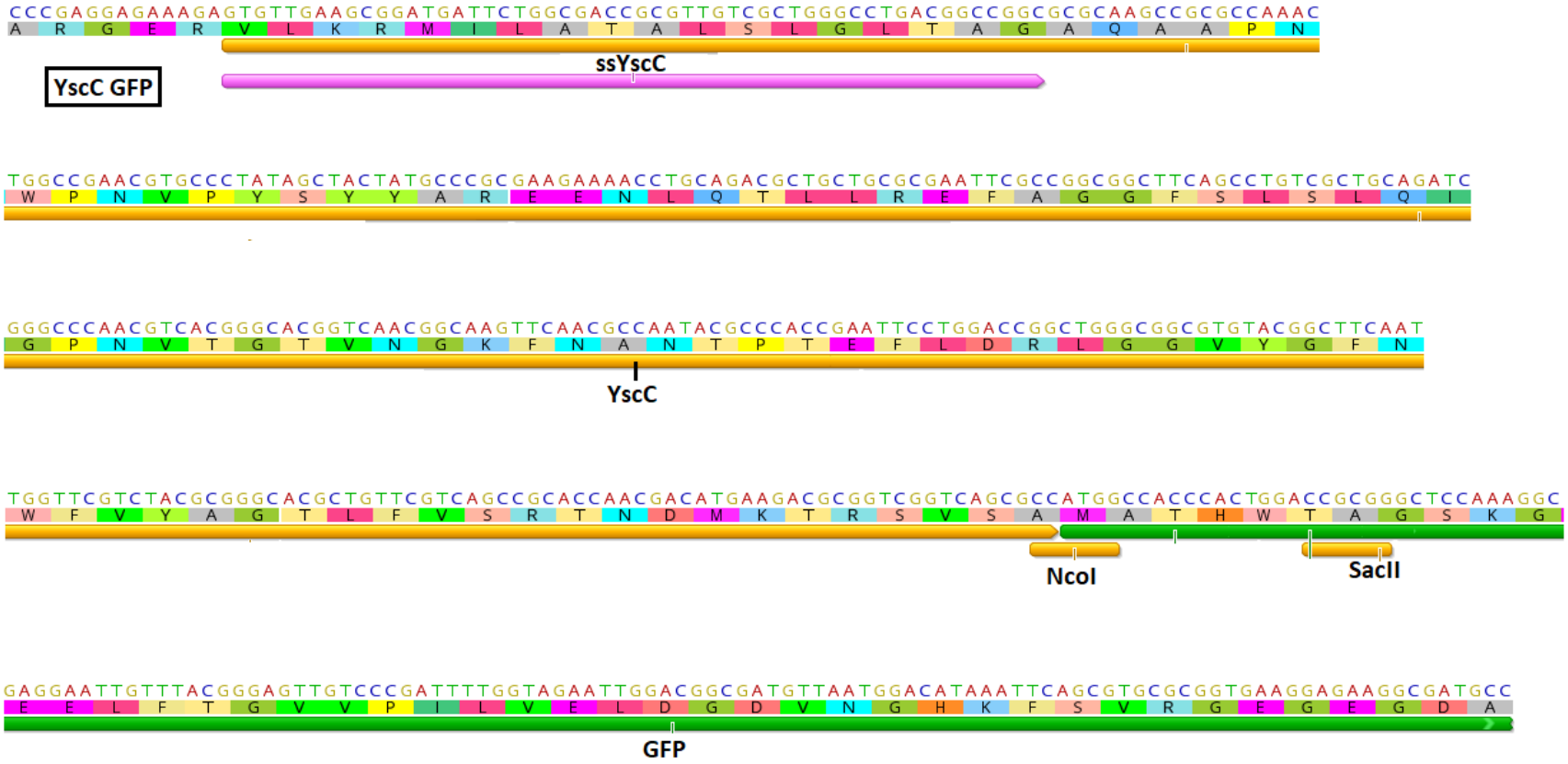
Figure 5.7 Structure and comparison of ssYscC

The tripartite structure of signal sequence of YscC is shown in different colours (top) and compared to conventional signal sequences (bottom). The non-cytoplasmic part (mature domain) and signal peptidase cleavage site is indicated (top). Signal sequences (ss) for Sec and SRP pathway dependent phage display is shown in relation to POI and C-terminal of display protein p3 (bottom). The signal sequence was analysed using Phobius (Stockholm Bioinformatics Centre) and signal-blast (University of Salzburg). The image is used with permission from (Steiner et al., 2006).

5.2.2 Testing YscC for better phage display of other stable (or fast folding) proteins

5.2.2.1 Phage display of highly stable and fast folding GFP with YscC

The signal sequence of YscC (ssYscC) was further explored to characterize the compatibility of phage display with other fast folding or stable proteins that are generally refractory to phage display. Green Fluorescent Protein (GFP) is one such protein which has an *E. Coli* optimized variant (GFPuv) which fluoresces under UV light (Crameri, Whitehorn, Tate, & Stemmer, 1996; Dammeyer & Tinnefeld, 2012; Fukuda, Arai, & Kuwajima, 2000; Paschke & Hohne, 2005) and an engineered fast folding variant (sfGFP)(Aronson et al., 2011; Dinh & Bernhardt, 2011; Fisher & DeLisa, 2008; Speck et al., 2011; Velappan et al., 2010). GFPuv (also called cycle 3 variant) emits green light (508 nm) when excited by UV light (395 nm) forms a stable protein and cannot be translocated to the periplasm with signal sequence pelB of pOB2 (i.e. via the conventional Sec pathway), with minimal translocation by signal sequences of gene III (Linton, Walsh, Sims, & Miller, 2012) and MBP (Feilmeier, Iseminger, Schroeder, Webber, & Phillips, 2000). Thus, constructs pOB2-GFPuv-B7 and YscC-GFPuv-B7 were constructed (Figure 5.8) to test for better phage display with the YscC fragment. These constructs were confirmed by DNA sequencing.



(The legend is provided in next page)

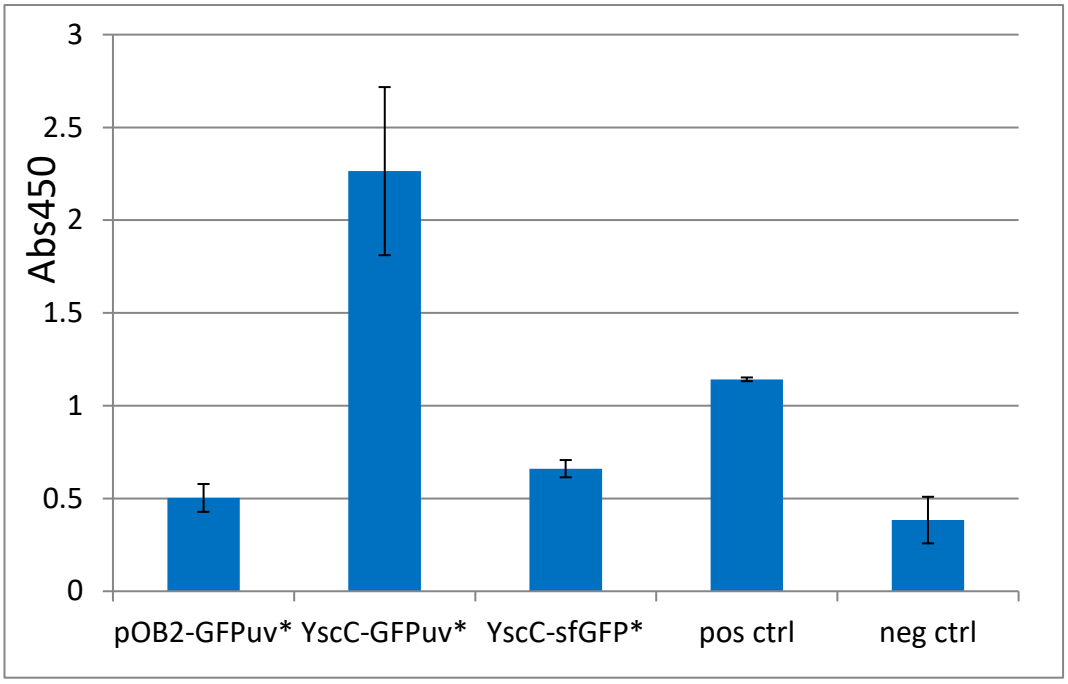


Figure 5.8 Phage display of GFPuv with YscC fragment

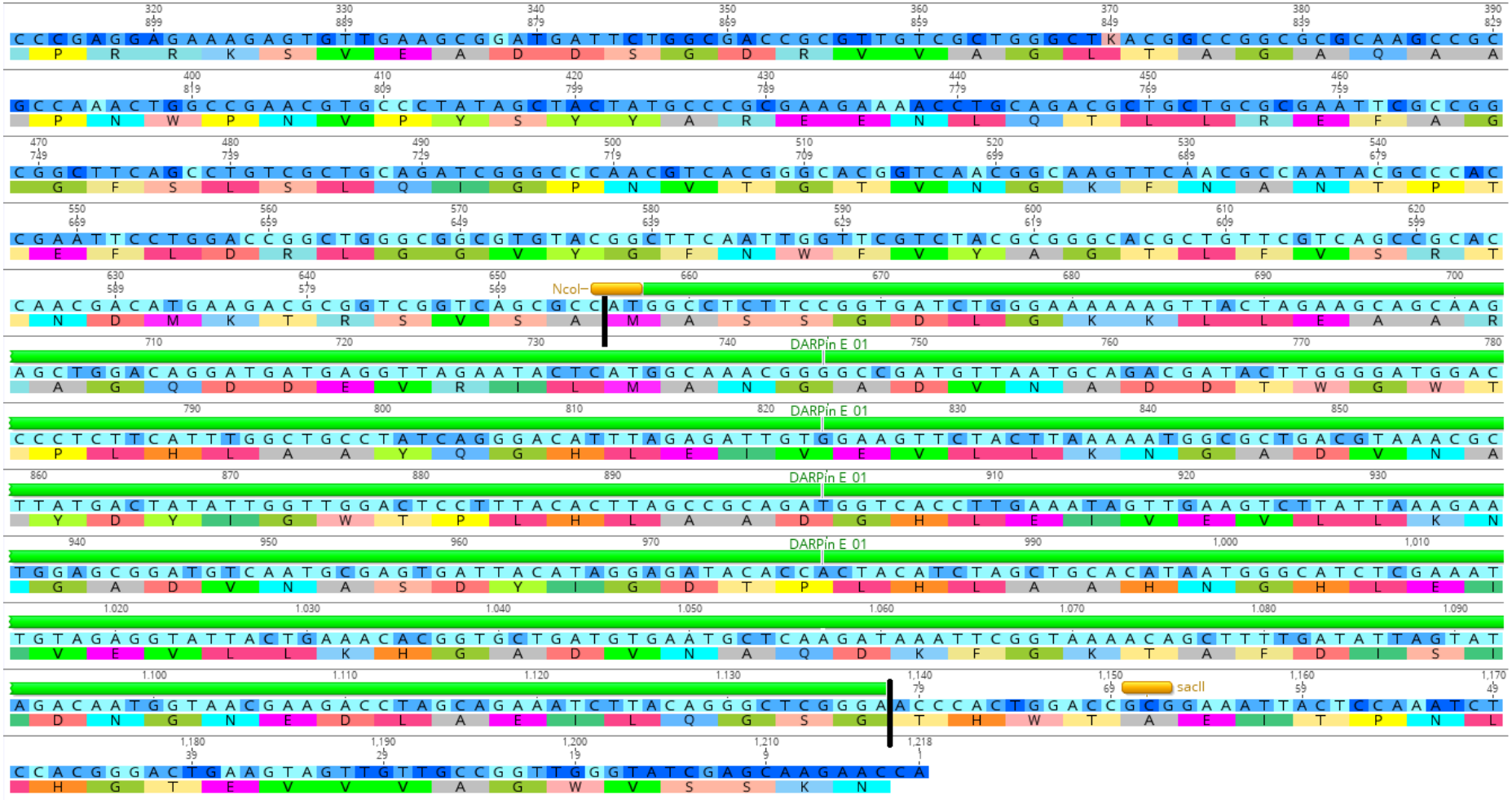
The bar graph (top) shows the phage ELISA of GFPuv (with B7* to detect P4) expressed on two different phagemid constructs (pOB2 and YscC) and YscC-sfGFP-B7. Anti-P4 antibody was used as the positive control to detect P4. YscC-B7 was used in an empty well as a negative control. The y-axis shows the absorbance (Abs450) at 450nm which is a function of the display level of B7 on the surface of phage. Note that pOB2 has its own signal sequence (PelB) and YscC has its own signal sequence (ssYscC). The DNA sequence (p.170) of YscC GFP showing previously annotated YscC (yellow, as in Figure 5.1) fused to 5'- end of gene for GFP (green). The restriction enzyme cut sites (Nco1 and Sac2) are also indicated. This experiment was performed in triplicates.

The phage ELISA of GFPuv with the two constructs shows that YscC produces up to 5 times higher absorbance (Abs450) than pOB2 confirming that the YscC produces a significant improvement in phage display of GFPuv.

However GFPuv has cysteines and is non-fluorescent if folding occurs in the periplasm due to formation of intermolecular disulfide bonds (similar to the OBody D7)(Aronson et al., 2011). The fast folding mutant super-folder GFP (sfGFP) folds fast enough to prevent cysteines from forming intermolecular disulfide bonds and is fluorescent in the periplasm. Thus, similar constructs were made with sfGFP as GFPuv (Figure 5.7) and efforts for phage display produced only marginally improved display with the YscC fragment. This was probably due to the incompatibility of the Sec system with folded proteins.

5.2.2.2 Phage display of highly stable but not fast folding DARPin with YscC

Further exploration of the signal sequence of YscC was limited to highly stable proteins like DARPins (with the exclusion of fast folding variants like sfGFP) (Binz et al., 2004; Stumpp et al., 2008; Wetzel, Settanni, Kenig, Binz, & Pluckthun, 2008). Designed Ankyrin Repeat Proteins (DARPins) are non-immunoglobulin based scaffolds that do not contain cysteines (Binz et al., 2005; Jost & Pluckthun, 2014) and are refractory to conventional (post-translational) Sec based translocation (Huber & Beckwith, 2006; Steiner et al., 2008; Steiner et al., 2006). Thus, a YscC based construct was designed similar to the GFP constructs (5.2.2.1, GFPuv and sfGFP) and phage ELISA was performed to assess phage display of the highly stable DARPin (E_01) with B7 as a detecting epitope (to P4, Figure 5.9). These constructs were confirmed by DNA sequencing using flanking primers. It is important to note that cloning of solely the signal sequence (ssYscC) was not achieved with GFP and DARPin, thus resorting to use of the entire YscC N-terminal fragment to these constructs.



(The legend is provided in next page)

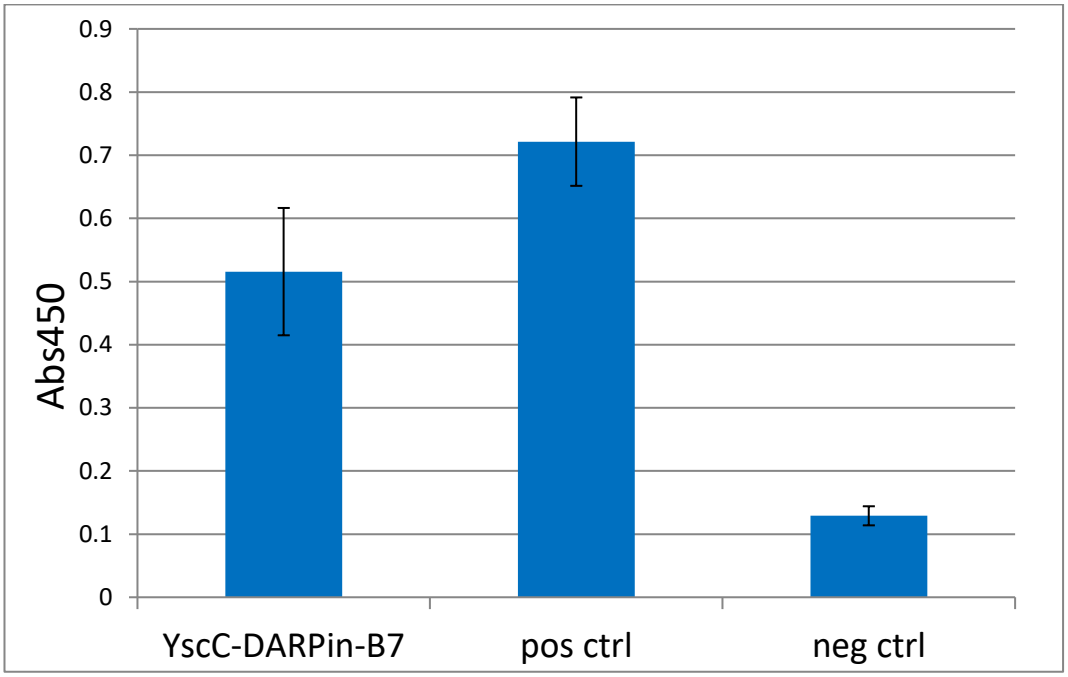


Figure 5.9 Phage display of DARPin with YscC

The bar graph (top) shows the phage ELISA of DARPin (with B7 to detect P4) expressed on phagemid construct YscC B7. Anti-P4 antibody was used as the positive control to detect P4. ssYscC B7 was used in an empty well as a negative control. The y-axis shows the absorbance (Abs450) at 450nm which is a function of the display level of B7 on the surface of phage. This experiment was performed in triplicates. The DNA sequence (p.173) of YscC DARPin showing annotated DARPin (green with black borders) fused at 5' end to YscC and 3' end to OBody B7. The restriction enzyme cut sites (Nco1 and Sac2, orange) are also indicated. Note that the location of restriction sites is different as compared to GFP.

Phage display of DARPin with YscC fragment did show moderate levels of display but needs to be repeated using only the signal sequence (ssYscC) since the functionless mature domain (N0) may interfere with Sec based translocation across the inner membrane.

5.3 Discussion

The original P4 binding OBody D7 was used as a template for the design of AM libraries. The YscC fragment was retained in the construction of AM libraries because it was hypothesized to aid phage display of OBodies. Phage selections using these libraries led to the isolation of a significant number of P4 binding OBodies possibly due to better phage display due to YscC fragment and predominantly from L4_DM due to larger nucleotide randomization (though not directly involved in P4 binding).

The influence of the YscC fragment was further characterized by constructing OBody B7 constructs with only the ssYscC and this was confirmed by experimentation: the signal sequence was responsible for the enhanced phage display of OBodies compared to the conventional pelB signal sequence, probably owing to efficient translocation of highly stable OBodies via the post-translational Sec pathway. The co-translational route of Sec pathway was ruled out as the signal sequence is not sufficiently hydrophobic to be recognized by the SRP (Fekkes & Driessen, 1999). This idea of efficient display of OBodies was further tested with other highly stable proteins such as GFP and DARPin which proved that ssYscC holds promise to efficiently display stable proteins other than OBodies. However, these experiments need to be repeated without the YscC fragment to explore the potential of the signal sequence in displaying stable proteins other than OBodies.

There are at least two other technologies that can efficiently display stable or fast folding proteins, SRP phage display and Tat phage display. SRP phage display uses the co-translational route of the Sec pathway and is not ideal for proteins that require cofactors prior folding (in the cytoplasm). Similarly Tat phage display exploits separate translocation of C- terminal pIII via the Sec pathway and POI via the Tat pathway followed by fusion in the periplasm via cysteine-based leucine zippers. However, this is incompatible with proteins containing cysteines due to risk of aberrant inter-molecular disulfide bonds. Thus, ssYscC which uses the conventional post-translational Sec pathway creates a niche which can overcome both the above limitations as it translocates unfolded proteins and also exports pIII fused to the POI without the requirement of cysteines.

The signal sequences can be modified to improve efficiency of translocation via the Sec pathway or stability of POI (Martoglio & Dobberstein, 1998; Singh et al., 2013). The signal sequences are in the folded form prior to translocation and α- helices are translocation competent (Knyazev, Kuttner, Zimmermann, Sobakinskaya, & Pohl, 2018). β-strand based proteins are poorly translocated via the bacterial Sec pathway unless they are preceded by an α- helix (Gonsberg et al., 2017). The next step would be to create an affinity maturation library where the signal sequence of YscC is randomized based on mRNA secondary structure, H-region and cleavage site (Jeiranikhameneh, Moshiri, Keyhan Falasafi, &

Zomorodipour, 2017). The fused stable protein that is refractory to phage display is used as a quality control to isolate ssYscC variants that efficiently translocate these proteins via Sec pathway and bind to established targets (Feilmeier et al., 2000; Heggeset et al., 2013).

6 Discussion

Protein based bio-sensors generally have higher sensitivity and specificity compared to small molecules sensors (A. J. Smith, 2015). Antibodies are the dominant proteins used for diagnostics and therapeutics in this context. They are expensive to produce and often require expensive storage and transport conditions. Their expensive production arises as there is a requirement for mammalian cell lines to facilitate correct folding and correct post- translational modifications (Li, Vijayasankaran, Shen, Kiss, & Amanullah, 2010). Different formats of antibodies and alternatives to antibodies from naturally occurring binding proteins have been engineered to get around these limitations (Gebauer & Skerra, 2009; Jost & Pluckthun, 2014; Nygren & Skerra, 2004). OBodies, based on the OB-fold, are a possible non-immunoglobulin scaffold as they fold into a stable β- barrel with high melting temperature (T_m) and can be produced in *E. coli* with an adaptive binding surface leading to the ability to produce bio-sensors that can be stored and transported at room temperature and produced in large amounts in bacteria thus reducing the costs of production.

OBodies (~13 kDa) also have a size advantage over antibodies (150 kDa) making it more applicable for intracellular applications like carrying a chemical payload or probe for radioimaging. Small molecule targets are mostly hydrophobic or polar and have to rely on limited interactions for binding to sensors unlike large protein targets (Cooper, 2000). A smaller biosensor like OBodies will be an advantage over a large biosensor like antibodies due to lower steric clashes (Du et al., 2016).

Progesterone (P4) is a ligand for an extra-cellular receptor and is similar to the environment of P4 binding OBodies displayed on the surface of phagemid particles during phage selections (Tanenbaum et al., 1998). Anti-P4 antibodies are ubiquitous in diagnostic testing but they are expensive to manufacture and require cold storage to prevent the antibody from aggregation/denaturation (Isobe & Nakao, 2003). Non-immunoglobulin scaffolds are good alternatives to produce low cost and thermostable bio-sensors with anticalins based on the lipocalin fold the only known non-immunoglobulin scaffold to bind small molecules (Arne Skerra, 2001; A. Skerra, 2008).

The OBody is another similar scaffold which has shown potential in being a non-immunoglobulin based small molecule bio-sensor with thermostability and ease of manufacturing. Aptamers based on single stranded oligonucleotides (DNA or RNA) are other alternatives to antibodies and non-immunoglobulin scaffolds as small molecule binding agents (Ruscito & DeRosa, 2016). However, aptamers are negatively charged, targeted by nucleases and are heat labile positioning non-immunoglobulin protein-based scaffolds as ideal small molecule bio-sensors.

Low levels of P4 in cow's milk, days 5 to 7 after artificial insemination has been shown to be correlated with low fertility (Stronge et al., 2005). Enzyme immunoassay (EIA) is the gold standard for testing P4 levels in cow's plasma or milk over radioimmunoassays (Isobe & Nakao, 2003). P4 binding OBodies are compatible with a similar system (ELISA) to detect P4 and can be a powerful alternative to EIA based on cumbersome antibodies.

6.1 Characterisation of OBody D7 from naïve library

An OBody (D7) was previously engineered to detect the small molecule progesterone (P4). Structural studies of this OBody with P4 would show the nature of binding and a rationale to improve binding affinity (K_d) by rational design. X-ray crystallography was planned to visualise protein crystals to obtain an atomic resolution model of the protein and binding characteristics in atomic detail. However, the OBody D7 failed to crystallize possibly owing to cysteines forming inter-molecular disulfide bonds and aggregation at high concentrations (Dale et al., 2003; Gosavi et al., 2008; Szilágyi et al., 2007). Therefore, mutants of the OBody D7 were designed to test the contribution of cysteines to binding with the hope of eliminating cysteines and increasing crystallizability. The loop (L4) mutants where cysteines were replaced with uncharged serines showed loss of binding to P4 in binding ELISA whereas similar β- sheet mutants did not affect binding but did show a relative decrease in dimerization when visualised by SDS-PAGE and size exclusion chromatography. Crystallization trials with these serine D7 mutants and MBP fused variants to increase crystallizability were not successful and it was planned to construct new OBody libraries from D7 to eliminate cysteines or increase binding affinity by directed evolution using phage display. Since D7 (K_d ~ 600 nM – 2.7 μM) shows at best 10 times weaker binding to P4 than an anti-P4 antibody (K_d ~ 60 nM), the construction of affinity

maturation libraries was considered the next step to make P4 binding OBodies competitive (Benvenuti & Mangani, 2007; Bondos & Bicknell, 2003; Clementi et al., 2012; Dufner et al., 2006).

The presence of cysteines in the P4 binding OBody D7 and the oxidising environment of the periplasm where it was assembled during phage display can be tested further by expressing large scale in an *E.coli* strain like AD494 (lacking *trxB*) or origami (lacking *trxB* and *gor*) from Novagen with an oxidising cytoplasm or signal sequence-based periplasmic targeting and expression, to test for retention of solubility and P4 binding from oxidation of cysteines unlike the reducing environment of cytoplasm.

6.2 Construction of AM libraries and characterisation of OBody B7

Construction of two affinity maturation libraries, each focusing on two locations containing cysteines on the hypothesized binding surface of OBody D7, one on β-sheets (pF) and the other on loop L4 were carried out (Arcus, 2002; Murzin, 1993; Steemson et al., 2014; Theobald et al., 2003). These libraries used appropriate D7 mutants as templates for randomisation of binding residues as a balanced strategy to improve binding affinity to P4 or eliminate cysteines without loss of binding affinity respectively. Phage display of these library clones and selected for P4 binding, isolated L4 based OBodies that were in the same range of binding affinity as OBody D7. Crystallization trials with the best P4 binding OBody from AM library (L4_DM), B7 which had similar binding affinity range as naïve OBody D7 (K_d ~ 700 nM when fused to MBP, standalone K_d ~ 2.6 – 3.8 μM, 4.2.6 and 4.2.4) and expressed well without cysteines showed that the OBodies bound to P4 in a novel way with a different binding surface. The formation of dimer during binding to P4 explains the possible tendency for dimerization of OBody D7 and lack of increase in binding affinity of OBodies from AM libraries. There was no consensus in the loop (L4) sequences of isolated OBodies from AM libraries confirming the lack of role of the conventional binding surface of the OB-fold in P4 binding. The lack of OBodies not binding to P4 (with the same P4 binding surface) may indicate the stability provided by certain loop sequences for OBody dimerization to bind P4 (Hashimoto & Panchenko, 2010). There was a correlation between solubility of protein and P4 binding as observed with D7

mutants (C81S, C85S) and AM Obodies that bound weakly, as they did not express well in *E. Coli.* However, this does not explain the loss of P4 binding with removal of loop 4 cysteines or the presence of cysteines in the opposite side of the P4 binding interface in OBody D7.

The AM bodies were tested for binding with SPR and interestingly the MBP fusion of B7 (which crystallized) showed a significant improvement in P4 binding possibly due to contributions from MBP- B7 linker to the hydrophobic packing of P4 molecules (Figure 4.20). This observation along with improvement of OBody D7 binding to P4 with decreasing pH shows that the unique binding surface has the potential to be improved further for strengthening binding affinity to P4. The OBody can be made as competitive as an antibody by construction of second generation (2^{nd}) AM libraries focused on randomisation of this unique P4 binding surface such that the figure of 8 model can be completed by binding of two P4 molecules between two OBodies. This can be achieved by improving hydrophobic packing of P4 molecules between the two OBodies and enhancing charged interactions between the carbonyl groups of P4 with that of nearby residues of participating OBodies.

The crystal packing of OBody B7 forming a dimer with the unique P4 binding surface can be confirmed by point mutations of the novel binding surface and testing retention of P4 binding and change in position of P4 in the crystal structure. The crystal packing of two OBodies is a possibility in its real environment on the end of each phagemid particle as it has five copies of pIII able to display atleast two adjacent pIII fusions with OBodies.

6.3 Characterisation of YscC

The selection of OBody D7 to ligand progesterone (P4) introduced a foreign gene insertion (the YscC fragment) from Yersinia encoding Ysc proteins required for assembly of Sec independent type III secretion system (Burkinshaw & Strynadka, 2014; Hueck, 1998; Zheng et al., 2007). This was incorporated into the construction of new affinity maturation libraries with sufficient diversity (pF_35 and L4_DM, 10^{12} members) using OE- PCR. The YscC fragment significantly improved the display of OBodies (~ 10^8 members/ml) compared to conventional

Sec based PelB signal sequence, where only one OBody (D7 with YscC fragment) was selected from naïve library selections.

Further exploration of the effects of the YscC fragment showed that it was the YscC signal sequence which enhanced phage display of stable proteins like OBodies possibly by improved translocation through the inner membrane using the same Sec system as PelB (Burkinshaw & Strynadka, 2014). This was further confirmed with improved phage display of other stable proteins such as GFP and DARPins, indicating a competitive technology to display cysteine-based proteins or rival Tat phage display and SRP display of stable proteins.

The improved phage display of OBodies by the signal sequence of YscC (InvG/EscC) can be explored further by improving the experimental conditions of phage ELISA. CsCl purification of rescued phagemid particles displaying the OBody can eliminate false positive results from fragments of *E.coli* containing OBody and virion proteins (like pVIII), formed from potential toxicity (Barbas, Burton, & Silverman, 2004). It is also important to titrate the phagemid particles displaying OBody with ssYscC and ssPelB to control for equal number of OBody displaying phagemid particles with both signal sequences, to rule out propagation based advantage in phagemid particle production (Vodnik et al., 2011).

6.4 Future experiments

The novel binding interface involved in binding to P4 has to be tested rigorously before optimising this surface for further improvement in P4 binding. Mutating some key amino acids like W472 can further increase our confidence in affecting the binding nature of the hydrophobic channels formed between OBodies. Next, the dimeric requirement of OBodies for P4 binding can be tested by adding large amino acids adjacent to the P4 binding interface to prevent dimerization and verify retention of P4 binding. Once the novel binding interface and requirement of dimerization for P4 binding can be validated, a second affinity maturation library can be designed with OBody B7 as a template.

Further improvement in P4 binding requires randomization of this novel binding surface (in B7) and phage selections to isolate stronger P4 binding OBodies. This will involve optimising availability of hydrophobic amino acids for aromatic packing and removal of charged or polar amino acids from the hydrophobic

channels that accommodate P4. Since there are no loops in this side of the OBody, most efforts of randomisation will be focused on β-sheet 4 and β-sheet 5. The best binding scenario of two OBodies held together by two P4 molecules in a crossed (figure of 8) pattern can also be made possible by modifying the residues in the P4 binding site of the OBodies such that the hydrophobic packing between P4 molecules and tryptophan 472 is improved with large hydrophobic amino acids and two P4 molecules made to bind better to two OBodies (if required) by hydrogen bond formation as shown in Figure 4.24. Addition of cysteines to fuse two OBodies together at the novel binding interface is also an option.

The inherent tendency of these OBodies to dimerize can be exploited to engineer bi-specific OBodies that bind to two different targets or stronger binding OBodies that bind to two different epitopes of the same target similar to other non-immunoglobulin scaffolds like Anticalins (Böhm, 2005; Nygren & Skerra, 2004) or DARPins (Boersma, Chao, Steiner, Wittrup, & Pluckthun, 2011; Jost & Pluckthun, 2014). Further characterization of these OBodies by SPR and crystallization trials will help in realizing OBodies as a robust alternative to antibodies to bind small molecules.

The 2GPATS library has not been as successful as 1GPATS library for selection of target (P4, LH, Wee1 kinase and MK13) binding OBodies or their crystallisation. It will be interesting to conduct phage selections with the 1GPATS naïve library against target P4 to determine whether OBodies binding to P4 at the conventional HEL binding surface can be isolated and crystallized, as the 2GPATS based OBodies may have an inherent tendency to dimerize hydrophobically.

With regards to the signal sequence of YscC, it is important to assess the massive improvement in phage display of only one type of stable protein (OB-folds) over other stable proteins (sfGFP, DARPin) with a similar fold called SH3 folds (Agrawal & Kishan, 2001).It is important to test other potential stable proteins for efficient phage display using this signal sequence. It is important to test different signal sequences (including Srp and Tat) with phage display of OBodies as translocation across the inner membrane may be a bottleneck preventing efficient display on the surface of phagemid particles.

Since signal sequences vary in translocation ability using the same translocation pathways, it may be possible to randomize the signal sequence and test for efficient phage display of a common stable protein (sfGFP) which can be detected by increased fluorescence (Feilmeier et al., 2000; Martoglio & Dobberstein, 1998; Singh et al., 2013). This affinity maturation of the signal sequence can also be combined with affinity maturation of the P4 binding OBody where improved display and binding may allow enrichment of stronger P4 binding OBodies.

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D, 66*(2), 213-221. doi:doi:10.1107/S0907444909052925

Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., & Urzhumtsev, A. (2013). Bulk-solvent and overall scaling revisited: faster calculations, improved results. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 4), 625-634. doi:10.1107/s0907444913000462

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., . . . Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr, 68*(Pt 4), 352-367. doi:10.1107/s0907444912001308

Afonine, P. V., Grosse-Kunstleve, R. W., Urzhumtsev, A., & Adams, P. D. (2009). Automatic multiple-zone rigid-body refinement with a large convergence radius. *Journal of Applied Crystallography, 42*(4), 607-615. doi:doi:10.1107/S0021889809023528

Agrawal, V., & Kishan, R. K. V. (2001). Functional evolution of two subtly different (similar) folds. *BMC Structural Biology, 1*, 5-5. doi:10.1186/1472-6807-1-5

Arcus, V. (2002). OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr Opin Struct Biol, 12*(6), 794-801.

Aronson, D. E., Costantini, L. M., & Snapp, E. L. (2011). Superfolder GFP is fluorescent in oxidizing environments when targeted via the Sec translocon. *Traffic, 12*(5), 543-548. doi:10.1111/j.1600-0854.2011.01168.x

Baneyx, F., & Mujacic, M. (2004). Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol, 22*(11), 1399-1408. doi:10.1038/nbt1029

Barbas, C. F., Burton, D. R., & Silverman, G. J. (2004). *Phage Display: A Laboratory Manual*: Cold Spring Harbor Laboratory Press.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography, 67*(Pt 4), 271-281. doi:10.1107/S0907444910048675

Benvenuti, M., & Mangani, S. (2007). Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nat Protoc, 2*(7), 1633-1651. doi:10.1038/nprot.2007.198

Bergquist, P. L., & Gibbs, M. D. (2007). Degenerate oligonucleotide gene shuffling. *Methods Mol Biol, 352*, 191-204. doi:10.1385/1-59745-187-8:191

Berkmen, M. (2012). Production of disulfide-bonded proteins in Escherichia coli. *Protein Expression and Purification, 82*(1), 240-251. doi:https://doi.org/10.1016/j.pep.2011.10.009

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., . . . Owens, R. J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res, 35*(6), e45. doi:10.1093/nar/gkm047

Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., . . . Pluckthun, A. (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol, 22*(5), 575-582. doi:10.1038/nbt962

Binz, H. K., Amstutz, P., & Pluckthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nat Biotechnol, 23*(10), 1257-1268. doi:10.1038/nbt1127

Bodmer, D. M., Tiefenauer, L. X., & Andres, R. Y. (1989). Antigen- versus antibody-immobilized ELISA procedures based on a biotinyl-estradiol conjugate. *Journal of Steroid Biochemistry, 33*(6), 1161-1166. doi:10.1016/0022-4731(89)90425-1

Boersma, Y. L., Chao, G., Steiner, D., Wittrup, K. D., & Pluckthun, A. (2011). Bispecific designed ankyrin repeat proteins (DARPins) targeting epidermal growth factor receptor inhibit A431 cell proliferation and receptor recycling. *J Biol Chem, 286*(48), 41273-41285. doi:10.1074/jbc.M111.293266

Bohling, S. D., Wittwer, C. T., King, T. C., & Elenitoba-Johnson, K. S. (1999). Fluorescence melting curve analysis for the detection of the bcl-1/JH translocation in mantle cell lymphoma. *Lab Invest, 79*(3), 337-345.

Böhm, H. J. (2005). Prediction of Non-bonded Interactions in Drug Design *Protein-Ligand Interactions* (pp. 3-20): Wiley-VCH Verlag GmbH & Co. KGaA.

Boivin, S., Kozak, S., & Meijers, R. (2013). Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr Purif, 91*(2), 192-206. doi:10.1016/j.pep.2013.08.002

Bondos, S. E., & Bicknell, A. (2003). Detection and prevention of protein aggregation before, during, and after purification. *Analytical Biochemistry, 316*(2), 223-231. doi:10.1016/s0003-2697(03)00059-9

Bornhorst, J. A., & Falke, J. J. (2000). [16] Purification of Proteins Using Polyhistidine Affinity Tags. *Methods in enzymology, 326*, 245-254.

Burkinshaw, B. J., & Strynadka, N. C. (2014). Assembly and structure of the T3SS. *Biochim Biophys Acta, 1843*(8), 1649-1663. doi:10.1016/j.bbamcr.2014.01.035

Carmen, S., & Jermutus, L. (2002). Concepts in antibody phage display. *Brief Funct Genomic Proteomic, 1*(2), 189-203.

Celej, M. S., Montich, G. G., & Fidelio, G. D. (2003). Protein stability induced by ligand binding correlates with changes in protein flexibility. *Protein Science : A Publication of the Protein Society, 12*(7), 1496-1506.

Cha-aim, K., Hoshida, H., Fukunaga, T., & Akada, R. (2012). Fusion PCR via Novel Overlap Sequences. In J. Peccoud (Ed.), *Gene Synthesis: Methods and Protocols* (pp. 97-110). Totowa, NJ: Humana Press.

Chan, V., Dreolini, L. F., Flintoff, K. A., Lloyd, S. J., & Mattenley, A. A. (2002). The effect of increasing plasmid size on transformation efficiency in Escherichia coli. *Journal of Experimental Microbiology and Immunology, 2*, 207-223.

Chen, J., Sawyer, N., & Regan, L. (2013). Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Science : A Publication of the Protein Society, 22*(4), 510-515. doi:10.1002/pro.2230

Chuang, C. C., Chen, C. Y., Yang, J. M., Lyu, P. C., & Hwang, J. K. (2003). Relationship between protein structures and disulfide-bonding patterns. *Proteins, 53*(1), 1-5. doi:10.1002/prot.10492

Ciric, M., Moon, C. D., Leahy, S. C., Creevey, C. J., Altermann, E., Attwood, G. T., . . . Gagic, D. (2014). Metasecretome-selective phage display approach for mining the functional potential of a rumen microbial community. *BMC Genomics, 15*(1), 356. doi:10.1186/1471-2164-15-356

Clementi, N., Mancini, N., Solforosi, L., Castelli, M., Clementi, M., & Burioni, R. (2012). Phage display-based strategies for cloning and optimization of monoclonal antibodies directed against human pathogens. *Int J Mol Sci, 13*(7), 8273-8292. doi:10.3390/ijms13078273

Cooper, G. (2000). Cell walls and the extracellular matrix. The cell: A molecular approach. Sunderland, MA: Sinauer Associates.

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Front Microbiol, 5*, 63. doi:10.3389/fmicb.2014.00063

Crameri, A., Whitehorn, E. A., Tate, E., & Stemmer, W. P. C. (1996). Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. *Nat Biotech, 14*(3), 315-319.

Dale, G. E., Oefner, C., & D'Arcy, A. (2003). The protein as a variable in protein crystallization. *Journal of Structural Biology, 142*(1), 88-97. doi:10.1016/s1047-8477(03)00041-8

Dammeyer, T., & Tinnefeld, P. (2012). Engineered fluorescent proteins illuminate the bacterial periplasm. *Comput Struct Biotechnol J, 3*, e201210013. doi:10.5936/csbj.201210013

Dauter, Z., & Jaskólski, M. (2016). Crystal pathologies in macromolecular crystallography. *Postepy biochemii, 62*(3), 401.

de Boer, H. A., Comstock, L. J., & Vasser, M. (1983). The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences of the United States of America, 80*(1), 21-25.

de Mol, N. J., & Fischer, M. J. E. (2008). Chapter 5 Kinetic and Thermodynamic Analysis of Ligand-Receptor Interactions: SPR Applications in Drug Development *Handbook of Surface Plasmon Resonance* (pp. 123-172): The Royal Society of Chemistry.

Diepold, A., & Wagner, S. (2014). Assembly of the bacterial type III secretion machinery. *FEMS Microbiol Rev, 38*(4), 802-822. doi:10.1111/1574-6976.12061

Dinh, T., & Bernhardt, T. G. (2011). Using superfolder green fluorescent protein for periplasmic protein localization studies. *J Bacteriol, 193*(18), 4984-4987. doi:10.1128/JB.00315-11

Du, X., Li, Y., Xia, Y.-L., Ai, S.-M., Liang, J., Sang, P., . . . Liu, S.-Q. (2016). Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *International Journal of Molecular Sciences, 17*(2), 144. doi:10.3390/ijms17020144

Dufner, P., Jermutus, L., & Minter, R. R. (2006). Harnessing phage and ribosome display for antibody optimisation. *Trends Biotechnol, 24*(11), 523-529. doi:10.1016/j.tibtech.2006.09.004

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2126-2132. doi:10.1107/s0907444904019158

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr, 62*(Pt 1), 72-82. doi:10.1107/S0907444905036693

Evans, P. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D, 67*(4), 282-292. doi:doi:10.1107/S090744491003982X

Evans, P., & McCoy, A. (2008). An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr, 64*(Pt 1), 1-10. doi:10.1107/S0907444907051554

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography, 69*(Pt 7), 1204-1214. doi:10.1107/S0907444913000061

Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H., & Phillips, G. J. (2000). Green Fluorescent Protein Functions as a Reporter for Protein Localization in Escherichia coli. *Journal of Bacteriology, 182*(14), 4068-4076.

Fekkes, P., & Driessen, A. J. M. (1999). Protein Targeting to the Bacterial Cytoplasmic Membrane. *Microbiology and Molecular Biology Reviews, 63*(1), 161-173.

Fiedler, M., & Skerra, A. (2014). Non-Antibody Scaffolds as Alternative Therapeutic Agents *Handbook of Therapeutic Antibodies* (pp. 435-474): Wiley-VCH Verlag GmbH & Co. KGaA.

Fisher, A. C., & DeLisa, M. P. (2008). Laboratory evolution of fast-folding green fluorescent protein using secretory pathway quality control. *PLoS One, 3*(6), e2351. doi:10.1371/journal.pone.0002351

FitzGerald, K. (2000). In vitro display technologies – new tools for drug discovery. *Drug Discovery Today, 5*(6), 253-258. doi:https://doi.org/10.1016/S1359-6446(00)01501-4

Fivash, M., Towler, E. M., & Fisher, R. J. (1998). BIAcore for macromolecular interaction. *Curr Opin Biotechnol, 9*(1), 97-101.

Friguet, B., Chaffotte, A. F., Djavadi-Ohaniance, L., & Goldberg, M. E. (1985). Measurements of the true affinity constant in solution of antigen-antibody complexes by enzyme-linked immunosorbent assay. *J Immunol Methods, 77*(2), 305-319.

Friguet, B., Chaffotte, A. F., Djavadi-Ohaniance, L., & Goldberg, M. E. (1995). Under proper experimental conditions the solid-phase antigen does not disrupt the liquid phase equilibrium when measuring dissociation constants by competition ELISA. *J Immunol Methods, 182*(1), 145-150.

Fuh, G., & Sidhu, S. S. (2000). Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. *FEBS Lett, 480*(2-3), 231-234.

Fukuda, H., Arai, M., & Kuwajima, K. (2000). Folding of Green Fluorescent Protein and the Cycle3 Mutant†. *Biochemistry, 39*(39), 12025-12032. doi:10.1021/bi000543l

Gagic, D., Ciric, M., Wen, W. X., Ng, F., & Rakonjac, J. (2016). Exploring the Secretomes of Microbes and Microbial Communities Using Filamentous Phage Display. *Front Microbiol, 7*, 429. doi:10.3389/fmicb.2016.00429

Gandham, S. H. A., Volk, D. E., & Gorenstein, D. G. (2015). Acquisition of data at multiple gains within a single thermal melt experiment using the Rotor-Gene Q instrument. *Journal of Analytical Science and Technology, 6*(1). doi:10.1186/s40543-015-0046-5

Gannon, P. M., Li, P., & Kumamoto, C. A. (1989). The mature portion of Escherichia coli maltose-binding protein (MBP) determines the

dependence of MBP on SecB for export. *Journal of Bacteriology, 171*(2), 813-818.

Garrett, R. H., & Grisham, C. M. (2004). *Biochemistry*. Pacific Grove, Calif.; London: Brooks/Cole.

Gaytan, M. O., Martinez-Santos, V. I., Soto, E., & Gonzalez-Pedrajo, B. (2016). Type Three Secretion System in Attaching and Effacing Pathogens. *Front Cell Infect Microbiol, 6*, 129. doi:10.3389/fcimb.2016.00129

Gebauer, M., & Skerra, A. (2009). Engineered protein scaffolds as next-generation antibody therapeutics. *Curr Opin Chem Biol, 13*(3), 245-255. doi:10.1016/j.cbpa.2009.04.627

Gonsberg, A., Jung, S., Ulbrich, S., Origi, A., Ziska, A., Baier, M., . . . Tatzelt, J. (2017). The Sec61/SecY complex is inherently deficient in translocating intrinsically disordered proteins. *J Biol Chem, 292*(52), 21383-21396. doi:10.1074/jbc.M117.788067

Gosavi, R. A., Mueser, T. C., & Schall, C. A. (2008). Optimization of buffer solutions for protein crystallization. *Acta Crystallogr D Biol Crystallogr, 64*(Pt 5), 506-514. doi:10.1107/S0907444908004265

Gotte, G., & Libonati, M. (2014). Protein Oligomerization. doi:10.5772/57489

Grossman, T. H., Kawasaki, E. S., Punreddy, S. R., & Osburne, M. S. (1998). Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene, 209*(1-2), 95-103.

Hagel, L., & Janson, J.-C. (1992). Chapter 6 Size-exclusion chromatography. *Journal of Chromatography Library, 51*, A267-A307. doi:http://dx.doi.org/10.1016/S0301-4770(08)61567-5

Hardy, F., Djavadi-Ohaniance, L., & Goldberg, M. E. (1997). Measurement of antibody/antigen association rate constants in solution by a method based on the enzyme-linked immunosorbent assay. *J Immunol Methods, 200*(1-2), 155-159.

Hashimoto, K., & Panchenko, A. R. (2010). Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc Natl Acad Sci U S A, 107*(47), 20352-20357. doi:10.1073/pnas.1012999107

Hawkins, R. E., Russell, S. J., & Winter, G. (1992). Selection of phage antibodies by binding affinity. *Journal of Molecular Biology, 226*(3), 889-896. doi:http://dx.doi.org/10.1016/0022-2836(92)90639-2

Hayhurst, A. (2000). Improved expression characteristics of single-chain Fv fragments when fused downstream of the Escherichia coli maltose-binding protein or upstream of a single immunoglobulin-constant domain. *Protein Expr Purif, 18*(1), 1-10. doi:10.1006/prep.1999.1164

Headd, J. J., Echols, N., Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Moriarty, N. W., . . . Adams, P. D. (2012). Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D Biol Crystallogr, 68*(Pt 4), 381-390. doi:10.1107/s0907444911047834

Heckman, K. L., & Pease, L. R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protocols, 2*(4), 924-932.

Heggeset, T. M., Kucharova, V., Naerdal, I., Valla, S., Sletta, H., Ellingsen, T. E., & Brautaset, T. (2013). Combinatorial mutagenesis and selection of improved signal sequences and their application for high-level production of translocated heterologous proteins in Escherichia coli. *Appl Environ Microbiol, 79*(2), 559-568. doi:10.1128/aem.02407-12

Heinrich, L., Tissot, N., Hartmann, D. J., & Cohen, R. (2010). Comparison of the results obtained by ELISA and surface plasmon resonance for the determination of antibody affinity. *J Immunol Methods, 352*(1-2), 13-22. doi:10.1016/j.jim.2009.10.002

Heras, B., & Martin, J. L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr, 61*(Pt 9), 1173-1180. doi:10.1107/S0907444905019451

Holland, I. B. (2004). Translocation of bacterial proteins--an overview. *Biochim Biophys Acta, 1694*(1-3), 5-16. doi:10.1016/j.bbamcr.2004.02.007

Hong, P., Koza, S., & Bouvier, E. S. P. (2012). Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *Journal of Liquid Chromatography & Related Technologies, 35*(20), 2923-2950. doi:10.1080/10826076.2012.743724

Hoogenboom, H. R. (2005). Selecting and screening recombinant antibody libraries. *Nat Biotechnol, 23*(9), 1105-1116. doi:10.1038/nbt1126

Huber, D., & Beckwith, J. (2006). Phage display extends its reach. *Nat Biotech, 24*(7), 793-794.

Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev, 62*(2), 379-433.

Hwang, I. (2014). Virus Outbreaks in Chemical and Biological Sensors. *Sensors, 14*(8), 13592.

Isobe, N., & Nakao, T. (2003). Direct enzyme immunoassay of progesterone in bovine plasma. *Animal Science Journal, 74*(5), 369-373. doi:10.1046/j.1344-3941.2003.00128.x

Jacobs, S. A., Diem, M. D., Luo, J., Teplyakov, A., Obmolova, G., Malia, T., . . . O'Neil, K. T. (2012). Design of novel FN3 domains with high stability by a consensus sequence approach. *Protein Eng Des Sel, 25*(3), 107-117. doi:10.1093/protein/gzr064

Jeiranikhameneh, M., Moshiri, F., Keyhan Falasafi, S., & Zomorodipour, A. (2017). Designing Signal Peptides for Efficient Periplasmic Expression of Human Growth Hormone in Escherichia coli. *J Microbiol Biotechnol, 27*(11), 1999-2009. doi:10.4014/jmb.1703.03080

Jeong, H., Kim, H. J., & Lee, S. J. (2015). Complete Genome Sequence of Escherichia coli Strain BL21. *Genome Announcements, 3*(2), e00134-00115. doi:10.1128/genomeA.00134-15

Jost, C., & Pluckthun, A. (2014). Engineered proteins with desired specificity: DARPins, other alternative scaffolds and bispecific IgGs. *Curr Opin Struct Biol, 27*, 102-112. doi:10.1016/j.sbi.2014.05.011

Kapust, R. B., & Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci, 8*(8), 1668-1674. doi:10.1110/ps.8.8.1668

Karplus, P. A., & Diederichs, K. (2012). Linking Crystallographic Model and Data Quality. *Science, 336*(6084), 1030-1033. doi:10.1126/science.1218231

Kim, S. J., Park, Y., & Hong, H. J. (2005). Antibody engineering for the development of therapeutic antibodies. *Mol Cells, 20*(1), 17-29.

Kimbrough, T. G., & Miller, S. I. (2002). Assembly of the type III secretion needle complex of Salmonella typhimurium. *Microbes Infect, 4*(1), 75-82.

Kleywegt, G. J., & Brünger, A. T. (1996). Checking your imagination: applications of the free R value. *Structure, 4*(8), 897-904. doi:http://dx.doi.org/10.1016/S0969-2126(96)00097-4

Knyazev, D. G., Kuttner, R., Zimmermann, M., Sobakinskaya, E., & Pohl, P. (2018). Driving Forces of Translocation Through Bacterial Translocon SecYEG. *J Membr Biol*. doi:10.1007/s00232-017-0012-9

Koide, A., Wojcik, J., Gilbreth, R. N., Hoey, R. J., & Koide, S. (2012). Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J Mol Biol, 415*(2), 393-405. doi:10.1016/j.jmb.2011.12.019

Kolaj, O., Spada, S., Robin, S., & Wall, J. G. (2009). Use of folding modulators to improve heterologous protein production in Escherichia coli. *Microbial Cell Factories, 8*(1), 9. doi:10.1186/1475-2859-8-9

Korotkov, K. V., Gonen, T., & Hol, W. G. (2011). Secretins: dynamic channels for protein transport across membranes. *Trends Biochem Sci, 36*(8), 433-443. doi:10.1016/j.tibs.2011.04.002

Kostakioti, M., Newman, C. L., Thanassi, D. G., & Stathopoulos, C. (2005). Mechanisms of protein export across the bacterial outer membrane. *J Bacteriol, 187*(13), 4306-4314. doi:10.1128/JB.187.13.4306-4314.2005

Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology, 372*(3), 774-797. doi:http://dx.doi.org/10.1016/j.jmb.2007.05.022

Lee, C. M. Y., Iorno, N., Sierro, F., & Christ, D. (2007). Selection of human antibody fragments by phage display. *Nat. Protocols, 2*(11), 3001-3008.

Leslie, A. G. W., & Powell, H. R. (2007). Processing diffraction data with mosflm. In R. J. Read & J. L. Sussman (Eds.), *Evolving Methods for Macromolecular Crystallography: The Structural Path to the Understanding of the Mechanismof Action of CBRN Agents* (pp. 41-51). Dordrecht: Springer Netherlands.

Li, F., Vijayasankaran, N., Shen, A., Kiss, R., & Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. *mAbs, 2*(5), 466-477. doi:10.4161/mabs.2.5.12720

Linton, E., Walsh, M. K., Sims, R. C., & Miller, C. D. (2012). Translocation of green fluorescent protein by comparative analysis with multiple signal peptides. *Biotechnol J, 7*(5), 667-676. doi:10.1002/biot.201100158

Lofblom, J., Feldwisch, J., Tolmachev, V., Carlsson, J., Stahl, S., & Frejd, F. Y. (2010). Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS Lett, 584*(12), 2670-2680. doi:10.1016/j.febslet.2010.04.014

Löfblom, J., & Frejd, F. (2011). *Alternative Scaffolds as Bispecific Antibody Mimetics*.

Marbach, A., & Bettenbrock, K. (2012). lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol, 157*(1), 82-88. doi:10.1016/j.jbiotec.2011.10.009

Martoglio, B., & Dobberstein, B. (1998). Signal sequences: more than just greasy peptides. *Trends in Cell Biology, 8*(10), 410-415. doi:10.1016/s0962-8924(98)01360-9

Marvin, D. A., Symmons, M. F., & Straus, S. K. (2014). Structure and assembly of filamentous bacteriophages. *Prog Biophys Mol Biol, 114*(2), 80-122. doi:10.1016/j.pbiomolbio.2014.02.003

Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol, 33*(2), 491-497.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr, 40*(Pt 4), 658-674. doi:10.1107/S0021889807021206

McCullum, E. O., Williams, B. A., Zhang, J., & Chaput, J. C. (2010). Random mutagenesis by error-prone PCR. *Methods Mol Biol, 634*, 103-109. doi:10.1007/978-1-60761-652-8_7

Mena, M. A., & Daugherty, P. S. (2005). Automated design of degenerate codon libraries. *Protein Engineering, Design and Selection, 18*(12), 559-561. doi:10.1093/protein/gzi061

Mitchell, J. C., Kerr, R., & Ten Eyck, L. F. (2001). Rapid atomic density methods for molecular shape characterization. *J Mol Graph Model, 19*(3-4), 325-330, 388-390.

Moon, A. F., Mueller, G. A., Zhong, X., & Pedersen, L. C. (2010). A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. *Protein Sci, 19*(5), 901-913. doi:10.1002/pro.368

Munoz, F. J., Santos, J. I., Arda, A., Andre, S., Gabius, H. J., Sinisterra, J. V., . . . Hernaiz, M. J. (2010). Binding studies of adhesion/growth-regulatory galectins with glycoconjugates monitored by surface plasmon resonance and NMR spectroscopy. *Org Biomol Chem, 8*(13), 2986-2992. doi:10.1039/b927139b

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., . . . Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D, 67*(4), 355-367. doi:doi:10.1107/S0907444911001314

Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D, 53*(3), 240-255. doi:doi:10.1107/S0907444996012255

Murzin, A. G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *Embo j, 12*(3), 861-867.

Nallamsetty, S., & Waugh, D. S. (2006). Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr Purif, 45*(1), 175-182. doi:10.1016/j.pep.2005.06.012

Nangola, S., Minard, P., & Tayapiwatana, C. (2010). Appraisal of translocation pathways for displaying ankyrin repeat protein on phage particles. *Protein Expr Purif, 74*(2), 156-161. doi:10.1016/j.pep.2010.08.010

Natale, P., Bruser, T., & Driessen, A. J. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. *Biochim Biophys Acta, 1778*(9), 1735-1756. doi:10.1016/j.bbamem.2007.07.015

Nian, S., Wu, T., Ye, Y., Wang, X., Xu, W., & Yuan, Q. (2016). Development and identification of fully human scFv-Fcs against Staphylococcus aureus. *BMC Immunology, 17*, 8. doi:10.1186/s12865-016-0146-z

Nieba, L., Krebber, A., & Pluckthun, A. (1996). Competition BIAcore for measuring true affinities: large differences from values determined from binding kinetics. *Anal Biochem, 234*(2), 155-165. doi:10.1006/abio.1996.0067

Niesen, F. H., Berglund, H., & Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc, 2*(9), 2212-2221. doi:10.1038/nprot.2007.321

Nishi, H., Hashimoto, K., Madej, T., & Panchenko, A. R. (2013). Evolutionary, physicochemical, and functional mechanisms of protein homooligomerization. *Prog Mol Biol Transl Sci, 117*, 3-24. doi:10.1016/B978-0-12-386931-9.00001-5

Nygren, P. A. (2008). Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J, 275*(11), 2668-2676. doi:10.1111/j.1742-4658.2008.06438.x

Nygren, P. A., & Skerra, A. (2004). Binding proteins from alternative scaffolds. *J Immunol Methods, 290*(1-2), 3-28. doi:10.1016/j.jim.2004.04.006

Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nat Rev Genet, 16*(7), 379-394. doi:10.1038/nrg3927

Pan, S. H., & Malcolm, B. A. (2000). Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). *Biotechniques, 29*(6), 1234-1238.

Parker, A. S., Griswold, K. E., & Bailey-Kellogg, C. (2011). Optimization of combinatorial mutagenesis. *J Comput Biol, 18*(11), 1743-1756. doi:10.1089/cmb.2011.0152

Parsons, S. (2003). Introduction to twinning. *Acta Crystallographica Section D, 59*(11), 1995-2003. doi:doi:10.1107/S0907444903017657

Paschke, M. (2006). Phage display systems and their applications. *Appl Microbiol Biotechnol, 70*(1), 2-11. doi:10.1007/s00253-005-0270-9

Paschke, M., & Hohne, W. (2005). A twin-arginine translocation (Tat)-mediated phage display system. *Gene, 350*(1), 79-88. doi:10.1016/j.gene.2005.02.005

Phue, J. N., Lee, S. J., Trinh, L., & Shiloach, J. (2008). Modified Escherichia coli B (BL21), a superior producer of plasmid DNA compared with Escherichia coli K (DH5alpha). *Biotechnol Bioeng, 101*(4), 831-836. doi:10.1002/bit.21973

Pollard, T. D. (2010). A guide to simple and informative binding assays. *Mol Biol Cell, 21*(23), 4061-4067. doi:10.1091/mbc.E10-08-0683

Qin, M., Wang, W., & Thirumalai, D. (2015). Protein folding guides disulfide bond formation. *Proceedings of the National Academy of Sciences of the United States of America, 112*(36), 11241-11246. doi:10.1073/pnas.1503909112

Rakonjac, J., Bennett, N. J., Spagnuolo, J., Gagic, D., & Russel, M. (2011). Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Curr Issues Mol Biol, 13*(2), 51-76.

Read, R. J., Adams, P. D., & McCoy, A. J. (2013). Intensity statistics in the presence of translational noncrystallographic symmetry. *Acta Crystallographica Section D: Biological Crystallography, 69*(Pt 2), 176-183. doi:10.1107/S0907444912045374

Reverberi, R., & Reverberi, L. (2007). Factors affecting the antigen-antibody reaction. *Blood Transfusion, 5*(4), 227-240. doi:10.2450/2007.0047-07

Rodi, D. J., & Makowski, L. (1999). Phage-display technology – finding a needle in a vast molecular haystack. *Current Opinion in Biotechnology, 10*(1), 87-93. doi:10.1016/s0958-1669(99)80016-0

Rondot, S., Koch, J., Breitling, F., & Dubel, S. (2001). A helper phage to improve single-chain antibody presentation in phage display. *Nat Biotech, 19*(1), 75-78.

Rosenberg, A., Griffin, K., Studier, F. W., McCormick, M., Berg, J., Novy, R., . . . Cloning, P. B. (1996). T7Select® Phage Display System: A powerful new protein display system based on bacteriophage T7. *NEWSLETTER*.

Rothe, A., Hosse, R. J., & Power, B. E. (2006). In vitro display technologies reveal novel biopharmaceutics. *FASEB J, 20*(10), 1599-1610. doi:10.1096/fj.05-5650rev

Ruigrok, Vincent J. B., Levisson, M., Eppink, Michel H. M., Smidt, H., & van der Oost, J. (2011). Alternative affinity tools: more attractive than antibodies? *Biochemical Journal, 436*(1), 1-13. doi:10.1042/bj20101860

Ruscito, A., & DeRosa, M. C. (2016). Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications. *Frontiers in Chemistry, 4*, 14. doi:10.3389/fchem.2016.00014

Saier, M. (2006). Protein secretion systems in Gram-negative bacteria. *MICROBE-AMERICAN SOCIETY FOR MICROBIOLOGY, 1*(9), 414.

Sali, A., Glaeser, R., Earnest, T., & Baumeister, W. (2003). From words to literature in structural proteomics. *Nature, 422*(6928), 216-225.

Schuch, R., & Maurelli, A. T. (2001). MxiM and MxiJ, base elements of the Mxi-Spa type III secretion system of Shigella, interact with and stabilize the MxiD secretin in the cell envelope. *J Bacteriol, 183*(24), 6991-6998. doi:10.1128/JB.183.24.6991-6998.2001

Sharff, A. J., Rodseth, L. E., Spurlino, J. C., & Quiocho, F. A. (1992). Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry, 31*(44), 10657-10663. doi:10.1021/bi00159a003

Sidhu, S. S., & Koide, S. (2007). Phage display for engineering and analyzing protein interaction interfaces. *Curr Opin Struct Biol, 17*(4), 481-487. doi:10.1016/j.sbi.2007.08.007

Singh, P., Sharma, L., Kulothungan, S. R., Adkar, B. V., Prajapati, R. S., Ali, P. S., . . . Varadarajan, R. (2013). Effect of signal peptide on stability and folding of Escherichia coli thioredoxin. *PLoS One, 8*(5), e63442. doi:10.1371/journal.pone.0063442

Skerra, A. (2001). 'Anticalins': a new class of engineered ligand-binding proteins with antibody-like properties. *Reviews in Molecular Biotechnology, 74*(4), 257-275. doi:10.1016/s1389-0352(01)00020-4

Skerra, A. (2008). Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *FEBS J, 275*(11), 2677-2683. doi:10.1111/j.1742-4658.2008.06439.x

Sliwiak, J., Jaskolski, M., Dauter, Z., McCoy, A. J., & Read, R. J. (2014). Likelihood-based molecular-replacement solution for a highly pathological crystal with tetartohedral twinning and sevenfold translational noncrystallographic symmetry. *Acta Crystallographica Section D: Biological Crystallography, 70*(Pt 2), 471-480. doi:10.1107/S1399004713030319

Smith, A. J. (2015). New horizons in therapeutic antibody discovery: opportunities and challenges versus small-molecule therapeutics. *J Biomol Screen, 20*(4), 437-453. doi:10.1177/1087057114562544

Smith, G. P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science, 228*(4705), 1315-1317.

Smith, G. P., & Petrenko, V. A. (1997). Phage Display. *Chem Rev, 97*(2), 391-410.

Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., & Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. *Protein Sci, 12*(7), 1313-1322. doi:10.1110/ps.0243403

Speck, J., Arndt, K. M., & Muller, K. M. (2011). Efficient phage display of intracellularly folded proteins mediated by the TAT pathway. *Protein Eng Des Sel, 24*(6), 473-484. doi:10.1093/protein/gzr001

Spreter, T., Yip, C. K., Sanowar, S., Andre, I., Kimbrough, T. G., Vuckovic, M., . . . Strynadka, N. C. (2009). A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nat Struct Mol Biol, 16*(5), 468-476. doi:10.1038/nsmb.1603

Ståhl, S., Kronqvist, N., Jonsson, A., & Löfblom, J. (2013). Affinity proteins and their generation. *Journal of Chemical Technology & Biotechnology, 88*(1), 25-38. doi:10.1002/jctb.3929

Steemson, J. D., Baake, M., Rakonjac, J., Arcus, V. L., & Liddament, M. T. (2014). Tracking molecular recognition at the atomic level with a new protein scaffold based on the OB-fold. *PLoS One, 9*(1), e86050. doi:10.1371/journal.pone.0086050

Steiner, D., Forrer, P., & Pluckthun, A. (2008). Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display. *J Mol Biol, 382*(5), 1211-1227. doi:10.1016/j.jmb.2008.07.085

Steiner, D., Forrer, P., Stumpp, M. T., & Pluckthun, A. (2006). Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat Biotechnol, 24*(7), 823-831. doi:10.1038/nbt1218

Stewart, E. J., Aslund, F., & Beckwith, J. (1998). Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. *Embo j, 17*(19), 5543-5550. doi:10.1093/emboj/17.19.5543

Stronge, A. J. H., Sreenan, J. M., Diskin, M. G., Mee, J. F., Kenny, D. A., & Morris, D. G. (2005). Post-insemination milk progesterone concentration and embryo survival in dairy cows. *Theriogenology, 64*(5), 1212-1224. doi:https://doi.org/10.1016/j.theriogenology.2005.02.007

Stumpp, M. T., Binz, H. K., & Amstutz, P. (2008). DARPins: a new generation of protein therapeutics. *Drug Discov Today, 13*(15-16), 695-701. doi:10.1016/j.drudis.2008.04.013

Szilágyi, A., Kardos, J., Osváth, S., Barna, L., & Závodszky, P. (2007). Protein Folding. In A. Lajtha & N. Banik (Eds.), *Handbook of Neurochemistry and Molecular Neurobiology: Neural Protein Metabolism and Function* (pp. 303-343). Boston, MA: Springer US.

Tanenbaum, D. M., Wang, Y., Williams, S. P., & Sigler, P. B. (1998). Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proceedings of the National Academy of Sciences, 95*(11), 5998-6003.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., . . . Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D, 64*(1), 61-69. doi:doi:10.1107/S090744490705024X

Theobald, D. L., Mitton-Fry, R. M., & Wuttke, D. S. (2003). Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct, 32*, 115-133. doi:10.1146/annurev.biophys.32.110601.142506

Thie, H., Schirrmann, T., Paschke, M., Dubel, S., & Hust, M. (2008). SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in E. coli. *N Biotechnol, 25*(1), 49-54. doi:10.1016/j.nbt.2008.01.001

Tobin, P. H., Richards, D. H., Callender, R. A., & Wilson, C. J. (2014). Protein Engineering: A New Frontier for Biological Therapeutics. *Current drug metabolism, 15*(7), 743-756.

Trivedi, M. V., Laurence, J. S., & Siahaan, T. J. (2009). The role of thiols and disulfides on protein stability. *Curr Protein Pept Sci, 10*(6), 614-625.

Valent, Q. A., Scotti, P. A., High, S., de Gier, J. W., von Heijne, G., Lentzen, G., . . . Luirink, J. (1998). The Escherichia coli SRP and SecB targeting pathways converge at the translocon. *Embo j, 17*(9), 2504-2512. doi:10.1093/emboj/17.9.2504

Van Der Merwe, P. A. (2001). Surface plasmon resonance. *Protein-ligand interactions: hydrodynamics and calorimetry, 1*, 137-170.

Vazquez-Lombardi, R., Phan, T. G., Zimmermann, C., Lowe, D., Jermutus, L., & Christ, D. (2015). Challenges and opportunities for non-antibody scaffold drugs. *Drug Discovery Today, 20*(10), 1271-1283. doi:https://doi.org/10.1016/j.drudis.2015.09.004

Velappan, N., Fisher, H. E., Pesavento, E., Chasteen, L., D'Angelo, S., Kiss, C., . . . Bradbury, A. R. (2010). A comprehensive analysis of filamentous phage display vectors for cytoplasmic proteins: an analysis with different fluorescent proteins. *Nucleic Acids Res, 38*(4), e22. doi:10.1093/nar/gkp809

Verma, D., Grigoryan, G., & Bailey-Kellogg, C. (2015). Structure-based design of combinatorial mutagenesis libraries. *Protein Sci, 24*(5), 895-908. doi:10.1002/pro.2642

Vodnik, M., Zager, U., Strukelj, B., & Lunder, M. (2011). Phage display: selecting straws instead of a needle from a haystack. *Molecules, 16*(1), 790-817. doi:10.3390/molecules16010790

Wang, W. (2005). Protein aggregation and its inhibition in biopharmaceutics. *Int J Pharm, 289*(1-2), 1-30. doi:10.1016/j.ijpharm.2004.11.014

Waugh, D. S. (2016). Crystal structures of MBP fusion proteins. *Protein Sci, 25*(3), 559-571. doi:10.1002/pro.2863

Weiser, A. A., Wittenbrink, N., Zhang, L., Schmelzer, A. I., Valai, A., & Or-Guil, M. (2011). Affinity maturation of B cells involves not only a few but a whole spectrum of relevant mutations. *International Immunology, 23*(5), 345-356. doi:10.1093/intimm/dxr018

Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K., & Pluckthun, A. (2008). Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. *J Mol Biol, 376*(1), 241-257. doi:10.1016/j.jmb.2007.11.046

Wilchek, M., Bayer, E. A., & Livnah, O. (2006). Essentials of biorecognition: the (strept)avidin-biotin system as a model for protein-protein and protein-ligand interaction. *Immunol Lett, 103*(1), 27-32. doi:10.1016/j.imlet.2005.10.022

Wilson, D. R., & Finlay, B. B. (1998). Phage display: applications, innovations, and issues in phage and host biology. *Canadian Journal of Microbiology, 44*(4), 313-329. doi:10.1139/w98-015

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., . . . Wilson, K. S. (2011). Overview of the CCP4 suite and current

developments. *Acta Crystallographica Section D, 67*(4), 235-242. doi:doi:10.1107/S0907444910045749

Wlodawer, A., Minor, W., Dauter, Z., & Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J, 275*(1), 1-21. doi:10.1111/j.1742-4658.2007.06178.x

Zhao, N., Schmitt, M. A., & Fisk, J. D. (2016). Phage display selection of tight specific binding variants from a hyperthermostable Sso7d scaffold protein library. *FEBS J, 283*(7), 1351-1367. doi:10.1111/febs.13674

Zheng, J., Li, N., Tan, Y. P., Sivaraman, J., Mok, Y. K., Mo, Z. L., & Leung, K. Y. (2007). EscC is a chaperone for the Edwardsiella tarda type III secretion system putative translocon components EseB and EseD. *Microbiology, 153*(Pt 6), 1953-1962. doi:10.1099/mic.0.2006/004952-0

Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N., & Adams, P. D. (2008). Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallographica Section D: Biological Crystallography, 64*(Pt 1), 99-107. doi:10.1107/S090744490705531X

7 Appendices

7.1 General Molecular Biology

7.1.1 Media composition

The recipe for the common media used are as follows:

7.1.1.1 2xYT (Yeast extract Tryptone) media

The recipe for this media is (BD-Difco # 244020)

16 g of pancreatic digest of casein

10 g of yeast extract

5 g of NaCl

pH 7 (in 1L) of UP H_2O

This was sterilized as 500 ml cultures in 2 L baffled flasks at 121°C for 20 minutes

7.1.1.2 Luria-Bertani (LB) media

The recipe for this buffer is as follows:

10 g bacto-tryptone

5 g yeast extract

10 g NaCl

pH 7 (in 1L) of UP H_2O

This was sterilized in a similar format as 2xYT.

7.1.1.3 M9 minimal media

This media was used to grow *E. coli* TG1 strains (without antibiotic resistance) over 36 hours at 37°C. The recipe is:

1x M9 salts

7 mM $MgSO_4$

0.3 mM thiamine

44 mM glucose

0.1 mM $CaCl_2$

Glucose was sterile filtered (0.2 μM, Pall acrodisc) and added separately after heat sterilization (autoclave) of the remaining ingredients of the recipe. M9 salts (5x) were made separately and autoclaved before use. The recipe for 5xM9 salts is:

233 mM Na_2HPO_4

110 mM KH_2PO_4

42.7 mM NaCl

9.3 mM NH_4Cl

7.1.1.4 SOC media

This media was used in recovery of EC *E. coli* after electroporation. The recipe is:

20 g/L Tryptone

5 g/L Yeast extract

10 mM NaCl

2.5 mM KCl

5 mM $MgS0_4$

10 mM $MgCl_2$

20 mM glucose

Glucose was sterile filtered and added separately after heat sterilization of the remaining ingredients of the recipe.

7.1.1.5 TSB media

10 % w/v PEG 8000

10 mM Mg SO4

10 mM MgCl2

This is made up to 95 % final volume in LB media and autoclaved and stored at 4 °C. DMSO is added to final 5 % v/v before use.

7.1.2 Antibiotics

The following antibiotics were used at 1/1000 stock concentrations in culture media for protein expression and phage selections. The stocks were stored at - 20°C.

1) Ampicillin/Carbenicillin 100 mg/ml (C100, for plasmids pOB2, pOPINM, pPROEX-HTb)
2) Kanamycin 50 mg/ml (K50, for plasmids pET28B, KM13, M13K07δIII)

7.1.3 Buffer composition

The recipe for the common buffers used are as follows:

7.1.3.1 Tris buffered saline (TBS)

20 mM Tris

150 mM NaCl

0.05% (v/v) Tween-20 (T20)

pH 7.5

This buffer was used for size exclusion chromatography (SEC) of OBodies, and general storage and incubation with ligand (P4) for crystal trials.

The following were added for immobilized metal affinity chromatography (IMAC) elution

20 mM Imidazole – binding buffer

1M imidazole – elution buffer

7.1.3.2 Phosphate buffered saline (PBS)

The buffer used for phage selections and occasional protein purification was Dulbecco's PBS (Sigma Aldrich # D5652, pH 7.2 – 7.8)

The following was added for washing steps in phage display and ELISA

0.1 % (v/v) T20

7.1.4 SDS-PAGE reagents

7.1.4.1 Resolving Gel (layer)

5.55 mL UP H_2O

16.50 mL 30% (w/v) acrylamide

7.5 mL 1.5 M Tris pH 8.8

0.3 mL 10% (w/v) SDS

0.15 ml 10% (w/v) ammonium persulfate (APS)

15 µL tetramethylethylenediamine (TEMED)

7.1.4.2 Stacking Gel (layer)

8.5 mL UP H_2O

2.125 mL 30% Acrylamide

1.6 mL 1.0 M Tris pH 6.8

0.125 mL 10% (w/v) SDS

63 µL 10% (w/v) APS

6.3 µL TEMED

7.1.4.3 4x SDS protein loading dye

10 mL 1 M Tris, pH 6.8

8 mL glycerol

16 mL 10% (w/v) SDS

4 mL β-mercaptoethanol

1 mL 1% bromophenol blue

7.1.4.4 1x Tris-glycine SDS (TGS) running buffer

25 mM Tris

250 mM glycine

1% SDS (w/v)

7.1.4.5 PAGE staining solution (Fairbanks A)

0.05 % (w/v) Coomassie Brilliant Blue R-250

25 % (v/v) IPA

10 % (v/v) Acetic acid

7.1.4.6 PAGE de-staining solution

10% Acetic acid

7.1.5 Agarose gel ingredients

7.1.5.1 1% agarose gel

50 ml TAE running buffer

0.5 g agarose

5 µl SYBR-safe DNA gel stain

7.1.5.2 1 x Tris-acetate EDTA (TAE) running buffer

40 mM Tris-acetate

1 mM EDTA

7.1.5.3 10x DNA loading dye

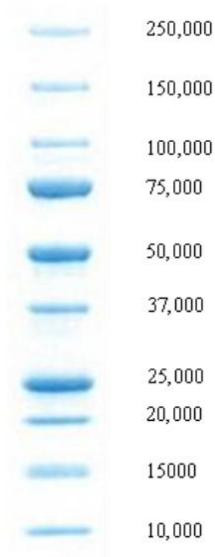
0.05% (w/v) Bromothymol Blue

0.25% (w/v) Xylene cyanol

30% (v/v) glycerol

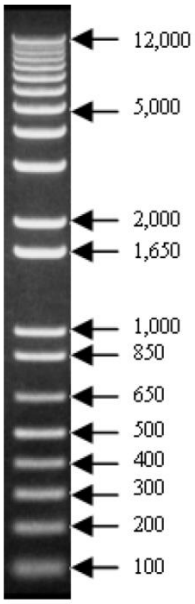
7.1.6 Electrophoresis standards

7.1.6.1 Precision Plus Protein standards



The protein standards stained by coomassie with numbers of bands indicating size of proteins in daltons (Da).

7.1.6.2 1kb Plus DNA Standards



The DNA standards stained by ethidium bromide with numbers of bands indicating size of nucleotides in base pairs (bp).

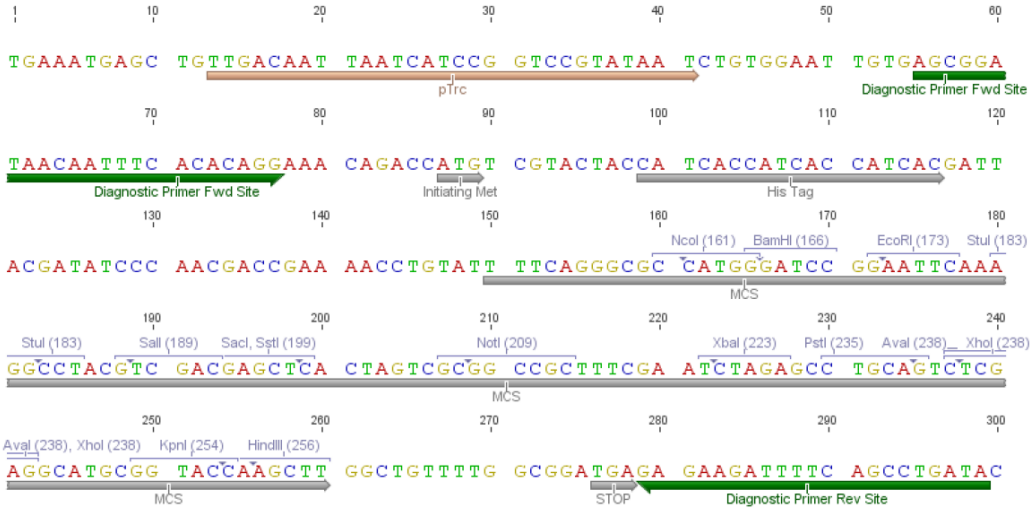
7.1.7 Bacterial strains

Name	*Genotype*
TG1	supE thi-1 Δ(lac-proAB) Δ(mcrB-hsdSM)5 (rK^- mK^-) [F´ traD36 proAB $lacI^qZ\Delta M15$]
DH5α	F^- thi-1 endA1 recA1 relA1 gyrA96 Φ80lacZΔM15

	Δ(lacZYA-argF)U169, hsdR17($rK^- mK^+$), λ–
Bl21 (DE3)	$F^–$ ompT gal dcm lon hsdSB($rB^- mB^-$) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])

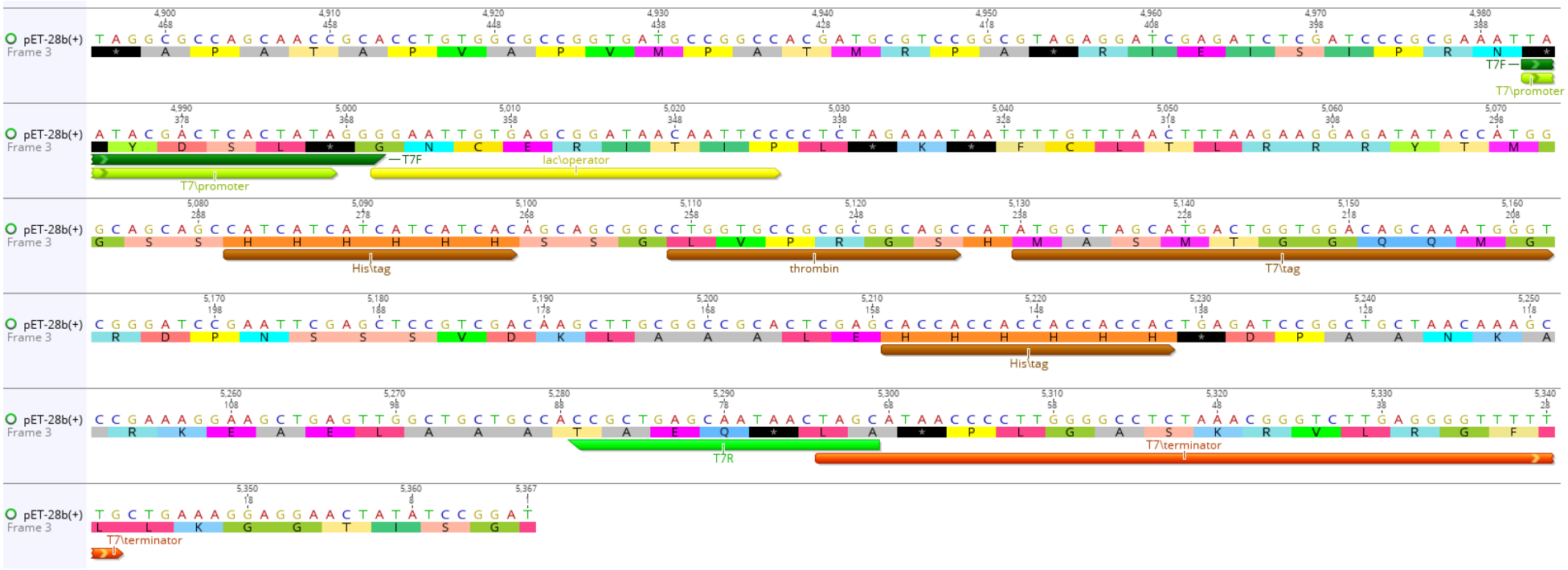
7.1.8 Plasmids

7.1.8.1 pPROEX-HTb



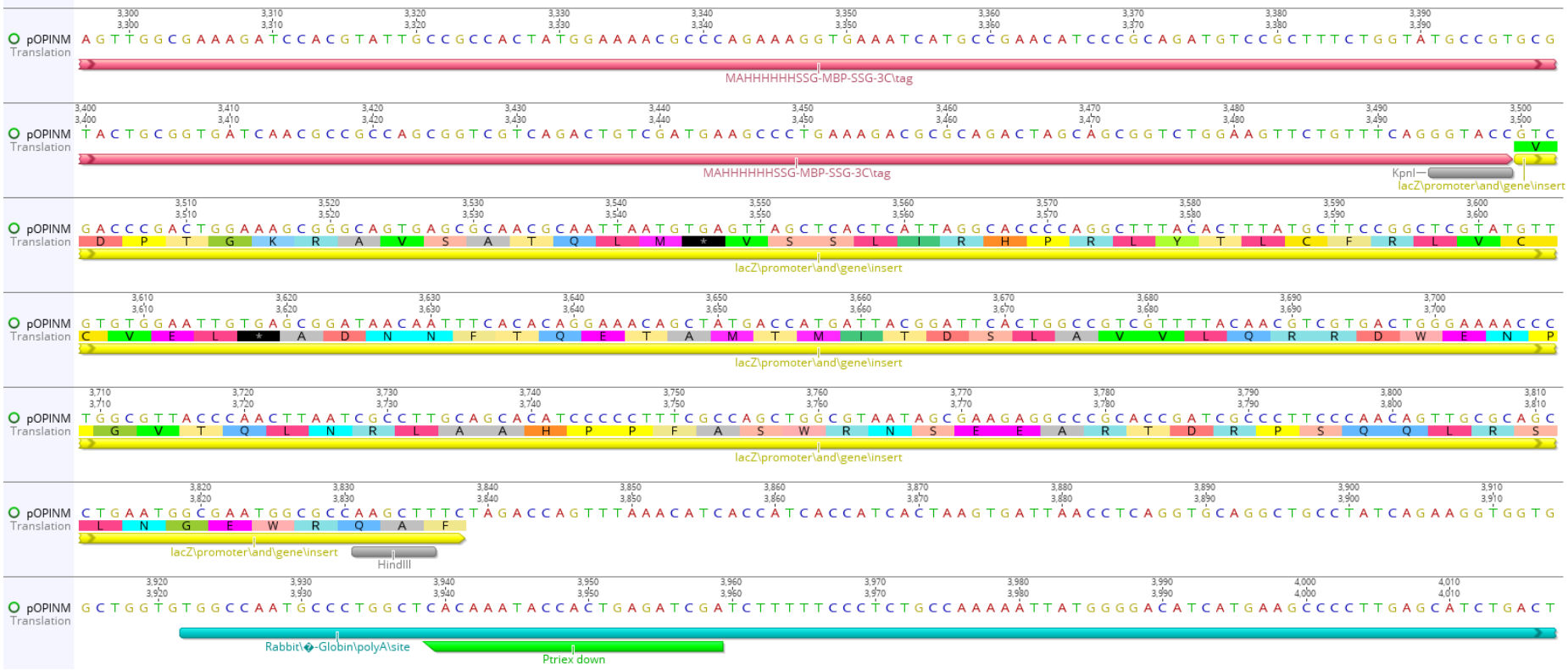
The multiple cloning site (MCS) of pPROEX-HTb indicating compatible restriction enzyme cut sites with N-terminal his-tag (grey) is shown. The forward and reverse diagnostic primers (green) and pTrc promoter (light brown) are also shown.

7.1.8.2 pET28b



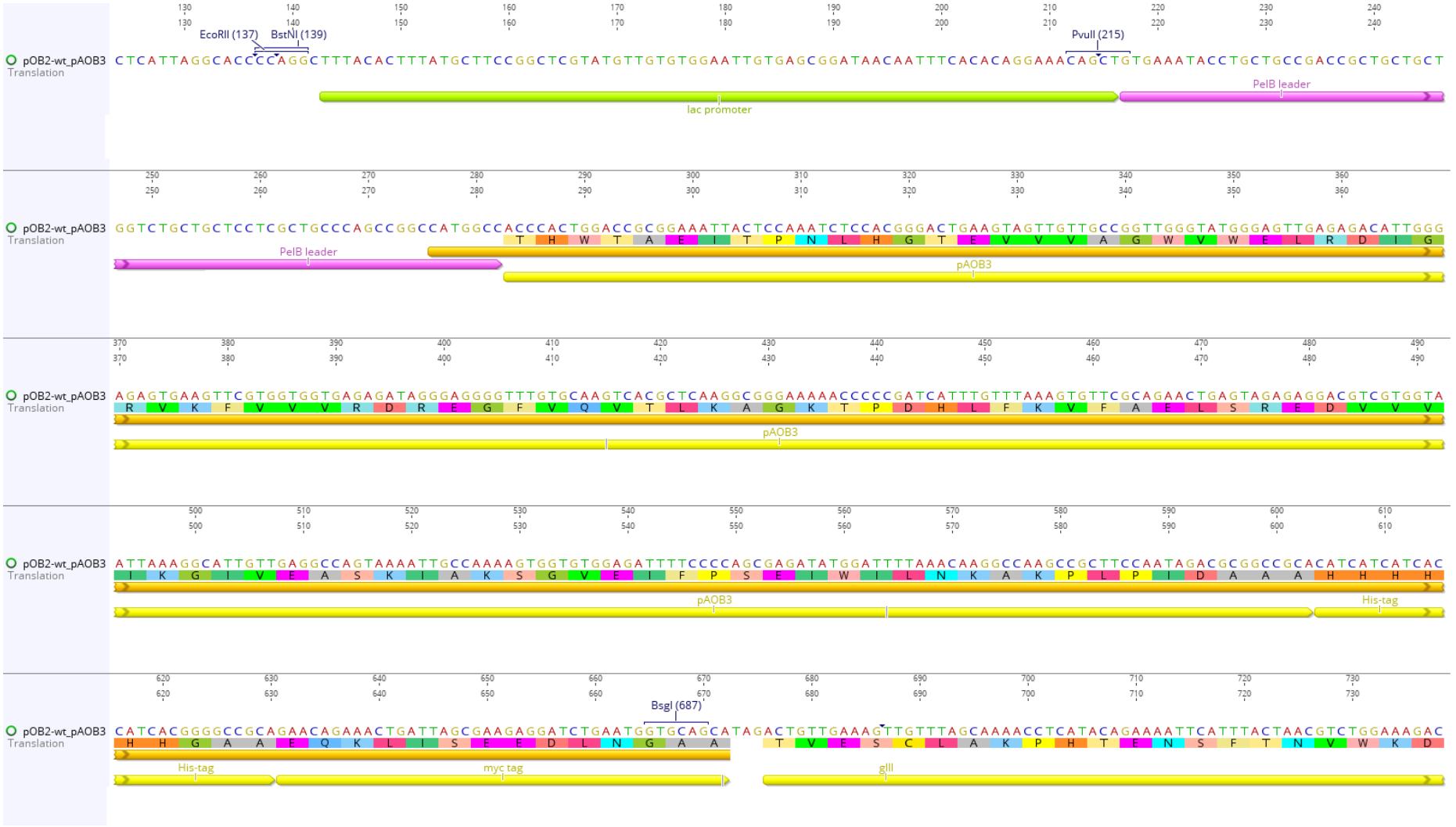
The multiple cloning site (MCS) of pET28b lies between N-terminal lac operon (yellow) and C-terminal his-tag (brown, adjacent to TGA stop codon). The N-terminal T7 promoter (light green), T7 forward primer (dark green), his-tag (brown) and C-terminal T7 reverse primer (bright green), T7 terminator (red) are also indicated.

7.1.8.3 pOPINM



The multiple cloning site (MCS) of pOPINM (yellow) between restriction enzyme cut sites KpnI and HindIII (grey) are shown. The N-terminal fusion protein MBP (red) and C-terminal sequencing primer pTriEx down (green) are also indicated.

7.1.8.4 pOB2



The OB-fold (pAOB3, yellow) is shown between pelB signal sequence (pink) and C-terminal his-tag (yellow). The resulting OBody is marked in orange. The N-terminal lac promoter (light green) and C-terminal gene (gIII) are also indicated.

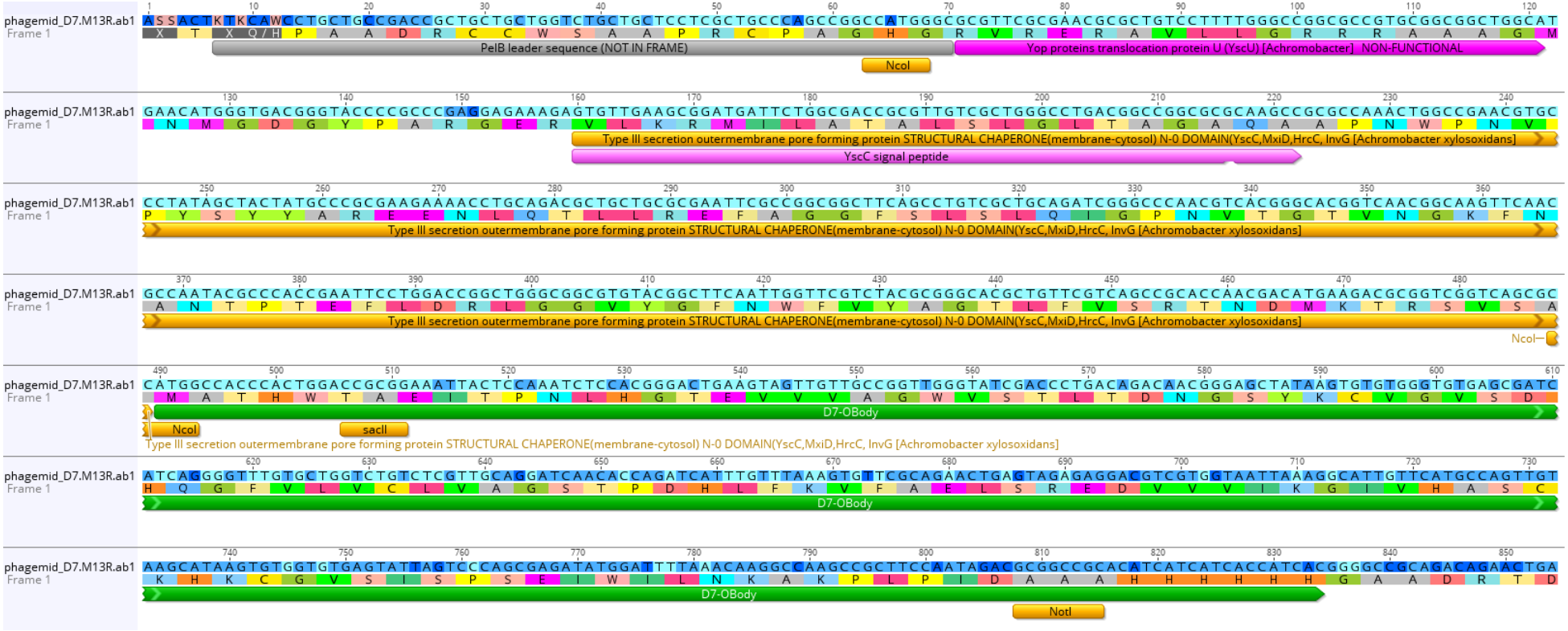
7.1.9 Primers

Name	DNA sequence	description
245	AAATACCTGCTGCCGACC	Diagnostic forward primer – pOB2
246	TTGTCGTCTTTCCAGACGTT	Diagnostic reverse primer – pOB2
D7F	CATGCCATGGCCACCCACTGGAC	forward cloning primer - pPROEX-HTb
D7R	ATAAGAATGCGGCCGCTCAGTCTATTGG	reverse cloning primer - pPROEX-HTb
C35S_F	GACAACGGGAGCTATAAGAGCGTGGGTGTGAGCGATCATC	C35S forward primer
C35S_R	GATGATCGCTCACACCCACGCTCTTATAGCTCCCGTTGTC	C35S reverse primer
C48S_F	CAGGGGTTTGTGCTGGTCAGCCTCGTTGCAGGATCAAC	C48S forward primer
C48S_R	GTTGATCCTGCAACGAGGCTGACCAGCACAAACCCCTG	C48S reverse primer
C81S_F	GGCATTGTTCATGCCAGTAGCAAGCATAAGTGTGGTGTG	C81S forward primer
C81S_R	CACACCACACTTATGCTTGCTACTGGCATGAACAATGCC	C81S reverse primer
C85S_F	CATGCCAGTTGTAAGCATAAGAGCGGTGTGAGTATTAGTCCCAGC	C85S forward primer
C85S_R	GCTGGGACTAATACTCACACCGCTCTTATGCTTACAACTGGCATG	C85S reverse primer
pOPINM_F (**KpnI**) pOPINM_F (LIC)	AAGTTCTGTTTCAGGG**TACC**ATGGCCACCCACTGG AAGTTCTGTTTCAGGGCCCGATGGCCACCCACTGG	forward cloning primer - pOPINM

pOPINM_R	ATGGTCTAGAAAGCTTTAGTCTATTGGAAGCGGC	reverse cloning primer - pOPINM
pF_rand_F	GCAGGATCAACACCAGATC	Randomization forward primer – pF_35
pF_rand_R	CTGGTGTTGATCCTGC**NNK**GAG**NNK**GAC**NNK**CACAAACCCCTGATGATC	Randomization reverse primer – pF_35 (randomized -red)
L4_rand_F	GGTGTGAGTATTAGTCCM	Randomization forward primer – L4_DM
L4_rand_R	GACTAATACTCACACC**NNKNNKNNKNNKNNKNNK**GGCATGAACAATGCC	Randomization reverse primer – L4_DM (randomized -red)
ssYscC_PstI	AACTGCAGATGGCCACCCACTGGACCGCGG	Primer to remove gene for YscC N0 domain using PstI restriction enzyme cut site
YscC_DARPin	GGGCTGCAGATGTCTTCCGGTGATCTGGG	Primer to clone DARPin E_01 in frame with YscC fragment
GFPuv_F	TCCCCGCGGCTAGCAAAGGAGAAGAACTTTTC	forward GFPuv cloning primer – pOB2
GFPuv_R	TTTTCCTTTTGCGGCCGCTTTGTAGAGCTCATCCATGCCATG	reverse GFPuv cloning primer – pOB2
GFPuv_OB_sacI	CCCGAGCTCTACAAAGCGGCCACCCACTGGACCGCGG	Primer to clone GFPuv between signal sequence and OBody

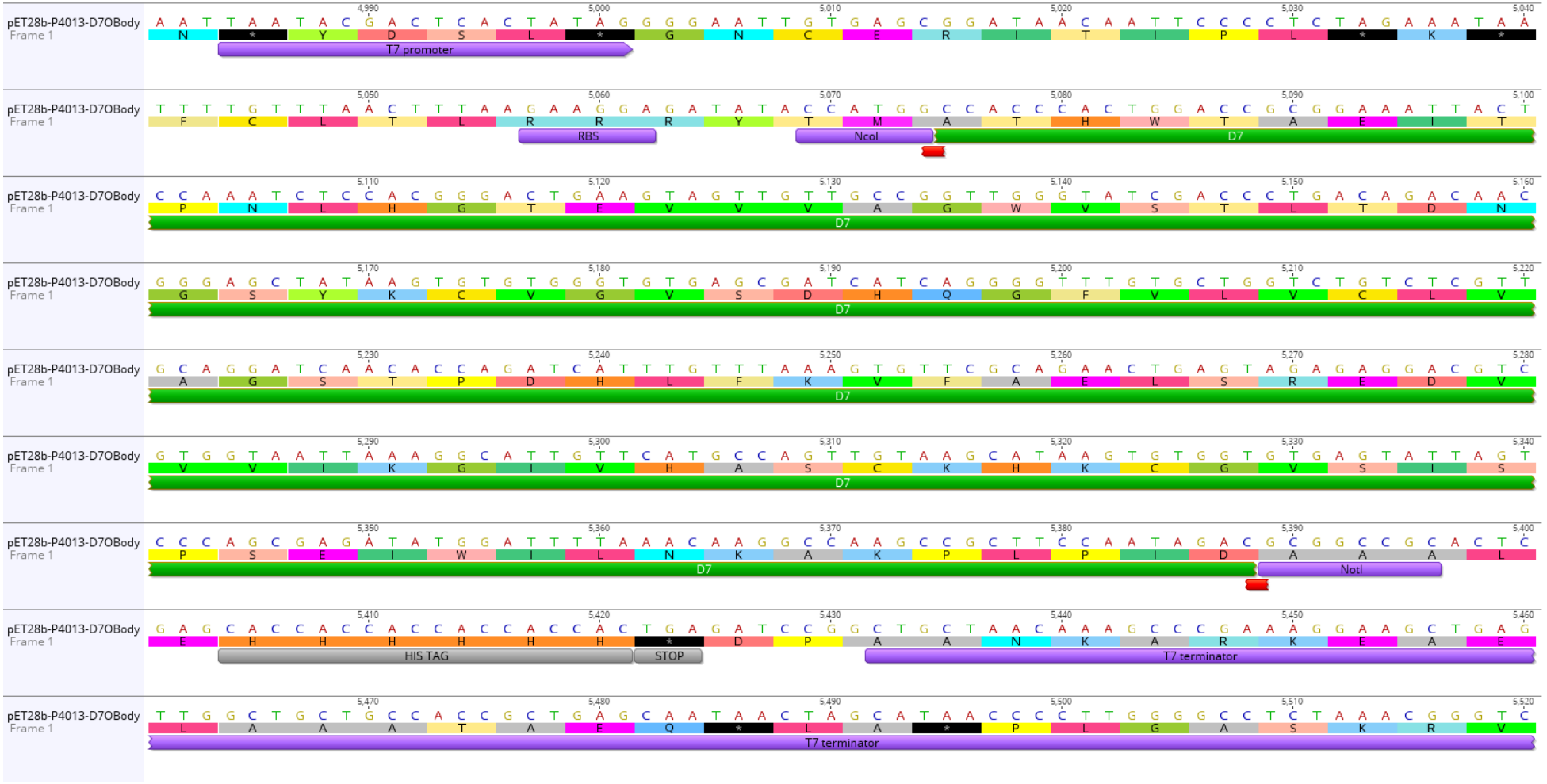
7.2 OBody sequences

7.2.1 Naïve library OBody D7 in phagemid (pOB2)



The naïve library OBody D7 (green) is shown with N-terminal YscC fragment (orange). The N-terminal signal sequence ssYscC (light pink) and important restriction enzyme cut sites are also indicated (Nco1, Sac2 and Not1). Note the insertion of YscC fragment between two Nco1 restriction sites.

7.2.2 Naïve library OBody D7 in cytoplasm expression plasmid (pET28b)

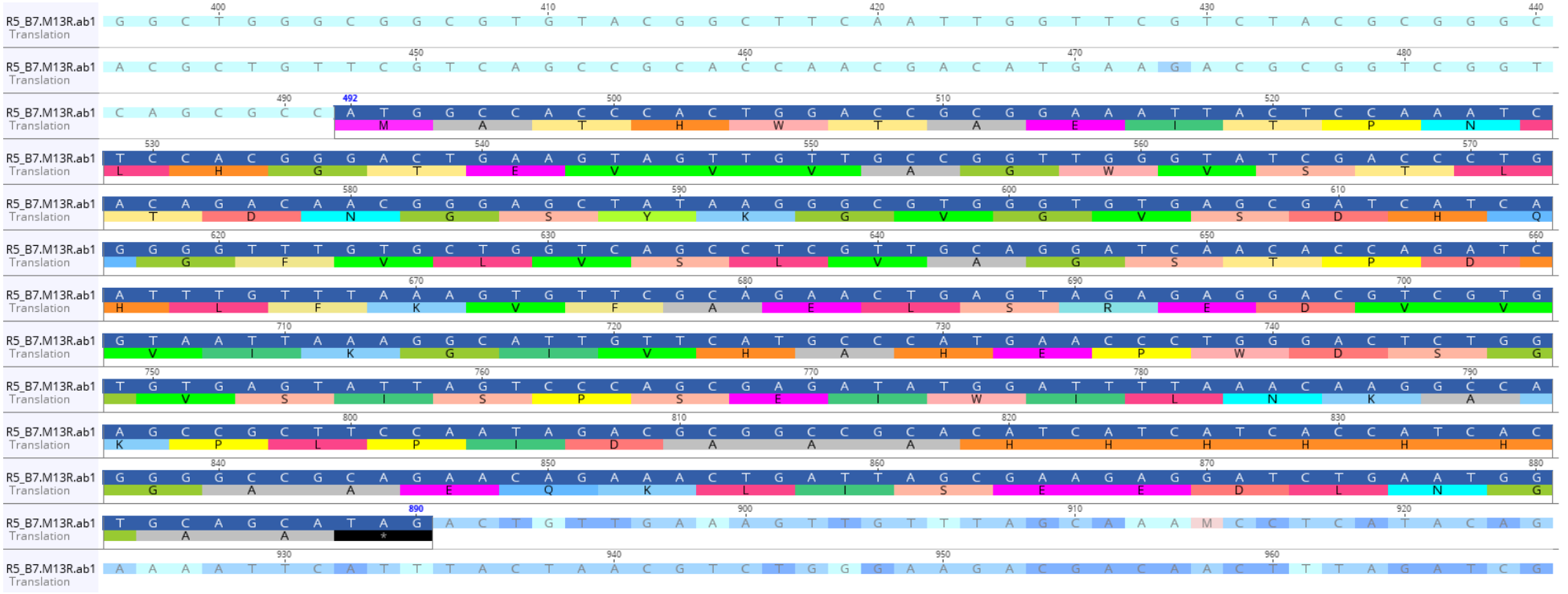


The naïve library OBody D7 (green) is shown without N-terminal YscC fragment. The important restriction enzyme cut sites are also indicated (Nco1 and Not1).

7.2.3 MBP fusion of OBody D7

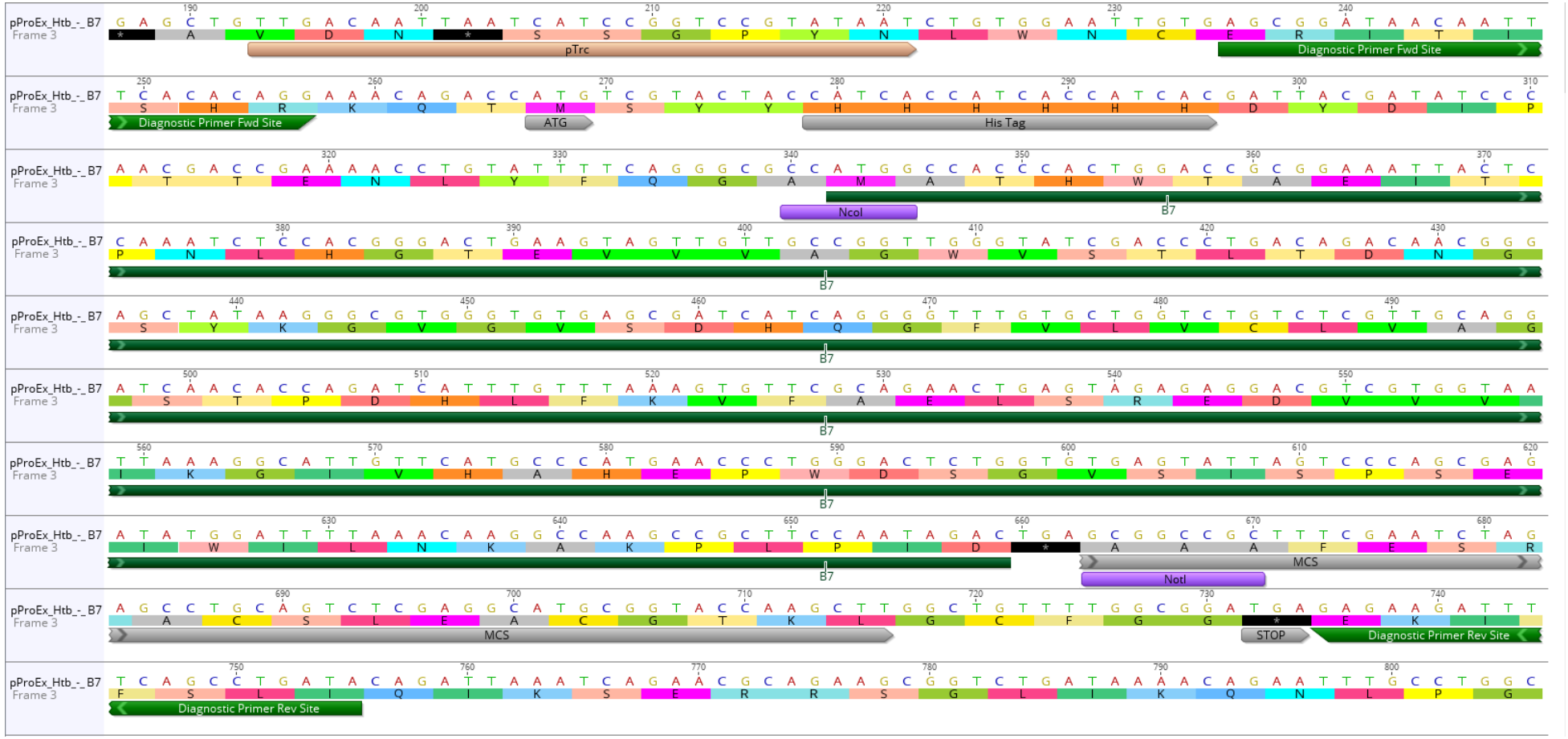
The naive library OBody D7 (green) is indicated showing nucleotide and amino acid sequences with N- terminal MBP (red). An important restriction site is indicated (HindIII, grey). Note that this fusion was created using ligation independent cloning (LIC).

7.2.4 Affinity maturation (AM) library OBody B7



The AM library OBody B7 (green) is highlighted showing nucleotide and amino acid sequences.

7.2.5 AM library OBody B7 in cytoplasm expression plasmid (pPROEX)

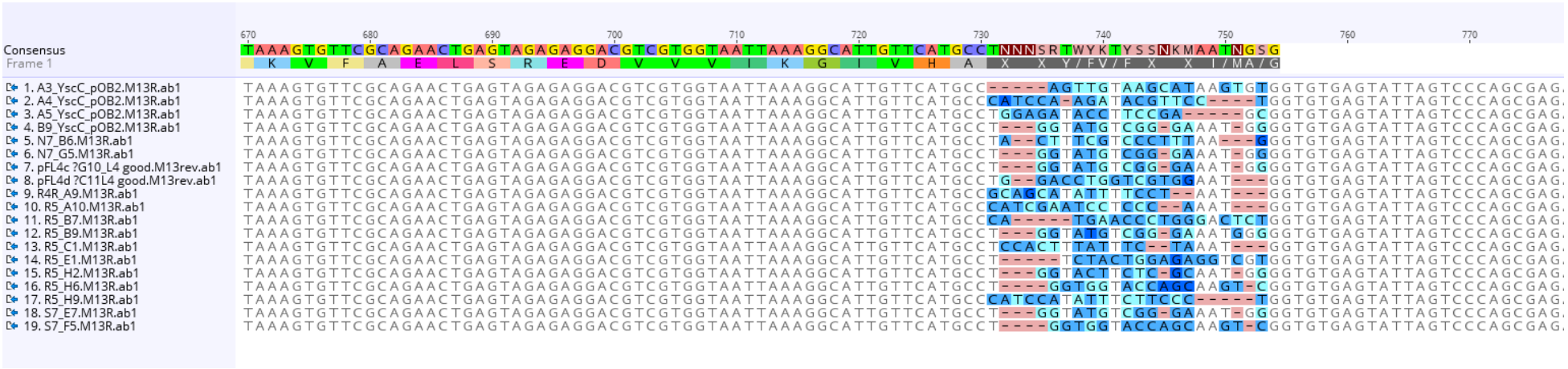


The AM library OBody B7 (green) is shown without N-terminal YscC fragment. Note that this plasmid has a N-terminal histag but uses the same restriction cut sites as pET28b (Nco1 and Not1, violet) as indicated.

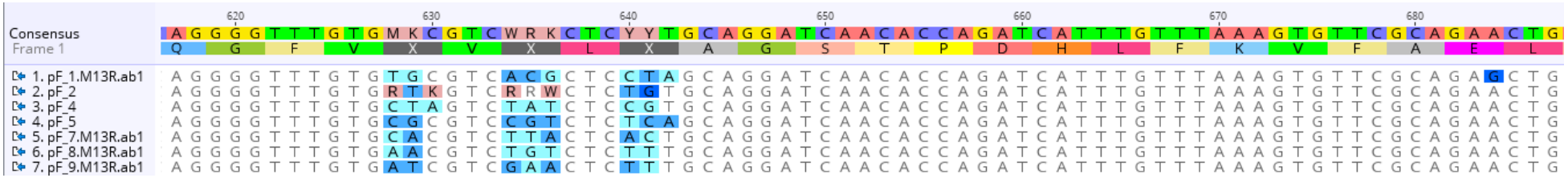
7.2.6 MBP fusion of OBody B7

The AM library OBody B7 (green) is indicated showing nucleotide and amino acid sequences with N- terminal MBP (red). The important restriction sites are indicated (KpnI and HindIII, grey).

7.2.7 AM library clones

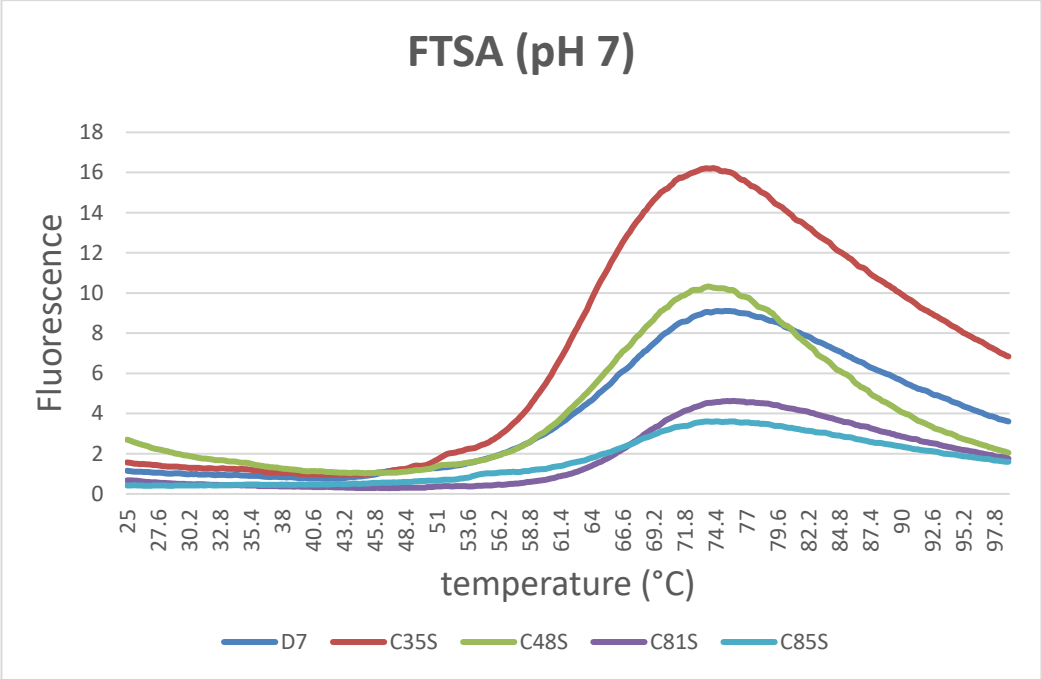


The DNA sequence alignment of OBodies from R4 to R7 of (L4_DM) AM library is shown focused on the randomized loop (colored). Note that the non-randomized areas of the OBodies have no point mutations.



The DNA sequence alignment of pF_35 OBodies from AM library is shown focused on the β-sheet 3 (colored). Note that the non-randomized areas of the OBodies have no point mutations.

7.3 FTSA of D7 mutants

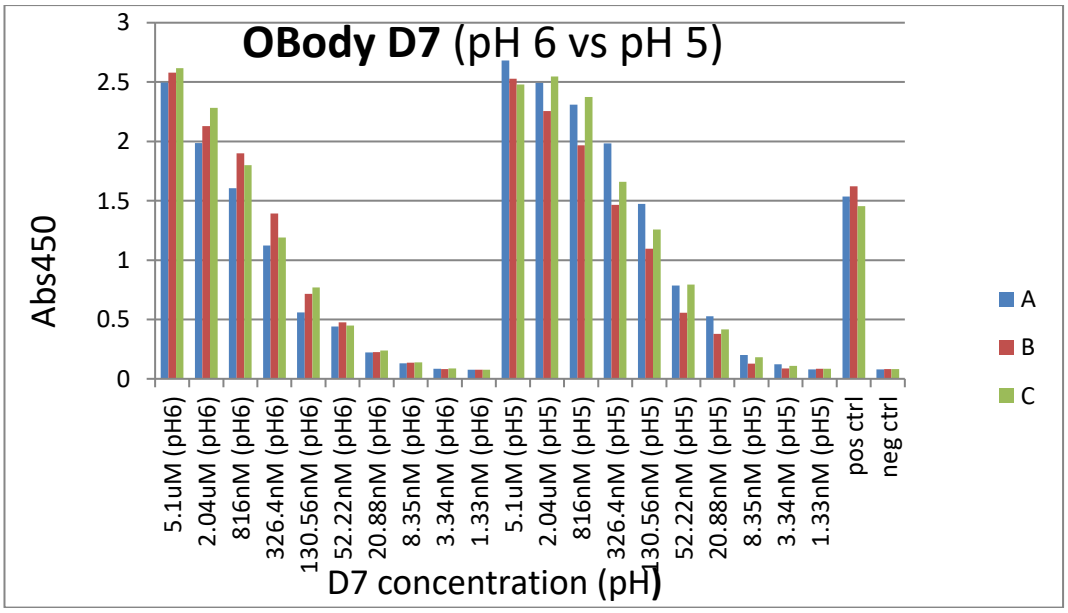


The raw thermal unfolding data of fluorescent thermal shift assays (FTSA) performed on labelled D7 mutants at pH7 is shown. The midpoint of the slope of these curves provide the respective melting temperatures (T_m).

7.4 ELISA of D7 and mutants

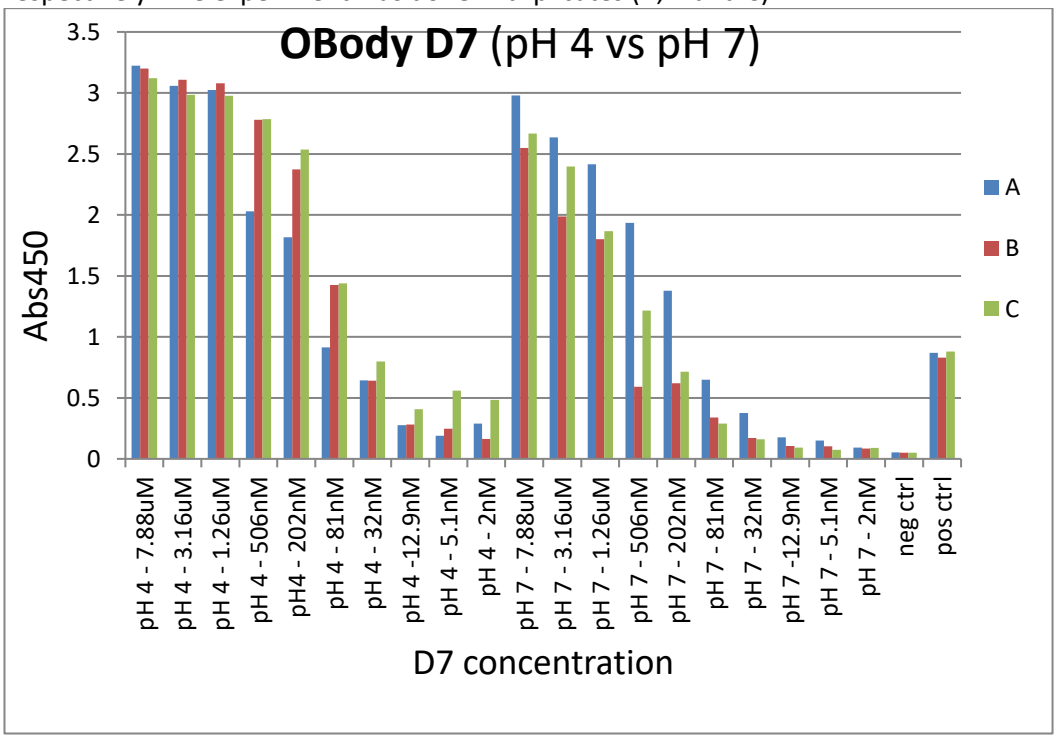
7.4.1 P4 immobilized ELISA: pH

7.4.1.1 pH 6 and pH 5



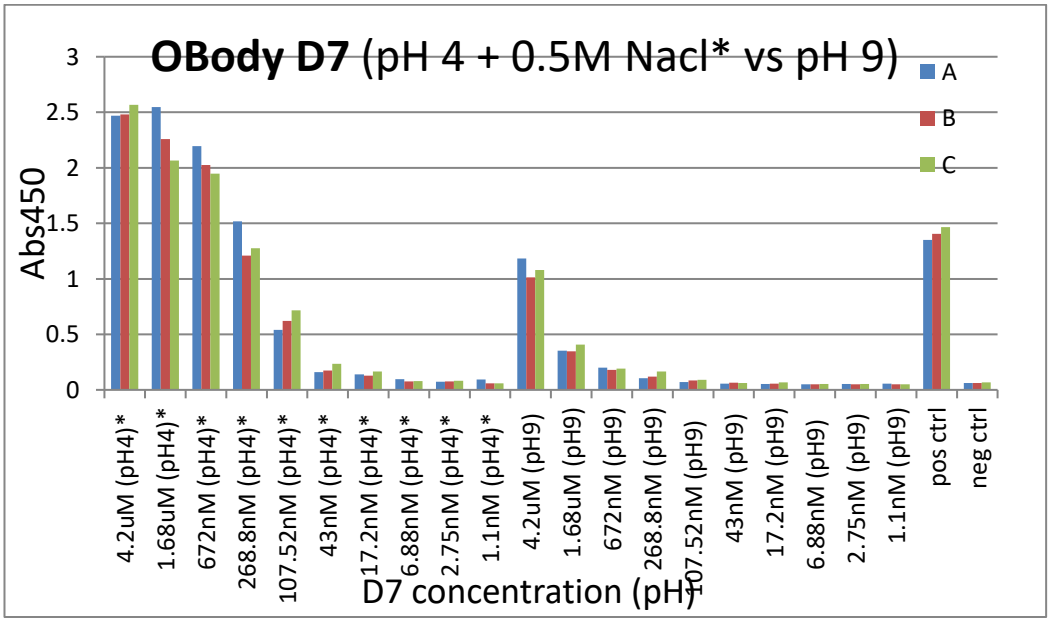
The raw data of ELISA of decreasing concentrations of OBody D7 with immobilized P4 at equilibrium is shown for pH 6 and pH 5. Biotinylated P4 was immobilized at a constant

concentration of 4.38 µM. Absorbance at 450nm (Y-axis) and D7 concentration in M (X-axis, pH) for the ELISA colorimetric reaction is shown. The controls used anti- P4 antibody to detect immobilized P4 and PBS for positive and negative control respectively. The experiment was done in triplicates (A, B and C).



The raw data of ELISA of decreasing concentrations of OBody D7 with immobilized P4 at equilibrium is shown for pH 4 and pH 7. Biotinylated P4 was immobilized at a constant concentration of 4.38 µM. Absorbance at 450nm (Y-axis) and D7 concentration in M (X-axis, pH) for the ELISA colorimetric reaction is shown. The controls used anti- P4 antibody to detect immobilized P4 and PBS for positive and negative control respectively. The experiment was done in triplicates (A, B and C).

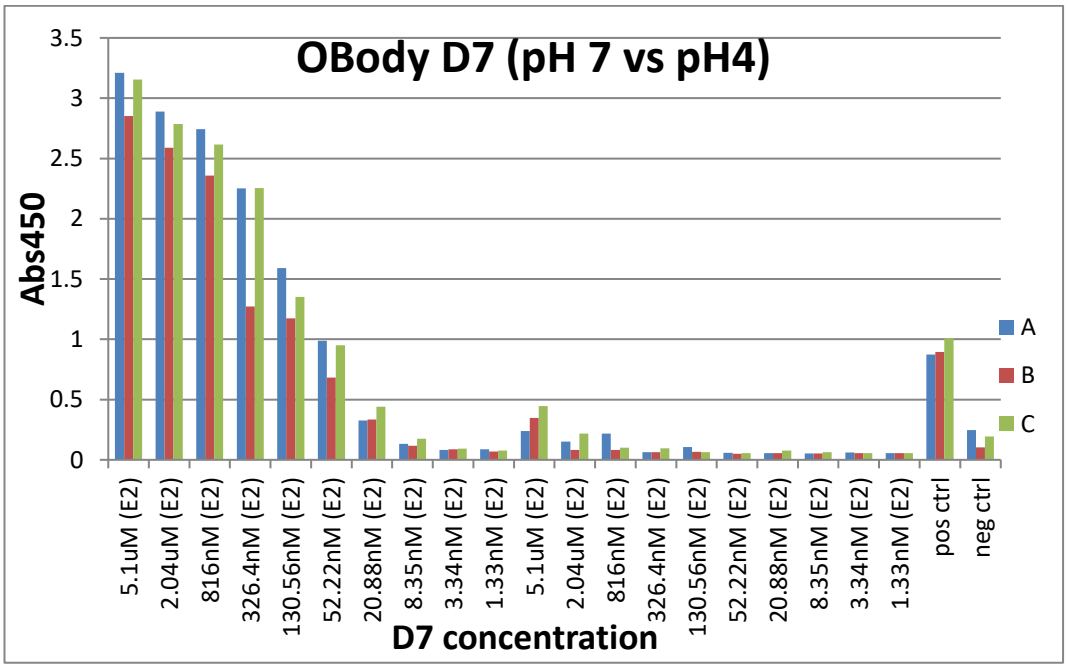
7.4.1.2 pH 4 with salt and pH 9



The raw data of ELISA of decreasing concentrations of OBody D7 with immobilized P4 at equilibrium is shown for pH 4 (with 0.5 M NaCl) and pH 9. Biotinylated P4 was immobilized at a constant concentration of 4.38 µM. Absorbance at 450nm (Y-axis) and D7 concentration in M (X-axis, pH) for the ELISA colorimetric reaction is shown. The controls used anti- P4 antibody to detect immobilized P4 and PBS for positive and negative control respectively. The experiment was done in triplicates (A, B and C). * refers to 0.5M NaCl being applicable only to pH 4.

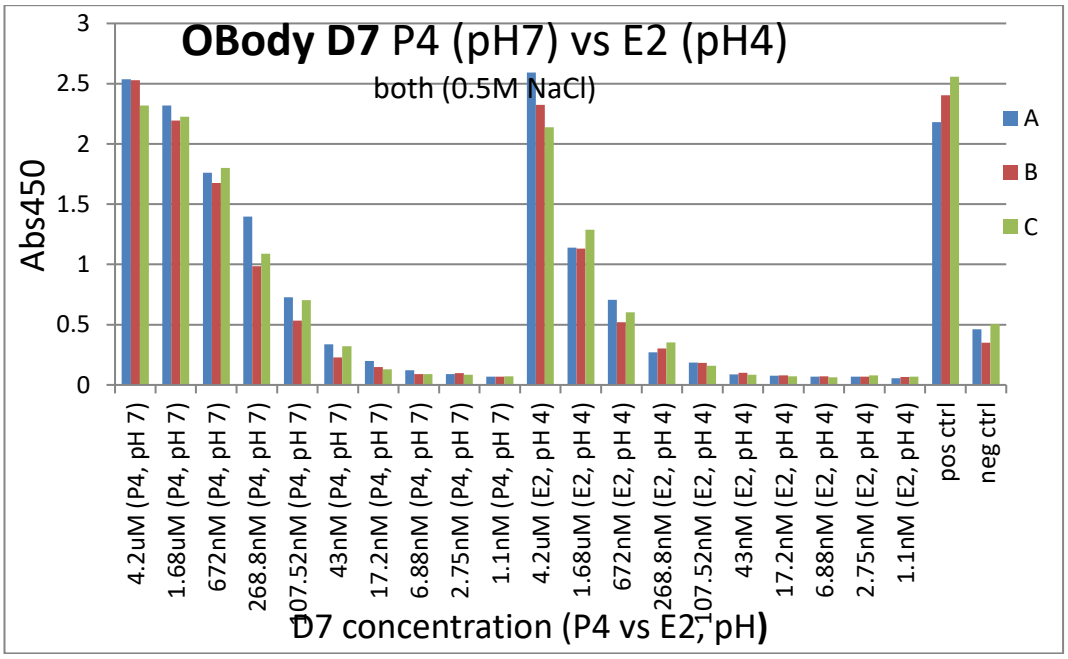
7.4.2 Antigen immobilized ELISA: P4 and E2

7.4.2.1 E2 (pH 7 vs pH 4)



The raw data of ELISA of decreasing concentrations of OBody D7 with immobilized E2 at equilibrium is shown for pH 7 and pH 4. Biotinylated E2 was immobilized at a constant concentration of 4.38 µM. Absorbance at 450nm (Y-axis) and D7 concentration in M (X-axis, E2) for the ELISA colorimetric reaction is shown. The controls used anti- P4 antibody to detect immobilized P4 and PBS for positive and negative control respectively. The experiment was done in triplicates (A, B and C).

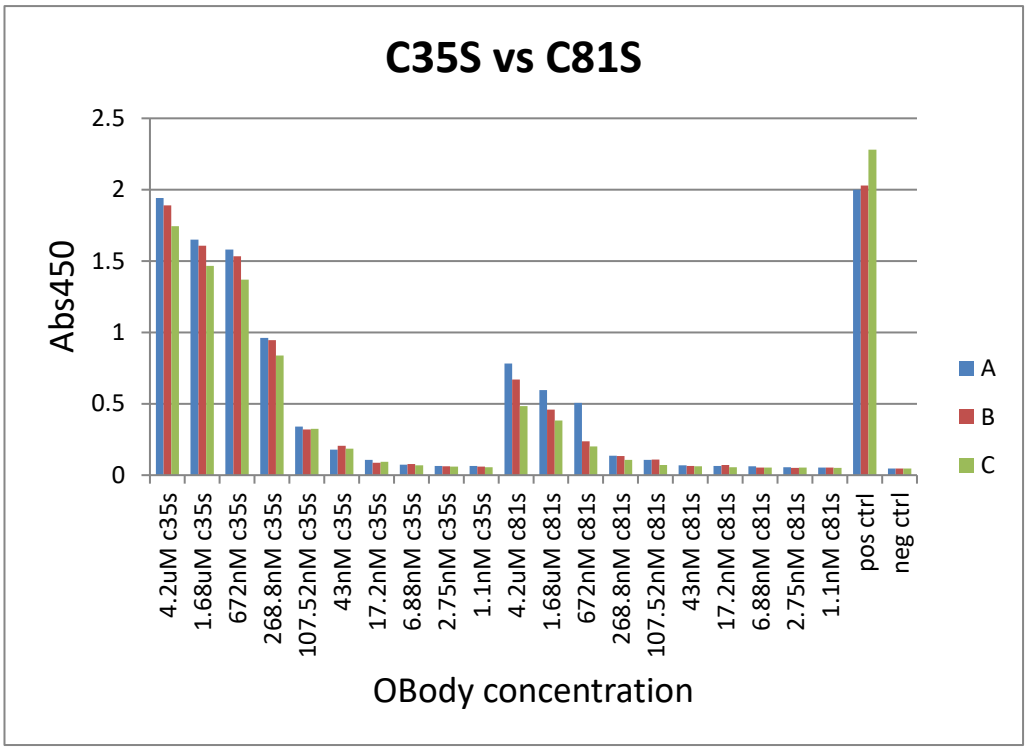
7.4.2.2 P4 (pH 7) and E2 (pH 4) both salt



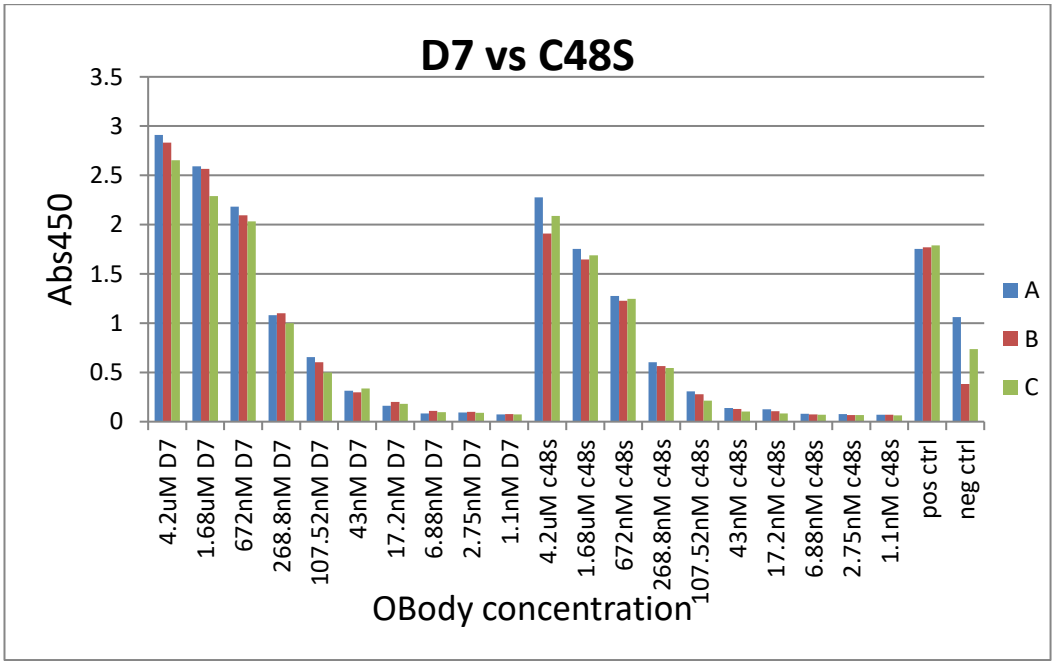
The raw data of ELISA of decreasing concentrations of OBody D7 with immobilized P4 or E2 at equilibrium is shown (with 0.5 M NaCl) at pH 7 and pH 4 respectively. Biotinylated P4 or E2 was immobilized at a constant concentration of 4.38 µM. Absorbance at 450nm (Y-axis) and D7 concentration in M (X-axis, P4 or E2, pH) for the ELISA colorimetric reaction is shown. The controls used anti- P4 antibody to detect immobilized P4 and PBS for positive and negative control respectively. The experiment was done in triplicates (A, B and C).

7.4.3 ELISA of D7 mutants

7.4.3.1 P4 immobilized ELISA

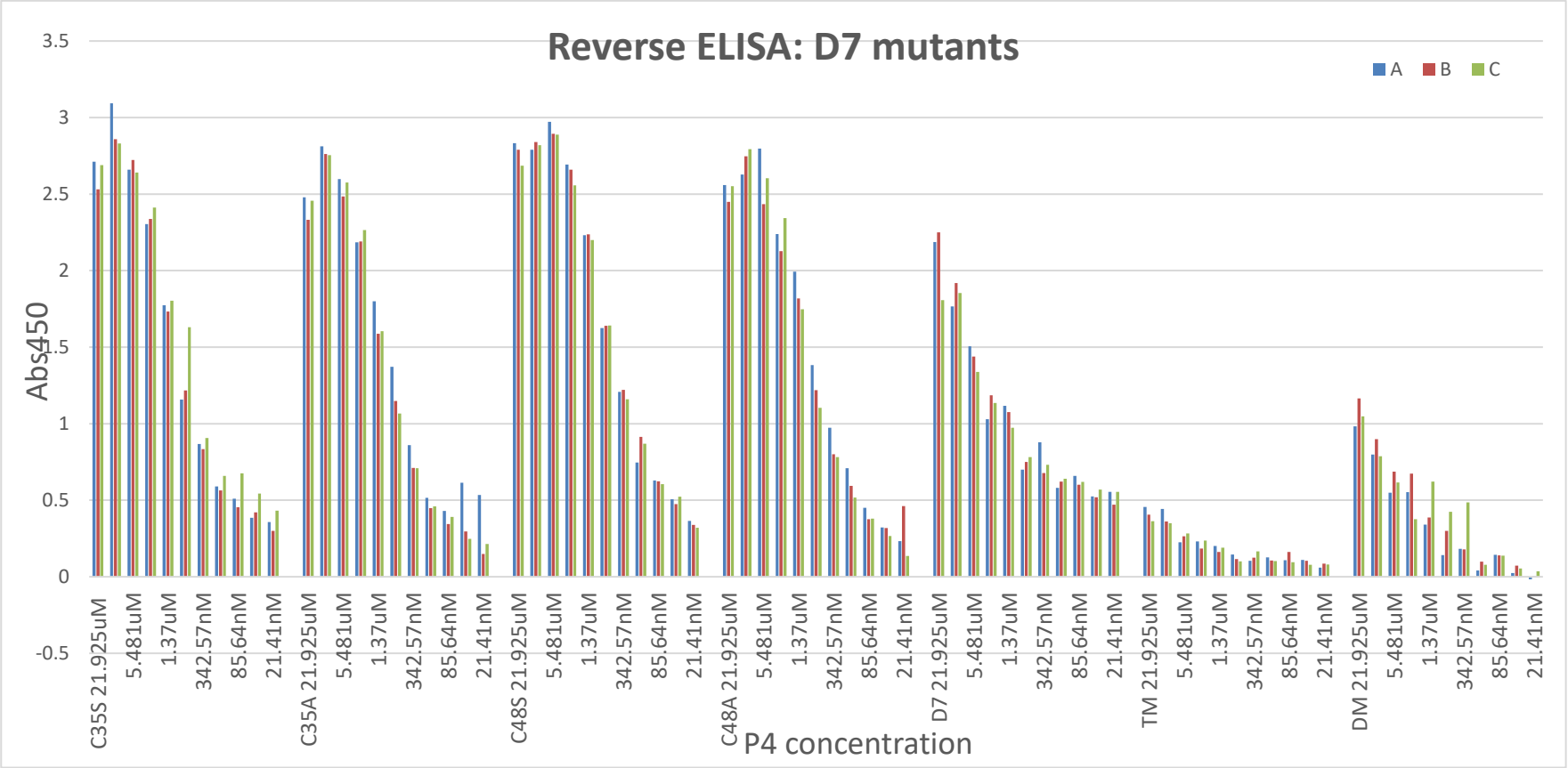


The raw data of ELISA of decreasing concentrations of OBody D7 mutants with immobilized P4 at equilibrium is shown for pH 7. Biotinylated P4 was immobilized at a constant concentration of 4.38 µM. Absorbance at 450nm (Y-axis) and D7 mutant concentration in M (X-axis, mutant) for the ELISA colorimetric reaction is shown. The controls used anti- P4 antibody to detect immobilized P4 and PBS for positive and negative control respectively. The experiment was done in triplicates (A, B and C).



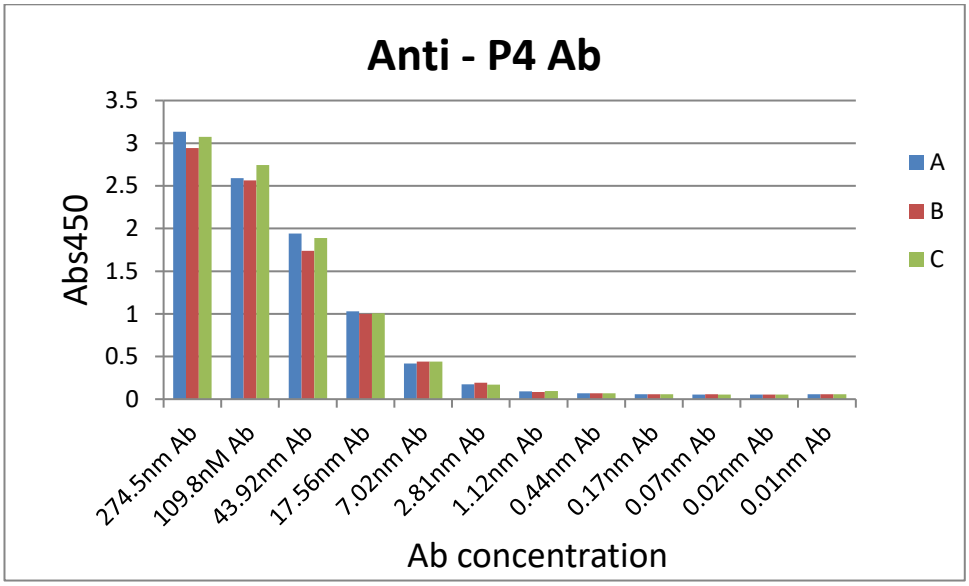
The raw data of ELISA of decreasing concentrations of OBody D7 mutants with immobilized P4 at equilibrium is shown for pH 7. Biotinylated P4 was immobilized at a constant concentration of 4.38 µM. Absorbance at 450nm (Y-axis) and D7 mutant concentration in M (X-axis, mutant) for the ELISA colorimetric reaction is shown. The controls used anti- P4 antibody to detect immobilized P4 and PBS for positive and negative control respectively. The experiment was done in triplicates (A, B and C).

7.4.3.2 OBody immobilized ELISA



The raw data of ELISA of immobilized OBody D7 mutants with decreasing concentrations of P4 at equilibrium is shown for pH 7. D7 mutants were immobilized at a constant concentration of 394.32 nM. Absorbance at 450nm (Y-axis) and P4 concentration in M (X-axis, OBody mutant) for the ELISA colorimetric reaction is shown. OBodies D7 and TM (C35S:C48S:C81S) were used (to detect immobilized P4) for positive and negative controls respectively. The experiment was done in triplicates (A, B and C). Note that TM (C35S:C48S:C81S) or triple mutant involving the loop (L4) has no binding to P4 and was thus used as negative control.

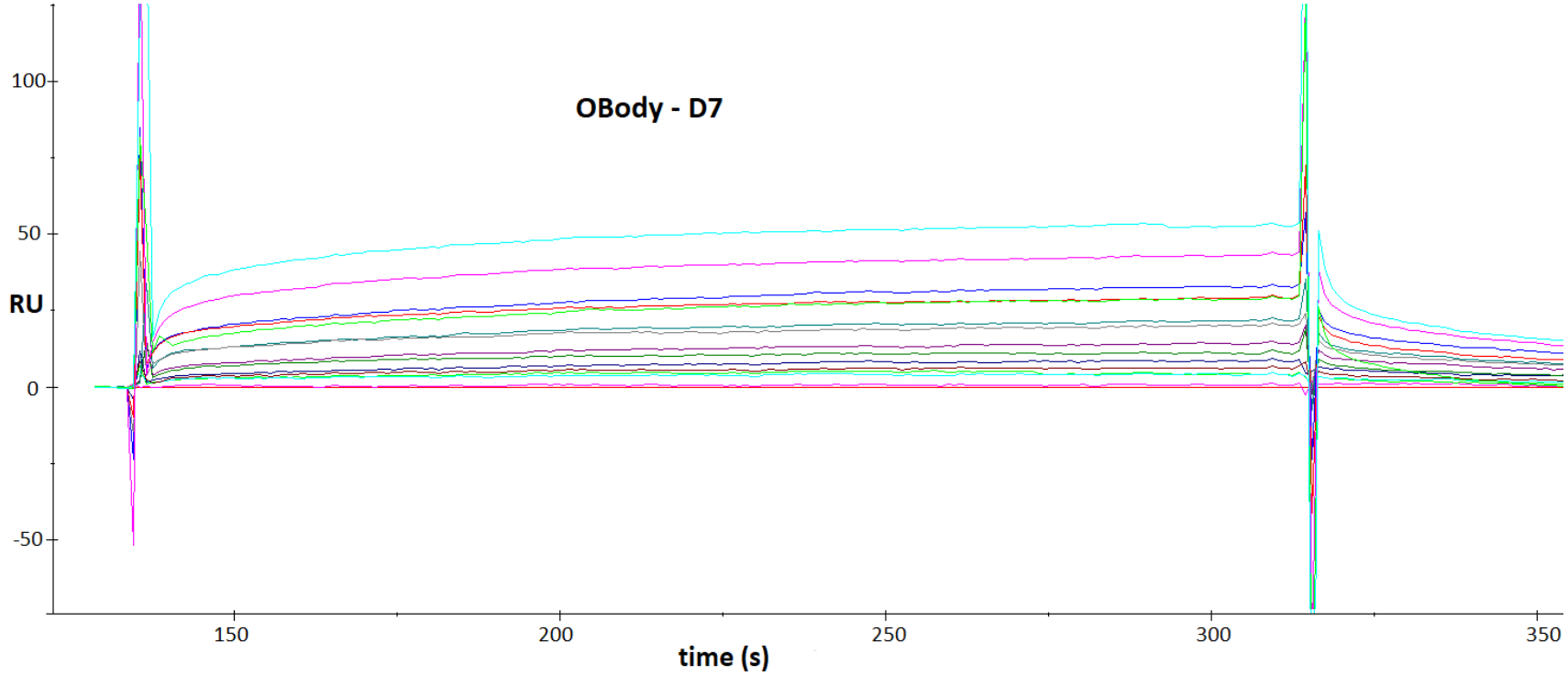
7.4.4 Antibody immobilized ELISA



The raw data of ELISA of decreasing concentrations of Anti -P4 antibody with immobilized P4 at equilibrium is shown for pH 7. Biotinylated P4 was immobilized at a constant concentration of 4.38 µM. Absorbance at 450nm (Y-axis) and anti- P4 Ab concentration in nM (X-axis) for the ELISA colorimetric reaction is shown. The experiment was done in triplicates (A, B and C).

7.5 SPR

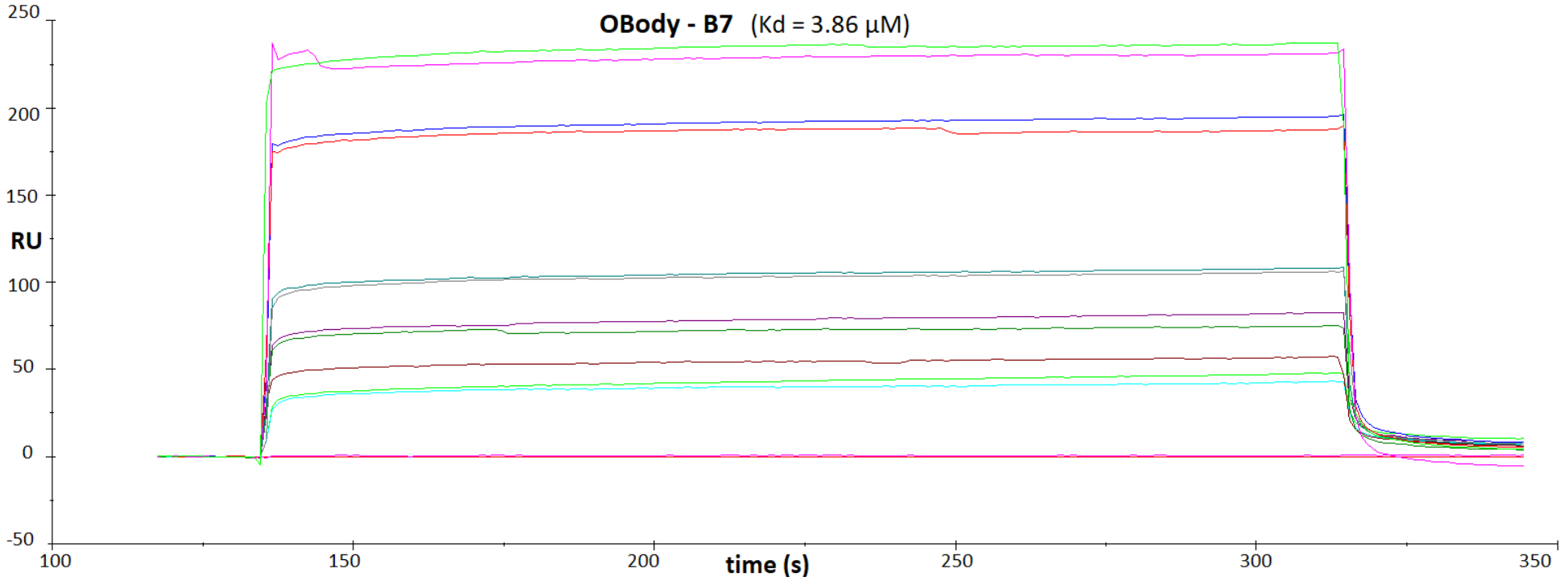
7.5.1 SPR of D7



The SPR generated sensorgram at equilibrium is shown for different concentrations of OBody D7 binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (Req, in RU) and the x-axis shows the duration of the experiment (in seconds). The rates of association

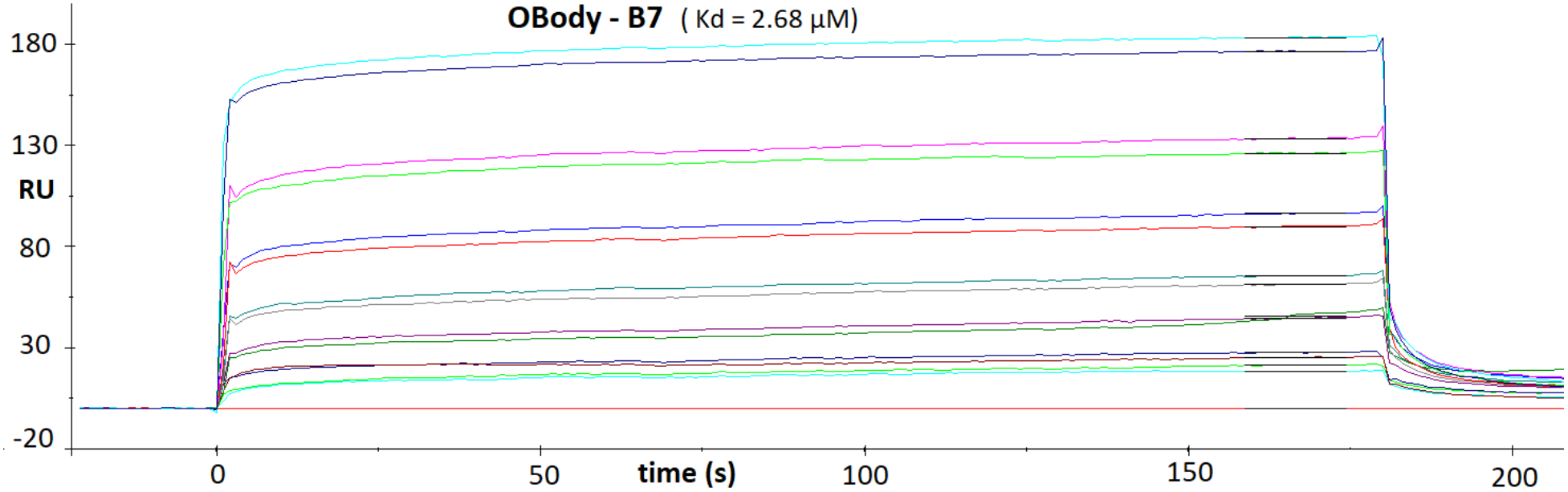
and dissociation of OBodies to P4 are rapid, therefore K_d was calculated at equilibrium. The different colours of the curves represent the different concentrations and duplicates of OBody D7.

7.5.2 SPR of B7



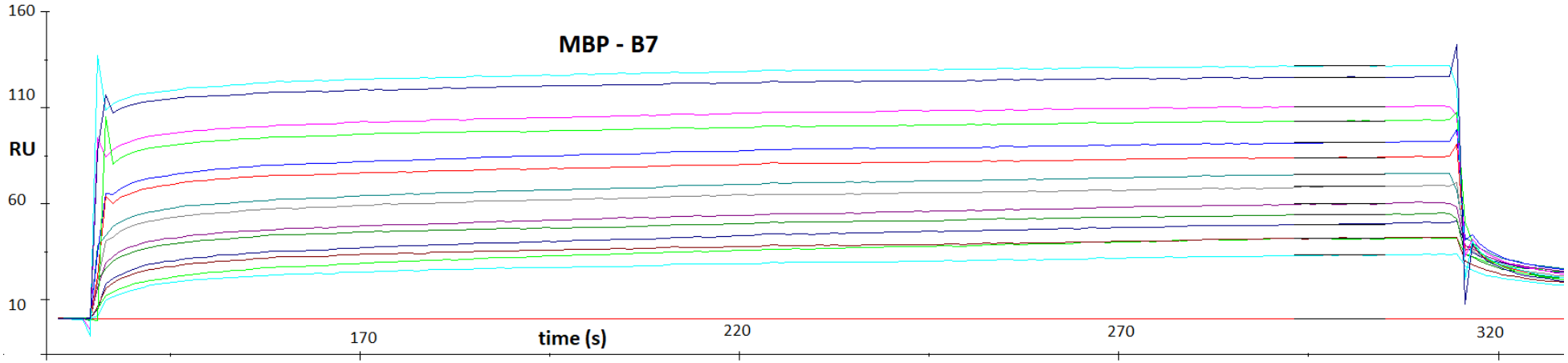
The SPR generated sensorgram at equilibrium is shown for different concentrations of OBody B7 binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (Req, in RU) and the x-axis shows the duration of the experiment (in seconds). The rates of association

and dissociation of OBodies to P4 are rapid, therefore K_d was calculated at equilibrium. The different colours of the curves represent the different concentrations and duplicates of OBody B7. Note that the gap in the sensorgram (100 – 150 RU) is due to omission of intermediate dilution of OBody B7.



The SPR generated sensorgram at equilibrium is shown for different concentrations of OBody B7 binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (Req, in RU) and the x-axis shows the duration of the experiment (in seconds). The rates of association and dissociation of OBodies to P4 are rapid, therefore K_d was calculated at equilibrium. The different colours of the curves represent the different concentrations and duplicates of OBody B7.

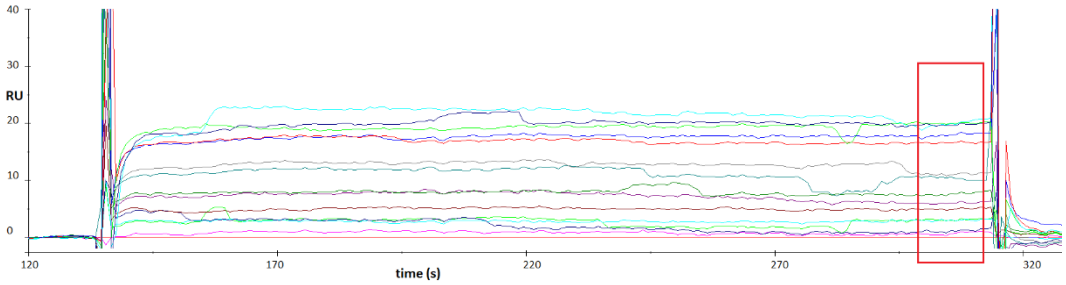
7.5.3 SPR of MBP-B7



The SPR generated sensorgram at equilibrium is shown for different concentrations of MBP fusion of OBody B7 binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (Req, in RU) and the x-axis shows the duration of the experiment (in seconds). The rates of association and dissociation of OBodies to P4 are rapid, therefore K_d was calculated at equilibrium. The different colours of the curves represent the different concentrations and duplicates of MBP-B7.

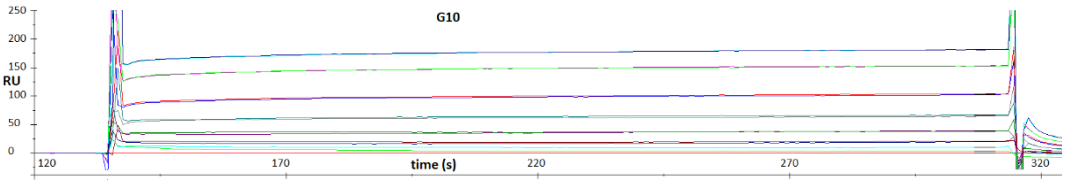
7.5.4 SPR of other OBodies

7.5.4.1 DM (C35S:C48S)



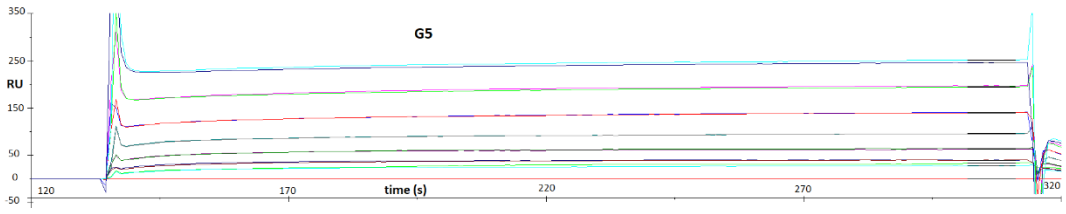
The SPR generated sensorgram at equilibrium is shown for different concentrations of D7 mutant DM binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (Req, in RU) and the x-axis shows the duration of the experiment (in seconds). The rates of association and dissociation of OBodies to P4 are rapid, therefore K_d was calculated at equilibrium. The different colours of the curves represent the different concentrations and duplicates of D7 mutant DM. Note that the data was used for analysis after reaching equilibrium (red box).

7.5.4.2 G10



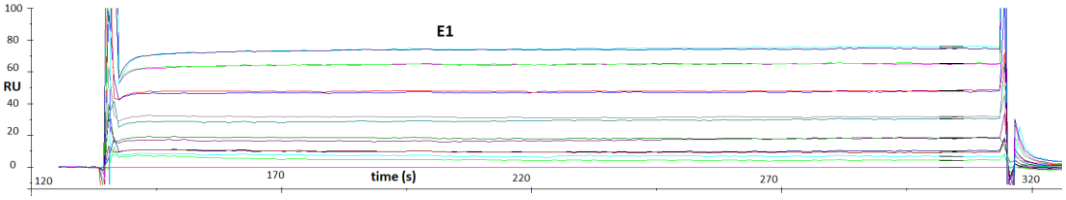
The SPR generated sensorgram at equilibrium is shown for different concentrations of OBody G10 binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (Req, in RU) and the x-axis shows the duration of the experiment (in seconds). The rates of association and dissociation of OBodies to P4 are rapid, therefore K_d was calculated at equilibrium. The different colours of the curves represent the different concentrations and duplicates of OBody G10.

7.5.4.3 G5



The SPR generated sensorgram at equilibrium is shown for different concentrations of OBody G5 binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (Req, in RU) and the x-axis shows the duration of the experiment (in seconds). The rates of association and dissociation of OBodies to P4 are rapid, therefore K_d was calculated at equilibrium. The different colours of the curves represent the different concentrations and duplicates of OBody G5. Note that G5 has the same loop (L4) sequence of G10 except presence of cysteines in the β- sheets.

7.5.4.4 E1



The SPR generated sensorgram at equilibrium is shown for different concentrations of OBody E1 binding to biotinylated P4 (ligand) immobilized on a SA chip. The y-axis shows the change in refractive index at equilibrium (Req, in RU) and the x-axis shows the duration of the experiment (in seconds). The rates of association and dissociation of OBodies to P4 are rapid, therefore K_d was calculated at equilibrium. The different colours of the curves represent the different concentrations and duplicates of OBody E1.

7.6 Diffraction datasets

Pathology	*Dataset*	*Resolution (Å)*	*Crystal condition*	*Space group*	*Unit cell parameters a, b, c, α, β, γ*	*Completeness (%)*
twinned	115_01	2.75	28% w/v PEG 4000, 0.1M Tris pH 8	P1 21 1	62.15, 109.49, 149.52, 90, 96.79, 90	87.6
twinned	115_07	2.82	28% w/v PEG 4000, 0.1M Tris pH 8	P1 21 1	62.38, 148.6, 109.7, 90, 90.08, 90	98.4
twinned	115_12	2.49	28% w/v PEG 4000, 0.1M Tris pH 8	P1 21 1	61.83, 142.76, 109.5, 90, 90.37, 90	84.4
	116_01	3.28	28% w/v PEG 4000, 0.1M Tris pH 8	P1 21 1	62.32, 150.83, 109.97, 90, 90.39, 90	99.5
	116_06	3.47	28% w/v PEG 4000, 0.1M Tris pH 8	P1 21 1	62.75, 150.73, 109.75, 90, 90.16, 90	98.6
tNCS	116_05	1.97	30% w/v PEG MME 2000, 0.1M Tris pH 8	P1 21 1	65.92, 193.9, 82.43, 90, 97.7, 90	95.3
tNCS	116_07	1.97	30% w/v	P1 21 1	65.78,	96.7

			PEG MME 2000, 0.1M Tris pH 8		195.67, 83.12, 90, 96.92, 90	
tNCS	118	1.83	30% w/v PEG MME 2000, 0.1M Tris pH 8	P1 21 1	65.43, 194.98, 82.96, 90, 97.13, 90	86.4